LECTURE NOTES 3 FOR CAMBRIDGE PART III COURSE ON "ELEMENTARY METHODS IN ANALYTIC NUMBER THEORY", LENT 2015

ADAM J HARPER

Abstract. These are rough notes covering the third block of lectures in the "Elementary Methods in Analytic Number Theory" course. In these lectures we use sieve weights to prove versions of Selberg's Fundamental Lemma/Symmetry Formula, and sketch how this leads to an elementary proof of the Prime Number Theorem. We also discuss the work of Goldston–Pintz–Yıldırım, of Zhang, and of Maynard and Tao on small gaps between primes.

(No originality is claimed for any of the contents of these notes. In particular, they borrow from the book of Montgomery and Vaughan [3].)

8. An elementary proof of the Prime Number Theorem

In this section we shall prove the following fundamental result.

Theorem 8.1 (Prime Number Theorem, Hadamard, de la Vallée Poussin, 1896). *We have*
$$\pi(x) := \#\{p \leq x : p \text{ prime}\} \sim \frac{x}{\log x} \quad \text{as } x \rightarrow \infty.$$
Equivalently, we have
$$\Psi(x) := \sum_{n \leq x} \Lambda(n) \sim x \quad \text{as } x \rightarrow \infty,$$
where $\Lambda(n)$ is the von Mangoldt function.

To see that the two statements are equivalent, note that
$$\begin{aligned}
\Psi(x) = \sum_{\sqrt{x} \leq p \leq x} \log p + O(\sqrt{x}\log x) &= \pi(x)\log x - \sum_{\sqrt{x} \leq p \leq x} \log(x/p) + O(\sqrt{x}\log x) \\
&= \pi(x)\log x + O(\sum_{1 \leq k \leq \frac{\log x}{2\log 2}+1} \sum_{x/2^k \leq p \leq x/2^{k-1}} k + \sqrt{x}\log x) \\
&= \pi(x)\log x + O(\frac{x}{\log x} \sum_{1 \leq k \leq \frac{\log x}{2\log 2}+1} \frac{k}{2^k} + \sqrt{x}\log x),
\end{aligned}$$
where the final equality uses Chebychev's estimate that $\sum_{p \leq x/2^{k-1}} 1 \ll x/(2^{k-1}\log(x/2^{k-1})) \ll x/(2^k \log x)$. Since the "big Oh" term is $O(x/\log x)$, we see that $\Psi(x) \sim x$ if and only if $\pi(x) \sim x/\log x$. We shall actually prove the theorem in the form $\Psi(x) \sim x$ as $x \rightarrow \infty$.

Date: 2nd April 2015.

Hadamard and de la Vallée Poussin's proofs used the Riemann zeta function, and for a long time it was thought that no really different proof was possible. In 1921, Hardy famously remarked that "A proof... not fundamentally dependent on the theory of functions, seems to me extraordinarily unlikely...". But in 1948-49, Erdős and Selberg found beautiful "elementary" proofs of the Prime Number Theorem.

One can think of the Riemann zeta function as providing a kind of precise averaged information about the distribution of the primes (actually a different average at each complex number s), from which one can deduce the Prime Number Theorem using complex analysis. For the elementary proofs we also need to start with some averaged information, which you have seen before on the first problem sheet.

Proposition 8.2. *Let $x \geq 1$ and let $T(x) := \sum_{n \leq x} \Psi(x/n)$. Then*

$$T(x) = \sum_{n \leq x} \log n = x \log x - x + O(\log(x+1)).$$

Note that $\sum_{n \leq x} \frac{x}{n} = x \log x + \gamma x + O(1)$, where γ is Euler's constant, so the Proposition is at least consistent with the Prime Number Theorem. Note also that we cannot hope to improve the "big Oh" term in the Proposition, since when x changes from being just smaller than an integer to just larger than an integer the left hand side will increase by about $\log x$, whereas $x \log x - x$ can change by an arbitrarily small amount.

Now if $(\lambda_n)_{n \in \mathbb{N}}$ is any real sequence, we have

$$\begin{aligned}\sum_{n \leq x} \Psi(x/n) \sum_{d|n} \lambda_d = \sum_{d \leq x} \lambda_d \sum_{m \leq x/d} \Psi(x/md) &= \sum_{d \leq x} \lambda_d T(x/d) \\ &= \sum_{d \leq x} \lambda_d \left(\frac{x \log(x/d)}{d} - \frac{x}{d} + O(\log(x/d+1)) \right).\end{aligned}$$

If $x \geq 2$, and if we choose $\lambda_1 = 1$, $\lambda_2 = -2$, and $\lambda_n = 0$ for all $n \geq 3$, then we obtain

$$\sum_{n \leq x} \Psi(x/n)(-1)^{n+1} = (x \log x - x) - 2((x/2)\log(x/2) - x/2) + O(\log x) = x \log 2 + O(\log x).$$

As you saw on the first problem sheet, this is sufficient to deduce Chebychev's bounds for $\Psi(x)$, or more precisely that

$$(\log 2)x + O(\log x) \leq \Psi(x) \leq (2 \log 2)x + O(\log x) \quad \forall x \geq 2.$$

Another obvious choice is to take $\lambda_n = \mu(n)$ for all n, since then $\sum_{d|n} \lambda_d = \sum_{d|n} \mu(d) = \mathbf{1}_{n=1}$, and we obtain

$$\Psi(x) = \sum_{d \leq x} \mu(d) \left(\frac{x \log(x/d)}{d} - \frac{x}{d} + O(\log(x/d+1)) \right).$$

But this is not really useful, because the contribution from the "big Oh" terms is $O(\sum_{d \le x} \log(x/d+1))$, which is $O(x)$. This is too big to be an error term in the Prime Number Theorem.

Our first key idea, motivated by the Selberg sieve, is to choose the weights λ_n to be a "smoothed out" version of the Möbius function $\mu(n)$. There are various such choices that will give an acceptable error term and still kill off most of the summands $\Psi(x/n)$ on the left hand side.

Proposition 8.3 (Selberg's Fundamental Lemma/ Symmetry Formula, 1949)**.** *Define the remainder $R(x) := \Psi(x) - x$. For any $x \ge 2$ we have*

$$\Psi(x)+\sum_{n\le x}\Psi(x/n)\frac{\Lambda(n)}{\log x} = 2x+O\left(\frac{x}{\log x}\right), \quad \textit{and} \quad R(x)+\sum_{n\le x}R(x/n)\frac{\Lambda(n)}{\log x} = O\left(\frac{x}{\log x}\right).$$

In addition, we have

$$R(x)+\sum_{n\le x}R(x/n)\frac{\Lambda(n)^2}{\log^2 x} - 2\sum_{\substack{p^k q^l \le x,\\ p\ne q}} R(x/p^k q^l)\frac{\log p \log q}{\log^2 x} = O\left(\frac{x}{\log x}\right),$$

where the second sum is over numbers that are a product of two powers of distinct primes.

Proof of Proposition 8.3. To prove the first two statements we choose $\lambda_d = \mu(d)\frac{\log(x/d)}{\log x}$ in the preceding discussion. On the one hand we have

$$\sum_{d|n}\lambda_d = \sum_{d|n}\mu(d)\frac{\log(x/n)+\log(n/d)}{\log x} = \mathbf{1}_{n=1} + \frac{\Lambda(n)}{\log x},$$

since $\sum_{d|n}\mu(d) = \mathbf{1}_{n=1}$ and $\sum_{d|n}\mu(d)\log(n/d) = \Lambda(n)$. Therefore we have

$$\Psi(x)+\sum_{n\le x}\Psi(x/n)\frac{\Lambda(n)}{\log x} = \sum_{n\le x}\Psi(x/n)\sum_{d|n}\lambda_d = \sum_{d\le x}\lambda_d\left(\frac{x\log(x/d)}{d} - \frac{x}{d} + O(\log(x/d+1))\right).$$

Here the contribution from the "big Oh" terms is

$$\ll \frac{1}{\log x}\sum_{d\le x}\log^2(x/d+1) \ll \frac{1}{\log x}\sum_{1\le k\le \frac{\log x}{\log 2}+1}\ \sum_{x/2^k\le d\le x/2^{k-1}} k^2 \ll \frac{x}{\log x}\sum_{1\le k\le\frac{\log x}{\log 2}+1}\frac{k^2}{2^k} \ll \frac{x}{\log x},$$

which is acceptable.

To determine the size of the other terms on the right hand side, we can employ a neat comparison trick. Note that if we define $U(x) := \sum_{n\le x}(\frac{x}{n} - (1+\gamma))$ for all x, where γ is Euler's constant, then we have $U(x) = x\log x - x + O(1)$. So if we run the same argument as above, with $\frac{x}{n} - (1+\gamma)$ replacing $\Psi(x/n)$ and $U(x)$ replacing $T(x)$, we obtain that

$$(x-(1+\gamma))+\sum_{n\leq x}(x/n-(1+\gamma))\frac{\Lambda(n)}{\log x}=\sum_{d\leq x}\lambda_d\left(\frac{x\log(x/d)}{d}-\frac{x}{d}+O(1)\right),$$

and therefore

$$\begin{aligned}\sum_{d\leq x}\lambda_d\left(\frac{x\log(x/d)}{d}-\frac{x}{d}\right) &= (x-(1+\gamma))+\sum_{n\leq x}(x/n-(1+\gamma))\frac{\Lambda(n)}{\log x}+O\left(\frac{x}{\log x}\right)\\ &= x+\sum_{n\leq x}\frac{x}{n}\frac{\Lambda(n)}{\log x}+O\left(\frac{x}{\log x}\right)=2x+O\left(\frac{x}{\log x}\right),\end{aligned}$$

by the estimates of Chebychev and Mertens (Facts 1 and 2 from Chapter 0). Inserting this above gives the desired equality

$$\Psi(x)+\sum_{n\leq x}\Psi(x/n)\frac{\Lambda(n)}{\log x}=2x+O\left(\frac{x}{\log x}\right),$$

or instead subtracting it gives that $R(x)+\sum_{n\leq x}R(x/n)\frac{\Lambda(n)}{\log x}=O\left(\frac{x}{\log x}\right)$.

To prove the other statement we instead choose $\lambda_d=\mu(d)(1-\frac{\log^2 d}{\log^2 x})$, for which we have

$$\sum_{d|n}\lambda_d=\sum_{d|n}\mu(d)-\frac{1}{\log^2 x}\sum_{d|n}\mu(d)(\sum_{t|d}\Lambda(t))^2=\mathbf{1}_{n=1}-\frac{1}{\log^2 x}\sum_{t_1,t_2|n}\Lambda(t_1)\Lambda(t_2)\sum_{\substack{d|n,\\ [t_1,t_2]|d}}\mu(d),$$

where $[t_1,t_2]$ denotes the least common multiple. Here the sum $\sum_{\substack{d|n,\\ [t_1,t_2]|d}}\mu(d)$ vanishes unless $[t_1,t_2]$ is squarefree and n has no prime factors other than those of $[t_1,t_2]$, so we see

$$\sum_{d|n}\lambda_d=\mathbf{1}_{n=1}+\mathbf{1}_{n=p^k}\frac{\log^2 p}{\log^2 x}-2\mathbf{1}_{n=p^kq^l,q\neq p}\frac{\log p\log q}{\log^2 x}.$$

Inserting this expression yields the result, the right hand side being handled as before by subtracting the corresponding expression for $x/n-(1+\gamma)$ and $U(x)$. □

Remark 8.4. It may not be clear that the choices of λ_d that we made when proving Proposition 8.3 really are like Selberg sieve weights. However, if we take $\mathcal{P}$ to be all primes and $g(d)=1/d$ in the "Optimal" Λ^2 construction (Lemma 2.5) then we get

$$\rho_d=\mu(d)\prod_{p|d}(1-\frac{1}{p})^{-1}\cdot\frac{1}{J}\sum_{\substack{t\leq\sqrt{D}/d,\\ t\text{ squarefree},(t,d)=1}}\frac{1}{t}\prod_{p|t}(1-\frac{1}{p})^{-1},\quad \text{where } J=\sum_{\substack{t\leq\sqrt{D},\\ t\text{ squarefree}}}\frac{1}{t}\prod_{p|t}(1-\frac{1}{p})^{-1}.$$

If we ignore the products over primes and coprimality and squarefreeness conditions (whose effects roughly cancel each other out), we see that

$$\rho_d\approx\mu(d)\frac{\sum_{t\leq\sqrt{D}/d}1/t}{\sum_{t\leq\sqrt{D}}1/t}\approx\mu(d)\frac{\log(\sqrt{D}/d)}{\log(\sqrt{D})},$$

as in Proposition 8.3.

Now we need to deduce the Prime Number Theorem from the Fundamental Lemma (Proposition 8.3). It will suffice to prove that $R(x) = o(x)$ as $x \to \infty$, so suppose we knew that $|R(t)| \leq \beta t + O(1)$ for all $t \leq x/2$, for some fixed $\beta > 0$. Then the Fundamental Lemma implies that

$$|R(x)| \leq \sum_{n \leq x} |R(x/n)| \frac{\Lambda(n)}{\log x} + O\left(\frac{x}{\log x}\right) \leq \beta \sum_{n \leq x} \frac{x}{n} \frac{\Lambda(n)}{\log x} + O\left(\frac{x}{\log x}\right) = \beta x + O\left(\frac{x}{\log x}\right),$$

where we again used the estimates of Chebychev and Mertens. If we could replace β here by any smaller constant then we would be able to prove the Prime Number Theorem by induction on x, so this obvious argument *just* fails (i.e. using the Selberg-type weights has removed *almost* enough terms for things to work out).

The second formula in the Fundamental Lemma fails in the same way, but it has the crucial advantage of a difference in signs between the two sums, so we have the possibility of extracting a little bit more cancellation. This is the second key idea in the elementary proof of the Prime Number Theorem (and is due originally, in various forms, to Erdős and Selberg).

Sketch proof of Theorem 8.1. Let x be large, and suppose $\beta > 0$ is such that $|R(t)| = |\Psi(t) - t| \leq \beta t + O(1)$ for all $t \leq x/2$. In view of Chebychev's estimates we may start by taking $\beta = 1/2$, and we will show that under these hypotheses we have

$$|R(x)| \leq (\beta - 0.007\beta^2)x + O(1).$$

Then inductively we can replace β by $\beta - 0.007\beta^2$ (for x large enough), and repeating the argument obtain that $R(t) = o(t)$ as $t \to \infty$.

By the Fundamental Lemma and the triangle inequality we have

$$|R(x)| \leq \sum_{0 \leq k \leq \frac{\log x}{\log 16}} \left| - \sum_{\frac{x}{16^{k+1}} < n \leq \frac{x}{16^k}} R(x/n) \frac{\Lambda(n)^2}{\log^2 x} + 2 \sum_{\substack{\frac{x}{16^{k+1}} < p^k q^l \leq \frac{x}{16^k}, \\ p \neq q}} R(x/p^k q^l) \frac{\log p \log q}{\log^2 x} \right| + O\left(\frac{x}{\log x}\right),$$

and if we just bounded every term by its absolute value, and used the assumption that $|R(t)| \leq \beta t$ for $t \leq x/2$, we would obtain (Exercise) that $|R(x)| \leq \beta x + O\left(\frac{x}{\log x}\right)$. So we need to find an extra saving of $0.007\beta^2 x$. In fact we will find an extra saving of order $\beta^2 x / \log x$ from each value $0 \leq k \leq (\log x)/2 \log 16$, say.

Now for any k, if we knew that $R(t) \leq \beta t/2$ for all $16^k \leq t \leq 4 \cdot 16^k$ then we could gain a term of order $\beta x / \log x$ in our estimate for the contribution from that k. Notice that even though our assumption on $R(t)$ is just a one-sided inequality, we get a saving in absolute value because we have two sums with different signs. Similarly, we could gain a term of order $\beta x/\log x$ if we knew that $R(t) \geq -\beta t/2$ for all $4 \cdot 16^k \leq t < 16^{k+1}$.

So suppose that $R(t_1) > \beta t_1/2$ for some $16^k \leq t_1 \leq 4 \cdot 16^k$ and $R(t_2) < -\beta t_2/2$ for some $4 \cdot 16^k \leq t_2 < 16^{k+1}$. Since $R(t) = \sum_{n \leq t} \Lambda(n) - t$ is right continuous, if we define

$$t_1 \leq \tilde{t_2} := \inf\{t \geq t_1 : R(t) \leq -\beta t/2\} \leq t_2$$

then we must have $R(\tilde{t_2}) \leq -\beta \tilde{t_2}/2$. Moreover, in fact for any $\frac{1-\beta/2}{1+\beta/2}\tilde{t_2} \leq t \leq \tilde{t_2}$ we have

$$R(t) \leq R(\tilde{t_2}) + (\tilde{t_2} - t) \leq (1-\beta/2)\tilde{t_2} - t \leq (1+\beta/2)t - t = \beta t/2,$$

and also $R(t) \geq -\beta t/2$ by definition of $\tilde{t_2}$. Here the interval $[\frac{1-\beta/2}{1+\beta/2}\tilde{t_2}, \tilde{t_2}]$ has length $\gg \beta 16^k$, and we have a saving of at least $\beta t/2$ in our bound for $|R(t)|$ (compared with the "trivial" bound $|R(t)| \leq \beta t + O(1)$), so now once again we can gain a term of order $\beta^2 x/\log x$. (Because the interval we are looking at is now fairly small we actually need to use Selberg's Fundamental Lemma again to estimate the sum of $\Lambda(n)$ and of $\log p \log q$ in the interval.) $\square$

9. Bounded gaps between primes

The twin prime conjecture is an ancient problem, but until recently it seemed fairly unattackable— the best known result was a 1973 theorem of Chen showing that $\gg x/\log^2 x$ of the primes $p \leq x$ are such that $p+2$ has one or two prime factors. Chen's proof applied sieve methods to the sequence of shifted primes $p+2$, and the fact that he could only show $p+2$ has at most two prime factors is a manifestation of the Parity Problem from section 4.

About ten years ago (but finally published more recently), a great advance was made by Goldston, Pintz and Yıldırım.

Theorem 9.1 (Goldston, Pintz and Yıldırım, 2009). *Suppose there is some $\theta > 1/2$ such that the following is true: for any fixed $A > 0$ and all large x,*

$$\sum_{q \leq x^\theta} \max_{(a,q)=1} \left| \pi(x;q,a) - \frac{1}{\phi(q)} \int_2^x \frac{dt}{\log t} \right| \ll_A \frac{x}{\log^A x}.$$

Then there exists a constant $C = C(\theta)$ such that the gap between primes is $\leq C$ infinitely often.

Note that the Bombieri–Vinogradov theorem is just insufficient to satisfy the hypotheses of the theorem (it allows any $\theta < 1/2$), but before the work of Goldston, Pintz and Yıldırım it was believed that even if one could take $\theta \approx 1$ (which is the maximum possible) this wouldn't imply bounded gaps. Goldston, Pintz and Yıldırım also showed unconditionally, using the Bombieri–Vinogradov theorem, that

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

where p_n denote the primes in increasing order. Note that the Prime Number Theorem implies the average size of $p_{n+1}-p_n$ is $\sim \log p_n$.

In the last couple of years, Zhang made a further breakthrough.

Theorem 9.2 (Zhang, 2014)**.** *There exists a constant C such that the gap between primes is $\leq C$ infinitely often.*

Zhang's argument is organised in the same basic way as Goldston, Pintz and Yıldırım's. His crucial new ingredient was a (slightly weaker) Bombieri–Vinogradov type theorem with θ just larger than $1/2$, which he proved with great effort and showed was still sufficient to obtain the bounded gaps consequence.

In this section we will discuss even more recent work of Maynard and Tao (independently), which reproves Zhang's theorem without needing his sophisticated Bombieri–Vinogradov type result. This work is again organised like Goldston, Pintz and Yıldırım's, but with a new and very powerful choice of weights. In fact, the work of Maynard and Tao shows that having the Bombieri–Vinogradov theorem for *any small positive* θ is sufficient to obtain bounded gaps.

Let $\mathcal{H} = \{h_1, ..., h_k\}$ be a set of non-negative integers such that for all primes p, the elements of $\mathcal{H}$ occupy at most $p-1$ residue classes mod p. Thus there is no congruence obstruction mod p to all of the numbers $(n-h)_{h\in\mathcal{H}}$ being prime, for some n. For example, we could take $\mathcal{H} = \{0, 2, 6\}$, although in practice one needs to choose k and the h_i somewhat larger to make the argument work.

The idea of Goldston, Pintz and Yıldırım is to compare the sizes of

$$S_1 := \sum_{N/2<n\leq N} w_n \quad \text{and} \quad S_2 := \sum_{N/2<n\leq N} w_n \sum_{h\in\mathcal{H}} \mathbf{1}_{n-h\in\mathcal{P}},$$

where N is large and the w_n are non-negative weights, and where $\mathcal{P}$ here denotes the set of all primes. If we can choose the weights w_n such that $S_2 > S_1$, then there must exist some $N/2 < n \leq N$ for which

$$\sum_{h\in\mathcal{H}} \mathbf{1}_{n-h\in\mathcal{P}} > 1,$$

in other words there must exist some n for which at least two of the numbers $(n-h)_{h\in\mathcal{H}}$ are prime. Note that if we add more elements to $\mathcal{H}$ then it should be easier to achieve $S_2 > S_1$, but the consequence is weaker because the two prime values in $(n-h)_{h\in\mathcal{H}}$ might be further apart (up to the maximum difference between elements of $\mathcal{H}$).

To give themselves the best chance of showing that $S_2 > S_1$, Goldston, Pintz and Yıldırım chose the weights w_n to be Selberg-type sieve weights for the sequence of products $(\prod_{h \in \mathcal{H}}(n-h))_{N/2<n\leq N}$. This choice was designed to concentrate the w_n on those values of n where $\prod_{h \in \mathcal{H}}(n-h)$ at least doesn't have too many prime factors. We shall be more general and suppose that $w_n = \sum_{d | \prod_{h \in \mathcal{H}}(n-h)} \lambda_d$, where the λ_d are a real sequence supported on those squarefree $d \leq D$. For convenience we also suppose that $D < N/2 - \max_{\mathcal{H}} h$.

Proposition 9.3. *With the above choice of w_n, we have*

$$S_1 = \frac{N}{2} \sum_{d \leq D} \frac{\lambda_d}{d} \prod_{p|d} \omega_{\mathcal{H}}(p) + O(\sum_{d \leq D} |\lambda_d| \prod_{p|d} \omega_{\mathcal{H}}(p)),$$

where $\omega_{\mathcal{H}}(p)$ denotes the number of residue classes mod p occupied by the elements of $\mathcal{H}$.

We also have

$$\begin{aligned} S_2 &= \sum_{h \in \mathcal{H}} \left(\int_{N/2-h}^{N-h} \frac{dt}{\log t} \right) \sum_{d \leq D} \frac{\lambda_d}{\phi(d)} \prod_{p|d} (\omega_{\mathcal{H}}(p) - 1) + \\ &\quad + O(\sum_{h \in \mathcal{H}} \sum_{d \leq D} |\lambda_d| \prod_{p|d} (\omega_{\mathcal{H}}(p) - 1) \max_{(a,d)=1} \left| \pi(N-h; d, a) - \pi(N/2-h; d, a) - \frac{1}{\phi(d)} \int_{N/2-h}^{N-h} \frac{dt}{\log t} \right|). \end{aligned}$$

Proof of Proposition 9.3. By definition we have

$$S_1 = \sum_{N/2<n\leq N} \sum_{d | \prod_{h \in \mathcal{H}}(n-h)} \lambda_d = \sum_{d \leq D} \lambda_d \sum_{\substack{N/2<n\leq N, \\ d | \prod_{h \in \mathcal{H}}(n-h)}} 1.$$

If d is squarefree (which we may assume, since otherwise $\lambda_d = 0$ anyway) then $d | \prod_{h \in \mathcal{H}}(n-h)$ if and only if each prime factor of d divides $\prod_{h \in \mathcal{H}}(n-h)$, which happens if and only if n lies in one of $\omega_{\mathcal{H}}(p)$ residue classes mod p for each prime $p \mid d$. By the Chinese Remainder Theorem, this is equivalent to asking for n to lie in one of $\prod_{p|d} \omega_{\mathcal{H}}(p)$ residue classes mod d. So overall we have

$$S_1 = \sum_{d \leq D} \lambda_d \prod_{p|d} \omega_{\mathcal{H}}(p) \left(\frac{N}{2d} + O(1) \right),$$

which gives the result.

We can rewrite the sum S_2 as

$$\sum_{h \in \mathcal{H}} \sum_{N/2-h<p\leq N-h} w_{p+h} = \sum_{h \in \mathcal{H}} \sum_{d \leq D} \lambda_d \sum_{\substack{N/2-h<p\leq N-h, \\ d | \prod_{g \in \mathcal{H}}(p+h-g)}} 1.$$

Similarly as in the estimation of S_1, we have $d | \prod_{g \in \mathcal{H}}(p+h-g)$ if and only if for each prime divisor q of d, there exists some $g \in \mathcal{H}$ such that $p \equiv g - h \bmod q$. The numbers $(g-h)_{g \in \mathcal{H}}$ occupy $\omega_{\mathcal{H}}(q)$ residue classes mod q, but one of those is the zero class and we cannot have $p \equiv 0 \bmod q$, since $p > N/2 - h$ is prime whilst $q \leq D < N/2 - \max_{\mathcal{H}} h$. This means in total there are $\prod_{q|d}(\omega_{\mathcal{H}}(q) - 1)$ residue classes mod d that p is allowed to occupy, and each of these residue classes is coprime to d.

Finally, we deduce that

$$\begin{aligned} S_2 &= \sum_{h\in\mathcal{H}}\sum_{d\leq D} \lambda_d \prod_{q|d}(\omega_{\mathcal{H}}(q) - 1)\left(\frac{1}{\phi(d)}\int_{N/2-h}^{N-h}\frac{dt}{\log t} + \right. \\ &\quad \left. + O\left(\max_{(a,d)=1}\left|\pi(N-h;d,a) - \pi(N/2-h;d,a) - \frac{1}{\phi(d)}\int_{N/2-h}^{N-h}\frac{dt}{\log t}\right|\right)\right), \end{aligned}$$

which gives the result. □

To control the "big Oh" term in the estimate for S_2 one needs to use something like the Bombieri–Vinogradov theorem. Note that the larger the range of d that one can sum over in that theorem, the larger one can take D, and then one has more flexibility in choosing the λ_d (and therefore a greater chance of achieving $S_2 > S_1$).

Goldston, Pintz and Yıldırım choose their λ_d roughly such that

$$\sum_{d|\prod_{h\in\mathcal{H}}(n-h)} \lambda_d \approx \left(\sum_{\substack{d|\prod_{h\in\mathcal{H}}(n-h), \\ d\leq\sqrt{D}}} \mu(d)\frac{\log^{k+l}(\sqrt{D}/d)}{\log^{k+l}(\sqrt{D})} \right)^2.$$

These look a bit like the weights $\mu(d)\frac{\log(x/d)}{\log x}$, $\mu(d)(1 - \frac{\log^2 d}{\log^2 x})$ that we saw in the proof of Selberg's Fundamental Lemma (Proposition 8.3), and in fact they are a simplified version of the Selberg Λ^2 weights from a certain sieving problem. Since the set $\mathcal{H}$ has k elements, the classical thinking would suggest that one should take $l = 0$ here, so that the exponent of the logarithms is k. But Goldston, Pintz and Yıldırım discovered that taking $l \approx (1/2)\sqrt{k}$ works much better for obtaining $S_2 > S_1$. Making such a choice in Proposition 9.3, one finds after some calculation that one can show $S_2 > S_1$ provided one can choose $D = x^\theta$ for some $\theta > 1/2$, which leads to Theorem 9.1. See Chapter 7.13 of Friedlander and Iwaniec [1] for a detailed exposition of this.

Following the work of Goldston, Pintz and Yıldırım, many unsuccessful attempts were made to improve the choice of the λ_d so as to remove the need for a strong Bombieri–Vinogradov result in their theorem. The key idea in the recent (successful!) work of Maynard and of Tao is to change things one step further back in the argument, by replacing the weights w_n by something more "multidimensional". So rather than taking

$w_n = \sum_{d | \prod_{h \in \mathcal{H}} (n-h)} \lambda_d$, where the λ_d are to be chosen, they take (roughly speaking)

$$w_n = \sum_{d_1 | n-h_1} ... \sum_{d_k | n-h_k} \lambda_{d_1,...,d_k},$$

where the $\lambda_{d_1,...,d_k}$ are to be chosen. Notice that if one were to choose $\lambda_{d_1,...,d_k}$ as a function of the product $d = d_1 d_2 ... d_k$, then one would get back to something like the Goldston–Pintz–Yıldırım situation. But it turns out that choosing

$$\lambda_{d_1,...,d_k} \approx \mu(d_1...d_k) f(d_1, ..., d_k),$$

where f is a smooth function but *not* just a function of the product, works much better. After some calculation one finds that one can show $S_2 > S_1$ provided one can choose $D = x^{\theta}$ for any small $\theta > 0$, which leads unconditionally to bounded gaps between primes. See Maynard's paper [2] for a detailed exposition of this.

References

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

[2] J. Maynard. Small Gaps Between Primes. *Ann. of Math. (2)*, **181**, no. 1, pp 383-413. 2015

[3] H. L. Montgomery, R. C. Vaughan. *Multiplicative Number Theory I: Classical Theory.* First edition, published by Cambridge University Press. 2007

Jesus College, Cambridge, CB5 8BL

E-mail address: A.J.Harper@dpmms.cam.ac.uk