TOPICS IN ANALYTIC NUMBER THEORY

TOM SANDERS

1. Arithmetic functions

An *arithmetic function* is a function $f : \mathbb{N} \to \mathbb{C}$; there are many interesting and natural examples in analytic number theory. To begin with we consider what is perhaps the best known

$$\pi(n) := \sum_{x \leqslant n:} 1_{\mathbb{P}}(x),$$

the usual counting function of the primes. Various heuristic arguments suggest that one should expect x to be prime with probability $1/\log x$, and coupled with a body of numerical evidence this prompted Gauss to conjecture that

$$\pi(n) \sim \mathrm{Li}(n) := \int_1^n \frac{1}{\log x} dx \sim \frac{n}{\log n}.$$

It is the purpose of the first section of the course to prove this result.

As it stands π can be a little difficult to evaluate because the indicator function of the primes is not very smooth. To deal with this we introduce the *von Mangoldt function* Λ defined by

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

It will turn out that Λ is smoother than the indicator function of the primes and while the von Mangoldt function is supported on a larger set (namely all powers of primes), in applications this difference will be negligible. For now we note that our interest lies in the sum of the von Mangoldt function, defined to be

$$\psi(n) := \sum_{x \leqslant n} \Lambda(x),$$

and which is closely related to π as the next proposition shows. The key idea in the proof of the proposition is a technique called partial summation or, sometimes, Abel transformation. This is the observation that

$$\sum_{x \leqslant n} f(x)(g(x) - g(x-1)) = f(n)g(n) - \sum_{x \leqslant n-1} g(x)(f(x+1) - f(x)),$$

and may be thought of as a discrete analogue of integration by parts.

Proposition 1.1. *$\psi(n) \sim n$ if and only if $\pi(n) \sim n/\log n$.*

Last updated: 28th April, 2012.

Proof. First note that power of primes larger than 1 make a negligible contribution to $\psi(n)$:

$$\psi(n)=\sum_{x\leqslant n}\Lambda(x)=\sum_{p\leqslant n}\log p+O(\sqrt{n}\log n).$$

On the other hand by partial summation we have

$$\begin{aligned}\sum_{p\leqslant n}\log p &= \sum_{x\leqslant n}(\pi(x)-\pi(x-1))\log x\\ &= \pi(n)\log n+\sum_{x\leqslant n-1}\pi(x)(\log(x+1)-\log x)\\ &= \pi(n)\log n+\sum_{x\leqslant n-1}\frac{\pi(x)}{x}+O(\log n),\end{aligned}$$

whence

$$\psi(n)=\pi(n)\log n+\sum_{x\leqslant n}\frac{\pi(x)}{x}+O(\sqrt{n}\log n). \tag{1.1}$$

Now, if $\pi(x)\sim x/\log x$ for all $x\leqslant n$, then

$$\psi(n)\sim n+\sum_{x\leqslant n}o(1)\sim n.$$

Conversely, if $\psi(x)\sim x$ for all $x\leqslant n$ then

$$x\gtrsim(\pi(x)-\pi(x^{1/2}))\log x^{1/2}=\frac{1}{2}(\pi(x)-O(x^{1/2}))\log x.$$

It follows that $\pi(x)/x=o(1)$ which can be inserted into (1.1) again to get that

$$n\sim\psi(n)\sim\pi(n)\log n+\sum_{x\leqslant n}o(1)\sim\pi(n)\log n.$$

It follows that $\pi(n)\sim n/\log n$. □

In fact this argument leads to explicit error terms – not just asymptotic results – but this will not concern us since it turns out that the main contribution to the error term $|\pi(n)-n/\log n|$ comes from approximating $\mathrm{Li}(n)$ by $n/\log n$.

There is a natural convolution operation on arithmetic functions. Given two arithmetic functions f and g we define their *convolution* $f*g$ point-wise by

$$f*g(x)=\sum_{zy=x}f(z)g(y).$$

This convolution has an identity δ and it turns out that 1, the function that takes the value 1 everywhere, has an inverse: we define the *Möbius μ-function* by

$$\mu(x):=\begin{cases}(-1)^k & \text{if } x=p_1\dots p_k \text{ for distinct primes } p_1,\dots,p_k\\ 0 & \text{otherwise,}\end{cases}$$

with the usual convention that 1 has a representation as the empty product.

Theorem 1.2 (Möbius inversion)**.** *We have the identity*

$$\mu * 1 = \delta = 1 * \mu.$$

Proof. Suppose that n is a natural number so that by the fundamental theorem of arithmetic there are primes $p_1, \dots, p_l$ and naturals $e_1, \dots, e_l$ such that $n = p_1^{e_1} \dots p_l^{e_l}$. Since μ is only supported on square-free integers, if $\mu(d) \neq 0$ then $d | p_1 \dots p_l$. It follows that

$$1 * \mu(n) = \sum_{d|n} \mu(d) = \sum_{d|p_1 \dots p_l} \mu(d) = \sum_{S \subset [l]} (-1)^{|S|} = (1-1)^l = \delta(n).$$

The results follows by symmetry of convolution. □

Convolution has the effect of smoothing or concentrating in Fourier space which is desirable because it means that the function is easier to estimate. As an example of this we shall estimate the average value of the *divisor function* $\tau(n)$ (sometimes denoted $d(n)$) defined by

$$\tau(n) := \sum_{d|n} 1 = 1 * 1(n),$$

that is the number of divisors of n. Before we begin we recall that the *Euler constant* is

$$\gamma := \int_1^\infty \frac{\{x\}}{x \lfloor x \rfloor} dx,$$

where $\lfloor x \rfloor$ is the integer part of x and $\{x\} := x - \lfloor x \rfloor$. There are many open questions about the Euler constant, although they will not concern us. For our work it is significant only as the constant in the following elementary proposition.

Proposition 1.3. *We have the estimate*

$$\sum_{x \leqslant n} \frac{1}{x} = \log n + \gamma + O(1/n)$$

Proof. Given the definition of γ this is essentially immediate:

$$\begin{aligned} \sum_{x \leqslant n} \frac{1}{x} - \log n &= \sum_{x \leqslant n} \frac{1}{x} - \int_1^{n+1} \frac{1}{x} dx + O(1/n) \\ &= \int_1^{n+1} \left(\frac{1}{\lfloor x \rfloor} - \frac{1}{x} \right) dx + O(1/n) \\ &= \int_1^{n+1} \frac{\{x\}}{x \lfloor x \rfloor} dx + O(1/n) \\ &= \gamma + \int_{n+1}^\infty O(1/x^2) dx + O(1/n) = \gamma + O(1/n). \end{aligned}$$

The result is proved. □

We shall now use Dirichlet's hyperbola method to estimate the average value of the divisor function.

Proposition 1.4. *We have the estimate*

$$\sum_{x\leqslant n}\tau(x) = n\log n + (2\gamma-1)n + O(\sqrt{n}).$$

Proof. An obvious start is to note that

$$\sum_{x\leqslant n}\tau(x) = \sum_{ab\leqslant n}1 = \sum_{a\leqslant n}\lfloor\frac{n}{a}\rfloor = \sum_{a\leqslant n}\Big(\frac{n}{a}+O(1)\Big) = n\log n + O(n)$$

by Proposition 1.3. The weakness of this argument is that the approximation

$$\lfloor\frac{n}{a}\rfloor = \frac{n}{a}+O(1)$$

is not a strong statement when a is close to n – the error term is of comparable size to the main term. However, since $ab\leqslant n$ we certainly have that at least one of a and b is always at most $\sqrt{n}$. It follows from the inclusion-exclusion principle that

$$\sum_{ab\leqslant n}1 = 2\sum_{a\leqslant\sqrt{n}}\lfloor\frac{n}{a}\rfloor - \sum_{a,b\leqslant\sqrt{n}}1.$$

This is called the hyperbola method because it is a way of counting lattice points below the hyperbola $xy=n$. Now, as before we have that

$$\sum_{x\leqslant n}\tau(x) = 2\sum_{a\leqslant\sqrt{n}}\Big(\frac{n}{a}+O(1)\Big) - (\sqrt{n}+O(1))^2.$$

On the other hand by Proposition 1.3 we have that

$$\sum_{a\leqslant\sqrt{n}}\frac{n}{a} = \frac{1}{2}n\log n + \gamma n + O(1),$$

and the result follows on rearranging. □

Recalling Stirling's formula (or directly) we have that

$$\sum_{x\leqslant n}\log x = \log n! = n\log n - n + O(\log n),$$

thus if we put

$$\Delta(n) := \sum_{x\leqslant n}(\tau(x) - \log x - 2\gamma),$$

then we showed above that $\Delta(n) = O(\sqrt{n})$. With additional ideas of a rather different nature Voroni showed that $\Delta(n) = O(n^{1/3}\log n)$ and this has since been improved to $O(n^{\alpha})$ for some $\alpha < 1/3$. In the other directly Hardy and Landau showed that $\Delta(n) = \Omega(n^{1/4})$, but the true order of the error term is not known.

We now return to the von Mangoldt function. By the Fundamental Theorem of Arithmetic if $n\in\mathbb{N}$ then $n = p_1^{e_1}\dots p_l^{e_l}$ for some primes $p_1,\dots,p_l$ and naturals $e_1,\dots,e_l$. Now,

$$1*\Lambda(n) = \sum_{d|n}\Lambda(n) = \sum_{p^k|n}\log p = \sum_{i=1}^{l}e_i\log p_i = \log n.$$

Applying Möbius inversion then gives us an expression for Λ as a convolution:

$$\Lambda = \mu * 1 * \Lambda = \mu * \log .$$

This expression will allow us to relate ψ to the function M defined by

$$M(n) := \sum_{x \leqslant n} \mu(x).$$

The Möbius μ-function takes the values -1 and 1 (and 0) and so we have the trivial upper bound $|M(n)| \leqslant n$. Since μ does not display any obvious additive patterns we might hope that μ takes the values 1 and -1 in a random way. If it did it would follow from the central limit theorem that we would have square-root cancellation:

$$M(n) = O_\varepsilon(n^{1/2+\varepsilon}) \text{ for all } \varepsilon > 0.$$

The Riemann hypothesis is the conjecture that this is the case and it is far from being know. On the other hand we shall be able to show that there is some cancellation so that $M(n) = o(n)$ and this already implies the prime number theorem.

Proposition 1.5. *If $M(n) = o(n)$ then $\psi(n) \sim n$.*

Proof. We use the hyperbola method again: let B be a parameter to be optimised later, and then note that

$$\begin{aligned}
\psi(n) - n + 2\gamma &= \sum_{x \leqslant n} (\Lambda(x) - 1(x) + 2\gamma\delta(x)) \\
&= \sum_{ab \leqslant n} \mu(a)(\log b - \tau(b) + 2\gamma) \\
&= \sum_{b \leqslant B} M(n/b)(\log b - \tau(b) + 2\gamma) \\
&\quad + \sum_{a \leqslant n/B} \mu(a) \sum_{B < b \leqslant n/a} (\log b - \tau(b) + 2\gamma).
\end{aligned}$$

By hypothesis we have that

$$\sum_{b \leqslant B} M(n/b)(\log b - \tau(b) + 2\gamma) = o(n).O(B.(\log B + \tau(B) + 2\gamma)) = o(B^2 n)$$

which covers the first term. For the second we note that

$$\begin{aligned}
|\sum_{a \leqslant n/B} \mu(a) \sum_{B < b \leqslant n/a} (\log b - \tau(b) + 2\gamma)| &\leqslant \sum_{a \leqslant n/B} |\Delta(n/a) - \Delta(B)| \\
&= \sum_{a \leqslant n/B} O(\sqrt{n/a} + \sqrt{B})
\end{aligned}$$

by the bound for Δ given by Proposition 1.4. On the other hand by integral comparison we have

$$\sum_{a \leqslant n/B} 1/\sqrt{a} \leqslant \int_0^{n/B} \frac{1}{\sqrt{a}} da = O(\sqrt{n/B}),$$

whence

$$\sum_{a\leqslant n/B} O(\sqrt{n/a}+\sqrt{B}) = O(n/\sqrt{B}).$$

Combining what we have shown we see that

$$\psi(n) - n + 2\gamma = o(B^2 n) + O(n/\sqrt{B}).$$

It remains to choose B tending to infinity sufficiently slowly which gives the result. □

2. The Fourier transform

In this section we shall develop some of the basic ideas of Fourier analysis. There are many references for this material, with the classic being Rudin's book 'Fourier analysis on groups'.

The Fourier transform on $\mathbb{R}$, $\mathbb{Z}$, $\mathbb{T}$ and finite groups is well known and in the 20th Century some efforts were made to unify the theories and it was discovered that the natural setting was that of locally compact abelian groups; to begin with let G be such. Those unfamiliar with the theory of topological groups may just as well think of G as just being one of the aforementioned examples.

Fourier analysis is important because of its relationship with convolution: suppose that μ and ν are (complex Borel) measures on G. Then the convolution $\mu * \nu$ is the measure defined by

$$\mu * \nu(A) = \int 1_A(x+y) d\mu(x) d\nu(y).$$

This form of convolution is a generalisation of convolution of arithmetic functions as we shall see later. We norm the space of complex Borel measures in the usual way:

$$\|\mu\| := \int d|\mu| := \sup\{\int f d\mu : f \in C_0(G)\},$$

and write $M(G)$ for the space of complex Borel measures on G with finite norm. Convolution on this space interacts well with the norm in that we have Young's inequality:

$$\|\mu * \nu\| \leqslant \|\mu\|\|\nu\| \text{ for all } \mu, \nu \in M(G).$$

We shall write $\widehat{G}$ for the *dual group* of G, that is the locally compact abelian group of continuous homomorphisms $\gamma : G \to S^1$, where $S^1 := \{z \in \mathbb{C} : |z| = 1\}$, and are now in a position to record the Fourier transform. Given $\mu \in M(G)$ we define its Fourier transform in $\ell^\infty(\widehat{G})$ by

$$\widehat{\mu}(\gamma) := \int \overline{\gamma} d\mu \text{ for all } \gamma \in \widehat{G}.$$

It is easy to check that

$$\widehat{\mu * \nu}(\gamma) = \widehat{\mu}(\gamma)\widehat{\nu}(\gamma).$$

There is a special measure on G called Haar measure which is the unique (up to scaling) translation invariant measure on G which we shall denote μ_G. If $f \in L^1(\mu_G)$ we may thus define the Fourier transform of f by

$$\widehat{f}(\gamma) := \int \overline{\gamma} f d\mu_G.$$

The advantage of this is that we have an inverse theorem.

Theorem 2.1 (The Fourier inversion theorem)**.** *Suppose that G is a locally compact abelian group, $f \in L^1(\mu_G)$ and $\widehat{f} \in L^1(\mu_{\widehat{G}})$. Then*

$$f(x) = \int \widehat{f}(\gamma)\gamma(x) d\mu_{\widehat{G}}(\gamma) \text{ for a.e. } x \in G.$$

We have been deliberately vague about scaling the Haar measure on G; in all applications the scaling will be clear.

It may be instructive to consider an example. Suppose that $G = \mathbb{Z}/p\mathbb{Z}$ for some prime p. Then the characters on G are just the maps

$$x \mapsto \exp(2\pi i r x/p),$$

and so $\widehat{G}$ is isomorphic to $\mathbb{Z}/p\mathbb{Z}$ – the natural measure on each of the groups is different however. We shall think of G as endowed with normalised counting measure and $\widehat{G}$ as endowed with counting measure so that

$$\mu_G(A) := |A|/|G| \text{ and } \mu_{\widehat{G}}(\Gamma) := |\Gamma|.$$

Given a function $f : G \to \mathbb{C}$, we may think of it as a sum of canonical basis vectors $(\delta_x)_{x \in G}$ where

$$\delta_x(y) := \begin{cases} |G| & \text{if } y = x; \\ 0 & \text{otherwise.} \end{cases}$$

We then have the decomposition

$$f(x) = \int f(y)\delta_y(x) d\mu_G(y);$$

the Fourier transform changes this basis to the set of characters. Indeed

$$\widehat{f}(\gamma) = \int f \overline{\gamma} d\mu_G = \langle f, \gamma \rangle$$

and the inversion formula just says that

$$f = \sum_{\gamma \in \widehat{G}} \langle f, \gamma \rangle \gamma = \sum_{\gamma \in \widehat{G}} \widehat{f}(\gamma)\gamma = \int \widehat{f}(\gamma)\gamma d\mu_{\widehat{G}}(\gamma).$$

Now, given a function f it also induces a convolution operator

$$M_f : L^2(\mu_G) \to L^2(\mu_G); g \mapsto f * g$$

and the Fourier transform serves to diagonalize this operator. Indeed if we decompose g in the Fourier basis:

$$g = \sum_{\gamma\in\widehat{G}} \langle g, \gamma\rangle\gamma,$$

then it is easy to check that

$$M_f g = \sum_{\gamma\in\widehat{G}} \widehat{f}(\gamma)\langle g, \gamma\rangle\gamma.$$

Note, of course, that $\widehat{f}(\gamma) = \langle f, \gamma\rangle$ too but we use the above notation to make things more suggestive.

Many problems in mathematics involve convolution: in probability theory the law of the sum of two independent random variables is the convolution of their laws. For enumerative problems $1_A * 1_B(x)$ is the (scaled) number of ways of writing x as a sum $a+b$ where $a \in A$ and $b \in B$. Indeed, before the end of the course we shall find ourselves in a position where we want to examine

$$1_A * 1_A * 1_A(x)$$

for A the set of primes less than or equal to N and so we shall be pleased to have access to a basis which diagonalizes this expression.

Returning now to the more immediate concern of arithmetic functions we suppose that G is the reals under addition and f is an arithmetic function. Then we write m_f for the atomic measure defined by

$$m_f(A) := \sum_{n\in\mathbb{N}} f(n)1_A(\log n).$$

Shortly we shall damp our measures so that they are better behaved, but before that it is easy to see that if f and g are arithmetic functions then

$$m_f * m_g = m_{f*g}.$$

An important additional class of measures is the class of damped versions of the above. Indeed, suppose that $\sigma > 1$, and write $m_{f,\sigma}$ for the exponentially damped measure defined by

$$m_{f,\sigma}(A) := \sum_{n\in\mathbb{N}} f(n)1_A(\log n)\exp(-\sigma\log n).$$

Usefully (and not entirely coincidentally) we have the identity

$$m_{f,\sigma} * m_{g,\sigma} = m_{f*g,\sigma}$$

for all arithmetic functions f and g.

If $|f(n)| = \log^{O(1)} n$ then we see that $m_{f,\sigma}$ is finite so that, for example, $m_{1,\sigma}, m_{\mu,\sigma}, m_{\Lambda,\sigma} \in M(\mathbb{R})$; in particular we may take the Fourier transform and, in fact,

$$\widehat{m_{1,\sigma}}(t) = \zeta(\sigma + it) \text{ for all } t \in \mathbb{R},$$

is the classical ζ-function.

3. The prime number theorem

We now turn to proving the prime number theorem. We should like to examine the inner product

$$\sum_{x\leqslant n}\mu(x)=\int I_\sigma dm_{\mu,\sigma}$$

where

$$I_\sigma(x)=\begin{cases}\exp(\sigma x) & \text{if } x\leqslant \log n\\ 0 & \text{otherwise.}\end{cases}$$

Unfortunately the function I_σ is not smooth enough and we are not able to control the Fourier transform of $m_{\mu,\sigma}$ well enough.

We shall estimate the Fourier transform of $m_{\mu,\sigma}$ through $m_{1,\sigma}$ via the usual convolution identity:

$$\widehat{m_{\mu,\sigma}}(t)=1/\widehat{m_{1,\sigma}}(t).$$

First we note a simple estimate – we shall make further use of the basic method in Lemma 3.5 so it is worth becoming familiar now.

Lemma 3.1. *We have the estimate*

$$\widehat{m_{1,\sigma}}(t)=\frac{1}{\sigma-1+it}+O(\log(1+|t|)).$$

Proof. Naturally we proceed by integral comparison: let $T>1$ be a parameter to be optimised later.

$$\begin{aligned}\widehat{m_{1,\sigma}}(t) &= \int_1^\infty x^{-\sigma-it}dx+\int_1^\infty(\lfloor x\rfloor^{-\sigma-it}-x^{-\sigma-it})dx\\ &= \frac{1}{\sigma-1+it}+\int_1^T O(x^{-\sigma})dx+\int_T^\infty(\lfloor x\rfloor^{-\sigma-it}-x^{-\sigma-it})dx\\ &= \frac{1}{\sigma-1+it}+O(\log T)+\int_T^\infty(\lfloor x\rfloor^{-\sigma-it}-x^{-\sigma-it})dx.\end{aligned}$$

To estimate the second integral we just note that

$$\begin{aligned}\int_T^\infty(\lfloor x\rfloor^{-\sigma-it}-x^{-\sigma-it})dx &= \sum_{n=T}^\infty\int_0^1\frac{((1+\theta/n)^{\sigma+it}-1)}{(n+\theta)^{\sigma+it}}d\theta\\ &= \sum_{n=T}^\infty O(|t|/n^{\sigma+1})=O(|t|/T).\end{aligned}$$

The result follows on setting $T=1+|t|$. □

We encode the Fundamental Theorem of Arithmetic in an analytic way as follows.

Lemma 3.2 (The Euler Product formula)**.** *For $\sigma>1$, $t\in\mathbb{R}$ we have the equivalence*

$$\widehat{m_{1,\sigma}}(t)=\prod_p(1-p^{-\sigma-it})^{-1}.$$

Proof. Write

$$P_N := \prod_{p\leqslant N}(1-p^{-\sigma-it})^{-1} = \prod_{p\leqslant N}(1+p^{-1(\sigma+it)}+p^{-2(\sigma+it)}+\dots).$$

By the Fundamental Theorem of Arithmetic if $n\leqslant N$ then there are powers $e_1,\dots,e_r$ and primes $p_1,\dots,p_r\leqslant N$ such that $n=p_1^{e_1}\dots p_r^{e_r}$ whence (multiplying out P_N) we have

$$|P_N-\sum_{n=1}^{N}n^{-\sigma-it}|\leqslant\sum_{n=N+1}^{\infty}n^{-\sigma}.$$

On the other hand

$$|\widehat{m_{1,\sigma}}(t)-\sum_{n=1}^{N}n^{-\sigma-it}|\leqslant\sum_{n=N+1}^{\infty}n^{-\sigma},$$

so

$$|P_N-\widehat{m_{1,\sigma}}(t)|\leqslant 2\sum_{n=N+1}^{\infty}n^{-\sigma}=O((\sigma-1)^{-1}N^{1-\sigma})$$

by the triangle inequality and integral comparison. The lemma is proved on letting $N\to\infty$. □

We now record the number theoretic content of this section.

Lemma 3.3. *There is an absolute constant $c>0$ such that if $\sigma\in(1,1+c)$ then we have the estimates*

$$\|m_{\mu,\sigma}\|=(\sigma-1)^{-1}+O(1)$$

and

$$|\widehat{m_{\mu,\sigma}}(t)|=O((\sigma-1)^{-3/4}\log^{1/2}(2+|t|)).$$

Proof. First we note that

$$\|m_{\mu,\sigma}\|=\|m_{1,\sigma}\|=(\sigma-1)^{-1}+O(1)$$

by integral comparison. The second is where we have the main idea: begin by noting that

$$\begin{aligned}1&\leqslant(1+(1+p^{-it}+p^{it})^2p^{-\sigma})\\&=1+3p^{-\sigma}+2p^{-it-\sigma}+2p^{it-\sigma}+p^{-2it-\sigma}+p^{2it-\sigma}\\&=(1+O(p^{-2\sigma}))(1-p^{-\sigma})^{-3}|1-p^{-it-\sigma}|^{-4}|1-p^{-2it-\sigma}|^{-2}.\end{aligned}$$

It should be remarked that this is, perhaps, the most mysterious part of the proof of the prime number theorem and has defied good explanation.

Thus

$$\begin{aligned}1&=O(\prod_p(1-p^{-\sigma})^{-3}|1-p^{-it-\sigma}|^{-4}|1-p^{-2it-\sigma}|^{-2})\\&=O(\|m_{1,\sigma}\|^3|m_{1,\sigma}(t)|^4|m_{1,\sigma}(2t)|^2),\end{aligned}$$

since all the products are absolutely convergent and

$$\prod_p (1+p^{-2\sigma}) = O(1).$$

the result now follows from Lemma 3.1 for $|t| > 1/100$ say. On the other hand, by Lemma 3.1 if $|t| < 1/100$ then

$$\widehat{m_{\mu,\sigma}}(t) = \widehat{m_{1,\sigma}}(t)^{-1} = \left(\frac{1}{\sigma - 1 + O(1)} + O(1)\right)^{-1} = O(1),$$

provided $\sigma - 1$ is sufficiently small. The result is proved. □

The above lemma shows us that we have cancellation in $\widehat{m_{\mu,\sigma}}$ compared with its possible maximum – there is no real t such that $\mu(n) \approx \exp(it \log n)$. To see this, suppose that there was such. Then by the first part of the lemma

$$\begin{aligned}\widehat{m_{\mu,\sigma}}(t) &= \sum_{n=1}^{\infty} \frac{1}{n^\sigma}.\mu(n)\overline{\exp(it\log n)} \\ &\approx \sum_{n=1}^{\infty} \frac{1}{n^\sigma} = (\sigma-1)^{-1} + O(1).\end{aligned}$$

However, by the second part of the lemma we know that $|\widehat{m_{\mu,\sigma}}(t)|$ is much smaller than this maximum if $\sigma - 1$ is small (and $|t|$ is not too large).

On its own the above cancellation isn't enough for what we want, but it can be bootstrapped by differentiating.

Lemma 3.4. *We have the estimate*

$$\frac{d^k \widehat{m_{\mu,\sigma}}}{dt^k}(t) = (\sigma-1)^{-3(k+1)/4} O(k\log(2+|t|))^{O(k)}.$$

Before embarking on the proof proper it will be useful (in light of Lemma 3.1) to define an auxiliary function:

$$f(t) := (\sigma - 1 + it)\widehat{m_{1,\sigma}}(t).$$

The following calculation is straightforward.

Lemma 3.5. *We have the estimate*

$$f^{(r)}(t) = O(\log(2+|t|))^{r+1}(1+|t|).$$

Proof. We proceed as before:

$$\begin{aligned}f(t) &= (\sigma-1+it)\int_1^\infty x^{-\sigma-it}dx \\ &\quad +(\sigma-1+it)\int_1^\infty (\lfloor x\rfloor^{-\sigma-it} - x^{-\sigma-it})dx \\ &= 1 + (\sigma-1+it)\int_1^\infty (\lfloor x\rfloor^{-\sigma-it} - x^{-\sigma-it})dx.\end{aligned}$$

Differentiating this we see that the first term is just $O(1)$; to estimate the second we introduce an additional auxiliary function

$$g(t) := \int_1^\infty (\lfloor x\rfloor^{-\sigma-it} - x^{-\sigma-it})dx.$$

By the product rule and linearity of differentiation

$$f^{(r)}(t) = O(1) + O(rg^{(r-1)}(t)) + O(|t|g^{(r)}(t)). \tag{3.1}$$

Now, suppose that $q \geqslant 1$ is a natural. Since $\sigma > 1$ we have that

$$g^{(q)}(t) = \int_1^\infty (-i\log\lfloor x\rfloor)^q \lfloor x\rfloor^{-\sigma-it} - (-i\log x)^q x^{-\sigma-it}dx.$$

As before we split the integral in two according to whether or not x is larger or smaller than some parameter $T > 1$. In the first instance

$$\begin{aligned}\int_1^T (-i\log\lfloor x\rfloor)^q \lfloor x\rfloor^{-\sigma-it} - (-i\log x)^q x^{-\sigma-it}dx &= O(\int_1^T (\log x)^q x^{-\sigma}dx)\\ &= O(\log^{q+1} T).\end{aligned}$$

In the second instance

$$\int_T^\infty (-i\log\lfloor x\rfloor)^q \lfloor x\rfloor^{-\sigma-it} - (-i\log x)^q x^{-\sigma-it}dx$$

is equal to

$$\sum_{n=T}^\infty \int_0^1 (-i\log n)^q n^{-\sigma-it} - (-i\log(n+\theta))^q (n+\theta)^{-\sigma-it}d\theta.$$

Each summand in this expression is then

$$O\left(\frac{\log^q n}{n^\sigma}\right).\int_0^1 ((1+\theta/n)^{\sigma+it} - (1+\theta\exp(O(q))/n\log n))d\theta,$$

which is $O(\log n)^q|t|/n^{\sigma+1}$. Setting $T = 2+|t|$ and combining this we conclude that

$$g^{(q)}(t) = O(\log(2+|t|))^{q+1},$$

and the lemma follows on inserting this into (3.1) with $q = r$ and $q = r-1$. □

Corollary 3.6. *We have the estimate*

$$\frac{f^{(r)}(t)}{f(t)} = (\sigma-1)^{-3/4}O(\log(2+|t|))^{r+3/2}.$$

Proof. This is immediate from Lemma 3.3 and Lemma 3.5. □

This corollary will be used in conjunction with the following easy fact.

Lemma 3.7. *Suppose that $f : \mathbb{R} \to \mathbb{C}$ is a k-fold differentiable function. Then*

$$(1/f)^{(k)} = \frac{1}{f}\sum_{a_1+2a_2+\cdots+ka_k=k} \frac{(a_1+\cdots+a_k)!}{a_1!\dots a_k!}\left(-\frac{f^{(1)}}{1!f}\right)^{a_1}\cdots\left(-\frac{f^{(i)}}{i!f}\right)^{a_i}\cdots.$$

Proof. The proof is left as an exercise – it is an easy induction. □

Proof of Lemma 3.4. Now we can leverage our knowledge of the auxiliary function f to provide the estimate that we are looking for. First note that

$$\frac{d^k\widehat{m_{\mu,\sigma}}}{dt^k}(t) = (\sigma - 1 + it)(1/f)^{(k)}(t) + ki(1/f)^{(k-1)}(t).$$

On the other hand by the previous corollary and lemma we get immediately that

$$\frac{d^k\widehat{m_{\mu,\sigma}}}{dt^k}(t) = O(k!^2(\sigma-1)^{-3(k+1)/4}O(\log(2+|t|))^{O(k)})$$

as claimed. □

The last ingredient we require is a smoothed version of I_σ.

Lemma 3.8 (Smoothed Perron inversion)**.** *For $x \in \mathbb{R}$ and $\sigma > 1$ we have*

$$\frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{\exp(x(\sigma+it))}{(\sigma+it)^2}dt = \begin{cases} x & \text{if } x \geqslant 0 \\ 0 & \text{otherwise.} \end{cases}$$

Proof. This is a simple contour integral and we shall split into two cases. To begin with we suppose that $x > 0$ and let $\mathcal{C}$ be the rectangle with sides $S_1 := [\sigma - iT, \sigma + iT]$, $S_2 := [-S - iT, -S + iT]$, $S_3 = [-S - iT, \sigma - iT]$ and $S_4 = [-S + iT, \sigma + iT]$, so that S_1 and S_2 are parallel and S_3 and S_4 are parallel. The integral in which we are interested is

$$\lim_{T\to\infty}\int_{S_1}\exp(xs)s^{-2}ds.$$

First we note that $f(s) := \exp(xs)s^{-2}$ is holomorphic in and on $\mathcal{C}$ except at $s = 0$. The Laurant expansion around that point is given by

$$f(x) = s^{-2} + xs^{-1} + x^2/2! + x^3s/3! + \ldots,$$

whence the residue is x, and it follows from Cauchy's integral theorem that

$$\int_{\mathcal{C}} f(s)ds = 2\pi i x.$$

It remains to show that the integrals along S_2, S_3 and S_4 are negligible. By the ML-Lemma we see that

$$\begin{aligned} |\int_{S_2} f(s)ds| &\leqslant \sup_{s\in S_2} |\exp(xs)||s|^{-2}.|S_2| \\ &\leqslant \sup_{t\in[-T,T]} |\exp(x(-S+it))||S+it|^{-2}.2T. \end{aligned}$$

Since $x > 0$ we see that $|\exp(x(-S+it))| = |\exp(-xS)| \leqslant 1$. Additionally $|S+it|^{-2} \leqslant |S|^{-2}$, whence

$$|\int_{S_2} f(s)ds| \leqslant 2T/S^2.$$

Turning to the contributions along S_3 and S_4 we have, by similar arguments, that

$$|\int_{S_3} f(s)ds| \leqslant \exp(\sigma x)(\sigma+S)/T^2 \text{ and } |\int_{S_4} f(s)ds| \leqslant \exp(\sigma x)(\sigma+S)/T^2.$$

Thus

$$S_1 = 2\pi i x + O(T/S^2) + O_{\sigma,x}(S/T^2).$$

Setting $T = S$ and letting it tend to infinity gives us that

$$\int_{-\infty}^{\infty} \frac{\exp(x(\sigma+it))}{(\sigma+it)^2}dt = 2\pi i x$$

when $x > 0$.

The case $x \leqslant 0$ is covered similarly except that this time $\mathcal{C}$ is taken to be the rectangle with sides $S_1 := [\sigma - iT, \sigma + iT]$, $S_2 := [S - iT, S + iT]$, $S_3 = [\sigma - iT, S - iT]$ and $S_4 = [\sigma + iT, S + iT]$, and the integral in Cauchy's theorem is 0 because f is holomorphic in and on $\mathcal{C}$. The result is proved. □

Finally we can prove the main result of this section.

Theorem 3.9. *We have the estimate*

$$F(n) := \sum_{x \leqslant n} \mu(x) \log^k x \log^+(n/x) = O_k(n \log^{3(k+1)/4} n)$$

Proof. Start by noting from Lemma 3.8 that

$$\begin{aligned} F(n) &= \sum_x \mu(x) \log^k x \int_{-\infty}^{\infty} n^{\sigma+it} x^{-(\sigma+it)} (\sigma+it)^{-2} dt \\ &= i^{-k} \int_{-\infty}^{\infty} \frac{d^k \widehat{m_{\mu,\sigma}}}{dt^k}(t) \frac{n^{\sigma+it}}{(\sigma+it)^2} dt. \end{aligned}$$

Inserting the bound from Lemma 3.4 we conclude that

$$F(n) = (\sigma-1)^{-3(k+1)/4}.O_k(n^\sigma),$$

and this can be optimized by choosing $\sigma = 1 + 1/\log N$. The result is proved. □

As a corollary of this we have the following estimate for $M(n)$.

Corollary 3.10. *We have the estimate*

$$M(n) = o(n).$$

Proof. First we define an auxiliary function

$$H(n) := \sum_{x \leqslant n} \mu(x) \log^k x$$

to estimate. Let $m \leqslant n/2$ be a parameter to be optimized later and note that

$$\begin{aligned} F(n+m) - F(n) &= \sum_{x=n+1}^{n+m} \mu(x) \log^k x \log \frac{n+m}{x} + \log \frac{n+m}{n} H(n) \\ &= O(m \log^k n \log \frac{n+m}{n}) + \log \frac{n+m}{n} H(n). \end{aligned}$$

Of course by Theorem 3.9 we have

$$F(n+m) - F(n) = O_k(n \log^{3(k+1)/4} n).$$

Since $\log \frac{n+m}{n} = \Omega(m/n)$ (as $m \leqslant n/2$) it follows that

$$H(n) = O_k(m \log^k n + \frac{n^2}{m} \log^{3(k+1)/4} n)).$$

By judicious choice of m this means that

$$H(n) = O_k(n \log^{7(k+1)/8} n).$$

Crucially, if $k > 7$ then the above represents genuine cancellation in $H(n)$. It is now a short exercise in partial summation to get an estimate for M:

$$\begin{aligned} H(n) &= \sum_{x \leqslant n} \mu(x) \log^k x \\ &= \sum_{x \leqslant n} (M(x) - M(x-1)) \log^k x \\ &= M(n) \log^k n + \sum_{x \leqslant n-1} M(x)(\log^k x - \log^k (x+1)) \\ &= M(n) \log^k n + \sum_{x \leqslant n-1} O(x).O(kx^{-1} \log^{k-1} x) \\ &= M(n) \log^k n + O_k(n \log^{k-1} n). \end{aligned}$$

It follows for $k = 8$ that we have

$$M(n) = O(n/\log^{1/8} n)$$

and the result is proved. □

We should remark that with a little more care the error term can be made quite explicit and, in particular, larger than any power of log, that is

$$M(n) = O_A(n/\log^A n) \text{ for all } A > 0.$$

Combining this last result with Propositions 1.1 and 1.5 we get the Prime Number Theorem.

Theorem 3.11 (The Prime Number Theorem)**.** $\pi(n) \sim n/\log n$.

4. Dirichlet's theorem; primes in arithmetic progressions

In this section we shall introduce so called L-functions. In the end we shall use these to prove a version of the prime number theorem in arithmetic progressions, but to motivate their introduction we shall begin by proving Dirichlet's theorem on primes in arithmetic progressions.

It had been conjectured for some time before Dirichlet proved his result that if $(a,q)=1$ then there are infinitely many primes p with $p\equiv a \pmod q$; the hypothesis $(a,q)=1$ is clearly necessary since (a,q) divides all integers of the form $a \pmod q$. We make the assumption $(a,q)=1$ for the remainder of this section.

Dirichlet's starting point was Euler's proof of this infinitude of primes which we now record. First we recall the Euler product formula of Lemma 3.2: for $s=\sigma+it$ with $\sigma>1$ we have

$$\zeta(s):=\widehat{m_{1,\sigma}}(t)=\prod_p(1-p^{-\sigma-it})^{-1}.$$

If the number of primes were finite then

$$\prod_p(1-p^{-\sigma})^{-1}$$

would be bounded above by an absolute constant. However this product is equal to $\widehat{m_{1,\sigma}}(0)$, and

$$\zeta(\sigma)=\widehat{m_{1,\sigma}}(0)=\frac{1}{\sigma-1}+O(1)$$

by Lemma 3.1. Letting $\sigma\to 1$ leads to a contradiction which gives the proof.

To pick out the primes of the form $a \pmod q$ we shall use the Fourier transform on $\mathbb{Z}/q\mathbb{Z}^*$. To assist with understanding we shall describe the characters on finite abelian groups. By the (weak) structure theorem for finite abelian groups, any such group G may be decomposed into a product of cyclic groups

$$G\cong\prod_{i=1}^k G_i,$$

where $G_i=\mathbb{Z}/q_i\mathbb{Z}$ for some $q_i\in\mathbb{Z}$. As usual we write 0_G for the identity of a group G, but the reader should be warned that, for example, $0_{S^1}=1$.

Now, if $\gamma:G\to S^1$ is a homomorphism then,

$$\gamma_i:G_i\to S^1; x\mapsto\gamma(0_{G_1},\dots,0_{G_{i-1}},x,0_{G_{i+1}},\dots)$$

is also a homomorphism. Since G_i is cyclic there is a q_ith root of unity ω_i such that

$$\gamma_i(x)=\omega_i^x,$$

which the reader may care to check is well-defined. Since γ is a homomorphism it follows that

$$\gamma(x_1,\dots,x_k)=\prod_{i=1}^k\omega_i^{x_i}.$$

Moreover, any sequence $\omega_1, \dots, \omega_k$ with ω_i a q_ith root of unity defines a unique homomorphism as above so that, in particular, $|G| = |\widehat{G}|$.

Finally we assign counting measure to G so that

$$\widehat{f}(\gamma) = \sum_{x \in G} f(x)\overline{\gamma(x)},$$

and therefore endow $\widehat{G}$ with normalized counting measure so that the inversion formula becomes

$$f(x) = \frac{1}{|\widehat{G}|} \sum_{\gamma \in \widehat{G}} \widehat{f}(\gamma)\gamma(x)$$

for all $x \in G$.

Now, returning to our group $G = \mathbb{Z}/q\mathbb{Z}^*$, given a character $\chi \in \widehat{G}$ we extend it to $\mathbb{Z}$ in the obvious way:

$$\chi(x) := \begin{cases} \chi(x \pmod q) & \text{if } (x,q) = 1 \\ 0 & \text{otherwise;} \end{cases}$$

such a function is sometimes called a *Dirichlet character*. Crucially they inherit their orthogonality properties from characters on $\mathbb{Z}/q\mathbb{Z}^*$. In particular if χ and χ' are (Dirichlet) characters then

$$\frac{1}{\phi(q)} \sum_{x=1}^{q} \chi(x)\overline{\chi'(x)} = \langle \chi, \chi' \rangle_{L^2(\mathbb{Z}/q\mathbb{Z}^*)} = \begin{cases} 1 & \text{if } \chi = \chi' \\ 0 & \text{otherwise.} \end{cases}$$

By the inversion formula, $A := \{x \in \mathbb{Z} : x \equiv a \pmod q\}$, we have the important relation

$$1_A(x) = \frac{1}{\phi(q)} \sum_{\chi \in \widehat{\mathbb{Z}/q\mathbb{Z}^*}} \chi(x)\overline{\chi(a)}$$

since $|\widehat{\mathbb{Z}/q\mathbb{Z}^*}| = |\mathbb{Z}/q\mathbb{Z}| = \phi(q)$.

In anticipation of the above we consider $\widehat{m_{\chi,\sigma}}(t)$ – an L-function. For $s = \sigma + it$ and $\sigma > 1$ we shall write

$$L(s,\chi) := \widehat{m_{\chi,\sigma}}(t) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}.$$

For $\sigma > 1$ is follows that $L(s,\chi)$ is a uniform limit of analytic functions and hence analytic itself. Moreover we have an Euler-product formula as in Lemma 3.2:

Lemma 4.1. *For $\sigma > 1$ we have the equivalence*

$$L(s,\chi) = \prod_p (1 - \chi(p)p^{-s})^{-1}.$$

The proof is left as an exercise following from the fact that χ is induced by a homomorphism. Now, it is easy enough to check from the above that

$$\log L(s,\chi) = \sum_p \sum_{m=1}^{\infty} \frac{\chi(p^m)}{mp^{ms}}.$$

Note that we fixed a branch of log here which we can do since $L(s,\chi)$ is non-zero when $\sigma > 1$. This follows from the fact that the product formula for $L(s,\chi)$ converges absolutely in this range and the prodands (the terms in the product) are never zero.

Now, by the inversion formula we then have

$$\frac{1}{\phi(q)} \sum_{\chi} \overline{\chi(a)} \log L(s,\chi) = \sum_p \sum_{m : p^m \equiv a \pmod q} \frac{1}{mp^{ms}}. \tag{4.1}$$

We are going to consider this expression with $s \to 1^+$.

In the first instance we consider the term corresponding to the so called *principal character* χ_0, that is the character induced by the identity character on $\mathbb{Z}/q\mathbb{Z}^*$. It takes 1 at all integers coprime to q and is 0 elsewhere.

Lemma 4.2. *We have the equivalence*

$$L(s,\chi_0) = \zeta(s) \prod_{p|q} (1 - p^{-s}).$$

The proof of this is left as an exercise. Crucially, on combination with Lemma 3.1 we see that

$$\log L(s,\chi_0) \to \infty \text{ as } s \to 1^+.$$

We should like to show that the remaining contributions in (4.1) are bounded; doing this will lead to the Dirichlet's theorem. We shall split into two cases according to whether or not the character χ is complex valued. Before that, however, we have an analytic extension of $L(s,\chi)$.

Lemma 4.3. *For all non-principal characters χ, the function $L(s,\chi)$ can be extended analytically in the range $\sigma > 0$ and satisfies*

$$L(s,\chi) = s \int_1^{\infty} S(x) x^{-(s+1)} dx$$

in that range, where $S(x) = \sum_{n \leqslant x} \chi(n)$.

Proof. As usual we proceed by partial summation: suppose that $\sigma > 1$. Then

$$\begin{aligned} L(s,\chi) &= \sum_{n=1}^{\infty} \frac{\chi(n)}{n^{-s}} \\ &= \sum_{n=1}^{\infty} (S(n) - S(n-1))n^{-s} \\ &= \sum_{n=1}^{\infty} S(n)(n^{-s} - (n+1)^{-s}) \\ &= s\int_1^{\infty} S(x)x^{-(s+1)}dx. \end{aligned}$$

Thus $L(s,\chi)$ satisfies the equality for $\sigma > 1$. However, by orthogonality of characters we see that $S(x) = O(q)$ since χ is non-principal. Thus the right hand side is analytic in the range $\sigma > 0$ and so $L(s,\chi)$ may be continued to this range and the result is proved. □

It will be similarly convenient to have a meromorphic extension for the ζ function.

Lemma 4.4. *The function $\zeta(s)$ can be meromorphically extended to the plane $\sigma > 0$ with a simple pole at $s = 1$ so that*

$$\zeta(s) = \frac{s}{s-1} - s\int_1^{\infty} \{x\}x^{-s-1}dx.$$

The proof is left as an exercise which can be done by the method in Lemma 3.1.

It follows from Lemma 4.3 that $L(1,\chi) = O(1)$ for all non-principal χ. As it happens it turns out to be harder to show that it is non-zero. If χ is complex valued it is a little easier because it has a companion character $\overline{\chi}$.

Lemma 4.5. *For all complex valued characters χ we have*

$$L(1,\chi) \neq 0.$$

Proof. Non-negativity of the right hand side of (4.1) when $s = \sigma > 1$ and $a = 1$ gives

$$\sum_{\chi'} \log L(\sigma,\chi') \geqslant 0.$$

It follows that

$$\prod_{\chi'} |L(\sigma,\chi')| \geqslant 1,$$

whenever $\sigma > 1$. On the other hand for all non-principal χ' we have $L(1,\chi') = O(1)$. Furthermore, by Lemma 4.4 we have $\lim_{s\to 1^+} L(s,\chi_0)(s-1) = 1$. However, if $L(1,\chi) = 0$ then so does $L(1,\overline{\chi})$ whence

$$\lim_{s\to 1^+} L(s,\chi)L(s,\overline{\chi})(s-1)^{-2} = O(1).$$

This combines to contradict the lower bound proving the lemma. □

We now turn our attention to the harder job of dealing with real characters. The rather mysterious proof below is due to de la Vallée Poussin.

Proposition 4.6. *For all real valued characters χ we have*

$$L(1,\chi)\neq 0.$$

Proof. We assume that $L(1,\chi)=0$ so that $L(s,\chi)$ has a zero at $s=1$ and then we see that $L(s,\chi)L(s,\chi_0)$ is analytic in the region $\sigma>0$. Moreover, since $L(2s,\chi_0)\neq 0$ if $\sigma>1/2$ we have that

$$\psi(s):=\frac{L(s,\chi)L(s,\chi_0)}{L(2s,\chi_0)}$$

is analytic in that region. Additionally $L(2s,\chi_0)\to\infty$ as $s\to 1/2^+$ so

$$\lim_{s\to 1/2^+}\psi(s)=0. \tag{4.2}$$

Now, if $\sigma>1$ then we see that

$$\begin{aligned}\psi(s) &= \prod_{p\nmid q}\frac{(1-\chi(p)p^{-s})^{-1}(1-p^{-s})^{-1}}{(1-p^{-2s})^{-1}}\\ &= \prod_{p\nmid q}\frac{1+p^{-s}}{1-\chi(p)p^{-s}}\\ &= \prod_{p:\chi(p)=1}\frac{1+p^{-s}}{1-p^{-s}}.\end{aligned}$$

Since we know the product is absolutely convergent we can expand it out and we see that we get

$$\psi(s)=\sum_{n=1}^{\infty}a_n n^{-s}$$

for coefficients $(a_n)_n$ with $a_n\geqslant 0$ and $a_1=1$. It particular we have

$$\psi^{(m)}(2)=(-1)^m\sum_{n=1}^{\infty}a_n(\log n)^m n^{-2}=:(-1)^m b_m$$

for some coefficients $(b_m)_m$ with $b_m\geqslant 0$.

On the other hand we know that $\psi(s)$ is regular for $\sigma>1/2$ whence it has an expansion around 2 of radius at least $3/2$:

$$\psi(s)=\sum_{m=0}^{\infty}\frac{1}{m!}\psi^(m)(2)(s-2)^m=\sum_{m=0}^{\infty}\frac{b_m}{m!}(2-s)^m$$

whenever $|s-2|<3/2$. Thus for $1/2<\sigma<2$ we have

$$\psi(\sigma)\geqslant b_0\geqslant a_1\geqslant 1.$$

This contradicts (4.2), and the proof is complete. □

It now remains to prove Dirichlet's theorem as promised.

Theorem 4.7. *Suppose that a and q are coprime naturals. Then there are infinitely many primes p with $p \equiv a \pmod q$.*

Proof. We examine (4.1). The left hand side tends to infinity since $L(1,\chi)$ is finite and non-zero for all non-principle χ and $L(s,\chi_0) \to \infty$ as $s \to 1$. However the right hand side is just

$$\sum_{p \equiv a \pmod q} \frac{1}{p} + O(1).$$

The result follows. □

The above sort of arguments coupled with partial summation techniques yield asymptotics for

$$\sum_{p \equiv a \pmod q, p \leqslant N} \frac{1}{p}.$$

These do not tend to be terribly useful though and we should much prefer to have an analogue of the prime number theorem. We write

$$\psi(x;q,a) := \sum_{n \leqslant x, n \equiv a \pmod q} \Lambda(n)$$

for the weighted counting function of primes in arithmetic progressions.

Theorem 4.8 (Siegel-Walfisz)**.** *Suppose that $A > 0$, x is a natural, a and q are coprime naturals and $q \leqslant \log^A x$. Then we have the estimate*

$$\psi(x;q,a) = \frac{x}{\phi(q)} + O_A(x \log^{-A} x).$$

5. Weyl's inequality

In this section we shall introduce an approach due to Weyl for estimating the Fourier transform of certain sequences. Our starting point is the following easy fact: if $\alpha \in \mathbb{R}$ then $\{n\alpha\}$ can be made arbitrarily close to 0. In particular, for all $Q \in \mathbb{N}$ there is some $q \leqslant Q$ such that

$$\min\{|\alpha q - z| : z \in \mathbb{Z}\} =: \|\alpha q\| \leqslant Q^{-1}.$$

We shall use the proof of this fact later in a different context and so record it formally now.

Lemma 5.1 (Dirichlet's pigeon-hole principle)**.** *Suppose that $\alpha \in \mathbb{R}$ and $Q \in \mathbb{N}$. Then there is $q \in \mathbb{N}$ with $q \leqslant Q$ and $a \in \mathbb{Z}$ with $(a,q) = 1$ such that*

$$\left|\alpha - \frac{a}{q}\right| \leqslant \frac{1}{qQ}.$$

Proof. We can clearly achieve the hypothesis $(a,q)=1$ by cancellation so it will be sufficient to prove the conclusion without this hypothesis.

We consider the fractional parts $0,\{\alpha\},\dots,\{Q\alpha\}$. There are clearly two that are within $1/Q$ of each other by the pigeon-hole principle, say $\{r\alpha\}$ and $\{s\alpha\}$ with $r<s$. We put $q:=s-r$ so that

$$q\alpha = a+\{r\alpha\}-\{s\alpha\}$$

for some integer a. Thus

$$|q\alpha - a| \leqslant 1/Q$$

and the result follows. □

Now, what happens with $\{n^2\alpha\}$? The situation is much less clear. Weyl introduced an approach for dealing with this sort of problem which will inform our work with representation functions later on.

The basic question will boil down to estimating terms of the form

$$|\sum_{n=0}^{N}\exp(2\pi i n^2\alpha)|$$

If α is rational then the non-uniformity of quadratic residues in congruence classes will mean that we cannot expect much cancellation in this term. Indeed, suppose that $\alpha=1/3$. Then it is easy to see that

$$|\sum_{n=0}^{N}\exp(2\pi i n^2/3)| = |N/3+2\omega N/3+O(1)|$$

where ω is a primitive cube root of unity. If α is irrational or, rather, not close to a rational with small denominator then we shall see that we do get cancellation in this quantity. To estimate it we consider its square:

$$\begin{aligned}\left|\sum_{n=0}^{N}\exp(2\pi i\alpha n^2)\right|^2 &= \sum_{n,m=0}^{N}\exp(2\pi i\alpha(n^2-m^2))\\ &= \sum_{n,m=0}^{N}\exp(2\pi i\alpha(n-m)(n+m))\\ &= \sum_{u=0}^{N}\sum_{\substack{v=-u\\ v\equiv u \pmod 2}}^{u}\exp(2\pi i\alpha uv)\\ &\quad+\sum_{u=N+1}^{2N}\sum_{\substack{v=u-2N\\ v\equiv u \pmod 2}}^{2N-u}\exp(2\pi i\alpha uv).\end{aligned}$$

We shall estimate this through the inner sums. To this end we record the following decay estimate for the Fourier transform.

Lemma 5.2. *Suppose that $\theta \in \mathbb{R}$ and $s, t \in \mathbb{Z}$. Then we have the estimate*

$$|\sum_{n=s}^{t} \exp(2\pi i\theta n)| \leqslant \min\{t-s+1, \frac{1}{2\|\theta\|}\}.$$

Proof. Without loss of generality $s = 0$. There are $t+1$ terms in the sum and each term has modulus 1 so it follows that the sum is at most the $t+1$. For the second estimate we sum the geometric progression:

$$\sum_{n=0}^{t} \exp(2\pi i\theta n) = \frac{\exp(2\pi i\theta(t+1)) - 1}{\exp(2\pi i\theta) - 1}$$

provided $\|\theta\| \neq 0$; if it is then we certainly have the estimate. However

$$|\exp(2\pi i\theta) - 1| = |\exp(\pi i\|\theta) - \exp(-\pi i\|\theta)| = 2\sin\pi\|\theta\| \geqslant 4\|\theta\|,$$

whence we have the result. □

Returning to our earlier calculations we have

$$\left|\sum_{n=0}^{N} \exp(2\pi i\alpha n^2)\right|^2 \leqslant \sum_{u=0}^{N} |\sum_{\substack{v=-u \\ v\equiv u \pmod 2}}^{u} \exp(2\pi i\alpha uv)| + \sum_{u=N+1}^{2N} |\sum_{\substack{v=u-2N \\ v\equiv u \pmod 2}}^{2N-u} \exp(2\pi i\alpha uv)|,$$

but this is at most

$$\sum_{u=0}^{N} \min\{u+1, \frac{1}{2\|2\alpha u\|}\} + \sum_{u=N+1}^{2N} \min\{2N-u+1, \frac{1}{2\|2\alpha u\|}\},$$

which is, in turn, at most

$$\sum_{u=0}^{4N} \min\{N+1, \frac{1}{2\|\alpha u\|}\}.$$

The next lemma will let us show that $\|\alpha u\|$ stays away from 0 enough to provide some cancellation in the preceding.

Lemma 5.3. *Suppose that $\theta_1, \dots, \theta_k$ are δ-separated, i.e. $\|\theta_i - \theta_j\| \geqslant \delta$ for all $i \neq j$. Then*

$$\sum_{i=1}^{k} \min\{Q, \frac{1}{2\|\theta_i\|}\} = O((Q + \delta^{-1})\log Q).$$

Proof. We may certainly assume that $\theta_i \in [-1/2, 1/2]$, that at least half the sum comes from θ_i with $\theta_i \geqslant 0$ and that $0 \leqslant \theta_1 < \theta_2 < \cdots < \theta_l$. Thus

$$S := \sum_{i=1}^{k} \min\{Q, \frac{1}{2\|\theta_i\|}\} \leqslant 2\sum_{i=1}^{l} \min\{Q, \frac{1}{2\theta_i}\}.$$

Of course, in this case the sum is clearly maximised by taking $\theta_i = (i-1)\delta$ so that

$$S \leqslant 2\sum_{i=1}^{l} \min\{Q, \delta^{-1}/2(i-1)\} \leqslant 2QR + \delta^{-1}(\log l - \log R + O(1))$$

for any $R \geqslant 1$ by Proposition 1.3. Now $l \leqslant k \leqslant \delta^{-1}$ and we may certainly assume that $\delta < 1/Q$ and thus put $R = \delta^{-1}/Q$. We conclude that

$$S \leqslant O(\delta^{-1}\log Q),$$

and the result is proved. □

Combining what we have done we can prove the following.

Proposition 5.4 (Weyl's inequality)**.** *Suppose that $\alpha \in \mathbb{R}$ is such that $|\alpha - a/q| \leqslant 1/qQ$ for some naturals $q \leqslant Q$ and an integer a coprime to q. Then*

$$|\sum_{n=0}^{N} \exp(2\pi i\alpha n^2)| = O((N/\sqrt{q} + \sqrt{N} + \sqrt{q})\log N)$$

Proof. We begin by recalling that

$$\left|\sum_{n=0}^{N} \exp(2\pi i\alpha n^2)\right|^2 \leqslant \sum_{u=0}^{4N} \min\{N+1, \frac{1}{2\|\alpha u\|}\}.$$

Split the range of u up into $1 + O(N/q)$ intervals of length at most $\lfloor q/2\rfloor$, so that $I_1 := \{0, 1, \dots, \lfloor q/2\rfloor - 1\}$, $I_2 = \{\lfloor q/2\rfloor, \dots, 2\lfloor q/2\rfloor - 1\}$ etc. with a possibly shorter final interval. Now, suppose that $u, u' \in I_i$ are distinct. Then

$$\|\alpha(u-u')\| \geqslant \|a(u-u')/q\| - |u-u'|/qQ \geqslant 1/2q$$

by the triangle inequality whence $\{\alpha u : u \in I_i\}$ is a $1/2q$-separated set. It follows from Lemma 5.3 that

$$|\sum_{n=0}^{N} \exp(2\pi i\alpha n^2)|^2 = (1 + O(N/q)).O((N+q)\log N).$$

The result follows on taking square roots. □

We now turn to the question of how we use this information. The basic idea is that if $\|\alpha n^2\|$ is never small for $n \leqslant N$ then there is a large interval around the origin which it misses. This, in turn, implies that it must have a large Fourier coefficient which we know is not so if q is large; if q is small the result is easy.

To begin with we recall that if P is a large prime then the characters on $\mathbb{Z}/P\mathbb{Z}$ are just the maps

$$x \mapsto \exp(2\pi i x r/P),$$

so we identify $\widehat{\mathbb{Z}/P\mathbb{Z}}$ with $\mathbb{Z}/p\mathbb{Z}$ in the obvious way. Thus if $f : \mathbb{Z}/P\mathbb{Z} \to \mathbb{C}$ then

$$\widehat{f}(r) := \int f(x)\overline{\exp(2\pi i x r/P)}d\mathbb{P}_G(x).$$

The following lemma encodes the Fourier space localisation of an interval.

Lemma 5.5. *Suppose that P is a prime, $A \subset G := \mathbb{Z}/P\mathbb{Z}$ is a set of density α and $A \cap [-L, L] = \emptyset$. Then*

$$\sup_{0 \neq |r| \leqslant (P/L)^2} |\widehat{1_A}(r)| \geqslant \alpha L/2P.$$

Proof. We write $I := \{1, \dots, L\}$ so that $\operatorname{supp} 1_I * 1_{-I} \subset [-L, L]$. Thus

$$0 = \langle 1_A, 1_I * 1_{-I} \rangle_{L^2(G)} = \sum_r \widehat{1_A}(r) |\widehat{1_I}(r)|^2$$

by Plancherel's theorem. We isolate the trivial mode ($r = 0$) where

$$\widehat{1_A}(0)|\widehat{1_I}(0)|^2 = \alpha(L/P)^2$$

and apply the triangle inequality so that

$$\alpha L^2/P^2 \leqslant \sum_{r \neq 0} |\widehat{1_A}(r)||\widehat{1_I}(r)|^2.$$

On the other hand by Lemma 5.2 we have

$$|\widehat{1_I}(r)| \leqslant \frac{1}{P} \min\{L, 1/2\|r/P\|\} \leqslant \min\{L/P, 1/2|r|\}$$

if r is taken to lie in $(-P/2, P/2]$. Thus we see that

$$\begin{aligned} \alpha L^2/P^2 &\leqslant \sup_{0 \neq |r| \leqslant (P/L)^2} |\widehat{1_A}(r)| . \sum_{|r| < (P/L)^2} |\widehat{1_I}(r)|^2 + \alpha \sum_{|r| > (P/L)^2} 1/4|r|^2 \\ &\leqslant \sup_{0 \neq |r| \leqslant (P/L)^2} |\widehat{1_A}(r)| . (L/P) + \alpha (P/L)^{-2}/2 \end{aligned}$$

by Parseval's theorem and the fact that

$$\sum_{|r| > X} \frac{1}{|r|^2} \leqslant 2 \int_X^\infty x^{-2} dx \leqslant 2/X.$$

The result follows on some rearrangement. □

Finally we can prove our approximation theorem.

Theorem 5.6. *For all $\alpha \in \mathbb{R}$ and $N \in \mathbb{N}$ there is a positive integer $n \leqslant N$ such that $\|\alpha n^2\| \leqslant N^{-1/5+o(1)}$.*

Proof. By rational approximation it suffices to prove the above result for $\alpha = b/P$ where P is prime with $P > 2N$. Note that the elements $\{bn^2 : 1 \leqslant n \leqslant N\}$ are all distinct (mod P) – call this set A – and so we can apply the previous lemma to get that either

$$\|bn^2/P\| \leqslant L/P$$

for some $1 \leqslant n \leqslant N$ whence we shall be done on optimizing for L; or else

$$|\sum_{n=1}^N \exp(2\pi i r b n^2/P)| \geqslant NL/2P$$

for some $r \neq 0$ with $|r| \leqslant (P/L)^2$.

By Dirichlet's pigeon-hole principle there is some $q \leqslant N$ such that $(a,q)=1$ and

$$\left|\frac{rb}{P} - \frac{a}{q}\right| \leqslant \frac{1}{qN}.$$

If $q \leqslant M$ then

$$\|\alpha r^2 q^2\| \leqslant P^2 M/L^2 N$$

and we shall be done on optimising for M. Thus we shall take $q \geqslant M$ from hereon. In this case we apply Weyl's inequality so that

$$NL/2P = O((N/\sqrt{M} + \sqrt{M} + \sqrt{N})\log N).$$

Putting $L = P/N^\epsilon$ and $M = N^{1-\delta}$ we get that

$$\inf_{1\leqslant n\leqslant N} \|\alpha n^2\| \leqslant \max\{N^{-\epsilon}, N^{2\epsilon-\delta}\}$$

or else

$$N^{1-\epsilon} = O(N^{1/2+\delta/2+o(1)}).$$

Putting $\delta = 3\epsilon$ tells us that we can take $\epsilon = 1/5 - o(1)$ giving the result. □

6. Vinogradov's three-primes theorem

In this section we shall begin our work on proving Vinogradov's three-primes theorem. To begin with we sketch the plan. We write

$$\Lambda_N(x) := \begin{cases} \Lambda(x) & \text{whenever} x \leqslant N \\ 0 & \text{otherwise.} \end{cases}$$

The quantity

$$R_N(x) := \Lambda_N * \Lambda_N * \Lambda_N(x)$$

is a sort of weighted representation of x as a sum of powers of primes. In particular if we write P_N for the set of primes less than or equal to N then, as in §1, we have that

$$\begin{aligned} 1_{P_N} * 1_{P_N} * 1_{P_N}(x) &\geqslant \frac{1}{\log^3 N} \sum_{a+b+c=x} 1_{P_N}(a)\Lambda_N(a).1_{P_N}(b)\Lambda_N(b).1_{P_N}(c)\Lambda_N(c) \\ &\geqslant \frac{1}{\log^3 N}\Bigg(\sum_{a+b+c=x} \Lambda_N(a)\Lambda_N(b)\Lambda_N(c) \\ &\qquad -3\sum_{k\geqslant 2, p^k+b+c=x} (\log p)\Lambda_N(b)\Lambda_N(c)\Bigg) \\ &\geqslant (R_N(x) - O(N^{3/2}\log N))/\log^3 N, \end{aligned}$$

where we have implicitly used $x = O(N)$ and are only really interested in x N. Of course $1_{P_N} * 1_{P_N} * 1_{P_N}(x)$ is the number of ways of representing x as a sum of three primes so if

we can show that $R_N(x)$ is large then we shall have that x has a representation as the sum of three primes.

Hardy and Littlewood introduced the idea of using the Fourier transform to study $R_N(x)$ noting that

$$R_N(x) = \int_0^1 \widehat{\Lambda_N}(\theta)^3 \exp(-2\pi i x\theta) d\theta$$

by the inversion formula. They then split the range of integration into two sets: the major arcs, denoted

$$\mathfrak{M} := \{\theta \in \mathbb{T} : |\theta - a/q| \leqslant 1/qQ \text{ for some } q \leqslant Q_0 \text{ and } (a,q) = 1\},$$

and the minor arcs

$$\mathfrak{m} := \{\theta \in \mathbb{T} : |\theta - a/q| \leqslant 1/qQ \text{ for some } q > Q_0 \text{ and } (a,q) = 1\}.$$

By Dirichlet's pigeon-hole principle these sets cover $\mathbb{T}$. On $\mathfrak{M}$ we shall estimate $\widehat{\Lambda_N}(\theta)$ using the Siegel-Walfisz theorem; on $\mathfrak{m}$ we shall need to develop a Weyl type estimate due to Vaughn to show that $\widehat{\Lambda_N}(\theta)$ is small.

The reason for the names major and minor is that the major arcs contribute the main term to the integral form of $R_N(x)$ and the minor arcs an error term.

6.1. **The minor arcs.** As stated we shall estimate the minor arcs using a technique developed by Vaughn simplifying an earlier approach of Vinogradov quite considerably. In the first instance we want to introduce long intervals because we can estimate their Fourier transform using Lemma 5.2. To do this we need a trick of the form we used in Weyl's inequality to generate long intervals to sum over. In this instance we have the convolution identities of §1 to fall back on. In particular

$$\Lambda(x) = \mu * \log(x),$$

so that

$$\widehat{\Lambda_N}(\theta) = \sum_{n \leqslant N} \Lambda(n) \exp(2\pi i\theta n) = \sum_{ab \leqslant N} \mu(a) \log b \exp(2\pi i ab\theta).$$

Since log is smooth long exponential sums involving log have a lot of cancellation which we shall exploit. We shall let $X = N^{2/5}$ be a parameter and naturally split into two ranges:

$$S_1 := \sum_{a \leqslant X} \mu(a) \sum_{b \leqslant N/a} \log b \exp(2\pi i ab\theta).$$

and

$$S_2 := \sum_{X < a < N} \mu(a) \sum_{b \leqslant N/a} \log b \exp(2\pi i ab\theta).$$

In this sum S_2 we shall decompose log as $1*\Lambda$ (in the sense of Dirichlet convolution):

$$\begin{aligned}S_2 &= \sum_{b\leqslant X}\log b\sum_{X<a\leqslant N/b}\mu(a)\exp(2\pi ab\theta)\\ &= \sum_{cd\leqslant N}\Lambda(d)\sum_{X<a\leqslant N/cd}\mu(a)\exp(2\pi acd\theta)\\ &= \sum_{cd\leqslant N}\Lambda(d)\sum_{X<a\leqslant N/cd}\mu(a)\exp(2\pi acd\theta)\\ &= \sum_{d\leqslant N}\Lambda(d)\sum_{X<u\leqslant N/d}\exp(2\pi iud\theta)\sum_{a|u,X<a}\mu(a)\\ &= \sum_{d\leqslant N}\Lambda(d)\sum_{X<u\leqslant N/d}\exp(2\pi iud\theta)(\delta(u)-\sum_{a|u,a\leqslant X}\mu(a)).\end{aligned}$$

Thus

$$\begin{aligned}S_2 &= -\sum_{d\leqslant N}\Lambda(d)\sum_{X<u\leqslant N/d}\exp(2\pi iud\theta)\sum_{a|u,a\leqslant X}\mu(a)\\ &= -\sum_{X<u\leqslant N}\sum_{a|u,a\leqslant X}\mu(a)\sum_{d\leqslant N/u}\Lambda(d)\exp(2\pi iud\theta).\end{aligned}$$

If d is not too large then we get a long range of u over which to sum $\exp(2\pi iud)$ which will, again, lead to good cancellation. Thus we write $S_2=S_3+S_4$, where

$$S_3=-\sum_{X<u\leqslant N}\sum_{a|u,a\leqslant X}\mu(a)\sum_{d\leqslant\min\{X,N/u\}}\Lambda(d)\exp(2\pi iu\theta)$$

and

$$S_4=-\sum_{X<u\leqslant N}\sum_{a|u,a\leqslant X}\mu(a)\sum_{X<d\leqslant N/u}\Lambda(d)\exp(2\pi iud\theta).$$

The sum in S_3 can be rearranged ever so slightly to give

$$\begin{aligned}S_3 &= -\sum_{u\leqslant N}\sum_{a|u,a\leqslant X}\mu(a)\sum_{d\leqslant\min\{X,N/u\}}\Lambda(d)\exp(2\pi iud\theta)+\widehat{\Lambda_X}(\theta)\\ &= -\sum_{a\leqslant X}\mu(a)\sum_{v\leqslant N/a}\sum_{d\leqslant\min\{X,N/av\}}\Lambda(d)\exp(2\pi iavd\theta)+O(N^{2/5}).\end{aligned}$$

Writing

$$S_5:=-\sum_{a\leqslant X}\mu(a)\sum_{v\leqslant N/a}\sum_{d\leqslant\min\{X,N/av\}}\Lambda(d)\exp(2\pi iavd\theta)$$

we have shown Vaughn's identity that

$$\widehat{\Lambda_N}(\theta)=S_1+S_4+S_5+O(N^{2/5}).$$

Moreover, we expect that S_1 and S_5 will be relatively easy to estimate; S_4 is considerably harder.

To proceed with estimating these terms we shall use the following variant of Lemma 5.3.

Lemma 6.2. *Suppose that $\theta \in \mathbb{R}$ has $|\theta - a/q| \leqslant 1/qQ$ with $q \leqslant Q$ and $R \in \mathbb{N}$. Then*

$$\sum_{x=1}^{R} \min\{\frac{N}{x}, \frac{1}{\|\theta x\|}\} = O((q + R + N/q)\log N \log R).$$

Proof. We consider the sum in two parts. In the first instance we address the case when x is small:

$$S := \sum_{x=1}^{\lfloor q/2 \rfloor} \min\{\frac{N}{x}, \frac{1}{\|\theta x\|}\} \leqslant \sum_{x=1}^{\lfloor q/2 \rfloor} \frac{1}{\|\theta x\|}.$$

As in the proof of Weyl's inequality the numbers θx are $1/2q$-separated as x ranges $\{0, \dots, \lfloor q/2 \rfloor\}$ so

$$S \leqslant \sum_{x=1}^{\lfloor q/2 \rfloor} \frac{2q}{x} = O(q \log q).$$

Now we dyadically decompose the remaining range:

$$L_i := \sum_{x=2^i}^{2^{i+1}-1} \min\{\frac{N}{2^{i+1}}, \frac{1}{\|\theta x\|}\}.$$

As before we split the xs into intervals of length $\lfloor q/2 \rfloor$ so that by Lemma 5.3 we get

$$L_i = O(2^i/q).O(N/2^i + q)\log N = O(N/q + 2^i)\log N.$$

Summing over the range of i with $q/2 \leqslant 2^i \leqslant R$ we get the result. □

We shall now use this lemma to estimate the three terms S_1, S_4 and S_5.

Lemma 6.3. *With conditions as in Lemma 6.2 we have*

$$S_1 = \sum_{a \leqslant X} \mu(a) \sum_{b \leqslant N/a} \log b \exp(2\pi i ab\theta) = O((q + N/q)\log^3 N).$$

Proof. We begin by differentiating log through partial summation:

$$\begin{aligned}
\sum_{b \leqslant M} \log b \exp(2\pi i ab\theta) &= \sum_{b \leqslant M} \log b (\widehat{1_{[b]}}(a\theta) - \widehat{1_{[b-1]}}(a\theta)) \\
&= \widehat{1_{[M]}}(a\theta)\log M + \sum_{b \leqslant M-1} (\log(b+1) - \log b)\widehat{1_{[b]}}(a\theta) \\
&= O(\log M \sup_{b \leqslant M} |\widehat{1_{[b]}}(a\theta)|) + \sum_{b \leqslant M-1} \widehat{1_{[b]}}(a\theta)/b.
\end{aligned}$$

Now, by Lemma 5.2 we get that

$$\begin{aligned}
\sum_{b \leqslant M} \log b \exp(2\pi i ab\theta) &= O(\log M \min\{M, 1/\|a\theta\|\}) \\
&\quad + O(\min\{M, (\log M)/\|a\theta\|\}).
\end{aligned}$$

Thus

$$|S_1| \leqslant \sum_{x\leqslant X} O(\log N \min\{N/a, 1/\|a\theta\|\}).$$

The result then follows from Lemma 6.2. □

Next we estimate S_5 which we also expect to be fairly straightforward.

Lemma 6.4. *With conditions as in Lemma 6.2 we have*

$$\begin{aligned} S_5 &= -\sum_{a\leqslant X}\mu(a)\sum_{v\leqslant N/a}\sum_{d\leqslant \min\{X,N/av\}}\Lambda(d)\exp(2\pi iavd\theta) \\ &= O((q+N^{4/5}+N/q)\log N). \end{aligned}$$

Proof. Of course we start by reordering the summation:

$$S_5 = -\sum_{a\leqslant X}\mu(a)\sum_{d\leqslant X}\Lambda(d)\sum_{v\leqslant N/ad}\exp(2\pi iavd\theta).$$

Now the fact that X^2 is significantly smaller than N comes in to ensure that N/ad is large: by Lemma 5.2 we get that

$$|S_5| \leqslant \sum_{a\leqslant X}\sum_{d\leqslant X}\Lambda(d)\min\{N/ad, 1/\|ad\theta\|\}.$$

But by grouping the terms where $ad = u$ and noting non-negativity of the summand we get

$$|S_5| \leqslant \sum_{u\leqslant X^2}\sum_{ad=u}\Lambda(d)\min\{N/u, 1/\|u\theta\|\}.$$

Since $1 * \Lambda = \log$ we conclude that

$$S_5 = O(\log N \sum_{u\leqslant X^2}\min\{N/u, 1/\|u\theta\|\}).$$

The result follows by Lemma 6.2 again. □

Finally we turn to S_4. It will be useful to have a trivial estimate for the second moment of τ from §1.

Lemma 6.5. *We have the estimate*

$$\sum_{x\leqslant N}\tau(x)^2 = O(N\log^3 N).$$

Proof. It is easy to check that τ is sub-multiplicative so that $\tau(ab) \leqslant \tau(a)\tau(b)$. Then

$$\begin{aligned}\sum_{x\leqslant N}\tau(x)^2 &= \sum_{ab\leqslant N}\tau(ab)\\ &\leqslant \sum_{a\leqslant N}\tau(a)\sum_{b\leqslant N/a}\tau(b)\\ &\leqslant \sum_{a\leqslant N}\tau(a)O(\frac{N}{a}\log N)\\ &= O(N\log N).\sum_{a\leqslant N}\tau(a)/a)\\ &= O(N\log N).\sum_{bc\leqslant N}\frac{1}{bc} = O(N\log^3 N),\end{aligned}$$

by Propositions 1.4 and then 1.3. □

Lemma 6.6. *With conditions as in Lemma 6.2 we have*

$$\begin{aligned}S_4 &= -\sum_{X<u\leqslant N}\sum_{a|u,a\leqslant X}\mu(a)\sum_{X<d\leqslant N/u}\Lambda(d)\exp(2\pi iud)\\ &= O(\log^4 N(N/\sqrt{q}+N/\sqrt{X}+\sqrt{Nq})).\end{aligned}$$

Proof. We put $w(u) := \sum_{a|u,a\leqslant X}\mu(a)$ and note that

$$|w(u)| \leqslant \sum_{a|u}1 = \tau(u).$$

Rewrite the main sum as

$$S_4 = -\sum_{X<u\leqslant N}w(u)\sum_{X<d\leqslant N/u}\Lambda(d)\exp(2\pi iud\theta),$$

ready for splitting up the range of u. Dyadically decompose the range of values of u via the numbers

$$\mathcal{R} := \{X, 2X, \dots, 2^kX\}$$

where k is maximal such that $2^{k-1}X \leqslant N/X$. Thus

$$|S_4| \leqslant \sum_{R\in\mathcal{R}}T_R$$

where

$$T_R := \sum_{R<u\leqslant 2R}|w(u)|\left|\sum_{X<d\leqslant N/u}\Lambda(d)\exp(2\pi iud\theta)\right|.$$

Apply Cauchy-Schwarz to each to the inner sums to get that

$$\begin{aligned}|T_R|^2 &\leqslant \left(\sum_{R<u\leqslant 2R}|w(u)|^2\right)\\ &\quad\times\left(\sum_{R<u\leqslant 2R}\sum_{X<d_1,d_2\leqslant N/u}\Lambda(d_1)\Lambda(d_2)\exp(2\pi iu(d_1-d_2)\theta)\right).\end{aligned}$$

The first sum here is $O(R\log^3 N)$ by the previous lemma; it is the second that we hope to get some cancellation from. Reordering summation and the trivial logarithmic bound on Λ gives

$$|T_R|^2 = O(R\log^5 N)\sum_{X<d_1,d_2\leqslant N/R}\left|\sum_{R<u\leqslant\min\{2R,N/d_1,N/d_2\}}\exp(2\pi u(d_1-d_2)\theta)\right|.$$

Thus, by Lemma 5.2 we get

$$|T_R|^2 = O(R\log^5 N)\sum_{X<d_1,d_2\leqslant N/R}\min\{R,1/\|\theta(d_1-d_2)\}\}.$$

Now the number of representation of a number r in the form d_1-d_2 is at most N/R whence

$$\begin{aligned}|T_R|^2 &= O(R\log^5 N).\frac{N}{R}\sum_{r\leqslant N/R}\min\{R,1/\|\theta r\|\}\\ &= O(N\log^5 N).(1+O(N/qR)).(R+q)\log N,\end{aligned}$$

where the last line is by Lemma 5.3 applied in the usual. Combining all this we get that

$$|T_R|^2 = O(N\log^6 N).(N/q+R+q+N/R).$$

Inserting these back into our expression for S_4 and noting that $X<R\leqslant N/X$ we get

$$|S_4| = O((N/\sqrt{q}+N/\sqrt{X}+\sqrt{Nq})\log^4 N).$$

The result is proved. □

Combining the three previous lemmas with Vaughn's identity we get the following.

Lemma 6.7. *Suppose that $\theta\in\mathbb{R}$ has $|\theta-a/q|\leqslant 1/qQ$ for some $1\leqslant q\leqslant Q\leqslant N$ and $(a,q)=1$. Then*

$$|\Lambda_N(\theta)| = O(\log^4 N(N/\sqrt{q}+N^{4/5}+\sqrt{Nq})).$$

Notice that this motivated our choice of X: we wanted $X^2\sim N/\sqrt{X}$, hence $X\ N^{2/5}$.

Having got our Weyl-type inequality we can see what Q and Q_0 are going to be. Let $A,A_0>0$ be fixed to be thought of as large and put

$$Q := N/\log^{A+A_0}N \text{ and } Q_0 := \log^{A_0}N.$$

We now give the so called 'minor arcs' estimate.

Theorem 6.8. *We have the estimate*

$$\int_{\theta\in\mathfrak{M}^c} |\widehat{\Lambda_N}(\theta)|^3 d\theta = O(N^2 \log^{5-A_0/2} N).$$

Proof. By Parseval's theorem and the prime number theorem we have

$$\int_0^1 |\widehat{\Lambda_N}(\theta)|^2 d\theta = \sum_{p\leqslant N} (\log p)^2 = O(N\log N).$$

Then by Hölder's inequality and Lemma 6.7 we get the result. □

6.9. **The major arcs.** These are conceptually easier to work with although they will still take time. Our main tool is the Siegel-Walfisz theorem which we leverage though the Fourier transform. It will be convenient to write $e(\nu) := \exp(2\pi i\nu)$ and to begin with we estimate $\widehat{\Lambda_N}(a/q)$ for a/q a rational in lowest terms with small denominator. First, we record a short calculation.

Lemma 6.10. *For a and q integers such that $(a,q)=1$ we have*

$$\sum_{1\leqslant r\leqslant q,(r,q)=1} e(ra/q) = \mu(q).$$

Proof. Without loss of generality $a=1$. Now, writing

$$c(d) := \sum_{1\leqslant r\leqslant d,(r,d)=1} e(r/d),$$

we see that

$$\sum_{d|q} c(d) = \sum_{1\leqslant r\leqslant q} e(r/q) = \delta(q).$$

The result then follows by the Möbius inversion formula. □

Lemma 6.11. *For $q\leqslant Q_0$ and $B>0$ we have*

$$\widehat{\Lambda_N}(a/q) = \frac{\mu(q)}{\phi(q)} N + O_{A_0,B}(N\log^{A_0-B} N).$$

Proof. We do the obvious thing and group the terms into congruence classes:

$$\begin{aligned}\widehat{\Lambda_N}(a/q) &= \sum_{n\leqslant N} \Lambda_N(n) e(an/q)\\ &= \sum_{r=1}^{q} \sum_{n\leqslant N: n\equiv r \pmod q} \Lambda_N(n) e(ar/q).\end{aligned}$$

Of course if $(r,q)\neq 1$ then

$$\sum_{n\leqslant N: n\equiv r \pmod q} \Lambda_N(n) e(ar/q) = O(\log N),$$

thus

$$\begin{aligned}\widehat{\Lambda_N}(a/q) &= \sum_{1\leqslant r\leqslant q,(r,q)=1}\sum_{n\leqslant N:n\equiv r \pmod q}\Lambda_N(n)e(ar/q)+O(q\log N)\\ &= \sum_{1\leqslant r\leqslant q,(r,q)=1}e(ar/q)\psi(N;q,r)+O(q\log N).\end{aligned}$$

In this situation we apply the Siegel-Walfisz theorem to see that

$$\widehat{\Lambda_N}(a/q)=\sum_{1\leqslant r\leqslant q,(r,q)=1}\left(e(ar/q).\frac{N}{\phi(q)}+O_{A_0,B}(N\log^{A_0-B}N)\right).$$

The result follows by the previous lemma. □

As a corollary we get the major arcs estimate for $\widehat{\Lambda_N}(\theta)$. We write

$$\mathfrak{M}(a,q):=\left[\frac{a}{q}-\frac{1}{qQ},\frac{a}{q}+\frac{1}{qQ}\right].$$

Corollary 6.12. *Suppose that $\theta\in\mathfrak{M}(a,q)$ for some $q\leqslant Q_0$. Then for any $B>0$ we have that*

$$\widehat{\Lambda_N}(\theta)-\frac{\mu(q)}{\phi(q)}\widehat{1_{[N]}}(\theta-a/q)=O_{A,A_0,B}(N\log^{A+2A_0-B}N).$$

Proof. We examine the difference which we denote by D:

$$\begin{aligned}D &= \sum_{n\leqslant N}(\Lambda_N(n)e(a/q)-\frac{\mu(q)}{\phi(q)})e(n(\theta-a/q))\\ &= \sum_{n\leqslant N}((\widehat{\Lambda_n}(a/q)-\frac{\mu(q)}{\phi(q)}n))-(\widehat{\Lambda_{n-1}}(a/q)-\frac{\mu(q)}{\phi(q)}(n-1))))e(n(\theta-a/q))\\ &= (\widehat{\Lambda_N}(a/q)-\frac{\mu(q)}{\phi(q)}N)e(N(\theta-a/q))\\ &\quad+\sum_{n\leqslant N-1}(\widehat{\Lambda_n}(a/q)-\frac{\mu(q)}{\phi(q)}n))(e((n+1)(\theta-a/q))-e(n(\theta-a/q))).\end{aligned}$$

Since

$$(e((n+1)(\theta-a/q))-e(n(\theta-a/q)))=O(|\theta-a/q|)$$

we are done by the previous lemma. □

It follows immediately from the above estimate that

$$\widehat{\Lambda_N}(\theta)^3-\frac{\mu(q)^3}{\phi(q)^3}\widehat{1_{[N]}}(\theta-a/q)^3=O_{A,A_0,B}(N^3\log^{A+2A_0-B}N)$$

whenever $\theta\in\mathfrak{M}(a,q)$.

At this point we have the crucial fact that the sets $\mathfrak{M}(a,q)$ are disjoint since $q \leqslant Q_0$ and N is large so that $2Q_0 < Q$. Thus integrating the above shows us that

$$\int_{\mathfrak{M}} \widehat{\Lambda_N}(\theta)^3 e(-N\theta)d\theta = \sum_{q\leqslant Q_0} \sum_{\substack{(a,q)=1\\ 1\leqslant a\leqslant q}} \frac{\mu(q)^3}{\phi(q)^3} \int_{\mathfrak{M}(a,q)} \widehat{1_{[N]}}(\theta - a/q)^3 e(-N\theta)d\theta + \sum_{q\leqslant Q_0} \sum_{\substack{(a,q)=1\\ 1\leqslant a\leqslant q}} \mu(\mathfrak{M}(a,q)) O_{A,A_0,B}(N^3 \log^{A+2A_0-B} N).$$

The second integral is rather easy to estimate by Lemma 5.2 and Parseval's theorem:

$$\begin{aligned}\int_{\mathfrak{M}(a,q)} \widehat{1_{[N]}}(\theta - a/q)^3 e(-N\theta)d\theta &= e(-Na/q) \int_{-1/qQ}^{1/qQ} \widehat{1_{[N]}}(\theta')^3 e(-N\theta')d\theta' \\ &= e(-Na/q) \int_{\mathbb{T}} \widehat{1_{[N]}}(\theta')^3 e(-N\theta')d\theta' \\ &\quad + \sup_{\|\theta\|\geqslant 1/qQ} |\widehat{1_{[N]}}(\theta)| \int_{\mathbb{T}} |\widehat{1_{[N]}}(\theta)^2 d\theta \\ &= e(-Na/q) \int_{\mathbb{T}} \widehat{1_{[N]}}(\theta')^3 e(-N\theta')d\theta' \\ &\quad + O(N^2 \log^{-A} N).\end{aligned}$$

Of course

$$\int_{\mathbb{T}} \widehat{1_{[N]}}(\theta')^3 e(-N\theta')d\theta' = 1_{[N]} * 1_{[N]} * 1_{[N]}(N) = \binom{N-1}{2},$$

so

$$\int_{\mathfrak{M}} \widehat{\Lambda_N}(\theta)^3 e(-N\theta)d\theta = \sum_{q\leqslant Q_0} \sum_{\substack{(a,q)=1\\ 1\leqslant a\leqslant q}} \frac{\mu(q)^3}{\phi(q)^3} e(-aN/q) \binom{N-1}{2} + \sum_{q\leqslant Q_0} \sum_{\substack{(a,q)=1\\ 1\leqslant a\leqslant q}} \mu(\mathfrak{M}(a,q)) O_{A,A_0,B}(N^3 \log^{A+2A_0-B} N) + O(N^2 \log^{-A} N).$$

Since $Q = N/\log^{A+A_0} N$ and $Q_0 = \log^{A_0} N$ the second double sum in the error term is of size at most

$$\sum_{q\leqslant Q_0} \frac{2\phi(q)}{qQ} O_{A,A_0,B}(N^3 \log^{A+2A_0-B} N) = O_{A,A_0,B}(N^2 \log^{2A+4A_0-B} N).$$

Of course $\phi(q)=\Omega(q^{3/4})$ (and, in fact, a much stronger estimate is true) so that by integral comparison we have shown

$$\begin{aligned}\int_{\mathfrak{M}}\widehat{\Lambda_N}(\theta)^3e(-N\theta)d\theta &= \binom{N-1}{2}\sum_{q\leqslant Q_0}\sum_{\substack{(a,q)=1\\1\leqslant a\leqslant q}}\frac{\mu(q)^3}{\phi(q)^3}e(-Na/q)\\ &\quad +O_{A,A_0,B}(N^2(\log^{2A+4A_0-B}N+\log^{-A}N)).\end{aligned}$$

The truncation error in the double sum is also small:

$$|\sum_{q>Q_0}\sum_{\substack{(a,q)=1\\1\leqslant a\leqslant q}}\frac{\mu(q)^3}{\phi(q)^3}e(-Na/q)| \leqslant \sum_{q>Q_0}\frac{1}{\phi(q)^2}=O(Q_0^{1/2}).$$

In light of this we have

$$\begin{aligned}\int_{\mathfrak{M}}\widehat{\Lambda_N}(\theta)^3e(-N\theta)d\theta &= \binom{N-1}{2}\sum_{q=1}^{\infty}\sum_{\substack{(a,q)=1\\1\leqslant a\leqslant q}}\frac{\mu(q)^3}{\phi(q)^3}e(-Na/q)\\ &\quad +O_{A,A_0,B}(N^2(\log^{2A+4A_0-B}N+\log^{-A}N+\log^{-A_0/2})).\end{aligned}$$

Finally we combine this with Theorem 6.8 to get that $R_N(N)$ equals

$$\begin{aligned}&\binom{N-1}{2}\sum_{q=1}^{\infty}\sum_{\substack{(a,q)=1\\1\leqslant a\leqslant q}}\frac{\mu(q)^3}{\phi(q)^3}e(-Na/q)\\ &+O_{A,A_0,B}(N^2(\log^{2A+4A_0-B}N+\log^{-A}N+\log^{-A_0/2}+\log^{5-A_0/2}N)).\end{aligned}$$

By suitable choice of A, A_0 and B we have proved the following proposition.

Proposition 6.13. *For all $C>0$ we have*

$$R_N(N)=\binom{N-1}{2}\sum_{q=1}^{\infty}\sum_{\substack{(a,q)=1\\1\leqslant a\leqslant q}}\frac{\mu(q)^3}{\phi(q)^3}e(-Na/q)+O_C(N^2\log^{-C}N).$$

It is possible to simplify this expression slightly more in the style of Lemma 6.10.

Lemma 6.14. *For integers n and q we have*

$$\sum_{1\leqslant r\leqslant q,(r,q)=1}e(rn/q)=\frac{\mu(q/(n,q))\phi(q)}{\phi(q/(n,q))}.$$

Proof. This follows immediately from Lemma 6.10 on noting that

$$\sum_{1\leqslant r\leqslant q,(r,q)=1}e(rn/q)=\frac{\phi(q)}{\phi(q/(n,q))}\sum_{\substack{1\leqslant r'\leqslant q/(q,n)\\(r',q/(q,n))=1}}e(r'\frac{n/(n,q)}{q/(n,q)}).$$

□

We conclude that

$$R_N(N) = \binom{N-1}{2} \sum_{q=1}^{\infty} \frac{\mu(q)\mu(q/(q,N))}{\phi(q)^2\phi(q/(q,N))} + O_C(N^2 \log^{-C} N).$$

However, since ϕ and μ are multiplicative we get that

$$R_N(N) = \binom{N-1}{2} \prod_{p\nmid N} (1 + \frac{1}{(p-1)^3}) \prod_{p|N} (1 - \frac{1}{(p-1)^2}) + O_C(N^2 \log^{-C} N).$$

Vinogradov's theorem is an immediate corollary.

Theorem 6.15 (Vinogradov's theorem)**.** *Every sufficiently large odd number is a sum of three primes.*

Proof. All we need to note is that

$$\begin{aligned} \prod_{p\nmid N} (1 + \frac{1}{(p-1)^3}) \prod_{p|N} (1 - \frac{1}{(p-1)^2}) &\geqslant \prod_{p|N} (1 - \frac{1}{(p-1)^2}) \\ &\geqslant \exp(-2\sum_{p|N} \frac{1}{(p-1)^2}) \geqslant \exp(-4) \end{aligned}$$

since $1/(p-1)^2 \leqslant 1/4$ if p is not 2. □

Mathematical Institute, University of Oxford, 24-29 St. Giles', Oxford OX1 3LB, England

E-mail address: tom.sanders@maths.ox.ac.uk