MULTIPLICATIVE FUNCTIONS

ALED WALKER

Abstract. These notes were first written for a 16 lecture graduate course at the University of Cambridge, delivered remotely via Zoom, during Michaelmas Term 2020.

Preamble

Course content

The classical proof of the prime number theorem, using Cauchy's residue theorem from complex analysis, is a beautiful piece of mathematics. It is also a little troubling. How can we really claim to understand the properties of the integers under multiplication if we have to resort to such witchcraft as the residue theorem?

In recent years there has been a movement, spearheaded by Andrew Granville and Kannan Soundararajan but building on earlier work of many people, to consider an alternative approach to the subject, one which avoids the use of the residue theorem. This has involved turning the focus away from the primes themselves and focussing instead on multiplicative functions. The programme has had numerous successes, not just in reinterpreting pre-existing theorems but in proving extremely surprising new results on multiplicative functions themselves.

In this course we will try to cover the following topics from the modern theory of multiplicative functions (in greater or lesser detail, depending on time constraints):

- Long averages: pretentious multiplicative functions and Halász's theorem;
- Application to Dirichlet characters: Granville–Soundararajan's improvement on the Pólya–Vinogradov inequality;
- Short averages: the Matomäki–Radziwiłł theorem;
- Correlations: Tao's proof of the logarithmically-averaged Chowla conjecture.

Prerequisites

I will assume familiarity with basic undergraduate real and complex analysis, including a little harmonic analysis (essentially just the Fourier inversion formula). It is *not* a prerequisite to have previously attended a first course in analytic number theory, but it will certainly be helpful to have done so, not least for putting the results of this course into their full context.

Examples Sheets

There will be two examples sheets for this course. Many of the questions on these sheets will lead the student through the proofs of various 'standard' estimates from analytic number theory, to save us from having to discuss them in detail in the lectures themselves. There will also be some problem material built around further uses and properties of multiplicative functions. Some exercises embedded in the text will also appear on the examples sheets.

Literature

Here is a list of some relevant recent papers:

- Granville, A. and Soundararajan, K., *Large character sums: pretentious characters and the Pólya–Vinogradov theorem.* Journal of the American Mathematical Society, **20**(2), 357-384 (2007).
- Granville, A., Harper, A., and Soundararajan, K., *A more intuitive proof of a sharp version of Halász's theorem.* Proceedings of the American Mathematical Society, 146(10), pp. 4099-4104 (2018).
- Matomäki, K. and Radziwiłł, M., *Multiplicative functions in short intervals.* Annals of Mathematics 1015-1056 (2016).
- Soundararajan, K., *The Liouville function in short intervals [after Matomäki and Radziwiłł].* Séminaire Bourbaki (2016), p.68ème.
- Tao, T., *The logarithmically averaged Chowla and Elliott conjectures for two-point correlations.* Forum of Mathematics, Pi (Vo. 4). Cambridge University Press 2016.

Acknowledgements

In preparing this course I am greatly indebted to various earlier notes and papers of Ben Green, Andrew Granville, Dimitris Koukoulopoulos, Alexander Mangerel, Hugh Montgomery, Kannan Soundararajan, Terence Tao, and Bob Vaughan.

1. Lecture 1: Examples and Motivation

This course has been designed to cater for two rather different audiences. I suspect that about half of those watching will already have a strong grounding in classical analytic number theory; for this audience, only the most novel material – and any insights into how this material relates to the classical theory – will be of interest. However, I suspect that the other half will comprise mathematicians who do not have such a strong background in this particular field (some graduate students in combinatorics or probability, say); for this audience, a more formal definition-theorem-proof style of course might be the more appreciated.

I shall try to keep both audience entertained. But I ask for the forbearance of each: for the forbearance of the experts, while I build up some basic material for the novices, and for the forbearance of the novices, while I make some contextual remarks for the experts.

To begin, let me attempt to give a summary of every single proof in analytic number theory, in four easy steps...

(1) Seek to estimate $\sum_{n\leqslant X} f(n)$ for some function $f : \mathbb{N} \longrightarrow \mathbb{C}$.

(2) Show by means of some integral transform (Fourier transform, Mellin transform, Perron's formula etc.) that

$$\sum_{n\leqslant X} f(n) \approx \int \widetilde{f}(w,X)g(w,X)\,dx,$$

where $\widetilde{f}$ is some transform of f, and where g depends on the type of transformation used. (I am being deliberately imprecise about what sort of variable w is, and about the range of integration.)

(3) Understand this integral extremely well, e.g. work out the regions where the integrand is large, where it is small, where it is slowly varying, etcetera.

(4) Use this information to estimate the integral, and thereby to estimate the original sum $\sum_{n\leqslant X} f(n)$.

Easy! (Of course the real meat is in Step 3...).

Some classical gems follow the above schematic rather closely, e.g. estimating the number of primes p_1, p_2, p_3 for which $p_1 + p_2 + p_3 = N$.

(1) Seek to estimate $\sum_{p_1,p_2,p_3\leqslant N} f(p_1,p_2,p_3)$, where

$$f(p_1,p_2,p_3) = \begin{cases} 1 & \text{if } p_1+p_2+p_3+ = N \\ 0 & \text{otherwise.} \end{cases}$$

(2) By Fourier inversion one ends up with

$$\sum_{p_1,p_2,p_3\leqslant N} f(p_1,p_2,p_3) = \int_0^1 \Big(\sum_{p\leqslant N} e^{2\pi i\alpha p}\Big)^3 e^{-2\pi i\alpha N}\,d\alpha.$$

(3) One proves that $|\sum_{p\leqslant N} e^{2\pi i\alpha p}|$ is large if $\alpha \approx a/q$ for a rational number a/q with q small, and that $|\sum_{p\leqslant N} e^{2\pi i\alpha p}|$ is small otherwise.

(4) One adds up the contributions to the integral from $\alpha \approx a/q$, proving an asymptotic formula.

But how well does the rubric apply to that triumph of 19^{th} century mathematics, the Prime Number Theorem (henceforth written PNT)?

Theorem 1.1 (PNT, Hadamard–de la Vallée Poussin, 1896)**.**

$$\sum_{p\leqslant X} 1 = (1+o(1))\frac{X}{\log X}$$

as $X \to \infty$.

The proof was completed independently by Hadamard and de la Vallée Poussin, but builds on work and ideas of Euler, Gauss, Dirichlet, Riemann, Stieltjes, Jensen, Cahen, von Mangoldt, etc.. In its modern formulation, the standard proof proceeds via the following steps:

(1) Seek to estimate $\sum_{p\leqslant X} 1$.

(2) By Perron's formula/Mellin inversion one ends up with

$$(\log X)\sum_{p\leqslant X} 1 \approx \frac{1}{2\pi i}\int\limits_{1+\frac{1}{\log X}-iT}^{1+\frac{1}{\log X}+iT} \Big(-\frac{\zeta'(s)}{\zeta(s)}\Big)\frac{X^s}{s}\,ds,$$

where $\zeta : \mathbb{C} \longrightarrow \mathbb{C}\cup\{\infty\}$ is the Riemann zeta function (meromorphic, a single simple pole at $s=1$) and T is some threshold. Here we have a contour integral along the line $\Re s = 1 + 1/\log X$.

(3) Shift the contour using Cauchy's Theorem to the line $\Re s = 1 - \frac{c}{\sqrt{\log X}}$ for some small $c > 0$, and show thereby that

$$\frac{1}{2\pi i}\int\limits_{1+\frac{1}{\log X}-iT}^{1+\frac{1}{\log X}+iT} \Big(-\frac{\zeta'(s)}{\zeta(s)}\Big)\frac{X^s}{s}\,ds \approx X\,\mathrm{Res}_{s=1}\Big(-\frac{\zeta'(s)}{\zeta(s)}\Big) + \frac{1}{2\pi i}\int\limits_{1-\frac{c}{\sqrt{\log X}}-iT}^{1-\frac{c}{\sqrt{\log X}}+iT} \Big(-\frac{\zeta'(s)}{\zeta(s)}\Big)\frac{X^s}{s}\,ds.$$

One shows that this second integral is small by showing that the integrand is small (by such observations as $|X^s| = X^{\Re s} = X^{1-\frac{c}{\sqrt{\log X}}} = o(X)$).

(4) Show that $\mathrm{Res}_{s=1}\Big(-\frac{\zeta'(s)}{\zeta(s)}\Big) = 1$.

But isn't Step 3 cheating?! We haven't really 'understood' the initial integral in the traditional sense, i.e. finding where the integrand is large, where it is small, where it is slowly varying. Rather, we have used the magic of Cauchy's Theorem to show that the initial integral is related to a different integral that we can understand much more easily.

To show that the Cauchy step works one needs to know that there are no more poles of $-\zeta'(s)/\zeta(s)$ in the region of interest, and this means understanding something about the zeros of $\zeta(s)$. This is of course highly non-trivial! However, all of the methods we currently have for doing this involve understanding certain things about $\zeta(s)$ where $\Re(s) > 1$ and then converting this understanding into certain (rather weak) information about the zeros. So why not cut out the middle-man (i.e. the zeros) and just focus on properly understanding the original integrand?

Now, to business...

Definition 1.2 (Multiplicative functions)**.** *We say that a function* $f : \mathbb{N} \longrightarrow \mathbb{C}$ *satisfying* $f(1) = 1$ *is:*

- multiplicative *if* $f(mn) = f(m)f(n)$ *for all* $m, n \in \mathbb{N}$ *for which* $\gcd(m, n) = 1$*;*
- completely multiplicative *if* $f(mn) = f(m)f(n)$ *for all* $m, n \in \mathbb{N}$*.*

At first reading, a natural question arises: the second of these notions seems to be the more natural of the two (indeed, f is completely multiplicative if and only if it is a homomorphism of monoids $f : (\mathbb{N}, \times) \longrightarrow (\mathbb{C}, \times)$), so why don't we call *those* functions 'multiplicative'? This is a good question. For now, we hope the reader will be satisfied with the short and standard answer: that is, it turns out that multiplicative functions form a richer class of objects than completely multiplicative functions, with better closure properties, and many of the functions that occur naturally (and historically) in number theory are multiplicative but not completely multiplicative.

Examples: (CM) = completely multiplicative, p_i is always prime.

(1) $n \mapsto \delta(n)$, where $\delta(n) := \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{otherwise.} \end{cases}$ (CM).

(2) $n \mapsto 1$. (CM).

(3) $n \mapsto n^\alpha$, where $\alpha \in \mathbb{C}$ is fixed. (CM).

(4) $n \mapsto \begin{cases} 1 & \text{if } n = m^k \text{ for some } m \in \mathbb{N} \\ 0 & \text{otherwise,} \end{cases}$ where $k \in \mathbb{N}$ is fixed.

(5) $n \mapsto \mu(n)$, where $\mu : \mathbb{N} \longrightarrow \{-1, 0, +1\}$ is the *Möbius function* defined by $\mu(1) = 1$ and

$$\mu(n) = \begin{cases} 0 & \text{if } p^2 | n \text{ for some prime } p \\ -1 & \text{if } n = p_1 \dots p_k \text{ with } p_i \text{ distinct and } k \text{ odd} \\ 1 & \text{if } n = p_1 \dots p_k \text{ with } p_i \text{ distinct and } k \text{ even.} \end{cases}$$

(6) $n \mapsto \lambda(n)$, where $\lambda : \mathbb{N} \longrightarrow \{-1, 1\}$ is the *Liouville function* defined by $\lambda(1) = 1$ and

$$\lambda(n) := \begin{cases} -1 & \text{if } n = p_1 \dots p_k \text{ with } k \text{ odd} \\ 1 & \text{if } n = p_1 \dots p_k \text{ with } k \text{ even.} \end{cases} \quad \text{(CM).}$$

(7) $n \mapsto \varphi(n)$, where $\varphi(n)$ is the *Euler* φ *function* defined by

$$\varphi(n) = |\{m \leqslant n : \gcd(m, n) = 1\}|.$$

(8) $n \mapsto \tau(n)$, where the *divisor function* τ is defined by

$$\tau(n) = |\{d \leqslant n : d | n\}|$$

(9) More generally, $n \mapsto \tau_k(n)$, where

$$\tau_k(n) = |\{d_1, \dots, d_k \leqslant n : d_1 d_2 \dots d_k = n\}|.$$

(10) $n \mapsto \sigma_\alpha(n)$, where for $\alpha \in \mathbb{C}$ we define $\sigma_\alpha(n) = \sum_{d|n} d^\alpha$.

(11) $n \mapsto \chi(n)$, where $\chi : \mathbb{N} \longrightarrow \mathbb{C}$ is a Dirichlet character. We'll be covering Dirichlet characters in detail later in the course, but in case you don't already know, a Dirichlet character (with modulus q), is a group homomorphism $\chi : (\mathbb{Z}/q\mathbb{Z})^\times \longrightarrow \mathbb{C}$, which is then extended to a function on $\mathbb{Z}$ by defining:
- $\chi(n) = 0$ if $\gcd(n, q) > 1$;
- $\chi(n + q) = \chi(n)$ for all $n \in \mathbb{Z}$.

The Legendre symbol $(\frac{n}{p})$ is one such Dirichlet character, with modulus p. (CM).

(12) $n \mapsto \frac{1}{4}|\{(x, y) \in \mathbb{Z}^2 : x^2 + y^2 = n\}|$.

(13) $n \mapsto s_y(n)$, where

$$s_y(n) = \begin{cases} 1 & \text{if } p|n \Rightarrow p \leqslant y \\ 0 & \text{otherwise.} \end{cases} \quad \text{(CM)}$$

is the indicator function of the *y-friable* numbers (also called the *y-smooth* numbers).

(14) $n \mapsto a(n)$, where $a(n)$ is the n^{th} Fourier coefficient of a normalised Hecke cusp eigenform, e.g. when $a(n)$ is the coefficient of q^n in the expansion of the Ramanujan Δ-function $\Delta(q) = q\prod_{n=1}^{\infty}(1 - q^n)^{24}$.

(15) $n \mapsto f(n)$, where, having chosen $f(1) = 1$ and arbitrary complex numbers $f(p^k)$ for all prime p and $k \in \mathbb{N}$, we define $f(n) = f(p_1^{k_1}) \dots f(p_m^{k_m})$, when $n = p_1^{k_1} \dots p_m^{k_m}$.

Exercise 1.3. Prove that all of the above examples, save for (12) and (14), are multiplicative (resp. completely multiplicative) as appropriate.

Exercise 1.4. (†). Prove that examples (12) and (14) are multiplicative.

We let

$$\begin{aligned} \mathcal{M} &:= \{f : f \text{ is multiplicative}\} \\ \mathcal{M}_0 &:= \{f \in \mathcal{M} : |f(n)| \leqslant 1 \text{ for all } n\} \\ \mathcal{M}_k &:= \{f \in \mathcal{M} : |f(n)| \leqslant \tau_k(n) \text{ for all } n\}. \end{aligned}$$

One can think of these classes of functions in terms of the sizes of $f(p)$, i.e. $f \in \mathcal{M}_k$ implies $|f(p)| \leqslant k$ for all primes p. Arguments that apply to functions in $\mathcal{M}_0$ can usually be adapated to functions in $\mathcal{M}_k$ with enough effort, but for simplicity we will almost exclusively work with $\mathcal{M}_0$ in this course.

In the coming lectures we will put a metric structure on $\mathcal{M}_0$, from which many rich properties will emerge.

New from old

- if $f \in \mathcal{M}$ then $|f| \in \mathcal{M}$.
- if $f \in \mathcal{M}$ and f is real-valued then $\text{sgn}(f) \in \mathcal{M}$.
- if $f, g \in \mathcal{M}$ (resp. completely multiplicative), then $fg \in \mathcal{M}$ (resp. completely multiplicative).
- A particularly important construction in analytic number theory is the *Dirichlet convolution* of two functions $f, g : \mathbb{N} \longrightarrow \mathbb{C}$, denoted by $f \star g : \mathbb{N} \longrightarrow \mathbb{C}$ and defined

to be

$$(f \star g)(n) = \sum_{d|n} f(d) g\Big(\frac{n}{d}\Big) = \sum_{ab=n} f(a)g(b).$$

Dirichlet convolution is often a way of creating a seemingly more complicated function out of simpler functions, or of decomposing a given functions into simpler components. For example, $\tau = 1 \star 1$.

The operator $\star$ interacts particularly well with multiplicative functions (see Exercises below).

Exercise 1.5.

(a) Let $f, g : \mathbb{N} \longrightarrow \mathbb{C}$. If $f, g \in \mathcal{M}$, show that $f \star g \in \mathcal{M}$ too. Show that this statement is false if we replace 'multiplicative' by 'completely multiplicative' throughout.
(b) Show that $1 \star \mu = \delta$.
(c) Show that $\mathbb{C}^{\mathbb{N}}$ can be given the structure of a commutative unital ring, with addition given by pointwise $+$, multiplication given by $\star$, and multiplicative identity δ. Show that $\mathcal{M}$ is contained within the multiplicative group of units.

Exercise 1.6 (Some Dirichlet convolution identities)**.**

(a) Show that $\varphi = \mu \star \mathrm{id}$, where $\mathrm{id}(n) := n$ for all n.
(b) Show that $\mu \star 1_{\square} = \lambda$, where $1_{\square}$ is the indicator function of the squares.
(c) Show that $\tau^3 \star 1 = (\tau \star 1)^2$.

Generating functions

Let $f : \mathbb{N} \longrightarrow \mathbb{C}$ be a function. We can associate a *Dirichlet series* $F(s)$ to f by writing the formal sum

$$F(s) := \sum_{n \leqslant 1} \frac{f(n)}{n^s}$$

for $s \in \mathbb{C}$. If $f \in \mathcal{M}$ this sum may be factorised (formally) as an *Euler product*

$$F(s) = \prod_p \Big(1 + \frac{f(p)}{p^s} + \frac{f(p^2)}{p^{2s}} + \frac{f(p^3)}{p^{3s}} + \cdots\Big).$$

This is just the fundamental theorem of arithmetic in a different guise.

The following facts are sometimes established in a first analytic number theory course.

Theorem 1.7 (Convergence of Dirichlet series)**.** *Let $F(s) = \sum_{n \geqslant 1} f(n)/n^s$ be a Dirichlet series. If $F(s_0)$ converges (resp. absolutely converges) for some complex number $s_0 = \sigma_0 + it_0$ then $F(s)$ converges (resp. absolutely converges) uniformly in compact subsets of the half-plane $\sigma > \sigma_0$. In particular, $F(s)$ defines a holomorphic function there.*

From this theorem we see that there is an abscissa of convergence

$$\sigma_c = \sigma_c(F) := \inf\{\sigma \in \mathbb{R} : \exists t \in \mathbb{R} \text{ such that } F(\sigma + it) \text{ converges}\}$$

and an abscissa of absolute convergence

$$\sigma_a = \sigma_a(F) := \inf\{\sigma \in \mathbb{R} : F(\sigma) \text{ converges absolutely}\},$$

and $F(s)$ is holomorphic in the region $\Re(s) > \sigma_c$.

Exercise 1.8. Prove that $\sigma_c \leqslant \sigma_a \leqslant \sigma_c + 1$. Also, show that $\sigma_c < \infty$ iff there is some $\theta \in \mathbb{R}$ for which $f(n) = O(n^\theta)$ for all $n \in \mathbb{N}$.

Theorem 1.9 (Euler products)**.** *Let f be a multiplicative function and $s \in \mathbb{C}$. Then the series $\sum_{n\geqslant 1} f(n)/n^s$ converges absolutely if and only if the double series $\sum_p \sum_{k\geqslant 1} f(p^k)/p^{ks}$ converges absolutely. When they both converges absolutely, we have*

$$\sum_{n=1}^{\infty} \frac{f(n)}{n^s} = \prod_p \Big(1 + \frac{f(p)}{p^s} + \frac{f(p^2)}{p^{2s}} + \cdots\Big).$$

Note that this theorem is only true in the domain of *absolute convergence*, as can be seen by considering the Dirichlet series $\sum_{n\leqslant 1}(-1)^{n-1}/n^s$ (for which $\sigma_c = 0$, but $\sigma_a = 1$ and the Euler product formula only holds when $\Re s > 1$).

Dirichlet series interact well with Dirichlet convolution.

Theorem 1.10 (Dirichlet convolution and Dirichlet series)**.** *If $f, g, h : \mathbb{N} \longrightarrow \mathbb{C}$ with Dirichlet series $F(s)$, $G(s)$, $H(s)$ respectiviely and abscissas of absolute convergence $\sigma_a(F), \sigma_a(G), \sigma_a(H)$, then*

$$h = f \star g \text{ if and only if } H(s) = F(s)G(s) \text{ for all } \Re s > \max(\sigma_a(F), \sigma_a(G)) \text{ and } \sigma_a(H) \leqslant \max(\sigma_a(F), \sigma_a(G)).$$

Very few schoolchildren dream of growing up to discover more properties of multiplicative functions. But at least some schoolchildren *do* dream of growing up to discover more properties about the primes. It turns out, fortunately, that these are not entirely unrelated endeavours.

Proposition 1.11 (Equivalent formulation of PNT, Landau 1906)**.** *The following are elementarily equivalent:*

(1) PNT;
(2) $\frac{1}{X}\sum_{n\leqslant X} \mu(n) = o(1)$ as $X \to \infty$;
(3) $\frac{1}{X}\sum_{n\leqslant X} \lambda(n) = o(1)$ as $X \to \infty$.

This proposition recasts PNT as a pseudorandomness principle for the Liouville function and the Möbius function, namely that integers are just as likely to have an even number of prime factors as they are to have an odd number of prime factors.

We'll give the proof of Proposition 1.11 in the next lecture. For now, let us chart the parallel history of prime numbers and of mutliplicative functions. To do this properly, we should quickly introduce another object from classical analytic number theory, the *von Mangoldt function* $\Lambda(n)$, which is a function supported on prime powers and defined by

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

This is **not** a multiplicative function. But it does enjoy good properties with respect to Dirichlet convolution, which it why it earns its central role in the story.

Exercise 1.12. *Prove the following identities:*

- $1 \star \Lambda = \log$;
- $\Lambda = \mu \star \log$;
- $\Lambda = -(1 \star \mu \log)$

There is a Dirichlet series explanation for these identities of course, namely that

$$\sum_{n\geqslant 1} \frac{\Lambda(n)}{n^s} = -\frac{\zeta'(s)}{\zeta(s)} \qquad \text{when } \Re s > 1,$$

where $\zeta(s) = \sum_{n\geqslant 1} 1/n^s$ is the Riemann zeta function.

Proposition 1.13 (Log-weights)**.** *The following are elementarily equivalent:*

(1) PNT;

(2) $\sum_{n\leqslant X} \Lambda(n) = (1+o(1))X$ *as* $X \to \infty$.

Proof. Observe first of all that

$$\sum_{n\leqslant X} \Lambda(n) = \sum_{k\leqslant \log X/\log 2} \sum_{p\leqslant X^{1/k}} \log p = \sum_{p\leqslant X} \log p + O(X^{1/2}\log^2 X).$$

So (2) is equivalent to the asymptotic $\sum_{p\leqslant X} \log p = (1+o(1))X$ as $X\to\infty$.

Furthermore, for all $\varepsilon > 0$ we have

$$\sum_{p\leqslant X} 1 \geqslant \frac{1}{\log X}\sum_{p\leqslant X}\log p \geqslant (1-\varepsilon)\sum_{X^{1-\varepsilon}<p\leqslant X} 1 \geqslant (1-\varepsilon)\sum_{p\leqslant X} 1 - O(X^{1-\varepsilon}).$$

(1) $\Rightarrow$ (2): From the above inequality and PNT we have

$$\frac{X}{\log X}(1+o(1)) \geqslant \frac{1}{\log X}\sum_{p\leqslant X}\log p \geqslant (1-\varepsilon)\frac{X}{\log X} - O(X^{1-\varepsilon}),$$

so

$$1 + o(1) \geqslant \frac{1}{X}\sum_{p\leqslant X}\log p \geqslant 1 - \varepsilon - O(X^{-\varepsilon}\log X).$$

If X is large enough in terms of ε we have

$$1 + 2\varepsilon \geqslant \frac{1}{X}\sum_{p\leqslant X}\log p \geqslant 1 - 2\varepsilon.$$

Since ε was arbitrary, we conclude that $\frac{1}{X}\sum_{p\leqslant X}\log p \sim 1$ as required.

(2) $\Rightarrow$ (1): A very similar argument. □

Returning to our main theme, let us compare the known results on prime numbers and on the Möbius function.

Prime Number Theorem	**Möbius cancellation**
1896	**1897**
Hadamard, de la Vallée Poussin	**von Mangoldt**
$\frac{1}{X}\sum_{n\leqslant X}\Lambda(n) = 1 + o(1)$ as $X\to\infty$	$\frac{1}{X}\sum_{n\leqslant X}\mu(n) = o(1)$ as $X\to\infty$

Assuming the Riemann Hypothesis (RH), one also gets similar estimates.

Prime Number Theorem on RH	**Möbius cancellation on RH**
1885	**1912**
Stieltjes	**Littlewood**
$\frac{1}{X}\sum_{n\leqslant X}\Lambda(n) = 1 + O_\varepsilon(X^{-\frac{1}{2}+\varepsilon})$ for all $\varepsilon > 0$.	$\frac{1}{X}\sum_{n\leqslant X}\mu(n) = O_\varepsilon(X^{-\frac{1}{2}+\varepsilon})$ for all $\varepsilon > 0$.

In shorter intervals, there *used* to also be a close relationship between the best known results in each case.

Prime in short intervals
1972
Huxley

$\frac{1}{X^\theta} \sum_{X<n\leqslant X+X^\theta} \Lambda(n) = 1 + o_\theta(1)$ as $X \to \infty$, provided $\theta > 7/12$.

Möbius cancellation in short intervals
1976
Ramachandra

$\frac{1}{X^\theta} \sum_{X<n\leqslant X+X^\theta} \mu(n) = o_\theta(1)$ as $X \to \infty$, provided $\theta > 7/12$.

NB: Heath-Brown slightly improved both of these results in 1988.

The best known results in 'almost all' short intervals also used to be comparable.

Primes in almost all short intervals on RH
1943
Selberg

For all $\varepsilon > 0$, if $h > \log^{2+\varepsilon} X$ then $\int_X^{2X} \Big| \sum_{x<n\leqslant x+h} \Lambda(n) - h \Big|^2 dx = o_\varepsilon(Xh^2)$ as $X \to \infty$.

Möbius cancellation in almost all short intervals on RH
2008-ish
Gao

There exists $A > 0$ such that if $h > \log^A X$ then $\int_X^{2X} \Big| \sum_{x<n\leqslant x+h} \mu(n) \Big|^2 dx = o_\varepsilon(Xh^2)$ as $X \to \infty$.

There is even a comparison between some of the conjectures.

Twin prime conjecture asymptotic

$\sum_{n\leqslant X} \Lambda(n)\Lambda(n+2) \sim CX$ as $X \to \infty$, for some explicit constant C.

Chowla conjecture

$\frac{1}{X} \sum_{n\leqslant X} \mu(n)\mu(n+1) \to 0$ as $X \to \infty$.

But the world is now irrevocably altered! By finding new ways of directly studying the relevant Dirichlet series, *without* having to proceed via the zeros of $\zeta(s)$, our understanding of multiplicative functions has pulled ahead of our understanding of the primes.

Theorem 1.14 (Matomäki–Radziwiłł, 2014)**.** *If $h = h(X) \to \infty$ as $X \to \infty$, then*

$$\int_X^{2X} \Big| \sum_{x<n\leqslant x+h} \mu(n) \Big|^2 dx = o(Xh^2).$$

This is a stunning result, and surprised everyone when it was released. We will prove a (slightly weaker) form of this theorem towards the end of this course.

Theorem 1.15 (Matomäk–Teräväinen, 2019)**.** *If $\theta > 11/20$ then*

$$\frac{1}{X^\theta} \sum_{X<n\leqslant X+X^\theta} \mu(n) = o_\theta(1).$$

Note that $11/20 = 0.55 < 0.5833... = 7/12$, so this result is an improvement over the bound of Ramachandra from earlier. People outside the field might struggle to get quite so excited about this result – although it is exciting! – so we won't discuss it further in this course.

Theorem 1.16 (Tao, 2015)**.**

$$\frac{1}{\log X} \sum_{n\leqslant X} \frac{\mu(n)\mu(n+1)}{n} = o(1)$$

as $X \to \infty$.

This is Tao's 'logarithmically averaged Chowla conjecture' (note that the trivial bound for the left-hand side is 1). The proof uses pretty much everything that we will cover in this course, and a lot more besides; proving it will be the climax of this course.

2. Lecture 2: Mean values and 1-pretentious functions

From last lecture, we owe the reader a proof of Proposition 1.11, which gave some elementary equivalences of the primes number theorem.

Proof of Proposition 1.11. (2) $\Rightarrow$ (1): From Proposition 1.13, it suffices to show that

$$\sum_{n\leqslant X}(\Lambda-1)(n)=o(X).$$

We proceed to write the left-hand side as a Dirichlet convolution of functions involving μ. Indeed

$$\begin{aligned}\Lambda&=\mu\star\log\\ 1&=\mu\star 1\star 1=\mu\star\tau,\end{aligned}$$

so

$$\Lambda-1=\mu\star(-\tau+\log).$$

One may prove (see Example Sheet 1), that

$$\frac{1}{B}\sum_{b\leqslant B}(-\tau+\log)(b)=-2\gamma+O\Big(\frac{1}{\sqrt{B}}\Big),$$

where $\gamma=0.5772156649...$ is the *Euler–Mascheroni constant*. In particular

$$\frac{1}{B}\sum_{b\leqslant B}(-\tau+2\gamma+\log)(b)=O\Big(\frac{1}{\sqrt{B}}\Big).$$

In general it is always more convenient to work with fucntions that have asymptotic mean value 0 than otherwise, and so this motivates our writing

$$\begin{aligned}\Lambda-1&=\mu\star(-\tau+\log)\\ &=\mu\star(-\tau+2\gamma+\log)+\mu\star 2\gamma\\ &=\mu\star(-\tau+2\gamma+\log)+2\gamma\delta.\end{aligned}$$

For each of writing we define $f:=-\tau+2\gamma+\log$. Then

$$\begin{aligned}\sum_{n\leqslant X}(\Lambda-1)(n)&=O(1)+\sum_{ab\leqslant X}\mu(a)f(b)\\ &=O(1)+\sum_{ab\leqslant X:b\leqslant Y}\mu(a)f(b)+\sum_{ab\leqslant X:b>Y}\mu(a)f(b)\\ &:=O(1)+\Sigma_Y^{\leqslant}+\Sigma_Y^{>},\end{aligned}$$

where $Y\geqslant 1$ is a parameter to be chosen. We have

$$\Sigma_Y^{\leqslant}=\sum_{b\leqslant Y}f(b)\sum_{a\leqslant X/b}\mu(a)\leqslant\sum_{b\leqslant Y}|f(b)|o\Big(\frac{X}{b}\Big)=o_Y(X).$$

Furthermore

$$\Sigma_Y^{>}=\sum_{a\leqslant X/Y}\mu(a)\sum_{Y<b\leqslant X/a}f(b)\leqslant\sum_{a\leqslant X/Y}O\Big(\sqrt{\frac{X}{a}}\Big)=O\Big(\frac{X}{\sqrt{Y}}\Big).$$

If $Y=Y(X)\to\infty$ sufficiently slowly, we deduce that $\Sigma_Y^{\leqslant}+\Sigma_Y^{>}=o(X)$ as required.

The technique of splitting the divisors by a parameter Y is called the 'Dirichlet hyperbola method', and it is a ubiquitous idea in the field.

(1) $\Rightarrow$ (2): Since

$$\Lambda(n) = \sum_{d|n} \mu(d) \log(n/d) = (\log n) \sum_{d|n} \mu(d) - \sum_{d|n} \mu(d) \log d = -\sum_{d|n} \mu(d) \log d$$

we have the identity

$$\Lambda = -(1 \star \mu \log).$$

So

$$\mu \star (\Lambda - 1) = -\mu \log - \delta.$$

This identity is useful because it can be used to relate sums of the function $\Lambda - 1$ to sums of the Möbius function. Adding $\mu(n) \log X$ to both sides, summing over n, and rearranging, we get

$$\sum_{n \leqslant X} \mu(n) = -\frac{1}{\log X} \sum_{n \leqslant X} (\mu \star (\Lambda - 1))(n) + \frac{1}{\log X} \sum_{n \leqslant X} \mu(n) \log\Big(\frac{X}{n}\Big) - \frac{1}{\log X}. \qquad (1)$$

So it suffices to control the sums on the right-hand side.

By the triangle inequality,

$$\Big|\frac{1}{\log X} \sum_{n \leqslant X} \mu(n) \log\Big(\frac{X}{n}\Big)\Big| \leqslant \frac{1}{\log X} \sum_{n \leqslant X} \Big|\log\Big(\frac{X}{n}\Big)\Big| = O\Big(\frac{X}{\log X}\Big),$$

see Examples Sheet 1. Furthermore,

$$\Big|\frac{1}{\log X} \sum_{n \leqslant X} (\mu \star (\Lambda - 1))(n)\Big| \leqslant \frac{1}{\log X} \sum_{d \leqslant X} \Big|\sum_{n \leqslant X/d} (\Lambda - 1)(n)\Big| = \frac{1}{\log X} \sum_{d \leqslant X} g\Big(\frac{X}{d}\Big)$$

for some function g for which $g(x) = o(x)$, by PNT. This means that

$$\frac{1}{\log X} \sum_{d \leqslant X} g\Big(\frac{X}{d}\Big) = o(X),$$

(see Examples Sheet 1). Therefore, from (1), we derive

$$\sum_{n \leqslant X} \mu(n) = o(X),$$

which is (2).

The equivalence of (2) $\Leftrightarrow$ (3) is on Examples Sheet 1 (Hint: use the relation $\mu \star 1_{\square} = \lambda$). $\square$

Later on in this lecture we will need to make use of two elementary estimates.

Theorem 2.1 (Chebyschev)**.** *For all $X \geqslant 2$ one has*

$$X \ll \sum_{n \leqslant X} \Lambda(n) \ll X,$$

(and this may be proved in a simple elementary way).

Theorem 2.2 (Mertens)**.** *For all $X \geqslant 2$ one has*

$$\prod_{p \leqslant X} \Big(1 - \frac{1}{p}\Big) = (1 + o(1)) \frac{e^{-\gamma}}{\log X},$$

$$\sum_{p \leqslant X} \frac{1}{p} = \log\log X + O(1),$$

and

$$\sum_{p \leqslant X} \frac{\log p}{p} = \log X + O(1),$$

where γ is the Euler–Mascheroni constant (and these may be proved in a simple elementary way).

A few exercise on Examples Sheet 1 will lead you through proofs of these statements.

Proposition 1.11 related PNT to a computation of the asymptotic mean values of μ and λ. For the rest of this lecture (and for several following), we will investigate these mean values more systematically. The culmination of our efforts will be Halász's theorem, which provides a sharp characterisation of those multiplicative functions that have asymptotic mean value 0 (and will, amongst other things, imply the prime number theorem).

If $f : \mathbb{N} \longrightarrow \mathbb{C}$ is a function we will define

$$M_X(f) := \frac{1}{X}\sum_{n\leqslant X} f(n)$$

and

$$M_{X,\log}(f) := \frac{1}{\log X}\sum_{n\leqslant X}\frac{f(n)}{n}.$$

Let us start with a simple, but nonetheless instructive, example.

Lemma 2.3 (Density of squarefrees)**.**

$$M_X(\mu^2) = \frac{6}{\pi^2} + O(X^{-1/2}).$$

This main term, though seemingly exotic, is not really so surprising. Indeed, a number n is squarefree iff $p^2 \nmid n$ for all primes p. The proportion of numbers n for which $p^2 \nmid n$ is $(1 - 1/p^2)$, so, assuming independence, the proportion of numbers n for which $p^2 \nmid n$ for all primes p is

$$\prod_p \Big(1 - \frac{1}{p^2}\Big),$$

which is

$$\prod_p \Big(1 + \frac{1}{p^2} + \frac{1}{p^4} + \frac{1}{p^6} + \cdots\Big)^{-1} = \Big(\sum_{n\geqslant 1}\frac{1}{n^2}\Big)^{-1} = \frac{6}{\pi^2}$$

by a standard limit. Here is a rigorous proof:

Proof. Write $\mu^2 = 1 \star g$, where $g = \mu^2 \star \mu$, i.e.

$$g(n) = \begin{cases} \mu(m) & \text{if } n = m^2 \\ 0 & \text{otherwise.} \end{cases}$$

So

$$\sum_{n\leqslant X}\mu^2(n) = \sum_{d\leqslant X} g(d)\sum_{e\leqslant X/d} 1 = X\sum_{d\leqslant X}\frac{g(d)}{d} + O(\sum_{d\leqslant X}|g(d)|).$$

Since $|g| \leqslant 1_\square$, the error term is $O(X^{1/2})$. Now

$$\sum_{d\leqslant X}\frac{g(d)}{d} = \sum_{d\leqslant X^{1/2}}\frac{\mu(d)}{d^2} = \sum_{d\geqslant 1}\frac{\mu(d)}{d^2} + O(X^{-1/2}) = \prod_p\Big(1-\frac{1}{p^2}\Big) + O(X^{-1/2}) = \frac{6}{\pi^2} + O(X^{-1/2})$$

as above. □

Not all multiplicative functions have a well-defined asymptotic mean value. To show this, it will be useful for us to introduce an extremely widely-applicable estimation device known as 'partial summation'.

Lemma 2.4 (Partial summation)**.** *Let $F \in C^1([a_1, a_2])$ and let $f : \mathbb{N} \longrightarrow \mathbb{C}$ be a function. Then*

$$\sum_{a_1<n<a_2} F(n)f(n) = F(a_2) \sum_{a_1<n<a_2} f(n) - \int_{a_1}^{a_2} F'(t) \sum_{a_1<n<t} f(n)\, dt.$$

Proof. One has

$$\begin{aligned}\int_{a_1}^{a_2} F'(t) \sum_{a_1<n<t} f(n)\, dt &= \int_{a_1}^{a_2} F'(t) \sum_{a_1<n<a_2} f(n)1_{n<t}\, dt \\ &= \sum_{a_1<n<a_2} f(n) \int_{n}^{a_2} F'(t)\, dt \\ &= \sum_{a_1<n<a_2} f(n)(F(a_2) - F(n)),\end{aligned}$$

and then the lemma follows by rearranging. □

Lemma 2.5 (No asymptotic mean value)**.** *If $t \in \mathbb{R}$ and $X \geqslant 2$, then*

$$M_X(n \mapsto n^{it}) = \frac{X^{it}}{1+it} + O(X^{-1}(1 + |t| \log X)).$$

Proof. The result is trivial for $t = 0$, so we may assume that $t \neq 0$. Therefore, by partial summation,

$$\begin{aligned}\sum_{n\leqslant X} n^{it} = X^{1+it} - it \int_1^X y^{it-1} \sum_{n\leqslant y} 1\, dy &= X^{1+it} - it \int_1^X y^{it} dy + O(|t| \log X) \\ &= \frac{X^{1+it}}{1+it} + O\Big(\frac{|t|}{|1+it|} + |t| \log X\Big) \\ &= \frac{X^{1+it}}{1+it} + O(1 + |t| \log X)\end{aligned}$$

as claimed. □

In this case we see that

$$\lim_{X\to\infty} |M_X(n \mapsto n^{it})| = \frac{1}{|1+it|},$$

so in particular is well-defined. But the argument of $M_X(n \mapsto n^{it})$ changes with X, so there is no asymptotic mean value. (From now on, for notational ease, we will also use n^{it} to refer to the function $n \mapsto n^{it}$.

Here are the following broad questions that we will go some way to addressing:

(1) What is a good guess for $M_X(f)$ and for $M_{X,\log}(f)$?

(2) Under what conditions does this guess provably hold?

(3) When does $M_X(f) \to 0$ as $X \to \infty$?

Let us first consider how $M_X(f)$ and $M_{X,\log}(f)$ are related.

Lemma 2.6. *Let $f : \mathbb{N} \longrightarrow \mathbb{C}$ and suppose that $\lim_{X\to\infty} M_X(f) = C_f$ for some constant C_f. Then $\lim_{X\to\infty} M_{X,\log}(f) = C_f$ as well.*

Proof. This is just summation by parts. Indeed

$$\begin{aligned} M_{X,\log}(f) &= \frac{1}{\log X}\sum_{n\leqslant X}\frac{f(n)}{n} \\ &= \frac{1}{\log X}M_X(f) + \frac{1}{\log X}\int_1^X \frac{tM_t(f)}{t^2}\,dt \\ &= O_f(1/\log X) + \frac{C_f}{\log X}\int_1^X \frac{1}{t}\,dt + \frac{1}{\log X}\int_1^X \frac{o_{t\to\infty,f}(1)}{t}\,dt \\ &= C_f + o_{X\to\infty,f}(1) \end{aligned}$$

as required. (See Examples Sheet for the final estimation step). □

So proving an asymptotic formula for $M_{X,\log}(f)$ is strictly easier than proving an asymptotic formula for $M_X(f)$.

If $g \in \mathcal{M}$ is real-valued and $g(n) \geqslant 0$ for all n, we could use Rankin's trick to upper-bound $M_X(g)$ by $M_{X,\log}(g)$. Indeed

$$M_X(g) = \frac{1}{X}\sum_{n\leqslant X} g(n) \leqslant \frac{1}{X}\sum_{n\leqslant X} g(n)\frac{X}{n} \leqslant (\log X)M_{X,\log}(g).$$

However, by using the multiplicativity of g it turns out that one can greatly improve this bound.

Lemma 2.7. *Let $g \in \mathcal{M}_k$ and suppose that g is real-valued and $g(n) \geqslant 0$ for all n. Then*

$$M_X(g) \ll_k M_{X,\log}(g).$$

Proof when g is completely multiplicative and $g \in \mathcal{M}_0$. Since $1 \star \Lambda = \log$ we have

$$\sum_{n\leqslant X} g(n)\log n = \sum_{ab\leqslant X} g(ab)\Lambda(b) \leqslant \sum_{a\leqslant X} g(a)\sum_{b\leqslant X/a}\Lambda(b) \ll X\sum_{a\leqslant X}\frac{g(a)}{a},$$

using Chebyschev's elementary bound $\sum_{n\leqslant Y}\Lambda(n) \ll Y$. This yields

$$\log X\sum_{n\leqslant X} g(n) = \sum_{n\leqslant X} g(n)\log n + \sum_{n\leqslant X} g(n)\log\frac{X}{n} \ll X\sum_{a\leqslant X}\frac{g(a)}{a}$$

as $\log\frac{X}{n} \leqslant \frac{X}{n}$, which is the lemma. □

One notes that the proof actually works for all positive sub-multiplicative functions, i.e. for which $g(ab) \leqslant g(a)g(b)$. For example, the Euler φ function.

Proof for $g \in \mathcal{M}_k$. This is only moderately more complicated. The central idea is the same. Since $g(p) \leqslant k$, one has

$$\begin{aligned} \sum_{n\leqslant X} g(n)\log n &= \sum_{n\leqslant X} g(n)\sum_{p^r\| n}\log(p^r) \\ &= \sum_{p\leqslant X} g(p)\log p \sum_{\substack{n\leqslant X/p\\(n,p)=1}} g(n) + \sum_{r\geqslant 2}\sum_{p\leqslant X^{1/r}} g(p^r)\log(p^r)\sum_{\substack{n\leqslant X/p^r\\(n,p)=1}} g(n) \\ &\leqslant \sum_{p\leqslant X} g(p)\log p \sum_{n\leqslant X/p} g(n) + \sum_{r\geqslant 2}\sum_{p\leqslant X^{1/r}} g(p^r)\log(p^r)\sum_{n\leqslant X/p^r} g(n) \end{aligned}$$

The first term may be estimated as above, giving a bound

$$\leqslant \sum_{n\leqslant X} g(n) \sum_{p\leqslant X/n} g(p)\log p \ll kX \sum_{n\leqslant X} \frac{g(n)}{n}.$$

The second term, bounding trivially, gives

$$\leqslant X \sum_{r\geqslant 2} \sum_{p\leqslant X^{1/r}} \frac{\tau_k(p^r)\log(p^r)}{p^r} \sum_{n\leqslant X/p^r} \frac{g(n)}{n} \leqslant X \sum_{n\leqslant X} \frac{g(n)}{n} \Big(\sum_p \sum_{r\geqslant 2} \frac{\tau_k(p^r)\log(p^r)}{p^r} \Big)$$
$$\ll_k X \sum_{n\leqslant X} \frac{g(n)}{n}$$

as well. So

$$\sum_{n\leqslant X} g(n)\log n \ll_k X \sum_{n\leqslant X} \frac{g(n)}{n},$$

and then we may complete the proof as before. □

NB: By considering this proof, one may see the justification for why another class of multiplicative functions that was often considered, historically, was those positive $h \in \mathcal{M}$ for which there existed constant $p_1 > 0$ and $0 < p_2 < 2$ for which $h(p^m) \leqslant p_1 p_2^m$.

Next, let us consider what might be a good guess for $M_X(f)$, by trying to generalise the argument that we used for μ^2.

Observe how, with μ^2, we were able to write $\mu^2 = 1 \star g$ where $\sum_{n\leqslant X} |g(n)|$ was small, and this enabled us to compute the mean value. Proceeding in general then, writing $f = 1 \star g$ we would get

$$XM_X(f) = \sum_{d\leqslant X} g(d) \sum_{e\leqslant X/d} 1 = X \sum_{d\leqslant X} \frac{g(d)}{d} + O(\sum_{n\leqslant X} |g(d)|).$$

Hoping (!) that the error is small, we have a main term of

$$X \sum_{d\leqslant X} \frac{g(d)}{d} \approx X \prod_{p\leqslant X} \sum_{m\geqslant 0} \frac{g(p^m)}{p^m} = X \prod_{p\leqslant X} \Big(1-\frac{1}{p}\Big)\Big(1+\frac{f(p)}{p}+\frac{f(p^2)}{p^2}+\cdots\Big)$$
$$:= X\mathcal{P}(f;X).$$

So that's a first guess, and it works for $f = \mu^2$. The following is a generalisation, that we will spend the rest of the lecture proving.

The next theorem is just a formalised version of the above observations.

Theorem 2.8 (Wintner)**.** *Let $f \in \mathcal{M}_0$ be real-valued and suppose that*

$$\sum_p \frac{1-f(p)}{p} < \infty. \tag{2}$$

Then

$$M_X(f) = \mathcal{P}(f) + o_f(1),$$

where

$$\mathcal{P}(f) := \lim_{X\to\infty} \mathcal{P}(f;X)$$

is well-defined, and is only equal to zero if $f(2^i) = -1$ for all $i \geqslant 1$.

We will call such $f \in \mathcal{M}_0$ for which (2) holds *strongly* 1*-pretentious*, in the sense that the function f 'pretends' to be the constant function 1 on primes. Observe that μ^2 is strongly 1-pretentious, so Wintner's theorem immediately implies $M_X(\mu^2) = \frac{6}{\pi^2} + o(1)$ as in Lemma 2.3.

Before proceeding to the proof, we need one auxiliary result.

Exercise 2.9. *Show that if (a_n) is a sequence of complex numbers with $a_n \neq -1$ for all n, then $\prod_{n\geqslant 1}(1+a_n)$ converges absolutely, to a non-zero limit, if and onlf if $\sum_{n\geqslant 1}|a_n|$ converges.*

Proof. Writing $f = 1 \star g$ we have

$$M_X(f) = \frac{1}{X}\sum_{d\leqslant X} g(d) \sum_{e\leqslant X/d} 1 = \sum_{d\leqslant X}\frac{g(d)}{d} + O(M_X(|g|)). \tag{3}$$

Since $|g| \in \mathcal{M}_2$ we have

$$M_X(|g|) \ll M_{X,\log}(|g|) \ll \frac{1}{\log X}\sum_{n:s_X(n)=1}\frac{|g(n)|}{n} \ll \frac{1}{\log X}\prod_{p\leqslant X}\Big(1+\frac{|g(p)|}{p}+\sum_{m\geqslant 2}\frac{|g(p^m)|}{p^m}\Big).$$

Observe that $|g(p)| = |(\mu\star f)(p)| = |1-f(p)| = 1-f(p)$ since $f\in\mathcal{M}_0$ is real-valued. Since $\sum_p |g(p)|/p < \infty$ by assumption, and

$$\sum_p\sum_{m\geqslant 2}\frac{|g(p^m)|}{p^m} < \infty$$

since $g\in\mathcal{M}_2$, we have that

$$\prod_{p\leqslant X}\Big(1+\frac{|g(p)|}{p}+\sum_{m\geqslant 2}\frac{|g(p^m)|}{p^m}\Big)$$

converges, and therefore the error term in (3) is $O((\log X)^{-1})$.

This same analysis shows that $\sum_{d\geqslant 1}\frac{g(d)}{d}$ converges absolutely, and therefore

$$\sum_{d>X}\frac{g(d)}{d} = o(1).$$

Hence, following on from (3), we have

$$M_X(f) = \sum_{d\geqslant 1}\frac{g(d)}{d} + o(1) = \prod_p\sum_{m\geqslant 0}\frac{g(p^m)}{p^m} + o(1) = \prod_p\Big(1-\frac{1}{p}\Big)\Big(\sum_{m\geqslant 0}\frac{f(p^m)}{p^m}\Big) + o(1),$$

since $g = \mu\star f$, where the Euler product converges absolutely.

So $\lim_{X\to\infty}\mathcal{P}(f;X) = \mathcal{P}(f)$, and moreover $\mathcal{P}(f)=0$ if and only if

$$\sum_{m\geqslant 0}\frac{f(p^m)}{p^m} = 0$$

for some prime p. However, note that

$$\sum_{m\geqslant 0}\frac{f(p^m)}{p^m} = 1+\sum_{m\geqslant 1}\frac{f(p^m)}{p^m} \geqslant 1-\sum_{m\geqslant 1}\frac{1}{p^m} = 1-\frac{1}{p-1} \geqslant 0,$$

with equality if and only if $p=2$ and $f(2^m)=-1$ for all $m\geqslant 1$. This settles the theorem. □

NB: Note that the multiplicative function $f(n) = (-1)^{n-1}$ has zero mean value because of this theorem.

Delange generalised this result to all $f \in \mathcal{M}_0$.

Next, we shall consider positive multiplicative functions again and prove a result in the opposite direction.

Theorem 2.10. *Let $g \in \mathcal{M}_k$ and assume that g is real-valued and $g(n) \geqslant 0$ for all n. Then*

$$M_X(g) \ll_k \exp\Big(-\sum_{p\leqslant X}\frac{1-g(p)}{p}\Big).$$

Therefore, if $g \in \mathcal{M}_0$ too and

$$\sum_p \frac{1-g(p)}{p} = \infty$$

then $M_X(g) = o(1)$.

Proof. We know from Lemma 2.7 that $M_X(g) \ll_k M_{X,\log}(g)$. So

$$M_{X,\log}(g) = \frac{1}{\log X}\sum_{n\leqslant X}\frac{g(n)}{n} \ll \prod_{p\leqslant X}\Big(1-\frac{1}{p}\Big)\Big(1+\frac{g(p)}{p}+\sum_{m\geqslant 2}\frac{g(p^m)}{p^m}\Big)$$

by Merten's theorem. Now

$$\Big(1-\frac{1}{p}\Big)\Big(1+\frac{g(p)}{p}+\sum_{m\geqslant 2}\frac{g(p^m)}{p^m}\Big) = 1-\frac{1-g(p)}{p}+O_k(p^{-3/2}),$$

so from the Taylor series of $\log(1-x)$ we get

$$\log(M_{X,\log}(g)) = O_k(1) - \sum_{p\leqslant X}\frac{1-g(p)}{p}.$$

Thus

$$M_{X,\log}(g) \ll_k \exp\Big(-\sum_{p\leqslant X}\frac{1-g(p)}{p}\Big).$$

This gives the theorem. □

In summary then, if $0 \leqslant g(n) \leqslant 1$ for all n then we have a satisfactory theory for calculating the asymptotics of $M_X(g)$, involving the convergence or otherwise of $\sum_p (1-g(p))/p$.

Wirsing generalised this result to all real valued $f \in \mathcal{M}_0$.

Theorem 2.11 (Wirsing)**.** *Let $f \in \mathcal{M}_0$ be real-valued. If*

$$\sum_p \frac{1-f(p)}{p} = \infty$$

then $M_X(f) = o(1)$.

This theorem is clearly much deeper than anything we have done in this lecture, as by putting $f = \mu$ one immediately deduces PNT. Unsuprisingly, then, Wirsing uses PNT at a critical point in his proof.

However, we do already have the tools to prove the same theorem for logarithmic averages (and this result will be the one we will need later on when considering Dirichlet characters). We may even work with complex valued functions.

Lemma 2.12 (Logarithmic averages of non-pretentious functions)**.** *Let $f \in \mathcal{M}_0$. Then*

$$M_{X,\log}(f) \ll \exp\Big(\frac{1}{2}\sum_{p\leqslant X}\frac{1-\Re f(p)}{p}\Big).$$

Proof. Write $f \star 1 = g$, where $g \in \mathcal{M}_2$. Then

$$\sum_{n\leqslant X} g(n) = \sum_{d\leqslant X} f(d) \sum_{m\leqslant X/d} 1 = X\sum_{d\leqslant X}\frac{f(d)}{d} + O(X).$$

Therefore

$$M_{X,\log}(f) = \frac{1}{\log X}M_X(g) + O\Big(\frac{1}{\log X}\Big) \leqslant \frac{1}{\log X}M_X(|g|) + O\Big(\frac{1}{\log X}\Big).$$

By Lemma 2.7, we have

$$\frac{1}{\log X}M_X(|g|) \ll \frac{1}{\log X}M_{X,\log}(|g|) \leqslant \frac{1}{\log^2 X}\prod_{p\leqslant X}\sum_{m=0}^{\infty}\frac{|g(p^m)|}{p^m}.$$

But this is equal to

$$\frac{1}{\log^2 X}\prod_{p\leqslant X}\Big(1+\frac{|f(p)+1|}{p}+O\Big(\frac{1}{p^{3/2}}\Big)\Big) \ll \prod_{p\leqslant X}\Big(1-\frac{1}{p}\Big)^2\Big(1+\frac{|f(p)+1|}{p}\Big).$$

This is

$$\ll \prod_{p\leqslant X}\Big(1-\frac{2-|f(p)+1|}{p}\Big) \ll \exp\Big(-\sum_{p\leqslant X}\frac{2-|f(p)+1|}{p}\Big)$$

since $1-x \leqslant e^{-x}$ for $x \geqslant 0$ (or use Taylor expansion of $\log(1-x)$).

One has the inequalities

$$\frac{1}{2}(1-\Re z) \leqslant 2-|z+1| \leqslant 1-\Re z$$

for all $|z| \leqslant 1$ (exercise), and so we have the final upper bound of

$$\exp\Big(-\frac{1}{2}\sum_{p\leqslant X}\frac{1-\Re f(p)}{p}\Big).$$

The $O(1/\log X)$ error can also be absorbed into the implied constant, since $\sum_{p\leqslant X}\frac{1}{p} \sim \log\log X$. □

For these real valued functions, one sees from this whole lecture that it is important to know 'how closely f pretends to be 1' on the primes, in order to determine the nature of $M_X(f)$. Next time, we will introduce a general way of describing and manipulating this notion of 'closeness'.

3. Lecture 3: Granville–Soundararajan distance

Just before we turn to the central notion of the lecture, and indeed the entire course – Granville–Soundararajan distance – we feel that is important for us to introduce another small concept.

The function Λ_f

Let $f \in \mathcal{M}_k$ have Dirichlet series $F(s)$ (which has abscissa of absolute convergence at least 1). Define the function $\Lambda_f : \mathbb{N} \longrightarrow \mathbb{C}$ by

$$f \star \Lambda_f = f\log,$$

i.e. such that

$$\sum_{n\geqslant 1}\frac{\Lambda_f(n)}{n^s}=-\frac{F'(s)}{F(s)}.$$

The function $\Lambda_f(n)$ is a generalisation of the von Mangoldt function $\Lambda(n)$ (which, in this notation, is $\Lambda_1(n)$).

Exercise 3.1.

- Show that Λ_f is supported on prime powers, and that $\Lambda(p)=f(p)\log p$.
- Find an example of $f\in\mathcal{M}_2$ for which $\Lambda_f(n)$ grows exponentially in n.

The condition $|\Lambda_f(n)|\leqslant\kappa\Lambda(n)$ is another common one to see imposes on multiplicative functions. One may prove Lemma 2.7 very easily, say, under this assumption. We will refer to Λ_f at various points on the examples sheets.

For the rest of this lecture, we will be investigating the basic properties of the following function.

Definition 3.2 (Granville–Soundararajan distance)**.** *Let $f,g\in\mathcal{M}_0$ and $X\geqslant 2$. We then define the (Granville–Soundararajan) distance between f and g to be*

$$\mathbb{D}(f,g;X):=\Big(\sum_{p\leqslant X}\frac{1-\Re(f(p)\overline{g(p)})}{p}\Big)^{\frac12}.$$

We let $\mathcal{D}(f,g;\infty):=\lim_{X\to\infty}\mathcal{D}(f,g;X)$.

Related objects, which will occasionally be more natural to consider, are

$$\mathbb{D}_\alpha(f,g):=\Big(\sum_{p}\frac{1-\Re(f(p)\overline{g(p)})}{p^\alpha}\Big)^{1/2},$$

$$\mathbb{D}^*(f,g;X):=\Big(\sum_{k\geqslant 1}\sum_{p\leqslant X^{1/k}}\frac{1-\Re(f(p^k)\overline{g(p^k)})}{p}\Big)^{1/2},$$

and

$$\mathbb{D}^*_\alpha(f,g):=\Big(\sum_{k\geqslant 1}\sum_{p}\frac{1-\Re(f(p^k)\overline{g(p^k)})}{p^{k\alpha}}\Big)^{1/2}.$$

As we will presently show, all these quantities differ by at most $O(1)$, so they may be thought of as essentially the same notions.

Lemma 3.3 (Easy properties of G–S distance)**.** *Let $f,g\in\mathcal{M}_0$. Then*

- $\mathbb{D}(f,g;X)\leqslant\sqrt{2\log\log X}+O(1)$
- $\mathbb{D}(f,g;X)=\mathbb{D}(g,f;X)=\mathbb{D}(1,\overline{f}g;X)$
- *when $\alpha=1+\frac{1}{\log X}$ we have $|\mathbb{D}(f,g;X)-\mathbb{D}^*_\alpha(f,g)|=O(1)$.*

Proof. The first of these points is immediate from Mertens' theorems, and the second is immediate from the definition, so we turn to the third point. Note that

$$\sum_{k\geqslant 2}\sum_{p}\frac{1-\Re(f(p^k)\overline{g(p^k)})}{p^{k(1+1/\log X)}}\ll\sum_{p}\sum_{k\geqslant 2}\frac{1}{p^k}\ll\sum_{p}\frac{1}{p^2}\ll 1,$$

so it remains to show that $|\mathbb{D}(f,g;X)-\mathbb{D}_\alpha(f,g)|=O(1)$ when $\alpha=1+\frac{1}{\log X}$. Letting $a_p:=1-\Re(f(p)\overline{g(p)})$, observe that both

$$\sum_{p>X}\frac{a_p}{p^{1+1/\log X}}\ll\sum_{n>X}\frac{1}{n^{1+1/\log X}}\ll X^{-\frac{1}{\log X}}\ll 1$$

and

$$\sum_{p\leqslant X} a_p\Big(\frac{1}{p}-\frac{1}{p^{1+1/\log X}}\Big) \ll \sum_{p\leqslant X}\frac{p^{1/\log X}-1}{p^{1+\log X}} \ll \sum_{p\leqslant X}\frac{\log p}{p\log X} \ll 1$$

by Mertens estimates. This settles the lemma. □

Armed with these bounds, let us introduce some vocabulary with which we can talk about functions $f, g \in \mathcal{M}_0$ and the distance $\mathbb{D}$.

We will say that f is *weakly g pretentious* if $\mathbb{D}(f,g;X)^2 = o(\log\log X)$, and that f is (strongly) *g pretentious* if $\mathbb{D}(f,g;\infty)<\infty$.

Lemma 3.4 (Relation to Dirichlet series)**.** *If $f\in\mathcal{M}_0$ and $F(s):=\sum_{n\geqslant 1} f(n)/n^s$, then if $X\geqslant 2$ we have*

$$\Big|F\Big(1+\frac{1}{\log X}+it\Big)\Big| \asymp (\log X)\Big|\sum_{m\geqslant 0}\frac{f(2^m)}{2^{m(1+1/\log X+it)}}\Big| \times \exp(-\mathbb{D}(f,n^{it};X)^2).$$

The fact that one has to deal with the prime 2 separately is really just a technical annoyance. If $f(2^m)=0$ for $m\geqslant 2$ (if $|f|=\mu^2$, say), or if $f(2^m)=f(2)^m$ (if f is completely multiplicative, say) then we recover the more natural-looking relationship

$$\Big|F\Big(1+\frac{1}{\log X}+it\Big)\Big| \asymp (\log X)\exp(-\mathbb{D}(f,n^{it};X)^2).$$

In any circumstances, we always have

$$\Big|F\Big(1+\frac{1}{\log X}+it\Big)\Big| \ll (\log X)\exp(-\mathbb{D}(f,n^{it};X)^2).$$

Before proving the lemma, let us just review a couple of aspects of the theory of the complex logarithm when applied to $F(s)$. We know that $F(s)$ is a holomorphic function on the simply connected domain $\Re s>1$. By the convergence of the Euler product there, we know that $F(s)=0$ if and only if $1+\sum_{k\geqslant 1} f(p^k)/p^{ks}=0$ for some prime p. But since $f\in\mathcal{M}_0$ we have (writing $\sigma=\Re s$)

$$\Big|\sum_{k\geqslant 1}\frac{f(p^k)}{p^{ks}}\Big| \leqslant \sum_{k\geqslant 1}\frac{1}{p^{k\sigma}} = \frac{1}{p^\sigma-1} < 1$$

if $\sigma>1$ and $p\geqslant 2$, so we conclude that $F(s)\neq 0$ for $\Re s>1$.

This means that we can define the principal branch of the logarithm $\operatorname{Log} F(s)$ by, writing $s=\sigma+it$,

$$\operatorname{Log} F(s) = \int_{\Gamma_s}\frac{F'(z)}{F(z)}\,dz + \log F(\sigma),$$

where Γ_s is a straight-line contour from σ to $\sigma+it$ and log is the usual real logarithm.

Now, for any branch of the complex logarithm in general it is only the case that $\operatorname{Log}(z_1z_2) = \operatorname{Log}(z_1)+\operatorname{Log}(z_2)$ modulo $2\pi i$, since

$$\operatorname{Log} z = \log|z| + i\arg(z).$$

However, for us we have that $\operatorname{Log} F(s)$ and $\sum_p \operatorname{Log}(\sum_{m\geqslant 0} f(p^m)/p^{ms})$ are both holomorphic functions on $\{\Re s>1\}$ that agree on the real axis and always differ by a multiple of $2\pi i$. So

$$\log F(s) - \sum_p \operatorname{Log}\Big(\sum_{m\geqslant 0}\frac{f(p^m)}{p^{ms}}\Big)$$

is a continuous function from a connected set $\{\Re s > 1\}$ into a discrete set $2\pi i\mathbb{Z}$, so it must be constant (and is therefore identically 0). Hence

$$\operatorname{Log} F(s) = \sum_p \operatorname{Log}\Big(\sum_{m\geqslant 0}\frac{f(p^m)}{p^{ms}}\Big).$$

Proof of Lemma 3.4. By the theory of the complex logarithm we have

$$\Big|F\Big(1+\frac{1}{\log X}+it\Big)\Big| = \exp\Big(\Re\operatorname{Log} F\Big(1+\frac{1}{\log X}+it\Big)\Big).$$

Furthermore

$$\begin{aligned}\operatorname{Log} F\Big(1+\frac{1}{\log X}+it\Big) - \operatorname{Log}\Big(\sum_{m\geqslant 0}\frac{f(2^m)}{2^{m(1+\frac{1}{\log X}+it)}}\Big) &= \sum_{p\geqslant 3}\operatorname{Log}\Big(1+\sum_{m\geqslant 1}\frac{f(p^m)}{p^{m(1+\frac{1}{\log X}+it)}}\Big)\\ &= \sum_{p\geqslant 3}\sum_{m\geqslant 1}\frac{f(p^m)}{p^{m(1+\frac{1}{\log X}+it)}} + O(1)\end{aligned}$$

by the Taylor expansion for $\operatorname{Log}(1+x)$ when $|x|<1$, since

$$\sum_{p\geqslant 3}\sum_{k\geqslant 2}\frac{(-1)^{k-1}}{k}\Big(\sum_{m\geqslant 1}\frac{f(p^m)}{p^{m(1+\frac{1}{\log X}+it)}}\Big)^k \leqslant \sum_{p\geqslant 3}\sum_{k\geqslant 2}\frac{1}{(p-1)^k} \leqslant \sum_{p\geqslant 3}\frac{1}{(p-1)(p-2)} = O(1).$$

Note how we had to remove the prime 2 from this calculation in order to end up with a bounded quantity!

Continuing, we get

$$\begin{aligned}&= \sum_p\sum_{m\geqslant 1}\frac{f(p^m)p^{-itm}}{p^{m(1+\frac{1}{\log X})}} + O(1)\\ &= \log\log X + \sum_p\sum_{m\geqslant 1}\frac{-1+f(p^m)p^{-itm}}{p^{m(1+\frac{1}{\log X})}} + O(1),\end{aligned}$$

by the fact that

$$\sum_p\frac{1}{p^{1+1/\log X}} = \sum_{p\leqslant X}\frac{1}{p} + O(1) = \log\log X + O(1),$$

as in the proof of the preceding lemma. So

$$\exp\Big(\Re\operatorname{Log} F\Big(1+\frac{1}{\log X}+it\Big) - \Re\operatorname{Log}\Big(\sum_{m\geqslant 0}\frac{f(2^m)}{p^{m(1+\frac{1}{\log X}+it)}}\Big)\Big)$$

is

$$\begin{aligned}&\asymp \log X\exp\Big(\sum_{p\geqslant 3}\sum_{m\geqslant 1}\frac{-1+f(p^m)p^{-itm}}{p^{m(1+\frac{1}{\log X})}}\Big)\\ &\asymp \log X\exp(-\mathbb{D}_\alpha(f,n^{it})^2)\\ &\asymp \log X\exp(-\mathbb{D}(f,n^{it};X)^2)\end{aligned}$$

by the previous lemma. This settles our claim. □

Now we come to the critical point, namely that $\mathbb{D}$ is a pseudometric on $\mathcal{M}_0$, i.e. it satisfies a triangle inequality.

Lemma 3.5 (Triangle inequality)**.** *Let $f,g,h\in\mathcal{M}_0$ and $X\geqslant 2$. Then*

$$\mathbb{D}(f,g;X)+\mathbb{D}(g,h;X)\geqslant\mathbb{D}(g,h;X).$$

In words, this says that 'if f pretends to be g and g pretends to be h, then f pretends to be h'.

Proof. When $|f(p)|, |g(p)|, |h(p)| = 1$ for all primes p, the proof is easy. Indeed, in this case

$$1 - \Re(f(p)\overline{g(p)}) = \frac{1}{2}|f(p) - g(p)|^2,$$

so the triangle inequality is simply the triangle inequality for the normed space $\ell^2(m)$, where m is the atomic measure supported on primes $p \leqslant X$ with $m(p) = 1/2p$ for all $p \leqslant X$. Concretely,

$$\frac{1}{2}\sum_{p\leqslant X}\frac{|f(p) - h(p)|^2}{p}$$

is equal to

$$\frac{1}{2}\sum_{p\leqslant X}\frac{|f(p) - g(p)|^2 + |g(p) - h(p)|^2 + 2\Re(f(p) - g(p))\overline{(g(p) - h(p))}}{p}$$
$$\leqslant \frac{1}{2}\sum_{p\leqslant X}\frac{|f(p) - g(p)|^2 + |g(p) - h(p)|^2 + 2|f(p) - g(p)||g(p) - h(p)|}{p}$$
$$\leqslant \frac{1}{2}\sum_{p\leqslant X}\frac{|f(p) - g(p)|^2}{p} + \frac{1}{2}\sum_{p\leqslant X}\frac{|g(p) - h(p)|^2}{p} + \Big(\sum_{p\leqslant X}\frac{|f(p) - g(p)|^2}{p}\Big)^{1/2}\Big(\sum_{p\leqslant X}\frac{|g(p) - h(p)|^2}{p}\Big)^{1/2}$$

by Cauchy–Schwarz. This inequality is exactly

$$\mathbb{D}(f, h; X)^2 \leqslant \mathbb{D}(f, g; X)^2 + \mathbb{D}(g, h; X)^2 + 2\mathbb{D}(f, g; X)\mathbb{D}(g, h; X) = (\mathbb{D}(f, g; X) + \mathbb{D}(g, h; X))^2,$$

as required.

Now, for general $f, g, h \in \mathcal{M}_0$ one has the following drop-dead gorgeous proof due to Tao. For every $u \in \mathbb{C}$ with $|u| \leqslant 1$, there are points u_1, u_2 on the unit circle such that u is the midpoint of the chord with endpoints u_1, u_2. (If u is on the unit circle, take $u_1 = u_2$). Then, if $|u|, |z| \leqslant 1$,

$$\frac{1}{8}\sum_{i,j\leqslant 2}|u_i - z_j|^2 = \frac{1}{4}\sum_{i,j\leqslant 2}(1 - \Re(u_i\overline{z_j})) = 1 - \Re\Big(\frac{1}{2}(u_1 + u_2), \frac{1}{2}(z_1 + z_2)\Big) = 1 - \Re u\overline{z}.$$

Applying this with $u = f(p)$, $z = g(p)$, one may deduce the triangle inequality for $\mathbb{D}$ from another ℓ^2 triangle inequality. □

Lemma 3.6 (Multiplication inequality)**.** *If $f_1, f_2, g_1, g_2 \in \mathcal{M}_0$, then*

$$\mathbb{D}(f_1, g_1; X) + \mathbb{D}(f_2, g_2; X) \geqslant \mathbb{D}(f_1f_2; g_1g_2; X).$$

In words, this is saying 'if f_1 pretends to be g_1, and f_2 pretends to be g_2, then f_1f_2 pretends to be g_1g_2'.

Proof. This follows from the triangle inequality, via

$$\begin{aligned}\mathbb{D}(f_1, g_1; X) + \mathbb{D}(f_2, g_2; X) &= \mathbb{D}(1, \overline{f_1}g_1; X) + \mathbb{D}(\overline{g_2}f_1, 1; X) \geqslant \mathbb{D}(\overline{f_1}g_1, \overline{g_2}f_2; X)\\ &= \mathbb{D}(\overline{f_1f_2}, \overline{g_1g_2}; X) = \mathbb{D}(f_1f_2, g_1g_2; X).\end{aligned}$$

□

Lemma 3.7 (Special case)**.** *If $t \in \mathbb{R}$ and $X \geqslant 2$ then*

$$\mathbb{D}(1, n^{it}; X)^2 \geqslant \begin{cases}\log(1 + |t|\log X) - O(1) & \text{when } |t| \leqslant 1\\ \log\log X - \log\log(|t| + 2) - O(1) & \text{when } |t| \geqslant 1.\end{cases}.$$

Proof. These bounds are a consequence of the elementary theory of the Riemann zeta function. Indeed, for any $N \in \mathbb{N}$ and s with $\Re s > 1$ one has the approximation

$$\zeta(s) = \sum_{n=1}^{N} \frac{1}{n^s} + \frac{N^{1-s}}{s-1} - s \int_N^\infty \frac{\{y\}}{y^{s+1}}\, dy,$$

where $\{y\}$ denotes the fractional part. (This follows by partial summation.) Then, if $|s-1| \gg 1$ we have

$$|\zeta(s)| \ll \sum_{n=1}^{N} \frac{1}{n^{\Re s}} + N^{1-\Re s} + |s| \int_N^\infty \frac{1}{y^{\Re s+1}}\, dy.$$

Choosing $N = \lceil |s| + 1 \rceil$ and using the fact that $\Re s > 1$ we get

$$|\zeta(s)| \ll \log(|s|+2) + O(1) + \frac{|s|}{|s|+1} \ll \log(|s|+2)$$

in this range. By a similar analysis taking $N = 1$, one obtains the bound

$$|\zeta(s)| \ll \frac{1}{|s-1|}$$

if $|s-1| \ll 1$.

By Lemma 3.4,

$$\Big|\zeta(1 + \frac{1}{\log X} + it)\Big| \asymp \log X \exp(-\mathbb{D}(1, n^{it}; X)^2).$$

The lemma then follows by inserting the above bounds on $|\zeta(1 + \frac{1}{\log X} + it)|$. □

Our final lemma (for now!) shows that a single $f \in \mathcal{M}_0$ cannot simultaneously pretend to be $n^{i\alpha}$ and $n^{i\beta}$ for two different values of α and β.

Lemma 3.8 (Repulsion/uniqueness for n^{it})**.** *Let $f \in \mathcal{M}_0$, and let α, β be two real numbers with $\delta := |\alpha - \beta|$. Then*

$$(\mathbb{D}(f, n^{i\alpha}; X) + \mathbb{D}(f, n^{i\beta}; X))^2 \geqslant \begin{cases} \log(1 + \delta \log X) - O(1) & \text{if } \delta \leqslant 1; \\ \log\log X - \log\log(2+\delta) - O(1) & \text{if } \delta \geqslant 1. \end{cases}$$

Proof. By the triangle inequality we have

$$(\mathbb{D}(f, n^{i\alpha}; X) + \mathbb{D}(f, n^{i\beta}; X))^2 \geqslant \mathbb{D}(n^{i\alpha}, n^{i\beta}; X)^2 = \mathbb{D}(1, n^{i(\beta-\alpha)}; X)^2$$

and this can be lower-bounded by Lemma 3.7. The lemma then follows. □

Corollary 3.9. *Let $f \in \mathcal{M}_0$ be real-valued, and suppose that there is a real α for which*

$$\mathbb{D}(f, n^{it}; X)^2 = o_t(\log\log X).$$

Then $t = 0$.

In other words, a real-valued $f \in \mathcal{M}_0$ cannot be weakly n^{it}-pretentious for any complex multiplicative phase n^{it}.

Proof. Suppose that there is a $t \neq 0$ for which $\mathbb{D}(f, n^{it}; X)^2 = o_t(\log\log X)$. Since f is real-valued, we have $\mathbb{D}(f, n^{-it}; X) = \mathbb{D}(f, n^{it}; X) = o_t(\log\log X)$. Taking X large enough, this contradicts Lemma 3.8 for $\alpha = t$ and $\beta = -t$. □

We saw in the previous lecture that $\mathbb{D}(f,1;X)$ controlled the behaviour of the average $M_X(f)$ when $f \in \mathcal{M}_0$ was positive and real-valued, and controlled $M_{X,\log}(f)$ for all $f \in \mathcal{M}_0$. The theorem of Wirsing (which we quoted but didn't prove), stated that $\mathbb{D}(f,1;X)$ controlled $M_X(f)$ for all $f \in \mathcal{M}_0$.

In the 1960s Halász proved a vast generalisation of all of these results.

Theorem 3.10 (Halász's Theorem, qualitative version)**.** *Let $f \in \mathcal{M}_0$.*

(1) Suppose that there exists $t \in \mathbb{R}$ for which $\mathbb{D}(f, n^{it}; \infty) < \infty$. Then

$$M_X(f) = (1+o(1))\frac{X^{it}}{1+it}\prod_{p\leqslant X}\Big(1-\frac{1}{p}\Big)\Big(1+\frac{f(p)}{p^{1-it}}+\frac{f(p^2)}{p^{2-2it}}+\cdots\Big)$$

as $X \to \infty$.

(2) Suppose that $\mathbb{D}(f, n^{it}; \infty) = \infty$ for all $t \in \mathbb{R}$. Then

$$M_X(f) = o(1)$$

as $X \to \infty$.

This is a real dichotomy, because, although the infinite product

$$\prod_{p\leqslant X}\Big(1-\frac{1}{p}\Big)\Big(1+\frac{f(p)}{p^{1-it}}+\frac{f(p^2)}{p^{2-2it}}+\cdots\Big)$$

needn't converge, one can show (see Examples Sheet 1) that it never diverges to 0.

Lectures 4 and 5 will be devoted to the proof of this theorem, and to some corollaries.

To finish this lecture, we will show how the notion of $\mathbb{D}(f,g;X)$ can be used to recover the classical argument for showing a zero-free region for ζ.

Proof that $\zeta(\rho) \neq 0$ if $\Re\rho = 1$. For all $\gamma \in \mathbb{R}$, by the triangle inequality we have

$$\mathbb{D}(1, n^{2i\gamma}; X) = \mathbb{D}(n^{-i\gamma}, n^{i\gamma}; X) \leqslant \mathbb{D}(n^{-i\gamma}, \mu; X) + \mathbb{D}(\mu, n^{i\gamma}; X) = 2\mathbb{D}(\mu, n^{i\gamma}; X).$$

Now, suppose that $\gamma_0 \neq 0$ and that $\zeta(1+i\gamma_0) = 0$. In that case $1/\zeta(s)$ has a pole at $s = 1+i\gamma_0$, and therefore by considering the Laurent series of $\frac{1}{\zeta(s)}$ around the point $s = 1 + i\gamma_0$, if X is large enough we have

$$\Big|\frac{1}{\zeta(1+\frac{1}{\log X}+i\gamma_0)}\Big| \gg_{\gamma_0} \log X.$$

Therefore, by Lemma 3.4 (for the function $f = \mu$), we have

$$\log X \ll \log X \exp(-\mathbb{D}(\mu, n^{i\gamma_0}; X)^2).$$

So

$$\mathbb{D}(\mu, n^{i\gamma_0}; X) = O_{\gamma_0}(1),$$

i.e. μ strongly pretends to by $n^{i\gamma_0}$. By the above inequalities we have

$$\mathbb{D}(1, n^{2i\gamma_0}; X) = O_{\gamma_0}(1).$$

But then

$$\Big|\zeta\Big(1+\frac{1}{\log X}+i\gamma_0\Big)\Big| \asymp (\log X)\exp(-\mathbb{D}(1, n^{2i\gamma_0}; X)^2) \gg_{\gamma_0} \log X,$$

so $\zeta(s)$ has a pole at $s = 1 + 2i\gamma_0$, which it doesn't. □

The general philosophy of this proof (that if $\zeta(1+i\gamma_0)=0$ then $n^{i\gamma_0}$ points towards -1 a lot of the time, and so $n^{2i\gamma_0}$ points towards 1 a lot of them) was known to Hadamard. But $\mathbb{D}(f,g;X)$, with its triangle inequality, packages this observation extremely conveniently.

In a usual first analytic number theory course, the standard identity (due to Mertens) that is used to prove that $\zeta(s)$ has no zeros on the 1-line is

$$\zeta(\sigma)^3|\zeta(\sigma+it)|^4|\zeta(\sigma+2it)| \geqslant 1$$

when $\sigma > 1$. One can derive this inequality from the triangle inequality for $\mathbb{D}$ too (exercise).

4. Lecture 4: Preliminaries for Halász's Theorem, or 'The analytic number theorist's tool box'

In this lecture we will begin (but probably not finish) going through some of the preliminary results that we will use in the course of proving Halász's theorem. These results in themselves dance across various tools of the analytic number theorist's trade, and so they have some value in their own right (and not just as lemmas on the way to proving the first big theorem of these notes).

A simple sieve

The central idea in the modern proof of Halász's theorem is a decomposition of the sum $\sum_{n\leqslant X} f(n)$ as a certain triple sum, which then enables a factorisation of the associated Dirichlet series as a triple product. To effect such a decomposition, it will be useful to assume that the summation variable n possessed prime factors in various helpful intervals, i.e. to show that

$$\sum_{n\leqslant X} 1_{(n,\prod_{p\in I} p)=1}$$

is small, where I is the interval in which we hope to find a prime factor.

This is a sieving problem, and it can be attacked by any of the standard small sieves (Selberg sieve, β-sieve). However, to show some of the unexpected power of the estimates on the averages of multiplicative functions that we have proved already, let me present a rather different proof.

Lemma 4.1 (A simple sieve)**.** *For all $X \geqslant 1$, and for all $m \in \mathbb{N}$ uniformly,*

$$|\{n \leqslant X : (n,m) = 1\}| \ll X \prod_{\substack{p|m\\ p\leqslant X}} \Big(1 - \frac{1}{p}\Big).$$

Proof. Let $f \in \mathcal{M}_0$ be the completely multiplicative function given by

$$f(n) = \begin{cases} 1 & \text{if } (n,m) = 1\\ 0 & \text{otherwise.}\end{cases}$$

Then, as we have already established in Lemma CITE,

$$M_X(f) \ll M_{X,\log}(f) \ll \prod_{p\leqslant X} \Big(1 - \frac{1}{p}\Big)\Big(1 - \frac{f(p)}{p}\Big)^{-1}$$

by Mertens' theorem. But this is equal to

$$\prod_{\substack{p|m\\ p\leqslant X}} \Big(1 - \frac{1}{p}\Big).$$

as required. □

This lemma implies the following result (which is what we will actually use).

Lemma 4.2 (Conveniently placed prime factors)**.** *Let $\varepsilon > 0$. Then*

$$\sum_{\substack{n\leqslant X\\ \exists p\in[X^{\varepsilon},X] \text{ s.t. } p|n\\ \exists p\in[X^{\varepsilon^2},X^{\varepsilon}] \text{ s.t. } p|n}} 1 = X - O(\varepsilon X).$$

In other words, at least a proportion $1 - O(\varepsilon)$ of integers $n \leqslant X$ have a prime factor in the range $[X^{\varepsilon}, X]$ and in the range $[X^{\varepsilon^2}, X^{\varepsilon}]$.

Proof. This follows from the above lemma together with the estimate

$$\prod_{X^\varepsilon \leqslant p \leqslant X} \Big(1 - \frac{1}{p}\Big) \sim \frac{\log(X^\varepsilon)}{\log X} = \varepsilon.$$

□

The Brun–Titchmarsh inequality

How many prime numbers should there be in the interval $[Y, Y+X)$? Well, 'on average' one would expect around $X/\log Y$ primes (as the density of primes at scale Y is around $1/\log Y$). However, if we believe the Hardy–Littlewood k-tuples conjecture (which is a generalisation of the twin prime conjecture), then sometimes the interval $[Y, Y+X)$ should have as many as $X/\log X$ primes in it. It is a major (and difficult!) open problem to show that

$$\sup_Y \sum_{p \in [Y, Y+X)} 1 \leqslant (1+o(1))\frac{X}{\log X}$$

as $X \to \infty$. (Solving this problem would, among other things, rule out the existence of Siegel zeros). However, the Brun–Titchmarsh inequality shows that this conjecture is true up to a constant multiplicative factor.

Theorem 4.3 (Brun–Titchmarsh inequality, $q = 1$ version)**.** *For $X \geqslant 2$ we have*

$$\sup_Y \sum_{p \in [Y, Y+X)} 1 \ll \frac{X}{\log X}.$$

The best known value of the implied constant is 2, which was proved by Montgomery and Vaughan in the 1970s. Replacing 2 by 1.999 would already rule out the existence of Siegel zeros. There are similar results for primes in arithmetic progressions with common difference q, but we won't need those in this course.

Note that when Y is large compared to X then the Brun–Titchmarsh inequality cannot be obtained even by assuming RH, as that would only give an upper bound of $O_\varepsilon(Y^{1/2+\varepsilon})$.

The proof is again via sieve methods, which we will not expose in full generality here. However, we can present a proof using the Selberg sieve weights, emphasising the aspects that are to do with estimating sums of multiplicative functions.

Proof. First, we note that if $Y \leqslant X$ the desired bound follows from Chebyschev's estimates, so henceforth we may assume that $Y \geqslant X$. Now let $z = X^u$ for some small constant u, and for $d \leqslant z$ define the weight

$$\rho_d = \frac{d\mu(d)}{L(z)\varphi(d)} \sum_{\substack{q \leqslant z/d \\ (q,d)=1}} \frac{\mu^2(q)}{\varphi(q)},$$

where $L(z) := \sum_{q \leqslant z} \frac{\mu^2(q)}{\varphi(q)}$. We will not explain so thoroughly why ρ_d is a good weight to use here, although we will see in the course of the calculation that it has certain seemingly magical properties. In the full treatment of the Selberg sieve, this weight is derived as a solution to a certain quadratic optimisation problem. One may show that $\rho_d \approx \mu(d)\log(z/d)/\log z$ for small d, so one can think of it as an 'arithmetically smoothed' Möbius fucntion.

We note that $\rho_1 = 1$, and so if $Y \geqslant X$ say we have

$$\sum_{p \in [Y, Y+X)} 1 \leqslant \sum_{Y \leqslant n < Y+X} \Big(\sum_{\substack{d \mid n \\ d \leqslant z}} \rho_d\Big)^2,$$

since the only contribution to the sum when n is prime comes from $d = 1$. Expanding out the square and rearranging, we get an upper bound of

$$\sum_{d_1,d_2\leqslant z}\rho_{d_1}\rho_{d_2}\sum_{\substack{Y\leqslant n<Y+X\\ d_1,d_2|n}}1 = X\sum_{d_1,d_2\leqslant z}\frac{\rho_{d_1}\rho_{d_2}}{[d_1,d_2]}+O\Big(\Big(\sum_{d\leqslant z}|\rho_d|\Big)^2\Big),$$

where $[d_1, d_2]$ is the least common multiple of d_1 and d_2.

Bounding crudely (one can do better here with some effort) we have

$$|\rho_d|\leqslant\frac{d}{\varphi(d)}=\prod_{p|d}\Big(1-\frac{1}{p}\Big)^{-1}\leqslant\prod_{p\leqslant z}\Big(1-\frac{1}{p}\Big)^{-1}\ll\log z\leqslant\log X,$$

so the error term is $O(X^{2u}\log^2 X)$. Regarding the main term, since $(d_1, d_2)[d_1, d_2] = d_1d_2$ and $1\star\varphi = \mathrm{id}$, we have

$$\begin{aligned}\sum_{d_1,d_2\leqslant z}\frac{\rho_{d_1}\rho_{d_2}}{[d_1,d_2]}&=\sum_{d_1,d_2\leqslant z}\frac{\rho_{d_1}\rho_{d_2}}{d_1d_2}(d_1,d_2)\\&=\sum_{d_1,d_2\leqslant z}\frac{\rho_{d_1}\rho_{d_2}}{d_1d_2}\sum_{\substack{u|d_1\\u|d_2}}\varphi(u)\\&=\sum_{u\leqslant z}\varphi(u)\Big(\sum_{\substack{d\leqslant z\\u|d}}\frac{\rho_d}{d}\Big)^2.\end{aligned}$$

Analysing the inner sum above, by the multiplicativity of μ and φ we get

$$\begin{aligned}\sum_{\substack{d\leqslant z\\u|d}}\frac{\rho_d}{d}&=\frac{\mu(u)}{L(z)\varphi(u)}\sum_{\substack{d\leqslant z/u\\(d,u)=1}}\frac{\mu(d)}{\varphi(d)}\sum_{\substack{q\leqslant z/du\\(q,du)=1}}\frac{\mu^2(q)}{\varphi(q)}\\&=\frac{\mu(u)}{L(z)\varphi(u)}\sum_{\substack{n\leqslant z/u\\(n,u)=1}}\mu^2(n)\Big(\frac{\mu}{\varphi}\star\frac{\mu^2}{\varphi}\Big)(n),\end{aligned}$$

since if n is square free and coprime to u, and if $qd = n$, then $(d, u) = 1$, $(q, du) = 1$, and $\mu^2(d) = \mu^2(q) = 1$. But vice versa, if $(d, u) = 1$, $(q, du) = 1$, and $\mu^2(d) = \mu^2(q) = 1$, then if $qd = n$ we have $\mu^2(n) = 1$ and $(n, u) = 1$.

By direct calculation,

$$\mu^2(n)\Big(\frac{\mu}{\varphi}\star\frac{\mu^2}{\varphi}\Big)(n)=\delta(n).$$

So

$$\sum_{\substack{d\leqslant z\\u|d}}\frac{\rho_d}{d}=\frac{\mu(u)}{L(z)\varphi(u)},$$

and thus

$$\sum_{d_1,d_2\leqslant z}\frac{\rho_{d_1}\rho_{d_2}}{[d_1,d_2]}=\frac{1}{L(z)^2}\sum_{u\leqslant z}\frac{\mu^2(u)}{\varphi(u)}=\frac{1}{L(z)}.$$

So

$$\sum_{Y\leqslant p<Y+X}1\leqslant\frac{X}{L(z)}+O(X^{2u}\log^2 X).$$

If we had time to introduce some more sophisticated techniques for estimating sums of multiplicative functions we would be able to show that

$$L(z)\sim\log z.$$

As it is we may satisfy ourselves with the estimate

$$L(z)=\sum_{q\leqslant z}\frac{\mu^2(q)}{\varphi(q)}=\sum_{q\leqslant z}\frac{\mu^2(q)}{q}\prod_{p|q}\Big(1-\frac{1}{p}\Big)^{-1}=\sum_{q\leqslant z}\frac{\mu^2(q)}{q}\prod_{p|q}\Big(1+\frac{1}{p}+\frac{1}{p^2}+\cdots\Big)\geqslant\sum_{q\leqslant z}\frac{1}{q}\gg\log z$$

if $z\geqslant 2$.

Hence finally we have

$$\sum_{Y\leqslant p<Y+X}1\ll\frac{X}{\log z}+O(X^{2u}\log^2 X)\ll\frac{X}{\log X}$$

if we choose $u=1/3$, say. □

The same proof gives the following more precise estimate, which is what we will actually use later.

Theorem 4.4. *Let $\varepsilon>0$, $X\geqslant 2$, and let $S(X^\varepsilon)$ denote $\{n\in\mathbb{N}:p|n\Rightarrow p>X^\varepsilon\}$. Then*

$$\sup_{Y}\sum_{\substack{n\in S(X^\varepsilon)\\ Y\leqslant n<Y+X}}1\ll\varepsilon^{-O(1)}\frac{X}{\log X}.$$

Proof. Pick $\varepsilon=u$ in the above proof. □

Some facts about additive Fourier transforms

I mentioned in the preamble to these notes that we would assume some basic familiarity with the Fourier transform. However, it is always prudent to go over a few of the fundamental notions again – if only to fix normalisations!

Definition 4.5 (Fourier transform)**.** *If $f:\mathbb{R}\longrightarrow\mathbb{C}$ and $f\in L^1(\mathbb{R})$, then we define the Fourier transform $\widehat{f}:\mathbb{R}\longrightarrow\mathbb{C}$ to be the function*

$$\widehat{f}(\omega)=\int_{-\infty}^{\infty}f(x)e^{-2\pi i\omega x}\,dx.$$

This might not be the normalisation that you are used to, but it tends to be the more commonly used normalisation in analytic number theory, i.e we use phase functions $e^{2\pi i\omega x}$ rather than $e^{i\omega x}$.

Regarding the Fourier inversion formula, the following is a general result (but one that can be hard to apply).

Theorem 4.6 (Fourier inversion for L^1 functions)**.** *If f is continuous such that both $f,\widehat{f}\in L^1(\mathbb{R})$, then*

$$f(x)=\int_{-\infty}^{\infty}\widehat{f}(\omega)e^{2\pi ix\omega}\,d\omega.$$

As far as possible in this course we will work with 'nice' functions, e.g. Schwarz functions, which are smooth function all of whose derivatives vanish faster than any polynomial. In fact, a lot of the time we will work with the even smaller class of compactly supported smooth functions.

Theorem 4.7 (Existence of bump functions)**.** *The function*

$$g(x)=\begin{cases}e^{-\frac{1}{1-x^2}} & \text{if } |x|<1\\ 0 & \text{if } |x|\geqslant 1\end{cases}$$

is smooth, and supported on $[-1,1]$.

Proof. Exercise. □

These functions are pleasant to work with because their Fourier transforms are rapidly decaying; in fact the Fourier transform is an automorphism of the space of Schwartz functions.

Theorem 4.8 (Fourier transform of Schwarz functions)**.** *Define the Schwarz class $\mathcal{S}$ to be the set of all functions $f \in C^\infty(\mathbb{R})$ for which for all $j, m \in \mathbb{Z}_{\geqslant 0}$, $|f^{(j)}(x)| = o_{f,j,m}(|x|^{-m})$ as $|x| \to \infty$. Then $\widehat{f} \in \mathcal{S}$ too, and the Fourier inversion formula holds.*

Sketch proof. The function f and all its derivatives lie in $L^1(\mathbb{R})$. So integrating by parts j times yields $|\widehat{f}(\omega)| \ll_{j,f} (1 + |\omega|)^{-j}$. Taking $j \geqslant 2$ then we have $\widehat{f} \in L^1(\mathbb{R})$ too, so Fourier inversion holds.

Differentiating under the integral definition of $\widehat{f}(\omega)$, and then applying integration by parts as above, one may show that $\widehat{f} \in \mathcal{S}$. □

If $f, g \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, say, we define the additive convolution $(f * g) \in L^1(\mathbb{R})$ by the formula

$$(f * g)(x) = \int_{-\infty}^{\infty} f(x-y)g(y)\,dy.$$

It is a standard lemma that

$$\widehat{f \star g} = \widehat{f}\widehat{g}.$$

It will be occasionally convenient to use some properties of the Fourier transforms of specific functions.

Lemma 4.9 (Continuous Fejér kernel)**.** *If*

$$f(x) = \begin{cases} 1 - |x| & \text{if } |x| \leqslant 1 \\ 0 & \text{if } |x| \geqslant 1, \end{cases}$$

then

$$\widehat{f}(\omega) = \Big(\frac{\sin \pi\omega}{\pi\omega}\Big)^2.$$

Proof. Exercise. □

Lemma 4.10 (Fourier transform of an interval)**.** *Let $I \subset \mathbb{R}$ be any finite closed interval, with indicator function 1_I and length $|I|$. Then*

$$|\widehat{1_I}(\omega)| \ll \min(|I|, |\omega|^{-1}).$$

Proof. We always have

$$|\widehat{1_I}| = \Big|\int_{-\infty}^{\infty} 1_I(x)e^{-2\pi i x\omega}\,dx\Big| \leqslant \int_{-\infty}^{\infty} |1_I(x)|\,dx = |I|.$$

Furthermore, letting $I = [a, b]$, we get

$$|\widehat{1_I}| = \Big|\int_a^b e^{-2\pi i\omega x}\,dx\Big| = \frac{1}{2\pi|\omega|}|e^{-2\pi i\omega b} - e^{-2\pi i\omega a}| \ll |\omega|^{-1}.$$

□

A first mean value theorem for Dirichlet polynomials

We have already seen the relevance, in our discussion of Granville–Soundararajan distance, the relevance of the size of Dirichlet series $|F(1+\frac{1}{\log X}+it)|$. In more sophisticated arguments – like the proof of Halász's inequality to come, but most especially when we are considering shorter averages of multiplicative functions – one will also want to consider the mean value of objects such as $|F(1+\frac{1}{\log X}+it)|$ taken over a range of values of t.

We begin by introducing the prototypical result in the theory of such mean values.

Theorem 4.11 (Montgomery)**.** *For any coefficients $a_n \in \mathbb{C}$, we have*

$$\int_{-T}^{T} \Big| \sum_{n\leqslant N} \frac{a_n}{n^{it}} \Big|^2 dt = (2T + O(N)) \sum_{n\leqslant N} |a_n|^2.$$

How should one think of this result? Well, suppose that all $|a_n| \approx 1$. Then square-root cancellation in the sum $\sum_{n\leqslant N} a_n n^{-it}$ for every $t \in [-T, T]$ would yield an upper bound of $O(TN)$, i.e. $O(T\sum_{n\leqslant N}|a_n|^2)$. However a single unit interval $t \in [t_0, t_0 + 1)$ on which no cancellation occurs, i.e. on which $|\sum_{n\leqslant N} a_n n^{-it}| \approx N$, would yield an upper bound of N^2. This can of course happen, say when $a_n = n^{-it_0}$ for some fixed t_0. So, at this level of generality, Montgomery's mean value theorem is the best we can hope for. As the course progresses, we will see increasingly sophisticated approaches for handling mean-values of Dirichlet polynomials, which can be used to improve upon Montgomery's estimate in more specific situations, e.g. where the support of the coefficient sequence (a_n) is sparse.

Proof. One line summary of the proof: "introduce a smoother majorant for $1_{[-T,T]}$ into the integral, and then expand the square".

Let $\Psi \in C^\infty(\mathbb{R})$ be a smooth non-negative bump function supported on $[-1, 1]$, with $\int \Psi = 1$. Let $\Psi_N(x) := \frac{1}{N}\Psi(x/N)$. Then $\int \Psi_N = 1$ and Ψ_N is supported on $[-N, N]$.

Then consider the function

$$g_{N,T} := 1_{[-T-N,T+N]} * \Psi_N.$$

By construction we have

$$1_{[-T,T]}(x) \leqslant g_{N,T}(x) \leqslant 1_{[-T-2N,T+2N]},$$

$\widehat{g}_{N,T}(0) = 2T + 2N$ and

$$|\widehat{g_{N,T}}(\omega)| = |\widehat{1}_{[-T-N,T+N]}(\omega)||\widehat{\Psi_N}(\omega)| \ll |\omega|^{-1}|\widehat{\Psi}(N\omega)| \ll N^{-1}|\omega|^{-2},$$

say. (It may help to draw a sketch of the function $g_{N,T}$.)

Then

$$\int_{-T}^{T} \Big| \sum_{n\leqslant N} \frac{a_n}{n^{it}} \Big|^2 t \leqslant \int_{-\infty}^{\infty} g(t) \Big| \sum_{n\leqslant N} \frac{a_n}{n^{it}} \Big|^2 dt = \sum_{1\leqslant m,n\leqslant N} a_m \overline{a_n} G_{N,T}\Big(\frac{m}{n}\Big),$$

where

$$G_{N,T}(x) = \int_{-\infty}^{\infty} g_{N,T}(t) x^{-it}\, dt.$$

But $G_{N,T}(x) = \widehat{g_{N,T}}((\log x)/2\pi)$. Therefore, separating out the terms with $m = n$ from the others, we have an upper bound of

$$\widehat{g}_{N,T}(0) \sum_{n\leqslant N} |a_n|^2 + O\Big(\frac{1}{N} \sum_{1\leqslant m\neq n\leqslant N} \frac{|a_m||a_n|}{|\log(m/n)|^2}\Big).$$

We know $\widehat{g}_{N,T}(0) = 2T + 2N$, and also for $1 \leqslant m \neq n \leqslant N$ one has (by the mean value theorem)

$$|\log(m/n)| = |\log m - \log n| \gg \frac{|m-n|}{m+n} \gg \frac{|m-n|}{N}.$$

Therefore, we may derive an overall upper bound of

$$\begin{aligned}&(2T+2N)\sum_{n\leqslant N}|a_n|^2 + O\Big(N\sum_{1\leqslant m\neq n\leqslant N}\frac{|a_m||a_n|}{|m-n|^2}\Big)\\ &= (T+2N)\sum_{n\leqslant N}|a_n|^2 + O\Big(N\sum_{n\leqslant N}|a_n|^2\sum_{\substack{m\leqslant N\\ m\neq n}}\frac{1}{|m-n|^2}\Big),\end{aligned}$$

since $|a_m||a_n| \leqslant |a_n|^2+|a_m|^2$. The inner sum is convergent, and thus the theorem follows. □

In certain textbook handlings one is much more explicit with the smoothing used, for instance letting

$$g_{N,T}(t) = \begin{cases} 0 & \text{if } t \leqslant -N\\ 1+\frac{t}{N} & \text{if } -N < t \leqslant 0\\ 1 & \text{if } 0 < t \leqslant T\\ 1-\frac{t-T}{N} & \text{if } T < t \leqslant T+N\\ 0 & \text{if } t > T+N.\end{cases}$$

But in the handling above I wanted to emphasise how essentially any smoothing at scale N is adequate for the theorem.

5. Lecture 5: Proof of Halász's Theorem

Let me begin this lecture with an important acknowledgement: in this lecture we will be following some notes of Terence Tao rather closely (254A Notes 10, from his blog).

A more precise mean value theorem

We mentioned last time how Montgomery's mean value theorem is best possible in full generality. However, there are various other tricks that one can employ to obtain stronger mean value estimates in cases where one knows something more about the coefficients.

We start with a simple lemma for relating arithmetic sums with Dirichlet polynomials.

Lemma 5.1. *If $f, g : \mathbb{N} \longrightarrow \mathbb{C}$ are any functions with finite support, and $\psi : \mathbb{R} \longrightarrow \mathbb{C}$ is continuous function such that $\psi, \widehat{\psi} \in L^1(\mathbb{R})$, then*

$$\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\frac{f(n)}{n}\overline{\frac{g(m)}{m}}\psi((\log m-\log n)/2\pi)=\int_{-\infty}^{\infty}\sum_{n=1}^{\infty}\frac{f(n)}{n^{1+it}}\overline{\sum_{m=1}^{\infty}\frac{g(m)}{m^{1+it}}}\widehat{\psi}(t)\,dt.$$

Proof. Swapping the orders of summation on the right-hand side (which is allowed by Fubini's theorem), the proposition would follow from the identity

$$\psi((\log m-\log n)/2\pi)=\int_{-\infty}^{\infty}\widehat{\psi}(t)(m/n)^{it}\,dt.$$

But this is just the Fourier inversion formula. □

Using this lemma we can obtain the following mean value estimate.

Lemma 5.2 (Mean values/ sums over intervals)**.** *If $f : \mathbb{N}$ is a function with finite support, and $T \geqslant 10$, then*

$$\int_{-T}^{T}\Big|\sum_{n=1}^{\infty}\frac{f(n)}{n^{1+it}}\Big|^2\,dt\ll\sum_{d=1}^{\infty}\frac{1}{d}\Big|\frac{T}{d}\sum_{m:|m-d|\leqslant 100d/T}|f(m)|\Big|^2.$$

Proof. Let

$$\psi(x)=\begin{cases}\frac{\pi^2}{4}(1-|x|) & \text{if } |x|\leqslant 1\\ 0 & \text{if } |x|\geqslant 1.\end{cases}$$

Then, using the explicit Fourier transform pairs that we stated last lecture, we have

$$\widehat{\psi}(\omega)=\frac{\pi^2}{4}\Big(\frac{\sin\pi\omega}{\pi\omega}\Big)^2.$$

So $\widehat{\psi}(\omega)\geqslant 1_{[-1/2,1/2]}(\omega)$ for all $\omega\in\mathbb{R}$. Therefore

$$\int_{-T}^{T}\Big|\sum_{n=1}^{\infty}\frac{f(n)}{n^{1+it}}\Big|^2\,dt\leqslant\int_{-\infty}^{\infty}\Big|\sum_{n=1}^{\infty}\frac{f(n)}{n^{1+it}}\Big|^2\widehat{\psi}(t/2T)\,dt.$$

Apply the previous lemma with $g=f$ and the function $t\mapsto\widehat{\psi}(t/2T)$. We obtain

$$2T\sum_{n=1}^{\infty}\frac{f(n)}{n}\sum_{m=1}^{\infty}\frac{\overline{f(m)}}{m}\psi((\log m-\log n)T/\pi).$$

Since ψ is supported on $[-1,1]$, there is only a contribution from pairs m,n with $|\log(m/n)|\leqslant\pi/T$, which means that there is only a contribution from pairs m,n with

$$|m-n|\leqslant\pi(m+n)/T$$

(by the mean value theorem applied to log).

So one already has a restriction to pairs m, n which lie close together. The rest of proof involves massaging this observation into the form stated in the lemma (which will be particularly convenient for our application).

We have an upper bound of

$$\begin{aligned}
&\ll T \sum_{n,m:|n-m|\leqslant \pi(m+n)/T} \frac{|f(n)||f(m)|}{nm} \\
&\ll T \sum_{n,m:|n-m|\leqslant \pi(m+n)/T} \frac{|f(n)||f(m)|}{nm} \frac{T}{(m+n)} \sum_{\substack{d \\ |d-n|\leqslant \pi(m+n)/T \\ |d-m|\leqslant \pi(m+n)/T}} \\
&\ll T^2 \sum_{d=1}^{\infty} \sum_{\substack{m,n \\ |m-d|\leqslant \pi(m+n)/T \\ |n-d|\leqslant \pi(m+n)/T}} \frac{|f(n)||f(m)|}{nm(m+n)}.
\end{aligned}$$

Since $T \geqslant 10$ we have that all of $n, m, (m+n), d$ have the same order of magnitude, and this yields an upper bound of

$$\ll T^2 \sum_{d=1}^{\infty} \frac{1}{d^3} \sum_{\substack{n:|n-d|\leqslant 100d/T \\ m:|m-d|\leqslant 100d/T}} |f(n)||f(m)|.$$

This gives the lemma. □

We are ready to prove Halász's Theorem. However, there is a dilemma in exposition here. The swiftest way to prove the result would be to quote a standard device in analytic number theory known as Perron's formula. However, this formula uses Fourier inversion as applied to a discontinuous function, which some people (including me) find to be distasteful when other approaches exist that deal with Schwartz functions throughout.

So, I will be presenting a version that does not appeal to Perron's formula. Though we will have to expend a small amount of sweat in bandlimiting a certain weight function, I strongly believe this to be conceptually more satisfying than the alternative.

Proof of Halász's theorem. Recall that we have a function $f \in \mathcal{M}_0$ and we assume that for all $t \in \mathbb{R}$ we have $\mathbb{D}(f, n^{it}; \infty) = \infty$.

As it stands this statement has no uniformity in $|t|$, and so we begin with some general analysis in order to derive for free some such weak uniformity. We claim that there exists some (possibly very slowly growing) function $T = T(X)$ for which

$$\lim_{X\to\infty} \inf_{|t|\leqslant T(X)} \mathbb{D}(f, n^{it}; X) = \infty.$$

Indeed, let us first establish that for all fixed T we have $\lim_{X\to\infty} \inf_{|t|\leqslant T} \mathbb{D}(f, n^{it}; X) = \infty$. Were this not so, then there would exist some $K > 0$, some sequence $(t_m)_{m=1}^{\infty}$ with $t_m \in [-T, T]$, and some sequence $X_m \to \infty$ as $m \to \infty$ for which $\mathbb{D}(f, n^{it_m}; X_m) \leqslant K$ for all m. By passing to a subsequence, we may assume that the sequence t_m converges, say to t_*. Since $\mathbb{D}(f, n^{it_*}; X) \to \infty$ as $X \to \infty$, there is some X_* for which $\mathbb{D}(f, n^{it_*}; X) > 2K$ if $X \geqslant X_*$. But by continuity we have $\mathbb{D}(f, n^{it_m}; X_*) \to \mathbb{D}(f, n^{it_*}; X_*)$ as $m \to \infty$. Yet, if m is large enough so that $X_m > X_*$, we have

$$\mathbb{D}(f, n^{it_m}; X_*) \leqslant \mathbb{D}(f, n^{it_m}; X_m) \leqslant K.$$

So

$$\lim_{m\to\infty} \mathbb{D}(f, n^{it_m}; X_*) \leqslant K,$$

contradicting the definition of X_*.

Now, we show how to construct $T(X)$. For each $n \in \mathbb{N}$, let X_n denote the threshold such that $\inf_{|t|\leqslant n} \mathbb{D}(f, n^{it}; X) \geqslant n$ for all $X \geqslant X_n$. Then for $X \in [X_n, X_{n+1})$ define $T(X) = n$. Then $T(X) \to \infty$ as $X \to \infty$ and $\inf_{|t|\leqslant T(X)} \mathbb{D}(f, n^{it}; X) \geqslant n$ for all $X \geqslant X_n$. So we have proved our claim.

Next we describe the decomposition of the function f that is central to the argument. Let $\varepsilon > 0$ be another parameter to be chosen (it will tend to zero very slowly), and write

$$f_{small} = \begin{cases} f(n) & \text{if } p|n \Rightarrow p \in [1, X^{\varepsilon^2}] \\ 0 & \text{otherwise,} \end{cases}$$

$$f_{med} = \begin{cases} f(n) & \text{if } n \in [2, X] \text{ and } p|n \Rightarrow p \in (X^{\varepsilon^2}, X^{\varepsilon}] \\ 0 & \text{otherwise,} \end{cases}$$

and

$$f_{large} = \begin{cases} f(n) & \text{if } n \in [2, X] \text{ and } p|n \Rightarrow p \in (X^{\varepsilon}, X] \\ 0 & \text{otherwise,} \end{cases}$$

and let

$$f_{split} = f_{small} \star f_{med} \star f_{large}.$$

Note that $f_{small} \in \mathcal{M}_0$ too, but $f_{med}, f_{large}, f_{split} \notin \mathcal{M}_0$ owing to their values at 1.

We know that $f(n) = f_{split}(n)$ if $2 \leqslant n \leqslant X$ and n has at least one prime factor in $(X^{\varepsilon^2}, X^{\varepsilon}]$ and at least one prime factor in $(X^{\varepsilon}, X]$. We established last lecture that all but $O(\varepsilon X)$ of the natural numbers $n \leqslant X$ have this property. So, we have

$$M_X(f) = M_X(f_{split}) + O(\varepsilon).$$

Furthermore f_{split} is supported on X-friable numbers, and so $\sum_{n\geqslant 1} |f_{split}(n)|/n$ is convergent (as it has a finite Euler product).

Now

$$M_X(f_{split}) = \sum_{1\leqslant n\leqslant X} \frac{f_{split}(n)}{n} \psi(\log n - \log X),$$

where $\psi(u) = e^u 1_{[-\infty,0]}(u)$. We are going to truncate on the Fourier side at height $T = T(X)$, at a small cost in physical space. To that end, let W be a Schwarz function with $\int W = 1$ such that $\widehat{W}$ is a smooth bump function supported in $[-1, 1]$. Let $W_T(x) = TW(Tx)$, and define

$$\psi_T = \psi * W_T.$$

Draw a rough picture of this function, if it helps you! This trick is very similar to the trick we used to prove Montgomery's mean value theorem last lecture. The idea is that W_T is concentrated on the set $|x| \leqslant 1/T$, which corresponds (by the uncertainty principle) to the fact that $\widehat{W_T}(\omega)$ is concentrated on the set $|\omega| \leqslant T$.

Since $\int W_T = 1$ one may calculate

$$\begin{aligned} |(\psi - \psi_T)(u)| &\leqslant T \int_{-\infty}^{\infty} |\psi(u - v) - \psi(u)||W(Tv)|\, dv \\ &\leqslant \int_{-\infty}^{\infty} |\psi(u - \frac{v}{T}) - \psi(u)||W(v)|\, dv. \end{aligned}$$

We always have $|\psi - \psi_T(u)| \ll 1$. Furthermore, if $u \geqslant 1/T$ we have

$$|\psi - \psi_T(u)| \ll \int_{Tu}^{\infty} |W(v)| \ll (Tu)^{-10}.$$

If $-2 \leqslant u \leqslant 1/T$ we use the Lipschitz constant bound $|e^{u-vT^{-1}} - e^u| \ll vT^{-1}$ to obtain the bound $|\psi - \psi_T(u)| \ll T^{-1}$. Finally, if $u < -2$ we get $|\psi - \psi_T(u)| \ll T^{-1}e^{u/2} \ll T^{-1}|u|^{-10}$.

All told, this means that $\sum_{1 \leqslant n \leqslant X} \frac{f_{split}(n)}{n} \psi(\log n - \log X)$ is

$$= \sum_{n \geqslant 1} \frac{f_{split}(n)}{n} \psi_T(\log n - \log X) + E_1 + E_2 + E_3 + E_4,$$

where E_1, E_2, E_3, and E_4 are four error terms, according to the various ranges of n.

All of the error terms may be shown to be $O(1/T)$. To give the gory details, E_1 corresponds to $n \geqslant Xe^{1/T}$, so

$$E_1 \ll \sum_{n \geqslant Xe^{1/T}} \frac{1}{nT^{10}(\log n - \log X)^{10}} \ll \sum_{k=1}^{\infty} \sum_{Xe^{k/T} \leqslant n \leqslant Xe^{(k+1)/T}} \frac{1}{nk^{10}} \ll \frac{1}{T}.$$

E_2 corresponds to $Xe^{-1/T} \leqslant n \leqslant Xe^{1/T}$, and we simply get

$$E_2 \ll \sum_{Xe^{-1/T} \leqslant n \leqslant Xe^{1/T}} \frac{1}{n} \ll \frac{1}{T}.$$

E_3 corresponds to $Xe^{-2} \leqslant n \leqslant X$, and we get

$$E_3 \ll \sum_{Xe^{-2} \leqslant n \leqslant Xe^{-1/T}} \frac{1}{n} \cdot \frac{1}{T} \ll \frac{1}{T}.$$

The final range E_4 corresponds to $1 \leqslant n \leqslant Xe^{-2}$, and we have the bound

$$E_4 \ll \frac{1}{T} \sum_{1 \leqslant n \leqslant Xe^{-2}} \frac{1}{n(\log X - \log n)^{10}} \ll \frac{1}{T} \sum_{k=2}^{\infty} \sum_{Xe^{-k-1} \leqslant n \leqslant Xe^{-k}} \frac{1}{nk^{10}} \ll \frac{1}{T}.$$

$$O\Big(\sum_{n \geqslant 1} \frac{1}{n} \cdot \frac{1}{(1 + T|\log n - \log X|)^{100}} \Big).$$

So

$$\sum_{1 \leqslant n \leqslant X} \frac{f_{split}(n)}{n} \psi(\log n - \log X) = \sum_{n \geqslant 1} \frac{f_{split}(n)}{n} \psi_T(\log n - \log X) + O(1/T).$$

Then, applying the Fourier inversion theorem to ψ_T, we have

$$\begin{aligned} \sum_{n \geqslant 1} \frac{f_{split}(n)}{n} \psi_T(\log n - \log X) &= \sum_{n \geqslant 1} \frac{f_{split}(n)}{n} \int_{-\infty}^{\infty} \widehat{\psi_T}(t) e^{2\pi i (\log n - \log X)t} \, dt \\ &= \int_{-T}^{T} \Big(\sum_{n \geqslant 1} \frac{f_{split}(n)}{n^{1-2\pi it}} \Big) X^{-2\pi it} \widehat{\psi_T}(t) \, dt, \end{aligned}$$

since $\widehat{\psi_T}(t) = \widehat{\psi}\widehat{W_T}$ and by construction $\widehat{W_T}(\omega) = \widehat{W}(\omega/T)$ is supported on $[-T, T]$. Also observe that we may use Fubini's theorem to swap the orders of summation and integration, since $\sum_{n \leqslant 1} |f_{split}(n)|/n$ converges.

Bounding $|\widehat{\psi_T}(t)| \ll 1$, and using $|X^{-2\pi it}| = 1$, we get an upper bound of

$$O\Big(\int_{-T}^{T} \Big|\sum_{n\geqslant 1} \frac{f_{split}(n)}{n^{1-it}}\Big|\, dt\Big).$$

This is

$$\ll \int_{-T}^{T} \Big|\sum_{n\geqslant 1} \frac{f_{small}(n)}{n^{1-it}}\Big| \Big|\sum_{n\geqslant 1} \frac{f_{med}(n)}{n^{1-it}}\Big| \Big|\sum_{n\geqslant 1} \frac{f_{large}(n)}{n^{1-it}}\Big|\, dt$$

$$\ll \sup_{|t|\leqslant T} |F_{small}(1+it)| \Big(\int_{-T}^{T} |F_{med}(1+it)|^2\, dt\Big)^{1/2} \Big(\int_{-T}^{T} |F_{large}(1+it)|^2\, dt\Big)^{1/2}$$

by the Cauchy-Schwarz inequality, were F_* denotes the Dirichlet series associated with the relevant functions.

From previous remarks we have

$$\begin{aligned}\sup_{|t|\leqslant T} |F_{small}(1+it)| &\ll (\log X^{\varepsilon^2}) \exp(-\inf_{|t|\leqslant T} \mathbb{D}(f, n^{it}; X^{\varepsilon^2}))\\ &\ll \varepsilon^2 (\log X) \exp(-\inf_{|t|\leqslant T} \mathbb{D}(f, n^{it}; X)) \exp(-2\log\varepsilon)\\ &\ll (\log X) \exp(-\inf_{|t|\leqslant T} \mathbb{D}(f, n^{it}; X)).\end{aligned}$$

By the mean value theorem from the start of this lecture, for $* = med$ or $large$, we have

$$\int_{-T}^{T} |F_*(1+it)|^2\, dt \ll \sum_{1\leqslant d\leqslant X} \frac{1}{d}\Big(\frac{T}{d} \sum_{\substack{n\geqslant 2\\ |n-d|\leqslant 100d/T\\ p|n\Rightarrow p\geqslant X^{\varepsilon^2}}} 1\Big)^2.$$

The inner sum is in fact supported on $n \geqslant X^{\varepsilon^2}$, so we can restrict the outer sum to $d \geqslant X^{\varepsilon^2}/2$, provided T is large enough. Then by the Brun-Titchmarsh inequality (as given in our previous Theorem 4.4), if $T \leqslant X^{\varepsilon^3}$ say this sum is

$$\ll \varepsilon^{-O(1)} \sum_{X^{\varepsilon}/2\leqslant d\leqslant X} \frac{1}{d}\Big(\frac{T}{d}\cdot\frac{d}{T}\cdot\frac{1}{\log X}\Big)^2 \ll \varepsilon^{-O(1)}\frac{1}{\log X}.$$

This saving of $(\log X)^{-1}$ is vital.

Bringing everything together, we have

$$\begin{aligned}|M_X(f)| &\ll \log X \exp(-\inf_{|t|\leqslant T} \mathbb{D}(f, n^{it}; X)) \cdot \frac{\varepsilon^{-O(1)}}{(\log x)^{1/2}} \cdot \frac{\varepsilon^{-O(1)}}{(\log x)^{1/2}} + \varepsilon + \frac{1}{T}\\ &\ll \varepsilon^{-O(1)} \exp(-\inf_{|t|\leqslant T} \mathbb{D}(f, n^{it}; X)) + \varepsilon + \frac{1}{T}.\end{aligned}$$

Since $T = T(X) \to \infty$ and $\exp(-\inf_{|t|\leqslant T} \mathbb{D}(f, n^{it}; X)) \to 0$, picking $\varepsilon = \varepsilon(X) \to 0$ suitably slowly we derive $M_X(f) = o(1)$ as $X \to \infty$, as claimed. This finishes the proof. □

Quantitative bounds

Unsurprisingly, by taking more care one can do better quantitatively in this argument. Montgomery and Tenenbaum were the first to prove a result of the strength

$$M_X(f) \ll (1 + D_{X,T})e^{-D_{X,T}} + \frac{1}{T},$$

where $D_{X,T} := \inf_{|t|\leqslant T} \mathbb{D}(f, n^{it}; X)^2$. Here T can range up to a power of $\log X$. The smallest that the right-hand side can ever be is $(X \log\log X)/(\log X)$, and it is this small for $f = \mu$, so we have established that

$$M_X(\mu) = O\Big(\frac{\log\log X}{\log X}\Big),$$

and hence the prime number theorem follows.

But with a terrible error term, compared to what is known by methods of complex analysis, namely

$$M_X(\mu) = O(e^{-c(\log X)^{3/5}/(\log\log X)^{1/5}}).$$

The problem here is that Halász's theorem is too general, in that it holds for all 1-bounded multiplicative functions. The statement of Montgomery and Tenenbaum is in fact sharp, in that there are multiplicative functions f for which $|M_X(f)|$ really is large as their upper bound. (This will be an exercise on the second examples sheet.)

A few years ago Dimitris Koukoulopoulos worked out how to combine ideas in pretentious number theory with a variety of other tricks to recover the best known bound in the prime number theorem using pretentious techniques. We won't discuss that work further in this course, however.

6. Lecture 6: Dirichlet characters

We now begin the second section of the course. Here, we will use some of the knowledge we have gained so far about the averages of multiplicative functions to understand cancellation in partial sums of Dirichlet characters. The main reference here is the paper of Granville–Soundararajan 'Large character sums: Pretentious characters and the Polya-Vinogradov Theorem' which I mentioned in the preamble.

Before we go any further, I need to introduce a certain amount of the classical theory of Dirichlet characters. The entirety of this lecture will most likely be familiar to any reader with a background in classical analytic number theory. Nonetheless, it is important that we review the fundamentals before reaching out into new territory.

Before anything else, let me introduce a time-saving piece of notation, namely the shorthand $e(\theta)$ for $e^{2\pi i\theta}$. This is a common shorthand in the field, and can help to clean up otherwise unwieldy expressions. For example, if $\theta \in \mathbb{Z}$ then automatically $e(\theta) = 1$.

Now let us begin in earnest by defining a few aspects of the theory of discrete Fourier transforms.

Definition 6.1 (Characters)**.** *A* character *on a finite abelian group G is any group homomorphism $\xi : G \to (\mathbb{C}, \times)$. We let $\widehat{G}$ denote the set of all characters on G.*

Observe that, since for any character ξ and $g \in G$ we have $1 = \xi(e_G) = \xi(g^{|G|}) = \xi(g)^{|G|}$, the set $\xi(G)$ is contained within the $|G|^{th}$-roots of unity. In particular $\xi(G)$ is a subset of the unit circle. Furthermore, if $\xi_1, \xi_2 \in \widehat{G}$ then $\xi_1\xi_2 \in G$, and this makes $\widehat{G}$ into a group under pointwise multiplication (with identity element given by the map $g \mapsto 1$, and the inverse of ξ just being the function $\overline{\xi}$). We will sometimes use 1 to denote the identity character (although once we move onto specialising to Dirichlet characters we will have a different, and better, notation).

Two groups will be of particular importance in this course, namely $\mathbb{Z}/q\mathbb{Z}$ under addition and $(\mathbb{Z}/q\mathbb{Z})^\times$ under multiplication (where this second group is the multiplicative group of units modulo q, i.e. the set $\{n \leqslant q : (n, q) = 1\}$ under multiplication modulo q).

The characters on $\mathbb{Z}/q\mathbb{Z}$ may be easily determined.

Lemma 6.2. *Let $q \geqslant 1$. Then there are exactly q characters on the group $\mathbb{Z}/q\mathbb{Z}$, and they are all of the form $\xi_a(n) = e(an/q)$, where $a \in \{1, \dots, q\}$.*

Proof. Let $\xi \in \widehat{\mathbb{Z}/q\mathbb{Z}}$. We know from our observation above that for all $x \in \mathbb{Z}/q\mathbb{Z}$ we have that $\xi(x)$ is a q^{th}-root of unity. So, writing $\{0, 1, \dots, q-1\}$ for the underlying set of $\mathbb{Z}/q\mathbb{Z}$, we may define $a \in \mathbb{Z}$ by letting $\xi(1) = e(a/q)$. Since 1 generates $\mathbb{Z}/q\mathbb{Z}$ additively, we have $\xi(n) = e(na/q)$ for all n, so ξ has the claimed form.

Conversely, note just by definition that any function of the form $n \mapsto e(na/q)$ is indeed a character on $\mathbb{Z}/q\mathbb{Z}$. □

Lemma 6.3. *Suppose that G_1 and G_2 are finite abelian groups, and that $G = G_1 \times G_2$. If χ_i is a character of G_i and $g \in G$ is written $g = (g_1, g_2)$ with $g_i \in G_i$, then $\xi(g) := \xi_1(g_1)\xi_2(g_2)$ is a character on G. Conversely, all characters in $\widehat{G}$ may be written in such a form in a unique way.*

Proof. The fact that $\xi_1(g_1)\xi_2(g_2)$ is a character just follows from the definitions, so the content is the converse statement. Let $\xi \in \widehat{G_1 \times G_2}$, and define $\xi_1 : G_1 \longrightarrow \mathbb{C}$ by $\xi_1(g_1) := \xi((g_1, e_{G_2}))$. Similarly, define $\xi_2 : G_2 \longrightarrow \mathbb{C}$ by $\xi_2(g_2) := \xi((e_{G_1}, g_2))$. It follows immediately that $\xi_1 \in \widehat{G_1}$ and $\xi_2 \in \widehat{G_2}$, and that $\xi = \xi_1\xi_2$.

Furthermore, if $\xi = \chi_1\chi_2$ for some $\chi_i \in \widehat{G_i}$ then by evaluating at (e_{G_1}, g_2) and (g_1, e_{G_2}) we see that $\chi_1 = \xi_1$ and $\chi_2 = \xi_2$. □

A particular consequence of this lemma is that $|\widehat{G_1 \times G_2}| = |\widehat{G_1}||\widehat{G_2}|$.

Lemma 6.4 (Duality and orthogonality)**.** *Let G be a finite abelian group. Then:*

- *G is (non-canonically) isomorphic to $\widehat{G}$;*
- *G is isomorphic to $\widehat{\widehat{G}}$ via the map $g \mapsto \{\xi \mapsto \xi(g)\}$.;*
- *if $\xi \in \widehat{G}$ and $\xi \not\equiv 1$ then $\sum_{g\in G} \xi(g) = 0$;*
- *if $g \in G \setminus \{e_G\}$ then $\sum_{\xi\in\widehat{G}} \xi(g) = 0$;*
- *if $\xi_1, \xi_2 \in \widehat{G}$ with $\xi_1 \neq \xi_2$ then $\sum_{g\in G} \xi_1(g)\overline{\xi_2(g)} = 0$.*

Proof. Recall the classical theorem that any finite abelian group is isomorphic to a direct product of cyclic groups, so $G \cong C_{n_1} \times C_{n_2} \times \cdots \times C_{n^r}$, say. Parts (1) and (3) of the lemma may be verified directly for cyclic groups and then follow for G using the previous structure theorem for $\widehat{G_1} \times \widehat{G_2}$.

For part (2), note that $g \mapsto \{\xi \mapsto \xi(g)\}$ is clearly an injective group homomorphism from G to $\widehat{\widehat{G}}$, but it is also surjective since $|G| = |\widehat{\widehat{G}}|$.

Part (4) follows from applying part (3) to the group $\widehat{G}$.

Part (5) follows from applying part (3) to the character $\xi_1\overline{\xi_2} \in \widehat{G}$. □

The upshot of all these small results is the fact that $\widehat{G}$ is a basis for the $\mathbb{C}$-vector space of functions $\{f : G \longrightarrow \mathbb{C}\}$, and furthermore that this basis is orthogonal with respect to the inner product $\langle f_1, f_2\rangle = \sum_{g\in G} f_1(g)\overline{f_2(g)}$.

Lemma 6.5 (Discrete Fourier Transforms)**.** *If G is a finite abelian group and $f : G \longrightarrow \mathbb{C}$, define the Fourier transform $\widehat{f} : \widehat{G} \longrightarrow \mathbb{C}$ by*

$$\widehat{f}(\xi) = \frac{1}{|G|}\sum_{g\in G} f(g)\overline{\xi(g)}.$$

Then for all $g \in G$ we have $f(g) = \sum_{\xi\in\widehat{G}} \widehat{f}(\xi)\xi(g)$. Furthermore, if f_1, f_2 are two such functions we have

$$\frac{1}{|G|}\sum_{g\in G} f_1(g)\overline{f_2(g)} = \sum_{\xi\in\widehat{G}} \widehat{f_1}(\xi)\widehat{f_2}(\xi).$$

Proof. Regarding the Fourier inversion statement, we have

$$\sum_{\xi\in\widehat{G}} \widehat{f}(\xi)\xi(g) = \sum_{\xi\in\widehat{G}} \Big(\frac{1}{|G|}\sum_{h\in G} f(h)\overline{\xi}(h)\Big)\xi(g) = \sum_{h\in G} f(h)\frac{1}{|G|}\sum_{\xi\in\widehat{G}} \xi(g)\overline{\xi(h)}.$$

However, $\xi(g)\overline{\xi(h)} = \xi(g)\xi(h^{-1}) = \xi(gh^{-1})$, so the inner sum is $|G|1_{g=h}$ by the previous lemma. Therefore the whole sum is $f(g)$ as claimed.

The Plancherel statement follows from a similar analysis. □

We are now ready for the central definition.

Definition 6.6 (Dirichlet characters)**.** *Let $q \geqslant 1$. Then a* Dirichlet character *is a character χ on the group $(\mathbb{Z}/q\mathbb{Z})^\times$, which we extend to a function on all of $\mathbb{Z}$ by defining $\chi(n) = 0$ if $(n, q) > 1$, and then extending q-periodically.*

The letter χ is traditionally used to denote Dirichlet characters. We say that q is a *modulus* of χ (i.e. χ is q-periodic). The Dirichlet character where $n \mapsto 1$ if $(n,q)=1$ and 0 otherwise is called the *principal character*, and is traditionally denoted by χ_0.

It is evident from the definition that $\chi \in \mathcal{M}_0$ for any Dirichlet character χ. These objects are an indispensable tool in studying the interaction between multiplicative number-theoretic properties and q-periodicity (e.g. primes in arithmetic progressions modulo q).

A few properties of Dirichlet characters follows quickly from the previous lemmas. Indeed, if χ_1 is a Dirichlet character with modulus q_1 and χ_2 is a Dirichlet character with modulus q_2 then $\chi_1\chi_2$ is a Dirichlet character with modulus $[q_1,q_2]$. Furthermore, if $q=q_1q_2$ with $(q_1,q_2)=1$ then if χ is a Dirichlet character mod q there exists unique Dirichlet characters χ_i mod q_i for which $\chi=\chi_1\chi_2$. This just follows from the Chinese remainder theorem fact that

$$(\mathbb{Z}/q\mathbb{Z})^\times \cong (\mathbb{Z}/q_1\mathbb{Z})^\times \times (\mathbb{Z}/q_2\mathbb{Z})^\times.$$

Regarding the structure of the set of Dirichlet characters modulus q, one cannot develop quite such a precise description as one could for the additive group $(\mathbb{Z}/q\mathbb{Z})$, but it is nearly as good. Indeed, by the Chinese Remainder Theorem again one has

$$(\mathbb{Z}/q\mathbb{Z})^\times \cong \otimes_{p^k \| q} (\mathbb{Z}/p^k\mathbb{Z})^\times.$$

From a first course in elementary number theory, one recalls the fact that if p is odd then $(\mathbb{Z}/p^k\mathbb{Z})^\times$ is a cyclic group of size p^k-p^{k-1}, generated by some primitive root $g_{p,k} \in (\mathbb{Z}/p^k\mathbb{Z})^\times$. Thus, by previous lemmas, characters on $(\mathbb{Z}/p^k\mathbb{Z})^\times$ are all of the form $g_{p,k}^n \mapsto e(an/(p^k-p^{k-1}))$ for some $a \in \mathbb{Z}$.

The multiplicative structure of the reduced residues modulo 2^k is more complicated. For $k=1$ or $k=2$ the group is cyclic (of order 1 or 2, respectively). For $k \geqslant 3$ then the group isn't cyclic, but one may show that it is generated by -1 and 5, i.e. for each n mod 2^k with $(n,2)=1$ there exist unique a mod 2 and b mod 2^{k-2} for which $n \equiv (-1)^a 5^b \bmod 2^k$. So

$$(\mathbb{Z}/2^k\mathbb{Z})^\times \cong C_2 \times C_{2^{k-2}},$$

and one may construct characters on this group using the direct product construction from earlier.

A quick example of the utility of characters

Since this section has been rather dense on definitions and lemmas so far, I wanted to quickly demonstrate a way in which Dirichlet characters show their worth. Suppose that we wanted to bound the sum

$$\sum_{\substack{n \leqslant X \\ n \equiv a \bmod q}} \lambda(n),$$

where q is considered fixed. (This turns out to be equivalent to proving an asymptotic for the number of primes in an arithmetic progressions modulo q.) Now, if $(a,q)=d$ we can factor out d to get

$$\lambda(d) \sum_{\substack{n \leqslant X/d \\ n \equiv a/d \bmod q/d}} \lambda(n),$$

so without loss of generality we may assume that a and q are coprime. But then

$$1_{n \equiv a \bmod q} = \frac{1}{\varphi(q)} \sum_{\chi \bmod q} \chi(na^{-1}) = \frac{\overline{\chi(a)}}{\varphi(q)} \sum_{\chi \bmod q} \chi(n),$$

using the identities we proved in the five part Lemma earlier, and the multiplicativity of χ. So

$$\Big|\sum_{\substack{n\leqslant X\\ n\equiv a \bmod q}}\lambda(n)\Big| \leqslant \frac{1}{\varphi(q)}\sum_{\chi \bmod q}\Big|\sum_{n\leqslant X}\lambda(n)\chi(n)\Big|.$$

This inner sum is the average of a multiplicative function, and can be shown to be $o_q(X)$ using Halász's theorem. One should remark at this point that it is not a trivial task to show that $\mathbb{D}(\lambda\overline{\chi}, 1; \infty) = \infty$ when χ is a real-valued character; for those of you who know what this statement means, the task is as difficult as showing that $L(1,\chi) \neq 0$. However, this being done, the divergent part of Halász's theorem can be applied.

The upshot is that, by expanding a q-period condition in terms of Dirichlet characters, we were able to utilise our understanding of sums of multiplicative functions in order to understand the original q-periodic function.

Another way of seeing this effect is via Dirichlet series. Indeed, one can show that for $\Re s > 1$ we have

$$\sum_{\substack{n\geqslant 1\\ n\equiv a \bmod q}}\frac{\Lambda(n)}{n^s} = \frac{\overline{\chi(a)}}{\varphi(q)}\sum_{\chi \bmod q}\sum_{n\geqslant 1}\frac{\Lambda(n)\chi(n)}{n^s} = -\frac{\overline{\chi(a)}}{\varphi(q)}\sum_{\chi \bmod q}\frac{L'(s,\chi)}{L(s,\chi)},$$

where

$$L(s,\chi) := \sum_{n\geqslant 1}\frac{\chi(n)}{n^s} = \prod_p\Big(1-\frac{\chi(p)}{p}\Big)^{-1}.$$

The distribution of Λ in arithmetic progression modulo q can thus be studied using the behaviour of the (in general) well-behaved Dirichlet series $L(s,\chi)$.

There are some specific examples of Dirichlet characters that you will probably have already met in previous number theory courses. Indeed, if $(\frac{n}{p})$ denotes the Legendre symbol modulo p, i.e.

$$\Big(\frac{n}{p}\Big) = \begin{cases} 0 & \text{if } p|n\\ -1 & \text{if } n \text{ is not a quadratic residue modulo } p\\ 1 & \text{if } n \text{ is a quadratic residue modulo } p,\end{cases}$$

then this is a Dirichlet character modulo p. Moreover it is a *real* Dirichlet character, in the sense that it takes real values. The real Dirichlet characters have an important (but still somewhat mysterious) role in the theory: we will mostly avoid them in this course.

Primitive Dirichlet characters

When q is prime, most elementary manipulations of the Dirichlet characters modulo q are straightforward. For composite q, an extra phenomenon occurs.

Definition 6.7. *Let χ be a Dirichlet character mod q, and let χ^* be a Dirichlet character mod d, for some $d|q$. We say that χ is* induced *by χ^* if*

$$\chi(n) = \begin{cases}\chi^*(n) & \text{if } (n,q)=1;\\ 0 & \text{otherwise.}\end{cases}$$

If χ is not induced by any such character χ^* modulo d, for any $d|q$ with $d<q$, then we say that χ is primitive.

Examples

Of course if χ_0 is the principal Dirichlet character modulo q, then χ_0 is induced by the unique Dirichlet character of period 1.

Slightly less trivially, consider say the character modulo 15 given by $\chi(1)=1$, $\chi(2)=-1$, $\chi(4)=1$, $\chi(7)=-1$, $\chi(8)=-1$, $\chi(11)=1$, $\chi(13)=-1$, $\chi(14)=1$ (then extended

periodically modulo 15, and with $\chi(n) = 0$ when $(n, 15) > 1$. Then one can see that actually χ is induced by the character χ^* modulo 5 given by $\chi^*(1) = 1$, $\chi^*(2) = -1$, $\chi^*(3) = -1$, $\chi^*(4) = 1$, i.e. χ is induced by the Legendre symbol modulo 5 (which is itself primitive).

Lemma 6.8. *For each Dirichlet character χ, there is a unique primitive Dirichlet character χ^* (possibly equal to χ) such that χ^* induces χ.*

Proof. Exercise (see Davenport). □

If the primitive character χ^* which induces χ is a Dirichlet character to modulus q, we call d the *conductor* of χ (and of χ^*).

Now we shall give a consequence of primitivity that will be useful when we analyse objects called Gauss sums below.

Lemma 6.9. *Let χ be a primitive Dirichlet character modulo q. Then if $d|q$ and $d < q$, for every integer a we have*

$$\sum_{\substack{n \leqslant q \\ n \equiv a \bmod d}} \chi(n) = 0.$$

Actually the second statement is equivalent to χ being primitive.

Proof. Since χ is primitive, χ is not induced by any character modulo d. Therefore there exist m, n such that $m \equiv n \bmod d$ but $\chi(m) \neq \chi(n)$ (and $\chi(mn) \neq 0$). (Indeed, otherwise the definition $\chi^*(m) = \chi(n)$ if $m \equiv n \bmod d$ and $\chi(n) \neq 0$ gives a well-defined character χ^* modulo d which induces χ .) Since $(m, q) = 1$ and $(n, q) = 1$, we may pick a c such that $cm \equiv n \bmod q$. Then $c \equiv 1 \bmod d$, $(c, q) = 1$, but $\chi(c) \neq 1$.

We can use this c to prove the lemma. Indeed, fix an integer a and observe that as k runs through a complete residue system mod q/d the numbers $n = ac + kcd$ run through all residue mod q for which $n \equiv a \bmod d$. Thus

$$\sum_{\substack{n \leqslant q \\ n \equiv a \bmod d}} \chi(n) = \sum_{k \leqslant q/d} \chi(ac + kcd) = \chi(c) \sum_{k \leqslant q/d} \chi(a + kd) = \chi(c) \sum_{\substack{n \leqslant q \\ n \equiv a \bmod d}} \chi(n).$$

Since $\chi(c) \neq 1$, we conclude that

$$\sum_{\substack{n \leqslant q \\ n \equiv a \bmod d}} \chi(n) = 0$$

as required. □

Gauss sums

Given a Dirichlet character $\chi \bmod q$, we define the *Gauss sum* $\tau(\chi)$ of χ to be

$$\tau(\chi) = \sum_{a \leqslant q} \chi(a) e(a/q).$$

The Gauss sum actually determines many of the additive Fourier coefficients of χ.

Lemma 6.10. *Suppose that χ is a Dirichlet character modulo q. If $(n, q) = 1$ then*

$$\chi(n)\tau(\overline{\chi}) = \sum_{a=1}^{q} \overline{\chi}(a) e(an/q).$$

Proof. When $(n, q) = 1$, the map $a \mapsto an$ permutes the residue classes modulo q, and hence

$$\sum_{a \leqslant q} \overline{\chi}(a) e(an/q) = \sum_{a \leqslant q} \overline{\chi}(an^{-1}) e(a/q) = \overline{\chi}(n^{-1}) \sum_{a \leqslant q} \overline{\chi}(a) e(a/q) = \chi(n)\tau(\overline{\chi})$$

as desired. □

Life is sweeter when χ is a primitive Dirichlet character.

Lemma 6.11. *Suppose that χ is a primitive Dirichlet character modulo q. Then*

$$\chi(n)\tau(\overline{\chi}) = \sum_{a=1}^{q} \overline{\chi}(a)e(an/q)$$

for all n, and $|\tau(\chi)| = \sqrt{q}$.

Proof. We need only consider the case $(n,q) > 1$, as the coprime case was dealt with in the previous lemma. Choose m and d so that $(m,d) = 1$ and $m/d = n/q$ (i.e. write the fraction n/q in lowest terms). Then by splitting into residue classes modulo d we get

$$\sum_{a \leqslant q} \chi(a)e(an/q) = \sum_{h \leqslant d} e(hm/d) \sum_{\substack{a \leqslant q \\ a \equiv h \bmod d}} \chi(a).$$

Since $d|q$ and $d < q$, the inner sum vanishes by a previous Lemma.

Regarding the size of the Gauss sum, taking absolute values and summing over n one gets

$$\varphi(q)|\tau(\chi)|^2 = \sum_{n \leqslant q} |\overline{\chi}(n)|^2|\tau(\chi)|^2 = \sum_{n \leqslant q} \Big| \sum_{a \leqslant q} \chi(a)e(an/q)\Big|^2 = \sum_{a,b \leqslant q} \chi(a)\overline{\chi}(b) \sum_{n \leqslant q} e((a-b)n/q)$$
$$= q \sum_{a \leqslant q} |\chi(a)|^2 = q\varphi(q).$$

Hence $|\tau(q)| = \sqrt{q}$. □

A corollary of this result is that the Fourier inversion formula takes a very pleasant form, namely if χ is a primitive Dirichlet character modulo q then

$$\chi(n) = \frac{1}{\tau(\overline{\chi})} \sum_{a \leqslant q} \overline{\chi}(a)e(an/q)$$

for all n.

Exercise 6.12. *Let χ be a Dirichlet character modulo q that is induced by the primitive character χ^* modulo d. Then $\tau(\chi) = \mu(q/d)\chi^*(q/d)\tau(\chi^*)$.*

I have already said that the special role of real Dirichlet characters will not be embarked on in this course. However, for your general education, let me state the fundamental theorem of such characters.

Theorem 6.13. *Let d be a fundamental discriminant, i.e. a non-zero integer such that either $d \equiv 1 \bmod 4$ and $\mu^2(d) = 1$ or $d = 4m$ where $m \equiv 2, 3 \bmod 4$ and $\mu^2(m) = 1$. Then let us define the Kronecker symbol $(\frac{d}{\cdot})$ to be the function from $\mathbb{Z}$ to $\{0, -1, +1\}$ given on primes by*

- $\left(\frac{d}{p}\right) = 0$ *when $p|d$;*
- $\left(\frac{d}{2}\right) = 1$ *when $d \equiv 1 \bmod 8$ and $\left(\frac{d}{2}\right) = -1$ when $d \equiv 5 \bmod 8$;*
- $\left(\frac{d}{p}\right)$ *is equal to the Legendre symbol for all odd primes;*
- $\left(\frac{d}{-1}\right) = \operatorname{sgn}(d)$*;*

and then extended so that $(\frac{d}{n})$ is a totally multiplicative function of n.

Then $n \mapsto (\frac{d}{n})$ is a real primitive Dirichlet character of conductor $|d|$, and moreover all such primitive Dirichlet characters are of this form.

7. Lecture 7: Repulsion theorems for Dirichlet characters

Last lecture we didn't quite have time to state and prove the Polya–Vinogradov inequality. To do so, let us start by recording a simple (but exceptionally useful!) lemma about exponential sums.

Lemma 7.1. *For all $\theta \in \mathbb{R}$,*

$$\sum_{n \leqslant N} e(n\theta) \ll \min(N, \|\theta\|^{-1}),$$

where $\|\theta\|$ is the distance from θ to the nearest integer.

Proof. Explicitly evaluating the geometric series we get

$$\sum_{n \leqslant N} e(n\theta) = \frac{e((N+1)\theta) - e(\theta)}{1 - e(\theta)} \ll \frac{1}{|1 - e(\theta)|} \ll \frac{1}{|e(-\theta/2) - e(\theta/2)|} \ll \frac{1}{|\sin \pi\theta|} \ll \frac{1}{\|\theta\|}$$

since $\sin \pi x \geqslant 2x$ for $x \in [0, 1/2]$.

That gives the bound $O(\|\theta\|^{-1})$. The other bound of N comes from the trivial argument

$$\Big| \sum_{n \leqslant N} e(n\theta) \Big| \ll \sum_{n \leqslant N} |e(n\theta)| \ll N.$$

This settles the lemma. □

If χ is a non-principal Dirichlet character to modulus q, by orthogonality of characters we have that $\sum_{n \leqslant q} \chi(n) = 0$. In particular (by q-periodicity) this implies that $\sum_{n \leqslant N} \chi(n) \leqslant q$. However, it turns out that we can do substantially better than this bound.

Theorem 7.2 (Pólya–Vinogradov Theorem)**.** *Let χ be a non-principal Dirichlet character mod q with $q \geqslant 2$. Then, for all $N \geqslant 1$ we have*

$$\sum_{n \leqslant N} \chi(n) \ll q^{1/2} \log q.$$

Proof. Suppose first that χ is a primitive character. Then by Fourier inversion we have

$$\sum_{n \leqslant N} \chi(n) = \frac{1}{\tau(\overline{\chi})} \sum_{a \leqslant q} \overline{\chi}(a) \sum_{n \leqslant N} e(an/q).$$

The inner sum is a geometric series, and we may bound it using the previous lemma. So

$$\sum_{n \leqslant N} \chi(n) \ll \frac{1}{|\tau(\overline{\chi})|} \sum_{a \leqslant q-1} \|a/q\|^{-1} \ll \frac{1}{\sqrt{q}} \cdot q \cdot \sum_{a \leqslant q-1} \frac{1}{a} \ll q^{1/2} \log q.$$

Now let us extend the proof to the general case. Suppose that χ is induced by the primitive character χ^* modulo d. We know that $d \geqslant 2$ and χ^* is non-principal, since χ is non-principal. Let r be the product of those primes that divide q but not d. Then

$$\begin{aligned}
\sum_{n \leqslant N} \chi(n) &= \sum_{\substack{n \leqslant N \\ (n,r)=1}} \chi^*(n) \\
&= \sum_{n \leqslant N} \chi^*(n) \sum_{k \mid (n,r)} \mu(k) \\
&= \sum_{k \mid r} \mu(k) \sum_{\substack{n \leqslant N \\ k \mid n}} \chi^*(n) \\
&= \sum_{k \mid r} \mu(k) \chi^*(k) \sum_{n \leqslant N/k} \chi^*(m).
\end{aligned}$$

We know that the absolute value of the inner sum is $\ll d^{1/2}\log d$. Furthermore, the number of divisors of r is at most $O(r^{1/2})$ (as can be seen easily, since if $ab = r$ then at least 1 of a and b is at most $r^{1/2}$). So

$$\sum_{n\leqslant N}\chi(n) \ll r^{1/2}d^{1/2}\log d \ll q^{1/2}\log q,$$

since $r \leqslant q/d$ and $d \leqslant q$. This settles the theorem. □

The Polya–Vinogradov inequality is a central result in the theory. There are many application to the local distribution of arithmetic objects, some of which will be on the examples sheet. For characters with even order, any improvement would have strong consequences for the position of the least quadratic non-residue – which seems to be a very difficult problem. However, for characters of odd order, it turns out that – remarkably – one can improve this inequality.

Theorem 7.3 (Granville–Soundararajan)**.** *Let χ be a Dirichlet character mod q, and let g be the order of χ. Suppose that $g \geqslant 3$ is odd. Then, for all $N \geqslant 1$ we ave*

$$\sum_{n\leqslant N}\chi(n) \ll_g q^{1/2}(\log q)^{1-\frac{\delta_g}{2}+o(1)},$$

where

$$\delta_g := 1 - \frac{g}{\pi}\sin\Big(\frac{\pi}{g}\Big).$$

In particular $\delta_g > 0$.

This theorem, proved in 2005, was the first major unconditional improvement to P–V since the inequality was first formulated in 1918! It has since been refined by Goldmakher and Lamzouri–Mangerel, who improved some of the machinery involving mean values of multiplicative functions in order to get a bound of $q(\log q)^{1-\delta_g+o(1)}$. We won't be able to get to this result, but, using the tools that we have developed on averages of multiplicative functions, we will be able to prove the Granville–Soundararajan result.

A final contextual remark: conditional on the Generalised Riemann Hypothesis, Montgomery showed back in the 70s that $\sum_{n\leqslant N}\chi(n) \ll q^{1/2}\log\log q$. This was known to be tight since the 1930s, when Paley came up with some examples (see examples sheet).

Our proof of Theorem 7.3 will be based on three main devices. One is the logarithmically-averaged Halasz's theorem that we introduced in Lecture 2. Another is a 'major arc, minor arc' fourier-analytic principal, familiar to any readers with some background in the circle method – that will covered next time. For today, we will concern ourselves with the third principal, namely a repulsion principal for $\mathbb{D}(\chi,\xi;X)$ when χ and ξ are certain Dirichlet characters.

We begin with a statement about prime numbers in arithmetic progressions, which we will use throughout.

Theorem 7.4 (Prime number theorem in APs, harmonic sum)**.** *Let $y \geqslant 2$, $q \leqslant (\log y)^A$, and $(a,q) = 1$. Then*

$$\sum_{\substack{q\leqslant p\leqslant y\\ p\equiv a \bmod q}}\frac{1}{p} = (1+o_A(1))\frac{\log\log y}{\varphi(q)}.$$

When $A < 2$, the implied constant is effective.

Such a theorem, or one like it, is a common staple of first courses in analytic number theory. As according to our general policy, we will not provide a proof in these notes. The reader is invited to consult Davenport for the details. However, the intuition for the result

is clear enough, namely that the primes are equidistributed across all the residues classes a modulo q in which one finds infinitely many primes, namely when $(a,q)=1$. Note that it is nonetheless important that we excluded the primes less than q from consideration, otherwise the sum of the left-hand could potentially be dominated by one large term coming from a single small prime.

For those readers with a little more background, the issue of effectivity comes as ever from the possibilities of an L-function $L(s,\chi)$ having a real zero very close to $s=1$. The class number formula yields the bound $L(1,\chi_d) \gg |d|^{-1/2}$, where χ_d is a real primitive character of conductor $|d|$, and this in turn leads to the effective bound that $1-\beta \gg_\varepsilon d^{-1/2-\varepsilon}$ for all $\varepsilon>0$. The 2 in this fraction $1/2$ leads to the effective range of $(\log y)^2$ in the theorem. Of course we have Siegel's result that $1-\beta \gg_\varepsilon d^{-\varepsilon}$, but this is ineffective, and leads to ineffective bounds in the theorem above.

In the applications later on in this section of the course, we will only be considering cases in which $q \leqslant \log y$. So, in fact, all the error terms will be effective.

We also need another simple observation about primitive characters.

Lemma 7.5. *If χ_1 and χ_2 are two distinct primitive characters, of conductors q_1 and q_2 respectively, then we cannot have the character $\chi_1\overline{\chi_2}$ be the principal character modulo $[q_1,q_2]$.*

Proof. The character χ_1 induces a character χ_1' modulo $[q_1,q_2]$, given by $\chi_1'(n)=\chi_\prime(n)$ if $(n,[q_1,q_2])=1$ and $\chi_1'(n)=0$ otherwise. Similarly the character χ_2 induces a character χ_2' modulo $[q_1,q_2]$. We find that $\chi_1' \neq \chi_2'$, since every character is induced by a unique primitive character. But this means directly that $\chi_1'\overline{\chi_2'}$ is non-principal, and hence that $\chi_1\chi_2$ is non-principal. □

Now let us use Theorem 7.4 to prove our first repulsion result for Dirichlet characters.

Lemma 7.6 (Successive repulsion)**.** *Let χ mod q be a primitive character and $y \geqslant 10$. Let us enumerate all the primitive characters $\{\psi_1,\dots,\psi_A\}$ with conductor at most $\log y$ so that the distances $\mathbb{D}(\chi,\psi_j;y)$ are arranged in ascending order. Then for each $1\leqslant j\leqslant A$ we have*

$$\mathbb{D}(\chi,\psi_j;y)^2 \geqslant \Big(1-\frac{1}{\sqrt{j}}+o(1)\Big)\log\log y$$

as $y\to\infty$.

In words, this lemma is saying that χ can only possibly pretend to be a single other primitive character with small conductor, namely ψ_1. All other primitive characters ψ_j are bounded away from χ.

Proof. Note that

$$\begin{aligned}\mathbb{D}(\chi,\psi_j;y)^2 &\geqslant \frac{1}{j}\sum_{k=1}^{j}\mathbb{D}(\chi,\psi_k;y)^2\\ &= \frac{1}{j}\sum_{p\leqslant y}\frac{1}{p}\sum_{k=1}^{j}(1-\Re\chi(p)\overline{\psi_k}(p))\\ &\geqslant \frac{1}{j}\sum_{p\leqslant y}\frac{1}{p}\Big(j-\Big|\sum_{k=1}^{j}\psi_k(p)\Big|\Big)\\ &= \log\log y+O(1)-\frac{1}{j}\sum_{p\leqslant y}\frac{1}{p}\Big|\sum_{k=1}^{j}\psi_k(p)\Big|.\end{aligned}$$

We will bound the second term above. Indeed, by the Cauchy–Schwarz inequalty we have that

$$\Big(\sum_{p\leqslant y}\frac{1}{p}\Big|\sum_{k=1}^{j}\psi_k(p)\Big|\Big)^2 \leqslant \Big(\sum_{p\leqslant y}\frac{1}{p}\Big)\Big(\sum_{p\leqslant y}\frac{1}{p}\Big|\sum_{k=1}^{j}\psi_k(p)\Big|^2\Big).$$

The first term on the right-hand side is equal to $\log\log y+O(1)$, which is $(1+o(1))\log\log y$. Expanding out the square in the second term, writing m_k for the conductor of ψ_k, we see that the second term is

$$\sum_{p\leqslant y}\frac{1}{p}\Big(j+\sum_{\substack{1\leqslant k,\ell\leqslant j\\ k\neq\ell}}\psi_k(p)\overline{\psi_\ell}(p)\Big),$$

which equals

$$\begin{aligned}
&= j(1+o(1))\log\log y+\sum_{\substack{1\leqslant k,\ell\leqslant j\\ k\neq\ell}}\sum_{\substack{a\leqslant[m_k,m_\ell]\\ (a,m_k)=(a,m_\ell)=1}}(\psi_k\overline{\psi_\ell})(a)\sum_{\substack{p\leqslant y\\ p\equiv a \bmod [m_k,m_\ell]}}\frac{1}{p}\\
&= j(1+o(1))\log\log y+\sum_{\substack{1\leqslant k,\ell\leqslant j\\ k\neq\ell}}\sum_{\substack{a\leqslant[m_k,m_\ell]\\ (a,m_k)=(a,m_\ell)=1}}(\psi_k\overline{\psi_\ell})(a)\sum_{\substack{[m_k,m_\ell]\leqslant p\leqslant y\\ p\equiv a \bmod [m_k,m_\ell]}}\frac{1}{p}+O(j^2\log_3 y)\\
&= j(1+o(1))\log\log y+\sum_{\substack{1\leqslant k,\ell\leqslant j\\ k\neq\ell}}\sum_{\substack{a\leqslant[m_k,m_\ell]\\ (a,m_k)=(a,m_\ell)=1}}(\psi_k\overline{\psi_\ell})(a)(1+o(1))\frac{\log\log y}{\varphi([m_k,m_\ell])}+O(j^2\log_3 y)
\end{aligned}$$

by the prime number theorem in arithmetic progressions as given earlier. Here $\log_3 y:=\log\log\log y$. Since $\psi_k\overline{\psi_\ell}$ is a non-principal character modulo $[m_k,m_\ell]$ (by our previous Lemma), we have that

$$\sum_{\substack{a\leqslant[m_k,m_\ell]\\ (a,m_k)=(a,m_\ell)=1}}(\psi_k\overline{\psi_\ell})(a)=0.$$

Therefore we have that the total contribution in the above sum is

$$j\log\log y+j^2o(\log\log y).$$

Hence

$$\begin{aligned}
\mathbb{D}(\chi,\psi_j;y)^2 &\geqslant (1+o(1))\log\log y-\frac{1}{j}((1+o(1))\log\log y)^{1/2}(j\log\log y+j^2o(\log\log y))^{1/2}\\
&\geqslant (1+o(1))\log\log y-\frac{1}{j}((1+o(1))\log\log y)^{1/2}(j^{1/2}(\log\log y)^{1/2}+jo((\log\log y)^{1/2}))\\
&\geqslant \Big(1-\frac{1}{\sqrt{j}}+o(1)\Big)\log\log y
\end{aligned}$$

as claimed. □

Note that for any Dirichlet character χ we have $\chi(-1)\in\{-1,1\}$, since $\chi(-1)^2=\chi((-1)^2)=\chi(1)=1$. We call a character χ *even* if $\chi(-1)=1$, and *odd* if $\chi(-1)=-1$. The second repulsion lemma shows that, at least in the case of odd order characters, a character χ cannot strongly pretend to be a character with small conductor of the opposite parity.

Lemma 7.7. *Let χ mod q be a primitive character of odd order g, and let $y\geqslant 10$. Suppose that ξ mod m is a character such that $\chi(-1)\xi(-1)=-1$. If $m\leqslant(\log y)^A$ then*

$$\mathbb{D}(\chi,\xi;y)^2\geqslant(\delta_g+o_A(1))\log\log y,$$

where $\delta_g=1-(g/\pi)\sin(\pi/g)$ as in the main theorem of this section, and where the error term is effective if $A<2$.

Proof. Since χ has odd order, $\chi(-1)=1$. Thus $\xi(-1)=-1$, and so ξ must have even order $k\geqslant 2$, say. We have

$$\begin{aligned}\mathbb{D}(\chi,\xi;y)^2&=\sum_{p\leqslant y}\frac{1-\Re(\chi(p)\overline{\xi(p)})}{p}\\&\geqslant\sum_{-\frac{k}{2}<\ell\leqslant\frac{k}{2}}\Big(\sum_{\substack{m\leqslant p\leqslant y\\ \xi(p)=e(\ell/k)}}\frac{1}{p}\Big)\min_{z^g=0,1}(1-\Re(ze(-\ell/k))),\end{aligned}$$

by splitting the sum according to the values of $\xi(p)$, and assuming that $\chi(p)=z$ is always the worst-case z.

Note that

$$\begin{aligned}\min_{z^g=0,1}(1-\Re(ze(-\ell/k)))&\geqslant\min_{h\in\mathbb{Z}}(1-\Re\Big(e\Big(\frac{h}{g}-\frac{\ell}{k}\Big)\Big)\\&=1-\max_{h\in\mathbb{Z}}\cos\Big(\frac{2\pi}{g}\Big(h-\frac{\ell g}{k}\Big)\Big)\\&=1-\cos\Big(\frac{2\pi}{g}\Big\|\frac{\ell g}{k}\Big\|\Big),\end{aligned}$$

where $\|\theta\|$ denotes the distance from θ to the nearest integer, as before. From the prime number theorem in arithmetic progressions we get

$$\begin{aligned}\sum_{\substack{m\leqslant p\leqslant y\\ \xi(p)=e(\ell/k)}}\frac{1}{p}&=\sum_{\substack{r\leqslant m\\ (r,m)=1\\ \xi(r)=e(\ell/k)}}\sum_{\substack{m\leqslant p\leqslant y\\ p\equiv r\bmod m}}\frac{1}{p}\\&=\sum_{\substack{r\leqslant m\\ (r,m)=1\\ \xi(r)=e(\ell/k)}}(1+o_A(1))\frac{\log\log y}{\varphi(m)}\\&=(1+o_A(1))\frac{\log\log y}{k}.\end{aligned}$$

Write in lowest terms $g/k=g^*/k^*\notin\mathbb{Z}$, since g and k have opposite parity. Reindexing ℓ, we have

$$\mathbb{D}(\chi,\xi;y)^2\geqslant(1+o_A(1))\frac{\log\log y}{k}\frac{k}{k^*}\sum_{-k^*/2<\ell\leqslant k^*/2}\Big(1-\cos\frac{2\pi\ell}{gk^*}\Big).$$

Summing the cosine series, using critically the fact that k^* is even so that we may write $-k^*/2+1$ as the least value of the summation variable ℓ, we get

$$\mathbb{D}(\chi,\xi;y)^2\geqslant(1+o_A(1))\Big(1-\frac{\sin(\pi/g)}{k^*\tan(\pi/gk^*)}\Big)\log\log y.$$

Since $k^*\tan(\pi/gk^*)\geqslant\pi/g$ we have

$$\mathbb{D}(\chi,\xi;y)^2\geqslant(1+o(1))\Big(1-\frac{g}{\pi}\sin\Big(\frac{\pi}{g}\Big)\Big)\log\log y=(1+o_A(1))\delta_g\log\log y$$

as required. □

Next time we will proceed to the Fourier analytic elements of the proof.

8. Lecture 8: Major arcs and minor arcs

Last time we studied the aspects of the proof of Theorem 7.3 that were concerned with the Granville–Soundararajan distance. Today, we will study the Fourier analytic aspects. For reasons of time, we will not be able to prove quite all the results that we need in full generality or precision. Rather, we will provide short and (relatively) clean proofs of similar statements, which are slightly weaker than we need.

Theorem 8.1 (Polya's Fourier expansion)**.** *If χ is a primitive non-principal character mod q, then we have the Fourier expansion*

$$\sum_{n\leqslant N}\chi(n)=\frac{\tau(\chi)}{2\pi i}\sum_{\substack{k\in\mathbb{Z}\\ 1\leqslant|k|\leqslant K}}\frac{\overline{\chi}(k)}{k}\Big(1-e\Big(-\frac{kN}{q}\Big)\Big)+O\Big(1+\frac{q\log q}{K}\Big).$$

We won't prove this result with the error term given here. To do so would entail us taking a small diversion into the realm of 'quantitative Fourier analysis for discontinuous functions', which I want to avoid in this primarily number-theoretic course. The key (qualitative) lemma is the following, which should be broadly familiar from undergraduate analysis.

Lemma 8.2. *Let $f\in L^1(\mathbb{R}/\mathbb{Z})$ be a piecewise continuously differentiable function. Then for all $\alpha\in\mathbb{R}/\mathbb{Z}$ we have*

$$\frac{f(\alpha^+)+f(\alpha^-)}{2}=\lim_{K\to\infty}\sum_{\substack{k\in\mathbb{Z}\\ |k|\leqslant K}}\widehat{f}(k),$$

where $f(\alpha^\pm)$ are the upper and lower limits of f at α, and where

$$\widehat{f}(k)=\int_0^1 f(\alpha)e(-k\alpha)\,d\alpha.$$

I hope that this is a familiar idea, namely that Fourier series converge to the mid-point of a jump discontinuity. For a quantitative version of this fact, see Appendix D of Montgomery-Vaughan.

Proof of Theorem 8.1, main term. Consider the function $f:\mathbb{R}/\mathbb{Z}\longrightarrow\mathbb{C}$ given by

$$f(\alpha):=\sum_{n\leqslant\alpha q}\chi(n).$$

This function is well-defined since $\sum_{n\leqslant q}\chi(n)=0$, and it is piecewise continuously differentiable (in fact it is piecewise constant, and has $\varphi(q)$ jump discontinuities). Then, for $k\neq 0$,

$$\begin{aligned}\widehat{f}(k)&=\int_0^1\sum_{n\leqslant\alpha q}\chi(n)e(-k\alpha)\,d\alpha\\ &=\sum_{n\leqslant q}\chi(n)\int_{n/q}^1 e(-k\alpha)\,d\alpha\\ &=\sum_{n\leqslant q}\chi(n)\Big(-\frac{1}{2\pi ik}\Big(1-e\Big(-\frac{kn}{q}\Big)\Big)\Big)\\ &=\frac{1}{2\pi ik}\sum_{n\leqslant q}\chi(n)e\Big(-\frac{kn}{q}\Big)\end{aligned}$$

$$= \frac{1}{2\pi i k}\overline{\chi}(-k)\tau(\chi)$$

since χ is primitive. Similarly we derive

$$\widehat{f}(0) = -\frac{1}{q}\sum_{n\leqslant q} n\chi(n).$$

One may derive an alternative expression for this 0^{th} Fourier coefficient. Indeed, the point 0 is a point of continuity for f, so from the lemma above we get

$$\begin{aligned} 0 = f(0) = \lim_{K\to\infty} \sum_{\substack{k\in\mathbb{Z}\\ |k|\leqslant K}} \widehat{f}(k) &= \widehat{f}(0) + \lim_{K\to\infty} \sum_{\substack{k\in\mathbb{Z}\\ 1\leqslant|k|\leqslant K}} \widehat{f}(k) \\ &= -\frac{1}{q}\sum_{n\leqslant q} n\chi(n) + \lim_{K\to\infty} \sum_{\substack{k\in\mathbb{Z}\\ 1\leqslant|k|\leqslant K}} \frac{1}{2\pi i k}\overline{\chi}(-k)\tau(\chi). \end{aligned}$$

Putting everything together, and using the preceding lemma again, when N/q is a point of continuity of f we have

$$\begin{aligned} \sum_{n\leqslant N}\chi(n) = f(N/q) &= \frac{\tau(\chi)}{2\pi i}\lim_{K\to\infty} \sum_{\substack{k\in\mathbb{Z}\\ 1\leqslant|k|\leqslant K}} \frac{\overline{\chi}(-k)}{k}\Big(e\Big(\frac{kN}{q}\Big) - 1\Big) \\ &= \frac{\tau(\chi)}{2\pi i}\lim_{K\to\infty} \sum_{\substack{k\in\mathbb{Z}\\ 1\leqslant|k|\leqslant K}} \frac{\overline{\chi}(k)}{k}\Big(1 - e\Big(\frac{-kN}{q}\Big)\Big) \end{aligned}$$

after relabelling k with $-k$. This gives the correct expression for the main term of the Fourier series claimed in the theorem. The error term comes from a quantitative version of the above argument. □

We will apply this theorem with $K = q$, deriving

$$\sum_{n\leqslant N}\chi(n) = \frac{\tau(\chi)}{2\pi i}\sum_{\substack{k\in\mathbb{Z}\\ 1\leqslant|k|\leqslant q}} \frac{\overline{\chi}(k)}{k}\Big(1 - e\Big(\frac{-kN}{q}\Big)\Big) + O(\log q).$$

Combining the contributions with k and $-k$, and using the multiplicativity of χ, we end up with

$$\sum_{n\leqslant N}\chi(n) = \frac{\tau(\chi)}{2\pi i}(1-\overline{\chi}(-1))\sum_{k\leqslant q}\frac{\overline{\chi}(k)}{k} - \frac{\tau(\chi)}{2\pi i}\sum_{k\leqslant q}\frac{\overline{\chi}(k)}{k}\Big(e\Big(-\frac{kN}{q}\Big) - \chi(-1)e\Big(\frac{kN}{q}\Big)\Big) + O(1+\log q).$$

The first term here

$$\frac{\tau(\chi)}{2\pi i}(1-\overline{\chi}(-1))\sum_{k\leqslant q}\frac{\overline{\chi}(k)}{k}$$

could be related to $L(1,\chi)$, but we won't need to do this as for us this term will immediately vanish. Indeed, specialising to the case were χ has odd order g, we know that $\chi(-1) = 1$ and so the term vanishes and we are left with

$$\sum_{n\leqslant N}\chi(n) = -\frac{\tau(\chi)}{2\pi i}\sum_{k\leqslant q}\frac{\overline{\chi}(k)}{k}\Big(e\Big(-\frac{kN}{q}\Big) - e\Big(\frac{kN}{q}\Big)\Big) + O(1+\log q).$$

Since Theorem 7.3 can easily be reduced to case when χ is primitive – as in the proof of the Polya–Vinogradov inequality – we have reduced matters to the following theorem:

Theorem 8.3. *Let χ mod q be a primitive non-principal Dirichlet character, with odd order $g \geqslant 3$. Then*

$$\sum_{k \leqslant q} \frac{\overline{\chi}(k)}{k} \Big(e\Big(-\frac{kN}{q} \Big) - e\Big(\frac{kN}{q} \Big) \Big) \ll (\log q)^{1-\delta_g/2+o(1)},$$

where $\delta_g = 1 - \frac{g}{\pi} \sin(\frac{\pi}{g})$.

Another observation is that δ_g is a decreasing function of g, and hence $(\log q)^{1-\delta_g/2} \geqslant (\log q)^{1-\delta_3/2} \geqslant (\log q)^{0.913\ldots}$. So any extra error terms of $(\log q)^c$ with $c < 0.913$ can be absorbed.

This is the end of the first part of the lecture. Note that the Fourier expansion above is also adequate for proving the Polya–Vinogradov inequality, since the trivial bound on the summation is $O(\log q)$, and as ever the Gauss sum satisfies $|\tau(\chi)| = \sqrt{q}$. But the sum in the statement of Theorem 8.3 enjoys some extra explicit structure – indeed, it is very close to a logarithmic average of $\overline{\chi}$, albeit twisted by an additive character. For the rest of this lecture, we discuss how to remove the additive character.

It turns out (as often happens!) that the behaviour of the sum

$$\sum_{k \leqslant q} \frac{\overline{\chi}(k)}{k} e\Big(\frac{kN}{q} \Big)$$

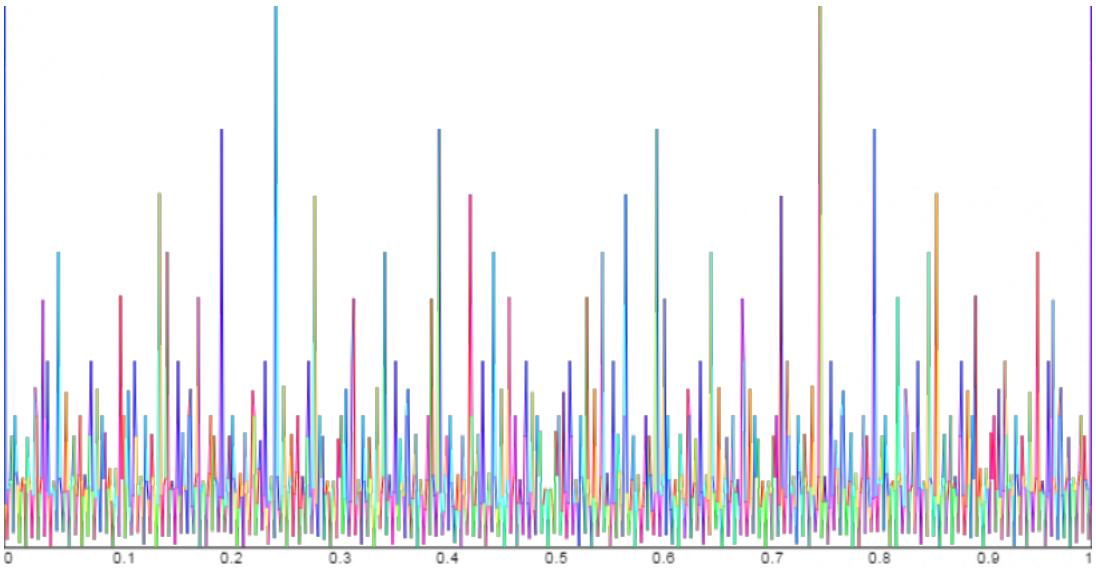
is determined by the quality of rational approximations to the phase N/q.

A brief diversion concerning major arcs and minor arcs

In a great many natural situations in analytic number theory, one is faced with estimating the the fourier coefficients of some arithmetic sequence, for example the exponential sum

$$\sum_{n \leqslant N} e(n^k \theta)$$

for some fixed natural number k, which comes up in the calculations involved in 'Waring's problem', i.e. when writing an integer as a sum of perfect k^{th} powers. Below is a graph of the absolute value of this sum as θ ranges from 0 to 1, in the case $k = 2$:



Notice how most of the time the exponential sum is very small, but very occasionally it is large. These points are when $\theta \approx b/r$ when $(b,r)=1$ and r is small. (Not all such points; witness the fact that there is no peak at $\theta = 1/2$). We call these small regions around fraction b/r to be 'major arcs', and everything else in $[0,1)$ to be a 'minor arc'. (The terminology dates back to Hardy–Littlewood, and comes from the point of view in which the phase $e(n^k\theta)$ is really a point on the unit circle, and one cuts up the circle into various arcs.)

The minor arcs tend to be the 'hard' ones in applications, so for ease of exposition we'll deal with those second. [Of course in actual research you tend to tackle the minor arcs *first*, as one feels that once the minor arcs have been controlled then everything else should fall into place.] To estimate matters in the major arc case, i.e. to estimate $\sum_{n\leqslant N} e(n^k b/r)$ when r is small, it often suffices to split this sum over n into the different residue classes modulo r. On these residue classes the phase $e(n^k b/r)$ is constant, and so doesn't affect the sum, and one can conclude by counting the number of k^{th} powers in arithmetic progressions modulo r.

The range over which one has good control of the arithmetic behaviour in arithmetic progressions modulo r is what determines how large r can be taken, i.e. how many major arcs you can take.

Let's put this method to the test in our context.

Lemma 8.4. *Let χ mod q be a primitive non-principal Dirichlet character, with odd order $g \geqslant 3$, and let $b \in \mathbb{Z}$ and $r \in \mathbb{N}$ with $(b,r)=1$, $r \leqslant (\log q)^{1/3}$. Then*

$$\max_{K\leqslant q}\Big|\sum_{k\leqslant K}\frac{\overline{\chi}(k)}{k}\Big(e\Big(-\frac{kb}{r}\Big)-e\Big(\frac{kb}{r}\Big)\Big)\Big| \ll_g (\log q)^{1-\delta_g/2+o(1)}.$$

Proof. We study

$$\sum_{k\leqslant K}\frac{\overline{\chi}(k)}{k}e\Big(\frac{kb}{r}\Big).$$

Splitting the sum according to the greatest common divisor of k and r,

$$\begin{aligned}\sum_{k\leqslant K}\frac{\overline{\chi}(k)}{k}e\Big(\frac{kb}{r}\Big) &= \sum_{d|r}\sum_{\substack{k\leqslant K\\(k,r)=d}}\frac{\overline{\chi}(k)}{k}e\Big(\frac{kb/d}{r/d}\Big)\\ &= \sum_{d|r}\frac{\overline{\chi}(d)}{d}\sum_{\substack{k\leqslant K/d\\(k,r/d)=1}}\frac{\overline{\chi}(k)}{k}e\Big(\frac{kb}{r/d}\Big). \qquad (4)\end{aligned}$$

Since $(kb, r/d)=1$ we get

$$\begin{aligned}e\Big(\frac{kb}{r/d}\Big) &= \frac{1}{\varphi(r/d)}\sum_{\substack{a \bmod r/d\\(a,r/d)=1}}e\Big(\frac{a}{r/d}\Big)\sum_{\psi \bmod r/d}\psi(a^{-1}kb)\\ &= \frac{1}{\varphi(r/d)}\sum_{a \bmod r/d}e\Big(\frac{a}{r/d}\Big)\sum_{\psi \bmod r/d}\overline{\psi}(a)\psi(kb),\end{aligned}$$

using the fact that for all $x \bmod r/d$ we have

$$1_{x=1} = \frac{1}{\varphi(r/d)}\sum_{\psi \bmod r/d}\psi(x).$$

By swapping the summations over a and ψ, we get

$$e\Big(\frac{kb}{r/d}\Big) = \frac{1}{\varphi(r/d)} \sum_{\psi \bmod r/d} \psi(kb) \sum_{a \bmod r/d} e\Big(\frac{a}{r/d}\Big)\overline{\psi}(a)$$
$$= \frac{1}{\varphi(r/d)} \sum_{\psi \bmod r/d} \psi(kb)\tau(\overline{\psi})$$

by the definition of the Gauss sum $\tau(\overline{\psi})$. Therefore

$$\sum_{\substack{k\leqslant K/d\\(k,r/d)=1}} \frac{\overline{\chi}(k)}{k} e\Big(\frac{kb}{r/d}\Big) = \frac{1}{\varphi(r/d)} \sum_{\psi \bmod r/d} \tau(\overline{\psi})\psi(b) \sum_{k\leqslant K/d} \frac{(\overline{\chi}\psi)(k)}{k}. \tag{5}$$

Recall the logarithmic Halasz theorem in Lecture 2/3 (Lemma 2.12) that stated that if $f \in \mathcal{M}_0$ then

$$M_{X,\log}(f) \ll \exp\Big(-\frac{1}{2}\sum_{p\leqslant X}\frac{1-\Re f(p)}{p}\Big),$$

provided $X \geqslant 2$. Therefore

$$\Big|\sum_{k\leqslant K/d} \frac{(\overline{\chi}\psi)(k)}{k}\Big| \ll 1 + (\log K/d)\exp(-\frac{1}{2}\mathbb{D}(\chi,\psi;K/d)^2)$$
$$\ll 1 + (\log K/d)\exp(-\frac{1}{2}\mathbb{D}(\chi,\psi;q)^2 + \sum_{K/d<p\leqslant q}\frac{1}{p})$$
$$\ll 1 + (\log q)\exp(-\frac{1}{2}\mathbb{D}(\chi,\psi;q)^2),$$

so we are in a position to start using our results on Dirichlet character repulsion.

Let $\xi \bmod m$ denote the primitive Dirichlet character with conductor below $(\log q)^{1/3}$ which minimises the distance $\mathbb{D}(\chi,\xi;q)$ amongst all such characters (if there is a tie, pick one). We split collection of characters ψ in the sum (5) according to whether ψ is induced by ξ or not.

Let $\psi \bmod r/d$ be induced by the primitive character ψ^*. Since ψ and ψ^* potentially differ only at primes $p|r/d$, we conclude that $\mathbb{D}(\chi,\psi;q)^2 = \mathbb{D}(\chi,\psi^*;q)^2 + O(\log\log\log q)$, since $r \leqslant (\log q)^{1/3}$. By appealing to Lemma 3.8, we observe that if $\psi^* \neq \xi$ then

$$\mathbb{D}(\chi,\psi;q)^2 \geqslant (1-1/\sqrt{2}+o(1))\log\log q,$$

and further there are at most $O(1)$ characters $\psi \bmod r/d$ for which $\mathbb{D}(\chi,\psi;q)^2 \leqslant \frac{2}{3}\log\log q$. Since $|\tau(\overline{\psi})| \leqslant \sqrt{r/d}$ we conclude that the contribution from all such characters is

$$\leqslant \frac{1}{\varphi(r/d)} \sum_{\substack{\psi \bmod r/d\\ \psi \text{ not induced by } \xi}} |\tau(\overline{\psi})|\Big|\sum_{k\leqslant K/d}\frac{(\overline{\chi}\psi)(k)}{k}\Big|$$
$$\ll \frac{\sqrt{r/d}}{\varphi(r/d)}(\log q)^{\frac{1}{2}+\frac{1}{2\sqrt{2}}+o(1)} + \sqrt{r/d}(\log q)^{2/3}.$$

Summing over all $d|r$ we end up with a total contribution in equation (4) of

$$\ll (\log q)^{\frac{1}{2}+\frac{1}{2\sqrt{2}}+o(1)} + \sqrt{r}(\log q)^{\frac{2}{3}+o(1)} \ll (\log q)^{0.9},$$

which (as we've already discussed) may be absorbed.

Now let us consider the overall contribution from characters ψ that are induced by ξ. If $m \nmid r/d$ then there are no such characters ψ. By using the Gauss-sum formula from Exercise 6.12, we obtain a contribution of

$$\xi(b)\tau(\overline{\xi}) \sum_{d|r/m} \frac{\overline{\chi}(d)}{d\varphi(r/d)} \mu\Big(\frac{r}{dm}\Big)\overline{\xi}\Big(\frac{r}{dm}\Big) \sum_{k \leqslant K/d} \frac{(\overline{\chi}\psi)(k)}{k},$$

where ψ in this sum is assumed to be the unique character mod r/d that is induced by ξ.

Now we combine this term with the contribution from $-b$, getting an overall term of

$$(1-\xi(-1))\xi(b)\tau(\overline{\xi}) \sum_{d|r/m} \frac{\overline{\chi}(d)}{d\varphi(r/d)} \mu\Big(\frac{r}{dm}\Big)\overline{\xi}\Big(\frac{r}{dm}\Big) \sum_{k \leqslant K/d} \frac{(\overline{\chi}\psi)(k)}{k}.$$

This vanishes unless $\xi(-1) = -1$, so for all ψ induced by ξ we have $\chi(-1)\psi(-1) = \chi(-1)\xi(-1) = -1$. Applying the logarithmic Halasz theorem again (Lemma 2.12), we get a contribution of

$$\ll \frac{\sqrt{m}}{\varphi(m)} (\log q)^{o(1)} (1 + \log q \exp(-\max_{\psi} \frac{1}{2}\mathbb{D}(\chi,\psi;q)^2),$$

where the max is taken over all the ψ induced by ξ, with modulus dividing r. Using the other repulsion lemma (namely Lemma 7.7, which is valid since the modulus of ψ is at most $(\log q)^{1/3}$ and $\chi(-1)\psi(-1) = -1$, we have an overall upper bound of

$$\ll \frac{\sqrt{m}}{\varphi(m)} (\log q)^{1-\delta_g/2+o(1)} \ll (\log q)^{1-\delta_g/2+o(1)},$$

which gives the lemma. □

This settles Theorem 8.3 in the case where N/q may be reduced a fraction b/r with small denominator. Next lecture we will need to strengthen this result to cover the cases in which $N/q \approx b/r$, rather than being exactly equal, but that will be possible.

Our final remaining large task is to settle Theorem 8.3 when N/q is not well-approximated by such a fraction b/r, in other words the 'minor arc' case.

Minor arcs and the principal of bilinear sums

The study and estimation of bilinear sums is an enormous area of analytic number theory. In this course we will only see a tiny aspect of the power of these techniques: the interested reader should look into Vaughan's identity, Heath-Brown's identity, Iwaniec's work on the bilinear form of the error term in the linear sieve, the large sieve, the Bombieri–Vinogradov theorem, etc. etc. etc.

Consider the sum

$$\sum_{\substack{M<m\leqslant 2M\\ N<n\leqslant 2N}} \alpha_m \beta_n g(mn),$$

where α_m and β_n are arbitrary coefficients that satisfy $|\alpha_m| \leqslant 1$ and $|\beta_n| \leqslant 1$ for all m, n, and $|g(mn)| \leqslant 1$ for all m, n. The trivial bound on such a sum would be MN, and indeed this bound could be achieved if g is completely multiplicative say, since then the sum factorises as

$$\Big(\sum_{M<m\leqslant 2M} \alpha_m g(m)\Big)\Big(\sum_{N<n\leqslant 2N} \beta_n g(n)\Big),$$

and we get no cancellation if we choose the weights $\alpha_m = \overline{g(m)}$ and $\beta_n = \overline{g(n)}$. However, if g does not exhibit multiplicative structure, then one can sometimes do better, even for completely general weights α_m, β_n.

Note: there is a somewhat baroque terminology surrounding the theory of bilinear sums. A sum of the form given here is sometimes called a 'Type II' sum. If one of the weights α_m satisfies $\alpha_m \equiv 1$, or more generally if α_m is a 'nice' function of m (e.g. smooth, or $\alpha_m = \log m$), such a sum is sometimes called a 'Type I' sum.

Note: One can think of such a bound in terms of the spectral norm of an M-by-N matrix whose coefficients are given by $g(mn)$.

We will be interested in Type II sums of the form

$$\sum_{\substack{M<m\leqslant 2M\\ N<n\leqslant 2N}} \alpha_m\beta_n e(nm\theta).$$

Clearly if $\theta = 0$ then we do not except to see any cancellation in the general case, as the weight function is constant 1, and therefore multiplicative. Similarly, if $\theta = b/r$ with r small then by splitting into arithmetic progressions modulo r we can split the sum into a small number of multiplicative pieces, so again we shouldn't expect a great deal of cancellation in general. However, if r is not so small then we may achieve a saving. Next time, we will prove the following bound.

Lemma 8.5 (Type II sum bound)**.** *Let $\theta \in \mathbb{R}$. Suppose that there exist $b \in \mathbb{Z}$ and $r \in \mathbb{N}$ with $(b, r) = 1$ for which*

$$|\theta - \frac{b}{r}| \leqslant \frac{1}{r^2}.$$

Then, for any coefficients $\alpha_m, \beta_n \in \mathbb{C}$ with $|\alpha_m|, |\beta_n| \leqslant 1$ we have

$$\sum_{\substack{M<m\leqslant 2M\\ N<n\leqslant 2N}} \alpha_m\beta_n e(nm\theta) \ll MNr^{-1/2} + M^{1/2}N(\log r)^{1/2} + MN^{1/2} + M^{1/2}N^{1/2}r^{1/2}(\log r)^{1/2}.$$

Compare this to the trivial bound of MN. We win if r is neither too small nor too large, and if neither M nor N is too small.

9. Lecture 9: Finishing minor arcs and improving Polya–Vinogradov

Proof. We have

$$\begin{aligned}
\Big|\sum_{\substack{M<m\leqslant 2M\\ N<n\leqslant 2N}} \alpha_m\beta_n e(mn\theta)\Big|^2 &\leqslant M\Big|\sum_{M<m\leqslant 2M}\sum_{\substack{N<n_1\leqslant 2N\\ N<n_2\leqslant 2N}} \beta_{n_1}\overline{\beta_{n_2}} e(m(n_1-n_2)\theta)\Big|\\
&\leqslant M\sum_{\substack{N<n_1\leqslant 2N\\ N<n_2\leqslant 2N}}\Big|\sum_{M<m\leqslant 2M} e(m(n_1-n_2)\theta)\Big|\\
&\leqslant M\sum_{\substack{N<n_1\leqslant 2N\\ N<n_2\leqslant 2N}} \min(M, \|(n_1-n_2)\theta\|^{-1})\\
&= M\sum_{-N+1\leqslant \ell\leqslant N-1} (N-|\ell|)\min(M, \|\ell\theta\|^{-1})\\
&\leqslant MN\sum_{-N+1\leqslant \ell\leqslant N-1} \min(M, \|\ell\theta\|^{-1}).
\end{aligned}$$

Now, split the range of ℓ summation into $O(\frac{N}{r}+1)$ ranges of length at most r. Observe that for any ℓ_0 we have

$$\sum_{\ell_0\leqslant\ell\leqslant\ell_0+r-1} \min(M, \|\ell\theta\|^{-1}) = \sum_{0\leqslant k\leqslant r-1} \min(M, \|\ell_0\theta + k\theta\|^{-1})$$

$$= \sum_{0 \leqslant k \leqslant r-1} \min(M, \|\ell_0\theta + \frac{kb}{r} + O(1/r)\|^{-1}).$$

Since $(b,r) = 1$ we have that kb ranges through a complete set of residues modulo r. Therefore at most $O(1)$ of the terms k satisfy $\|\theta\ell_0 + \frac{kb}{r} + O(1/r)\| \leqslant 1/r$. Therefore the previous sum is

$$\ll M + \sum_{1 \leqslant |k| \leqslant r/2} \Big(\frac{k \bmod r}{r}\Big)^{-1} \ll M + r\log r.$$

Putting it all together we get an exponential sum bound of

$$\ll MN(\frac{N}{r} + 1)(M + r\log r),$$

which is

$$\ll \frac{M^2N^2}{r} + MN^2\log r + M^2N + MNr\log r.$$

Taking square-roots gives the lemma. □

We can apply this Type II bound to control the exponential sum of a multiplicative function, using a similar decomposition trick to the one we used in the proof of Halasz's theorem.

Lemma 9.1. *Let $f \in \mathcal{M}_0$ be a multiplicative function. Let $\theta \in \mathbb{R}$, and choose $b \in \mathbb{Z}$ and $r \in \mathbb{N}$ with $(b,r) = 1$ and*

$$\Big|\theta - \frac{b}{r}\Big| \leqslant \frac{1}{r^2}.$$

Then for all $X \geqslant 2$ we have

$$\sum_{n \leqslant X} f(n)e(n\theta) \ll \frac{X\log\log X}{(\log X)^{1/2}} + \frac{X}{r^{1/2}} + \frac{X(\log r)^{1/2}}{(\log X)^{100}} + X^{1/2}r^{1/2}(\log r)^{1/2}$$

and

$$\sum_{n \leqslant X} \frac{f(n)}{n} e(n\theta) \ll (\log X)^{1/2+o(1)} + \frac{\log X}{r^{1/2}} + (\log\log X)(\log r)^{1/2} + \log r.$$

Furthermore,

$$\sum_{r \leqslant n \leqslant X} \frac{f(n)}{n} e(n\theta) \ll (\log X)^{1/2+o(1)} + \frac{\log X}{r^{1/2}} + (\log\log X)(\log r)^{1/2} + (\log r)^{1/2}.$$

The bounds here are reasonably disgusting, in part owing to the fact that we have sought the simplest proof, rather than the cleanest final statement. To make sense of them, as always one should consider the bounds in the context of the trivial upper bounds of X and $\log X$ respectively. In the first case, if r is a small power of $\log X$ (e.g. $r \approx (\log X)^{1/3}$) then we will save a small power of $\log X$ over the trivial bound.

Montgomery–Vaughan proved an essentially optimal version of the first case of this lemma, in which the $X(\log\log X)(\log X)^{-1/2}$ term is replaced by a $X(\log X)^{-1}$ term. Note that this bound is unimprovable, at least in the case of general 1-bounded multiplicative functions, since potentially it could be that the multiplicative function f satisfies $f(p) = e(-p\theta)$ for all primes p in the range $X/2 < p \leqslant X$, and $f(p^k) = 0$ otherwise.

Proof. Let $\varepsilon > 0$ to be chosen later and, in a similar move to the one we used in the proof of Halasz's theorem, we write

$$f_{small}(n) = \begin{cases} f(n) & \text{if } p|n \Rightarrow p \leqslant X^{\varepsilon} \\ 0 & \text{otherwise} \end{cases}$$

and

$$f_{large}(n) = \begin{cases} f(n) & \text{if } n \in [2, X] \text{ and } p|n \Rightarrow p > X^{\varepsilon} \\ 0 & \text{otherwise.} \end{cases}$$

Let $f_{split} = f_{small} \star f_{large}$. Then we know that

$$\begin{aligned} \sum_{n \leqslant X} f(n)e(n\theta) &= \sum_{n \leqslant X} f_{split}(n)e(n\theta) + O(\varepsilon X) \\ &= \sum_{ab \leqslant X} f_{small}(a) f_{large}(b) e(ab\theta) + O(\varepsilon X). \end{aligned}$$

By the preceding lemma, we have that for any M, N,

$$\begin{aligned} &\sum_{\substack{M < a \leqslant 2M \\ N \leqslant b < 2N}} f_{small}(a) f_{large}(b) e(ab\theta) \\ &\qquad \ll MNr^{-1/2} + M^{1/2}N(\log r)^{1/2} + MN^{1/2} + M^{1/2}N^{1/2}r^{1/2}(\log r)^{1/2}. \end{aligned}$$

So it remains to split up the summation range $ab \leqslant X$ into suitable rectangles $\{M < a \leqslant 2M,\ N < b \leqslant 2N\}$, together with a small error. From the limited support of f_{large} we can assume that $b > X^{\varepsilon}$. To that end, let I denote the greatest natural number i for which $2^i X^{\varepsilon} \leqslant X(\log X)^{-100}$. Now, from the preceding lemma, for any $i, j \geqslant 0$ we have

$$\begin{aligned} &\sum_{2^i X^{\varepsilon} < b \leqslant 2^{i+1} X^{\varepsilon}} f_{large}(b) \sum_{2^j < a \leqslant \min(2^{j+1}, X/b)} f_{small}(a) e(ab\theta) \\ &\ll X^{\varepsilon} 2^{i+j} r^{-\frac{1}{2}} + 2^{i+\frac{j}{2}} X^{\varepsilon} (\log r)^{\frac{1}{2}} + 2^{\frac{i}{2}+j} X^{\frac{\varepsilon}{2}} + 2^{\frac{i+j}{2}} X^{\frac{\varepsilon}{2}} r^{\frac{1}{2}} (\log r)^{\frac{1}{2}}. \end{aligned}$$

(Note that we can treat an incomplete interval $2^j < a \leqslant X/b$ like a dyadic interval $2^j < a \leqslant 2^{j+1}$ simply by extending the weight function to be 0 on the rest of the range $X/b < a \leqslant 2^{j+1}$.)

Summing over all $j \geqslant 0$ such that $2^j \leqslant X/(2^{i+1}X^{\varepsilon})$ and over all i in the range $0 \leqslant i \leqslant I$, we end up with a bound of

$$Xr^{-1/2} + X(\log X)^{-100}(\log r)^{1/2} + X^{1-\varepsilon/2} + X^{1/2}r^{1/2}(\log r)^{1/2}.$$

Note that for all such i and j, and with a and b in the dyadic summation ranges given above, $ab \leqslant X$.

What remains is to bound the contribution from the outstanding pairs a, b with $ab \leqslant X$ that we haven't accounted for yet, namely those pairs for which $b \geqslant X(\log X)^{-100}$. In this instance we can use the Brun–Titchmarsh theorem to deduce that

$$\begin{aligned} \sum_{X(\log X)^{-100}/2 < b \leqslant X} |f_{large}(b)| \sum_{a \leqslant X/b} |f_{small}(a)||e(ab\theta)| &\ll \sum_{\substack{X(\log X)^{-100}/2 < b \leqslant X \\ p|b \Rightarrow p \geqslant X^{\varepsilon}}} \frac{X}{b} \\ &\ll \sum_{k=0}^{200 \log\log X} \sum_{\substack{2^{-k+1}X < b \leqslant 2^{-k}X \\ p|b \Rightarrow p \geqslant X^{\varepsilon}}} \frac{X}{b} \\ &\ll X \log\log X (\log X^{\varepsilon})^{-1} \\ &\ll \frac{X \log\log X}{\varepsilon \log X}. \end{aligned}$$

So all in all, we have a bound of

$$\ll \varepsilon X + \frac{X \log\log X}{\varepsilon \log X} + \frac{X}{r^{1/2}} + \frac{X(\log r)^{1/2}}{(\log X)^{100}} + X^{1-\varepsilon/2} + X^{1/2}r^{1/2}(\log r)^{1/2}.$$

We can always absorb the $X^{1-\varepsilon/2}$ term, and so, picking $\varepsilon = (\log X)^{-1/2}$, we get a bound of

$$\ll \frac{X \log\log X}{(\log X)^{1/2}} + \frac{X}{r^{1/2}} + \frac{X(\log r)^{1/2}}{(\log X)^{100}} + X^{1/2} r^{1/2} (\log r)^{1/2}$$

as claimed in the first part.

To prove the second part, we apply partial summation. Then

$$\begin{aligned}
\sum_{n \leqslant X} \frac{f(n)}{n} e(n\theta) &= \sum_{n<r} \frac{f(n)}{n} e(n\theta) + \sum_{r \leqslant n \leqslant X} \frac{f(n)}{n} e(n\theta) \\
&\leqslant (\log r) + \frac{1}{X} \sum_{r \leqslant n \leqslant X} f(n) e(n\theta) + \int_r^X \frac{\sum_{r \leqslant n \leqslant t} f(n) e(n\theta)}{t^2}\, dt \\
&\ll 1 + \log r + \int_r^X \frac{\log\log t}{t(\log t)^{1/2}} + \frac{1}{t r^{1/2}} + \frac{(\log r)^{1/2}}{t(\log t)^{100}} + \frac{r^{1/2}(\log r)^{1/2}}{t^{3/2}}\, dt \\
&\ll \log r + (\log X)^{1/2+o(1)} + \frac{\log X}{r^{1/2}} + \log\log X (\log r)^{1/2} + (\log r)^{1/2} \\
&\ll \log r + (\log X)^{1/2+o(1)} + \frac{\log X}{r^{1/2}} + \log\log X (\log r)^{1/2}
\end{aligned}$$

as claimed. □

Proof of Theorem 8.3

Now, finally, we can enact the major arc/minor arc distinction promised from last lecture, and prove Theorem 8.3. It is not supposed to be obvious why the thresholds s and S work well in the analysis to follow. When trying to come up with such arguments yourself, you should run the argument first with arbitrary parameters, and then towards the end of this process you will get a feel for what a suitable choice will be (there is certainly some flexibility in the parameters).

Let $s = (\log q)^{1/3}$ and $S = \lfloor \exp((\log q)^{5/6}) \rfloor$. Let us recall a foundational result in the theory of diophantine approximation.

Lemma 9.2 (Dirichlet's approximation theorem)**.** *For all $\theta \in \mathbb{R}$, there exists $b \in \mathbb{Z}$ and $r \in \mathbb{N}$ for which $(b, r) = 1$, $r \leqslant S$ and*

$$\left| \theta - \frac{b}{r} \right| < \frac{1}{rS}.$$

Proof. Consider $0, \theta, 2\theta, \dots, S\theta \bmod 1$. By the pigeonhole principle, there exists distinct r_1, r_2 in the range $0 \leqslant r_1, r_2 \leqslant S$ and some $i \in \mathbb{Z}_{\geqslant 0}$ with $0 \leqslant i \leqslant S-1$ for which

$$r_1\theta, r_2\theta \bmod 1 \in [i/S, (i+1)/S).$$

Then $\| \, |r_1 - r_2|\theta \, \| = \|r_1\theta - r_2\theta\| < 1/S$. Letting $r = |r_1 - r_2|$ and letting b be the nearest integer to $r\theta$, we conclude that

$$|r\theta - b| < \frac{1}{S}.$$

This is nearly as required, except that b and r might have a non-trivial common factor. However, if $b/r = b^*/r^*$ in lowest terms, we have

$$\left| \theta - \frac{b^*}{r^*} \right| = \left| \theta - \frac{b}{r} \right| < \frac{1}{rS} \leqslant \frac{1}{r^* S}.$$

This settles the lemma. □

We say that $\theta \in [0,1]$ is in a *minor arc* if for all $r \leqslant S$ for which there exists a $b \in \mathbb{Z}$ with $(b,r)=1$ and $|\theta - b/r| \leqslant 1/rS$ we have $r > s$. Otherwise we say that θ is in a *major arc*.

Minor arc case

Suppose that N/q is in a minor arc. By Dirichlet's approximation theorem we can find some $r \leqslant S$ for which there exists a $b \in \mathbb{Z}$ with $(b,r)=1$ and $|N/q - b/r| \leqslant 1/rS \leqslant 1/r^2$. By the minor arc assumption, $s < r \leqslant S$. Therefore, by applying Lemma 9.1 to the phase $\theta := N/q$, we have that

$$\begin{aligned}\sum_{k\leqslant q}\frac{\overline{\chi}(k)}{k}e\Big(\frac{kN}{q}\Big) &\ll \log r + (\log q)^{1/2+o(1)} + \frac{\log q}{r^{1/2}} + \frac{\log\log q}{(\log r)^{1/2}}\\ &\ll \log S + (\log q)^{1/2+o(1)} + \frac{\log q}{s^{1/2}} + \frac{\log\log q}{(\log s)^{1/2}}\\ &\ll (\log q)^{5/6+o(1)} \ll (\log q)^{0.9},\end{aligned}$$

so may be absorbed.

Major arc case

Now suppose that N/q is in a major arc. Again we can find some $r \leqslant S$ for which there exists a $b \in \mathbb{Z}$ with $(b,r)=1$ and $|N/q - b/r| \leqslant 1/rS \leqslant 1/r^2$. By the major arc assumption, $r \leqslant s$. Following on from the work of last lecture, we will be done after the next lemma:

Lemma 9.3 (Major arc approximations)**.** *Under the major arc assumptions, we have*

$$\sum_{k\leqslant q}\frac{\overline{\chi}(k)}{k}e\Big(\frac{kN}{q}\Big) = \sum_{k\leqslant K}\frac{\overline{\chi}(k)}{k}e\Big(\frac{kb}{r}\Big) + O((\log q)^{1/2+o(1)}),$$

where $K = \min(q, |\frac{rN}{q} - b|^{-1})$.

Proof. If $K = q$, then the lemma is immediate, as

$$\begin{aligned}\Big|\sum_{k\leqslant q}\frac{\overline{\chi}(k)}{k}e\Big(\frac{kN}{q}\Big) - \sum_{k\leqslant q}\frac{\overline{\chi}(k)}{k}e\Big(\frac{kb}{r}\Big)\Big| &\leqslant \sum_{k\leqslant q}\frac{1}{k}\Big|e\Big(\frac{kN}{q}\Big) - e\Big(\frac{kb}{r}\Big)\Big|\\ &\ll \sum_{k\leqslant q}\frac{1}{k}k\Big|\frac{N}{q} - \frac{b}{r}\Big|\\ &\leqslant \sum_{k\leqslant q}\frac{1}{rq} \leqslant 1.\end{aligned}$$

Otherwise we have $K = |\frac{rN}{q} - b|^{-1} \geqslant S$. Then by the same argument as above we have

$$\sum_{k\leqslant K}\frac{\overline{\chi}(k)}{k}e\Big(\frac{kN}{q}\Big) = \sum_{k\leqslant K}\frac{\overline{\chi}(k)}{k}e\Big(\frac{kb}{r}\Big) + O(1),$$

so it remains to show that

$$\sum_{K<k\leqslant q}\frac{\overline{\chi}(k)}{k}e\Big(\frac{kN}{q}\Big) = O((\log q)^{1/2+o(1)}).$$

By Dirichlet's theorm again, we may pick $r_1 \leqslant K$ and find a b_1 with $(b_1, r_1) = 1$ and $|\frac{N}{q} - \frac{b_1}{r_1}| < 1/r_1K$. Then we cannot have $b/r = b_1/r_1$, since then we would have $1/rK = |\frac{N}{q} - \frac{b}{r}| = |\frac{N}{q} - \frac{b_1}{r_1}| < 1/r_1K$, contradicting the fact that $r = r_1$. So

$$\frac{1}{rr_1} \leqslant \Big|\frac{b}{r} - \frac{b_1}{r_1}\Big| \leqslant \Big|\frac{b}{r} - \frac{N}{q}\Big| + \Big|\frac{N}{q} - \frac{b_1}{r_1}\Big| \leqslant \frac{1}{rK} + \frac{1}{r_1K},$$

and this implies that $K \leqslant r_1 + r \leqslant r_1 + s$, so $K - s \leqslant r_1 \leqslant K$. Since $K \geqslant S$, we have $K/2 \leqslant r_1 \leqslant K$.

Thus, applying Lemma 9.1 with the approximation b_1/r_1, we get

$$\begin{aligned}\sum_{K<k\leqslant q} \frac{\overline{\chi}(k)}{k} e\Big(\frac{kN}{q}\Big) &\ll \sum_{r_1\leqslant k\leqslant q} \frac{\overline{\chi}(k)}{k} e\Big(\frac{kN}{q}\Big)\\ &\ll (\log q)^{1/2+o(1)} + \frac{\log q}{r_1^{1/2}} + \frac{\log\log q}{(\log r)^{1/2}} + (\log r_1)^{1/2}\\ &\ll (\log q)^{1/2+o(1)} + \frac{\log q}{K^{1/2}} + \frac{\log\log q}{(\log K)^{1/2}} + (\log K)^{1/2}\\ &\ll (\log q)^{1/2+o(1)}.\end{aligned}$$

This settles the lemma. □

Since $1/2 < 0.9$ we can absorb this $O((\log q)^{1/2+o(1)})$ error term. Finally, we have already established in Lemma 8.4 that

$$\max_{K\leqslant q} \Big| \sum_{k\leqslant K} \frac{\overline{\chi}(k)}{k} \Big(e\Big(-\frac{kb}{r}\Big) - e\Big(\frac{kb}{r}\Big)\Big)\Big| \ll (\log q)^{1-\delta_g/2+o(1)}.$$

Putting all the cases together, we conclude Theorem 8.3, and hence finally we have prove Theorem 7.3 of Granville and Soundararajan.

10. Almost-all short intervals

Introduction

After some intense and detailed lectures last week, and following the general feedback that has been given to the faculty about Part III courses, I have (slightly) changed my plan for the remaining lectures of this course. We will still cover the intended topics, but not in equal detail. Rather, I will spend 5-6 of the remaining lectures describing the groundbreaking work of Matomäki–Radziwiłł on the averages of multiplicative functions in short intervals. The final 1-2 lectures will be spent on sketching the main ideas from Tao's (equally groundbreaking) work on correlations of multiplicative functions – discussing the main ideas but not giving all details.

Without further ado, here is the spectacular theorem that we will be addressing for the next three weeks or so.

Theorem 10.1 (Matomäki–Radziwiłł 2015)**.** *Let $f \in \mathcal{M}_0$ be a real-valued multiplicative function, and let X be an asymptotic parameter tending to infinity. Let $H = H(X)$ be a function with $H \to \infty$ as $X \to \infty$ and $H(X) \leqslant X$. Then, for all but $o(X)$ natural numbers $x \in [X, 2X)$*

$$\Big|\frac{1}{H}\sum_{x<n\leqslant x+H} f(n) - \frac{1}{X}\sum_{X<n\leqslant 2X} f(n)\Big| = o(1).$$

The error term is independent of the choice of f.

In words, in 'almost-all' short intervals, the short-average of f agrees asymptotically with the long average of f.

There is an equivalent version with finite thresholds – we leave the proof of equivalence as an exercise.

Theorem 10.2. *Let $\varepsilon > 0$. Then there exist positive parameters $X_0(\varepsilon)$ and $H_0(\varepsilon)$, depending only on ε, for which the following is true: if $f \in \mathcal{M}_0$ is a real-valued multiplicative function, then for all $X \geqslant X_0(\varepsilon)$ and for all H in the range $H_0(\varepsilon) \leqslant H \leqslant X$ we have*

$$\Big|\frac{1}{H}\sum_{x<n\leqslant x+H} f(n) - \frac{1}{X}\sum_{X<n\leqslant 2X} f(n)\Big| \leqslant \varepsilon$$

for at least $(1-\varepsilon)X$ integer values of $x \in [X, 2X)$.

As you might expect, Matomäki–Radziwiłł proved an explicit quantitative version of this theorem, with explicit dependence of the $o(X)$ and $o(1)$ terms on X and on $H(X)$. However, in keeping with our general philosophy of this course, we are concerned more with qualitative mean-value results than quantitative results. The restriction to real-valued multiplicative functions is also not entirely necessary, and can be replaced by the assumption that $f \in \mathcal{M}_0$ is not n^{it} pretentious for any $t \neq 0$. It would be a very good way of testing your understanding of the material to go through the lectured proof for real-valued multiplicative functions and adapt it to non-pretentious complex-valued multiplicative functions.

As I mentioned in Lecture 1, this result is wildly stronger than anything that had been previously known, even assuming the Riemann hypothesis. It is also stronger than anything currently known about the distribution of $\Lambda(n)$ in short intervals, where the best unconditional result is

$$\Big|\frac{1}{H}\sum_{x<n\leqslant x+H} \Lambda(n) - \frac{1}{X}\sum_{X<n\leqslant 2X} \Lambda(n)\Big| = o(1)$$

for all but $o(X)$ natural numbers $x \in [X, 2X)$, provided $H(X) > X^{1/6+\varepsilon}$.

Our exposition here will broadly follow that of Soundararajan from his Bourbaki seminar of 2016. His exposition is masterful, but not optimally suited for the present purposes. It is just a little too long for our limited time – he proves the theorem 5 times, in increasing generality – and I also feel it is also important to add some extra details and further explanations for the student (and in particular a proof of the Turán–Kubilius inequality, and more details about L^2 bounds of Dirichlet polynomials, rather than deferring this to an examples sheet).

Some corollaries

The work of Matomäki–Radziwiłł – as well as being a powerful tool in other even more sophisticated arguments (see the final lectures on correlations of multiplicative functions) – admits some cute corollaries. Let us prove a couple of them.

Corollary 10.3 (Sign-changes)**.** *The Liouville function enjoys a positive proportion of sign changes: precisely, there exists a $\delta > 0$ such that for all large enough N there is a $K \geqslant \delta N$ and integers $1 \leqslant n_1 < n_2 < \cdots < n_K \leqslant N$ such that $\lambda(n_j)\lambda(n_{j+1}) < 0$ for all $1 \leqslant j \leqslant K-1$.*

Proof. This is almost immediate. Indeed, choose $\varepsilon > 0$ to be small. Letting $H = H(\varepsilon)$ be large enough, and assuming $N \geqslant N(\varepsilon) \geqslant H$ is large enough, by Theorem 10.2 and the prime number theorem we see that for all but εN integers $x \in (N, 2N]$ we have

$$\sum_{x<n\leqslant x+H} \lambda(n) \leqslant \varepsilon H.$$

Call such an interval *good* (otherwise call the interval *bad*). If $I := (x, x+H]$ is good then there exists some $n_x^+ \in I$ for which $\lambda(n_x^+) = 1$ and some $n_x^- \in I$ for which $\lambda(n_x^-) = -1$.

It remains to find a large collection of disjoint good intervals. To this end, consider the partition of (a subset of) $(N, 2N]$ into intervals of the form $(N + 2kH, N + 2(k+1)H]$ for $k \in \mathbb{Z}_{\geqslant 0}$ ranging from 0 to $N/100H$. If ε is small enough, it must be that $\gg N/H$ of the intervals $(N+2kH, N+2(k+1)H]$ contain a good interval. Indeed, if not then there would be $\gg H \cdot (N/H) = N$ bad intervals, contrary to assumption.

Since each good interval gives us at least one sign change, and the intervals above are disjoint by construction, we have $\gg N/H \gg_\varepsilon N$ sign changes in total. □

Corollary 10.4. *For every $\varepsilon > 0$, there is a constant $C(\varepsilon)$ such that for all large enough N, the interval $[N, N + C(\varepsilon)\sqrt{N}]$ contains an integer that is N^ε-friable.*

Contrast this result with what is known for primes. It is known by work of Baker–Harman–Pintz that we can find a prime in every interval of the form $[N, N + N^{0.525}]$ (provided N is large enough), but this 0.525 is extremely hard won! Note that the prime number theorem by itself can't even show that there always exists primes in the range $[N, N + N^{0.999}]$.

A result of strength $[N, N + N^{0.5+\varepsilon}]$ is only known under the Lindelhöf hypothesis (which is the statement that $|\zeta(1/2 + it)| \ll_\varepsilon |t|^\varepsilon$ for every $\varepsilon > 0$). And even the Riemann hypothesis only gets you to $[N, N + N^{1/2} \log N]$.

(Actually, Matomäki–Radziwiłł manage to prove an asymptotic for the number of N^ε-friables in intervals of size $H\sqrt{N}$ for any $H = H(N) \to \infty$, but we won't present the proof of this.)

To prove the above corollary we need to know something about the averages of the indicator function of friable numbers in long intervals. Many precise statements are known, but we satisfy ourselves with a weak and easy-to-prove result.

Lemma 10.5. *Let*

$$s_{N^\varepsilon}(n) = \begin{cases} 1 & \text{if } p|n \Rightarrow p \leqslant N^\varepsilon \\ 0 & \text{otherwise} \end{cases}$$

be the indicator function of the N^ε-friable numbers. Then

$$\sum_{N<n\leqslant 2N} s_{N^\varepsilon}(n) \gg_\varepsilon N.$$

Proof. Without loss of generality we assume that $\varepsilon = 1/k$ for some large integer k, and set $\delta = \varepsilon - \varepsilon^2/2$. If $n \in (N, 2N]$ has the form $mp_1 \dots p_k$, where $N^\delta \leqslant p_1 \leqslant \dots \leqslant p_k \leqslant N^\varepsilon$, then $s_{N^\varepsilon}(n) = 1$, since

$$m \leqslant 2N^{1-k\delta} = 2N^{\varepsilon/2}$$

implies that m is N^ε-friable too. Since $m < p_1$, such a number n has a unique representation in this form, up to possible reordering of the $p_1, \dots, p_k$ if some of these primes happen to be equal. Hence

$$\begin{aligned} \sum_{n\leqslant N} s_{N^\varepsilon}(n) &\geqslant \frac{1}{k!} \sum_{N^\delta\leqslant p_1\leqslant\dots\leqslant p_k\leqslant N^\varepsilon} \ \sum_{N/(p_1\cdots p_k)<m\leqslant 2N/(p_1\cdots p_k)} 1 \\ &\geqslant \frac{1}{k!} \sum_{N^\delta\leqslant p_1\leqslant\dots\leqslant p_k\leqslant N^\varepsilon} \Big(\frac{N}{p_1\dots p_k} - 1\Big). \end{aligned}$$

The number of terms in the sum is at most

$$\Big(\sum_{p\leqslant N^\varepsilon} 1\Big)^k \ll \Big(\frac{N^\varepsilon}{(\log(N^\varepsilon))}\Big)^k = o_\varepsilon(N)$$

by Chebyschev's bounds, so the contribution from the -1 terms can be ignored.

The rest of the sum is

$$\geqslant \frac{N}{(k!)^2}\Big(\sum_{N^\delta\leqslant p\leqslant N^\varepsilon} \frac{1}{p}\Big)^k.$$

The most basic form of Mertens estimate for $\sum_{p\leqslant X} 1/p = \log\log X + O(1)$ is a bit too weak to apply here (as the $O(1)$ terms could conspire against us), but we can conclude from partial summation and Chebyschev's bounds that

$$\sum_{N^\delta\leqslant p\leqslant N^\varepsilon} \frac{1}{p} \gg \sum_{j=0}^{J} \sum_{N^\varepsilon 2^{-j-1}<p\leqslant 2^{-j}N^\varepsilon} \frac{1}{p} \gg \sum_{j=0}^{J} \frac{2^{-j}N^\varepsilon}{\log N}\frac{1}{2^{-j}N^\varepsilon} \gg \frac{J}{\log N},$$

where

$$J \gg \log(N^\varepsilon/N^\delta) = (\varepsilon - \varepsilon + \varepsilon/2)\log N \gg_\varepsilon \log N.$$

So

$$\sum_{N^\delta\leqslant p\leqslant N^\varepsilon} \frac{1}{p} \gg_\varepsilon 1,$$

and this settles the lemma. □

Proof of corollary. Without loss of generality we may assume that ε is sufficiently small. Now, since $s_{N^\varepsilon}(n) \in \mathcal{M}_0$ and real-valued, we can apply Theorem 10.2 to this function. By the preceding lemma we have $\frac{1}{\sqrt{N}} \sum_{\sqrt{N}<n\leqslant 2\sqrt{N}} s_{N^\varepsilon}(n) \gg_\varepsilon 1$.

Now, if $N \geqslant N_0(\varepsilon)$ and $H(\varepsilon)$ is large enough, the exceptional set

$$\mathcal{E} := \{x \in [\sqrt{N}/2, 2\sqrt{N}] : \text{ the interval } [x, x + H(\varepsilon)] \text{ contains no } N^\varepsilon\text{-friable number}\},$$

has measure $|\mathcal{E}| \ll \varepsilon\sqrt{N}$. This is since for all $x \in \mathcal{E}$ we have, replacing x by the nearest integer to x if necessary,

$$\Big|\frac{1}{H(\varepsilon)}\sum_{x<n\leqslant x+H(\varepsilon)} s_{N^\varepsilon}(n) - \frac{1}{\sqrt{N}}\sum_{\sqrt{N}<n\leqslant 2\sqrt{N}} s_{N^\varepsilon}(n)\Big| \geqslant \Big|\frac{1}{\sqrt{N}}\sum_{\sqrt{N}<n\leqslant 2\sqrt{N}} s_{N^\varepsilon}(n)\Big| - O\Big(\frac{1}{H(\varepsilon)}\Big) \geqslant g(\varepsilon) > 0$$

for some function g. Now apply Theorem 10.2 with the value $g(\varepsilon)$.

If for some $x \in [\sqrt{N}, 2\sqrt{N}]$ we had both $x \notin \mathcal{E}$ and $N/x \notin \mathcal{E}$, then we would be able to find N^ε-friable numbers in $[x, x+H(\varepsilon)]$ and also in $[N/x, N/x+H(\varepsilon)]$ and their product would be in $[N, N+4H(\varepsilon)\sqrt{N}]$ (and therefore the corollary would be satisfied). Thus, were the corollary to fail, we would have to have

$$N^{1/2} \leqslant \int_{\sqrt{N}}^{2\sqrt{N}} (1_{\mathcal{E}}(x) + 1_{\mathcal{E}}(N/x))\, dx \leqslant 4|\mathcal{E}| \leqslant 4\varepsilon N^{1/2},$$

which is a contradiction if ε is small enough.

□

Converting to an L^2 estimate

Now let us start our journey towards Theorem 10.1. Let us assume that f is not 1-pretentious, i.e. that $\mathbb{D}(f, 1; \infty) = \infty$. Since f is real, we have the observation from Lecture 3 that $\mathbb{D}(f, n^{it}; \infty) = \infty$ for all $t \neq 0$. Hence, by Halasz's theorem we know that

$$\frac{1}{X}\sum_{X<n\leqslant 2X} f(n) = \frac{1}{X}\Big(\sum_{n\leqslant 2X} f(n) - \sum_{n\leqslant X} f(n)\Big) = 2M_{2X}(f) - M_X(f) = o(1).$$

Therefore our task is to show that for at least $X - o(X)$ integers $x \in (X, 2X]$ we have

$$\Big|\frac{1}{H}\sum_{x<n\leqslant x+H} f(n)\Big| = o(1)$$

as $X \to \infty$. If it helps you, for most of what we say you can assume that $f = \lambda$. Unfortunately, as we have already discussed when mentioning the quantitative aspects of Halasz's theorem at the end of Lecture 5, λ does enjoy some stronger estimates on its partial sums than are enjoyed by multiplicative functions in general. This means that one central idea to the general proof is not required (but that won't come up for another three lectures, so perhaps best not to worry at this point).

We will follow our 'four-step programme' from Lecture 1. For Step 1, we need to express the theorem we want to prove in terms of estimating a particular sum.

Lemma 10.6. *Suppose that for all $f \in \mathcal{M}_0$ that are real-valued and not 1-pretentious we have*

$$\frac{1}{X}\int_X^{2X} \Big|\frac{1}{H}\sum_{x<n\leqslant x+H} f(n)\Big|^2 dx = o(1).$$

Then Theorem 10.1 follows (for the same f).

The idea, of proving 'almost–all' results by deducing them from L^2 average, is an exceptionally common one (that I have used in several of my papers in fact!). In probabilistic language, we are making use of Chebyschev's inequality.

Proof. Let $\varepsilon > 0$, and let X and H be large enough so that

$$\frac{1}{X}\int_X^{2X} \Big|\frac{1}{H}\sum_{x<n\leqslant x+H} f(n)\Big|^2\, dx \leqslant \varepsilon^3.$$

Then let $\mathcal{E} \subset [X, 2X]$ be the set of exceptional $x \in [X, 2X]$ for which $\frac{1}{H}|\sum_{x<n\leqslant x+H} f(n)| \geqslant \varepsilon$. We then have that

$$|\mathcal{E}| \ll \varepsilon^3\varepsilon^{-2}X = \varepsilon X.$$

Theorem 10.2 follows immediately, and so Theorem 10.1 too. □

Our task is then to estimate

$$\frac{1}{X}\int_X^{2X} \Big|\frac{1}{H}\sum_{x<n\leqslant x+H} f(n)\Big|^2\, dx.$$

11. Dirichlet polynomials I

Transformation to an integral

Step 2 of our rubric is to transform the sum from Step 1 into an integral over 'frequency space'. We will make use of the following version of Parseval's theorem (which is very similar to the first lemma of Lecture 5).

Lemma 11.1 (Parseval). *Let $f : \mathbb{N} \longrightarrow \mathbb{C}$ be a finitely supported function, and define the associated Dirichlet polynomial $F(t) := \sum_{n \geqslant 1} f(n)n^{it}$. Let $T \geqslant 1$ be a real number. Then*

$$\int_0^\infty \Big| \sum_{xe^{-1/T} < n \leqslant xe^{1/T}} f(n) \Big|^2 \frac{dx}{x} = \frac{2}{\pi} \int_{-\infty}^\infty |F(t)|^2 \Big(\frac{\sin(t/T)}{t} \Big)^2 dt.$$

NB: the measure $\frac{dx}{x}$ is the natural translation invariant measure on the multiplicative group $(\mathbb{R}_{>0}, \times)$. This lemma is basically just writing out explicitly the convolution identity for Fourier transforms on this group.

Proof. For any real number u put

$$g(u) = \sum_{e^{u-\frac{1}{T}} < n \leqslant e^{u+\frac{1}{T}}} f(n).$$

Then

$$\begin{aligned} \widehat{g}(t) = \int_{-\infty}^\infty g(u)e(-tu)\,du &= \sum_n f(n) \int_{\log n - 1/T}^{\log n + 1/T} e(-tu)\,du \\ &= \sum_n \frac{f(n)}{-2\pi i t} (e(-t(\log n + 1/T)) - e(-t(\log n - 1/T))) \\ &= F(-2\pi t) \frac{\sin(2\pi t/T)}{\pi t}. \end{aligned}$$

We have $g \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, so by Plancherel's theorem and the substitution $x = e^u$ we have

$$\begin{aligned} \int_0^\infty \Big| \sum_{xe^{-1/T} < n \leqslant xe^{1/T}} f(n) \Big|^2 \frac{dx}{x} = \int_{-\infty}^\infty |g(u)|^2\,du = \int_{-\infty}^\infty |\widehat{g}(t)|^2\,dt &= \int_{-\infty}^\infty |F(-2\pi t)|^2 \Big(\frac{\sin(2\pi t/T)}{\pi t} \Big)^2 dt \\ &= \frac{2}{\pi} \int_{-\infty}^\infty |F(t)|^2 \Big(\frac{\sin(t/T)}{t} \Big)^2 dt \end{aligned}$$

after a change of variables. □

Now, this lemma was not quite what we wanted, because the interval $xe^{-1/T} \leqslant n \leqslant xe^{1/T}$ is 'multiplicatively small' rather than 'additively small'. But, by a suitable averaging argument, we can get round this issue.

Lemma 11.2 (An averaging argument). *Let X be sufficiently large, and suppose $f : \mathbb{N} \longrightarrow \mathbb{C}$ is a function supported on $[X, 3X]$. Define the associated Dirichlet polynomial $F(t) := \sum_{n \geqslant 1} f(n)n^{it}$. Then for any real number H in the range $1 \leqslant H \leqslant X/10$, we have*

$$\frac{1}{X} \int_0^\infty \Big| \frac{1}{H} \sum_{x < n \leqslant x+H} f(n) \Big|^2 dx \ll \int_{-\infty}^\infty |F(t)|^2 \min\Big(\frac{1}{X^2}, \frac{1}{t^2H^2} \Big)\, dt.$$

Proof. For ease of notation, we temporarily define $\mathcal{F}(x) := \sum_{n\leqslant X} f(n)$. Note that, by the triangle inequality, for any $\nu \in [2H, 3H]$ we have

$$\int_0^\infty |\mathcal{F}(x+H) - \mathcal{F}(x)|^2\, dx \ll \int_0^\infty (|\mathcal{F}(x+\nu) - \mathcal{F}(x)|^2 + |\mathcal{F}(x+H) - \mathcal{F}(x+\nu)|^2)\, dx.$$

Integrating this over all ν in the range $2H \leqslant \nu \leqslant 3H$, we obtain that

$$\begin{aligned}
&H\int_{-\infty}^{\infty} \Big|\sum_{x<n\leqslant x+H} f(n)\Big|^2\, dx\\
&= \int_{2H}^{3H}\int_{-\infty}^{\infty} \Big|\sum_{x<n\leqslant x+H} f(n)\Big|^2\, dx\, d\nu\\
&\ll \int_{2H}^{3H}\int_{-\infty}^{\infty} |\mathcal{F}(x+\nu) - \mathcal{F}(x)|^2\, dx\, d\nu + \int_{2H}^{3H}\int_{-\infty}^{\infty} |\mathcal{F}(x+\nu) - \mathcal{F}(x+H)|^2\, dx\, d\nu\\
&\ll \int_{2H}^{3H}\int_{-\infty}^{\infty} |\mathcal{F}(x+\nu) - \mathcal{F}(x)|^2\, dx\, d\nu + \int_{2H}^{3H}\int_{-\infty}^{\infty} |\mathcal{F}(x+\nu-H) - \mathcal{F}(x)|^2\, dx\, d\nu\\
&\ll \int_{-\infty}^{\infty}\int_{H}^{3H} |\mathcal{F}(x+\nu) - \mathcal{F}(x)|^2\, d\nu\, dx,
\end{aligned}$$

by translating the ν variable in the second integral. Now in the inner integral over ν we substitute $\nu = \delta x$. It follows, from the limited supported of f and the bound $H \leqslant X/10$, that the above integral is

$$\begin{aligned}
&\ll \int_{X/2}^{3X}\int_{H/10X}^{10H/X} |\mathcal{F}(x(1+\delta)) - \mathcal{F}(x)|^2 x\, d\delta\, dx\\
&= \int_{H/10X}^{10H/X}\int_{X/2}^{3X} |\mathcal{F}(x(1+\delta)) - \mathcal{F}(x)|^2 x\, dx\, d\delta\\
&\ll HX \max_{H/10X\leqslant\delta\leqslant 10H/X} \int_{X/2}^{3X} |\mathcal{F}(x(1+\delta)) - \mathcal{F}(x)|^2\, \frac{dx}{x}.
\end{aligned}$$

Now, appealing to the previous lemma with $T = 2/\log(1+\delta)$, the present lemma follows, noting that

$$\Big(\frac{\sin(t/T)}{t}\Big)^2 \ll \min(1/T^2, 1/t^2).$$

□

So, to prove the main theorem (Theorem 10.1) when $H(X) \leqslant X/10$ it suffices to show that

$$\int_{-\infty}^{\infty} |F(t)|^2 \min\Big(\frac{1}{X^2}, \frac{1}{t^2H^2}\Big)\, dt = o(1)$$

where $F(t) = \sum_{X \leqslant n \leqslant 3X} f(n) n^{it}$ for $f \in \mathcal{M}_0$ real-valued. (The constraint $H \leqslant X/10$ will not be significant in what follows).

Let us put this bound into context. Remember Montgomery's mean value theorem bound from Lecture 4 (Theorem 4.11), which gives

$$\int_{-T}^{T} |F(t)|^2\, dt \ll (T+X) \sum_n |f(n)|^2 \ll (T+X)X.$$

Remember how the TX term expressed a principle of 'square-root cancellation on average' – one is not going to be able to improve this part – and the the X^2 corresponds to the possibility that $|F(t)|$ could be as large as X for a few values of t. By using more information about the structure of $F(t)$, there is a chance that we can improve the X^2 part.

It should be noted that Montgomery's mean value theorem can be used to establish the bound

$$\int_{-\infty}^{\infty} |F(t)|^2 \min\Big(\frac{1}{X^2}, \frac{1}{t^2H^2}\Big)\, dt = O(1),$$

so all we are looking for is the smallest possible amount of further cancellation in this integral. Indeed,

$$\int_{-X/H}^{X/H} |F(t)|^2 \min\Big(\frac{1}{X^2}, \frac{1}{t^2H^2}\Big)\, dt = \frac{1}{X^2} \int_{-X/H}^{X/H} |F(t)|^2 \ll \frac{1}{X^2}((X/H)+X)X) \ll 1,$$

and

$$\begin{aligned}
\int_{|t| \geqslant X/H} |F(t)|^2 \min\Big(\frac{1}{X^2}, \frac{1}{t^2H^2}\Big)\, dt &= \frac{1}{H^2} \int_{|t| \geqslant X/H} \frac{|F(t)|^2}{t^2}\, dt \\
&\ll \frac{1}{H^2} \sum_{k=0}^{\infty} \frac{H^2}{(2^k X)^2} \int_{2^k X/H \leqslant t \leqslant 2^{k+1}X/H} |F(t)|^2\, dt \\
&\ll \frac{1}{H^2} \sum_{k=0}^{\infty} \frac{H^2}{(2^k X)^2} \cdot X\Big(X + \frac{2^{k+1}X}{H}\Big) \\
&\ll 1.
\end{aligned}$$

Note in both cases that the 'square-root cancellation on average' term is actually $O(1/H)$, and is therefore $o(1)$. Furthermore note that, by the same method, we may discount any contribution to the integral that comes from

$$|t| \geqslant Xw(X)/H,$$

where $w(X) \to \infty$ as $X \to \infty$. Thus, more or less, we have replace a short average of length H in physical space with a long average of length X/H in frequency space. Contrast this with the proof of Halasz's theorem, we had a long average of length X in physical space, which we replaced with a short average of length T (for some very slowly growing function

T) in frequency space.

The rest of the action of the proof then moves into Step 3 of our four-step plan – the hard step! – when we have to understand the integrand $|F(t)|^2$ extremely well. There will be two main tools here:

- a more precise mean value theorem than Montgomery's all-purpose estimate, which will enable us to extract nearly square-root cancellation on average for the quantity $|F(t)|$, where the average is taken over a possibly much smaller and much more irregular set $\mathcal{E} \subset [-T, T]$ rather than the full interval $t \in [-T, T]$;

- an approximate factorisation $F(t) \approx F_1(t)F_2(t)\dots F_k(t)$, with which we will play L^∞ and L^2 bounds off against each other (in a not dissimilar way to our proof of Halasz's theorem, albeit within a much more intricate argument).

The factorisation will be the topic of next lecture. For the rest of today, we will prove the following mean value theorem.

Theorem 11.3 (Halasz–Montgomery bound)**.** *Let $f : \mathbb{N} \longrightarrow \mathbb{C}$ be an arithmetic function, let $T \geqslant 2$, and let $\mathcal{E} \subset [-T, T]$ be a measurable subset with measure $|\mathcal{E}|$. Then*

$$\int_{\mathcal{E}} \Big| \sum_{n \leqslant N} f(n) n^{it} \Big|^2 \, dt \ll (N + |\mathcal{E}| T^{1/2} \log T) \sum_{n \leqslant N} |f(n)|^2.$$

This bound is stronger than Montgomery's all-purpose mean value theorem provided $|\mathcal{E}| \ll T^{1/2} (\log T)^{-1}$.

There are a few different ways to handle the proof of this result – we will choose the most concrete way (avoiding the use of the 'duality principle', as there are already rather a lot of ideas in this course as it is!).

Proof. Let $F_N(t) = \sum_{n \leqslant N} f(n) n^{it}$. Then

$$\begin{aligned} I := \int_{\mathcal{E}} |F_N(t)|^2 \, dt &= \int_{\mathcal{E}} \sum_{n \leqslant N} \overline{f(n)} n^{-it} F_N(t) \, dt \\ &\leqslant \sum_{n \leqslant N} |f(n)| \Big| \int_{\mathcal{E}} F_N(t) n^{-it} \, dt \Big|. \end{aligned}$$

Using Cauchy–Schwarz we obtain

$$\begin{aligned} I^2 &\leqslant \Big(\sum_{n \leqslant N} |f(n)|^2 \Big) \Big(\sum_{n \leqslant N} \Big| \int_{\mathcal{E}} F_N(t) n^{-it} \, dt \Big|^2 \Big) \\ &\leqslant \Big(\sum_{n \leqslant N} |f(n)|^2 \Big) \Big(\sum_{n \leqslant 2N} \Big(2 - \frac{n}{N} \Big) \Big| \int_{\mathcal{E}} F_N(t) n^{-it} \, dt \Big|^2 \Big). \end{aligned}$$

We have seen this device, of introducing a smoother cut-off function, when we proved Montgomery's estimate.

Expanding out the square, and swapping the orders of summation and integration, we observe that the second term of the above is equal to

$$\int_{t_1, t_2 \in \mathcal{E}} F_N(t_1) \overline{F_N}(t_2) \sum_{n \leqslant 2N} \Big(2 - \frac{n}{N} \Big) n^{i(t_2 - t_1)} \, dt_1 \, dt_2. \tag{6}$$

Now, assuming the bound

$$\sum_{n\leqslant 2N}\Big(2-\frac{n}{N}\Big)n^{it}\ll\frac{N}{1+|t|^2}+(1+|t|)^{1/2}\log(2+|t|),\tag{7}$$

we may conclude quickly. Indeed, since $|F_N(t_1)||F_N(t_2)|\leqslant|F_N(t_1)|^2+|F_N(t_2)|^2$ we obtain that (6) is

$$\begin{aligned}&\ll\int\limits_{t_1,t_2\in\mathcal{E}}|F_N(t_1)||\overline{F_N}(t_2)|\Big(\frac{N}{1+|t_1-t_2|^2}+(1+|t_1-t_2|)^{1/2}\log(2+|t_1-t_2|)\Big)\,dt_1\,dt_2\\&\ll\int\limits_{t_1\in\mathcal{E}}|F_N(t_1)|^2\int\limits_{t_2\in\mathcal{E}}\Big(\frac{N}{1+|t_2-t_1|^2}+T^{1/2}\log T\Big)\,dt_2\Big)dt_1\\&\ll(N+|\mathcal{E}|T^{1/2}\log T)I.\end{aligned}$$

So

$$I^2\ll I(N+|\mathcal{E}|T^{1/2}\log T)\sum_{n\leqslant N}|f(n)|^2,$$

and the theorem follows by dividing through by I. □

For the rest of the lecture we will be focussed on proving the bound (7). The techniques we will cover here are not going to appear again in future lectures, so if you like you can think of this material as purely supplementary. However, these techniques do speak of more general principles for bounding exponentials sums that you will come across as you learn more analytic number theory.

We did prove some bound on $\sum_{n\leqslant N}n^{it}$ before, way back in lecture 2, using a simple partial summation argument, namely

$$\sum_{n\leqslant N}n^{it}=\frac{N^{1+it}}{1+it}+O(1+|t|\log N).$$

This implies

$$\Big|\sum_{n\leqslant N}n^{it}\Big|\ll\frac{N}{1+|t|}+O(1+|t|\log N),$$

and with the extra smoothing given by the $(2-n/N)$ term one could establish the bound

$$\Big|\sum_{n\leqslant 2N}\Big(2-\frac{n}{N}\Big)n^{it}\Big|\ll\frac{N}{1+|t|^2}+O(1+|t|\log N)$$

by similar means (the fast decay of $(1+|t|^2)^{-1}$ versus $(1+|t|)^{-1}$ comes from the extra smoothing). However, this error term is in no way sufficient for our later purposes. We will use a stronger argument than just basic partial summation, one which is based on the poisson summation formula.

First some preparatory lemmas.

Lemma 11.4 (First derivative bound)**.** *Let $r,\theta:[a,b]\longrightarrow\mathbb{R}$ be functions such that $r(x)$ is differentiable and θ is twice-differentiable. Suppose that $r(x)/\theta'(x)>0$ and $(r(x)/\theta'(x))'<0$ for all $x\in[a,b]$. Then*

$$\Big|\int\limits_a^b r(x)e(\theta(x))\,dx\Big|\ll\frac{r(a)}{\theta'(a)}.$$

If $r(x)/\theta'(x) < 0$ and $(r(x)/\theta'(x))' < 0$ for all $x \in [a,b]$ then

$$\Big| \int_a^b r(x)e(\theta(x))\,dx \Big| \ll -\frac{r(b)}{\theta'(b)}.$$

Proof. By integration by parts,

$$\begin{aligned}\int_a^b r(x)e(\theta(x))\,dx &= \int_a^b \frac{r(x)}{\theta'(x)}\theta'(x)e(\theta(x))\,dx \\ &= \Big[\frac{r(x)e(\theta(x))}{2\pi i\theta'(x)}\Big]_a^b - \frac{1}{2\pi i}\int_a^b e(\theta(x))\Big(\frac{r(x)}{\theta'(x)}\Big)'\,dx\end{aligned}$$

Therefore, by the triangle inequality and the assumptions of the lemma, we have

$$\begin{aligned}\Big| \int_a^b r(x)e(\theta(x))\,dx\Big| &\leqslant \frac{r(b)}{2\pi\theta'(b)} + \frac{r(a)}{2\pi\theta'(a)} + \frac{1}{2\pi}\int_a^b \Big|\Big(\frac{r(x)}{\theta'(x)}\Big)'\Big|\,dx \\ &= \frac{r(b)}{2\pi\theta'(b)} + \frac{r(a)}{2\pi\theta'(a)} - \frac{1}{2\pi}\int_a^b \Big(\frac{r(x)}{\theta'(x)}\Big)'\,dx \\ &= \frac{r(b)}{2\pi\theta'(b)} + \frac{r(a)}{2\pi\theta'(a)} - \frac{1}{2\pi}\Big(\frac{r(b)}{\theta'(b)} - \frac{r(a)}{\theta'(a)}\Big) \ll \frac{r(a)}{\theta'(a)}\end{aligned}$$

which gives the first half of the lemma. The second part follows by the same argument. □

Montgomery gives a nice geometric description of this lemma. Namely, considering the function $Z(t) := \int_a^t r(x)e(\theta(x))\,dx$ as a curve in the complex plane, the conditions of the lemma imply that the radius of curvature is decreasing, and thus the curve $Z(t)$ spirals inwards (and so is bounded by the radius of the initial osculating circle).

Lemma 11.5 (Second derivative bound)**.** *Let $f : [a,b] \longrightarrow \mathbb{R}$, and suppose that $0 < \lambda \leqslant f''(x)$ for all $x \in [a,b]$. Then*

$$\Big| \int_a^b e(f(x))\,dx\Big| \ll \frac{1}{\sqrt{\lambda}}.$$

Proof. Let $\delta > 0$ be a parameter to be chosen later, and let $\mathcal{J} := \{x \in [a,b] : |f'(x)| \leqslant \delta\}$. Since f' is monotonic, $\mathcal{J}$ is an interval of length $\ll \delta/\lambda$. We estimate the contribution from this range trivially, namely

$$\Big| \int_{\mathcal{J}} e(f(x))\,dx\Big| \ll \frac{\delta}{\lambda}.$$

The remaining portion of the integral (which consists of at most two intervals) can be estimated using the first derivative bound in Lemma 11.4, taking $r(x) = 1$ and $\theta(x) = f(x)$. Using this bound we get

$$\Big|_a^b e(f(x))\,dx\Big| \ll \frac{\delta}{\lambda} + \frac{1}{\delta},$$

which is $O(1/\sqrt{\lambda})$ if we take $\delta = \sqrt{\lambda}$. The lemma is settled. □

Lemma 11.5 is the beginning of a general method for estimating exponential sums and integrals known as the 'method of stationary phase', or the 'saddle point method'. In its most basic form, the idea is that when $f'(x)$ is close to zero we have that $f(x)$ is slowly

varying, and thus the phases $e(f(x))$ all point in roughly the same direction. Thus we do not expect cancellation from this portion of the integral; but when $|f'(x)|$ is large, $f(x)$ varies rapidly and one can hope for cancellation over the phase $e(f(x))$.

Now we come to the real meat. We will use a version of the Poisson summation formula, which says that if $r \in L^1(\mathbb{R})$ and has finitely many jump discontinuities (or more generally has bounded variation) then

$$\sum_{n\in\mathbb{Z}} \frac{r(n^+)+r(n^-)}{2} = \lim_{K\to\infty} \sum_{-K}^{K} \widehat{r}(k),$$

where $r(n^+)$ and $r(n^-)$ denote the upper and lower limits respectively.

Lemma 11.6 (Truncated Poisson summation)**.** *Let $f : [a,b] \longrightarrow \mathbb{R}$ be a twice differentiable function and assume that $f''(x) > 0$ for all $x \in [a,b]$. Then, writing $\alpha = f'(a)$ and $\beta = f'(b)$, we have*

$$\sum_{a\leqslant n\leqslant b} e(f(n)) = \sum_{\alpha-\frac{1}{2}\leqslant k\leqslant\beta+\frac{1}{2}} \int_a^b e(f(x)-kx)\,dx + O(\log(2+\beta-\alpha)).$$

Proof. As an initial manoeuvre, by replacing $f(x)$ by $f(x)-Nx$ (where N is the integer such that $|N-(\alpha+\beta)/2| \leqslant 1/2$), one may assume without loss of generality that $|\alpha+\beta| \leqslant 1$, i.e. $\alpha \approx -\beta$. We may also adjust a and b so that $\|2\pi a\| \geqslant 1/100$ and $\|2\pi b\| \geqslant 1/100$.

Now, consider the function

$$r(x) = \begin{cases} e(f(x)) & \text{if } x \in [a,b] \\ 0 & \text{otherwise.} \end{cases}$$

Then $r \in L^1(\mathbb{R})$, and r is continuous apart from the discontinuities at a and b, so we can apply the Poisson summation formula to conclude that

$$\sum_{n\in\mathbb{Z}} \frac{r(n^+)+r(n^-)}{2} = \lim_{K\to\infty} \sum_{-K}^{K} \widehat{r}(k).$$

Then, the sum on the left-hand side is within $O(1)$ of the sum we wish to estimate, and

$$\int_a^b e(f(x)-kx)\,dx = \widehat{r}(k).$$

Therefore the lemma is proved, provided we can show that

$$\Big|\sum_{\substack{k\notin[\alpha-\frac{1}{2},\beta+\frac{1}{2}]\\ |k|\leqslant K}} \widehat{r}(k)\Big| \ll \log(2+\beta-\alpha)$$

for all sufficiently large k.

Integration by parts yields

$$\widehat{r}(k) = \frac{e(f(a)-ka)}{2\pi ik} - \frac{e(f(b)-kb)}{2\pi ik} + \frac{1}{k}\int_a^b f'(x)e(f(x)-kx)\,dx.$$

We can use the first derivative bound in Lemma 11.4 to estimate the second integral. Indeed, if $k > \beta$ then $f'(x)/(f(x)-kx)' = f'(x)/(f'(x)-k)$ has the opposite sign to $f'(x)$. Furthermore, we have

$$\Big(\frac{f'(x)}{(f(x)-kx)'}\Big)' = \Big(\frac{f'(x)}{f'(x)-k}\Big)' = \frac{-kf''(x)}{(f'(x)-k)^2} < 0$$

since $k > 0$ and $f''(x) > 0$. Splitting into two intervals (one on which $f'(x) < 0$ and one on which $f'(x) > 0$), we use Lemma 11.4 to get the bounds

$$\Big|\frac{1}{k}\int\limits_{\substack{x\in[a,b]\\ f'(x)>0}} f'(x)e(f(x)-kx)\,dx\Big| \ll \frac{-f'(b)}{k(f'(b)-k)} = \frac{\beta}{k(k-\beta)},$$

plus some similar terms. The sum of this contribution is at msot

$$\sum_{k>\beta+\frac{1}{2}} \frac{\beta}{k(k-\beta)} \ll \sum_{\beta+\frac{1}{2}<k<2\beta+2} \frac{\beta}{k(k-\beta)} + \sum_{k\geqslant 2\beta+2} \frac{\beta}{k(k-\beta)} \ll \sum_{k\leqslant\beta+2} \frac{1}{k} + \sum_{k\geqslant 2\beta+2} \frac{\beta}{k^2} \ll \log(2+\beta).$$

Performing an analogous argument for $k < \alpha$, one ends up with

$$\sum_{\substack{k\notin[\alpha-1,\beta+1]\\ |k|\leqslant K}} \widehat{r}(k) = e(f(b)) \sum_{\beta+1<k\leqslant K} \frac{\sin 2\pi kb}{\pi k} - e(f(a)) \sum_{\beta+1<k\leqslant K} \frac{\sin 2\pi ka}{\pi k} + O(\log(2+\beta-\alpha)).$$

But partial summation shows that the Fourier series

$$\begin{aligned}\sum_{L\leqslant k\leqslant M} \frac{\sin 2\pi kb}{\pi k} &= \frac{1}{\pi M}\Big(\sum_{L\leqslant k\leqslant M} \sin 2\pi kb\Big) + \int\limits_L^M \Big(\sum_{L\leqslant k\leqslant y} \sin 2\pi kb\Big) y^{-2}\,dy \\ &= O(1) + \int\limits_L^M O(1/\|2\pi b\|) y^{-2}\,dy = O(1).\end{aligned}$$

Arguing analogously for $\sin 2\pi ka$, this completes the lemma. □

This lemma is part of a general technique of exponential sums called the 'van der Corput B process' – have at look at the theory of 'exponent pairs' if you're interested.

Finally we can prove the estimate (7).

Proof. Without loss of generality, by taking complex conjugates we can assume that $t < 0$, and thus

$$\Big(\frac{t\log x}{2\pi}\Big)'' = -\frac{t}{2\pi x^2} > 0$$

for all $x > 0$. Also we may assume that $|t| \geqslant 2$, since the estimate (7) is trivial otherwise. Using partial summation, we get

$$\sum_{n\leqslant 2N} \Big(2 - \frac{n}{N}\Big) n^{it} = \frac{1}{N}\int\limits_1^{2N} \sum_{n\leqslant y} n^{it}\,dy,$$

and by using truncated poisson summation as in the previous lemma we obtain

$$\begin{aligned}\sum_{y/2<n\leqslant y} n^{it} &= \sum_{y/2<n\leqslant y} e\Big(\frac{t\log n}{2\pi}\Big) \\ &= \sum_{\alpha-\frac{1}{2}\leqslant k\leqslant\beta+\frac{1}{2}y/2} \int\limits^y e\Big(\frac{t\log x}{2\pi} - kx\Big)\,dx + O\Big(\log\Big(2+\frac{t}{4\pi y}\Big)\Big),\end{aligned}$$

where $\alpha = t/(4\pi y)$ and $\beta = t/(2\pi y)$.

By the second derivative bound (Lemma 11.5), we obtain

$$\int\limits_{y/2}^y e\Big(\frac{t\log x}{2\pi} - kx\Big)\,dx \ll \frac{y}{|t|^{1/2}}.$$

Now we just split into various regimes of estimation. When $y \geqslant |t|$, we only get a contribution from $k = 0$, and hence

$$\sum_{|t| \leqslant n \leqslant y} n^{it} = \int_{|t|}^{y} x^{it}\, dx + O(1) = \frac{y^{1+it}}{1+it} + O(1).$$

When $|t|^{1/2} \leqslant y \leqslant |t|$, we use the second derivative bound above to derive that

$$\sum_{y/2 < n \leqslant y} n^{it} \ll \frac{|t|}{y} \cdot \frac{y}{|t|^{1/2}} + O(\log |t|) \ll |t|^{1/2} + O(\log |t|).$$

Summing over dyadic ranges we obtain

$$\Big| \sum_{|t|^{1/2} \leqslant n \leqslant |t|} n^{it} \Big| \ll |t|^{1/2} \log |t|.$$

Finally when $y \leqslant |t|^{1/2}$ we can bound the sum trivially as

$$\Big| \sum_{n \leqslant y} n^{it} \Big| \ll |t|^{1/2}.$$

Putting everything together, we get

$$\begin{aligned} \Big| \frac{1}{N} \int_{1}^{2N} \sum_{n \leqslant y} n^{it}\, dy \Big| &\ll |t|^{1/2} \log |t| + \frac{1}{N} \Big| \int_{|t|}^{2N} \frac{y^{1+it}}{1+it}\, dy \Big| \\ &\ll |t|^{1/2} \log |t| + \frac{1}{N} \Big| \frac{N^{2+it}}{(1+it)(2+it)} \Big| \\ &\ll |t|^{1/2} \log |t| + \frac{N}{1+|t|^2} \end{aligned}$$

as required. □

Thus we have finally proved the Halasz–Montgomery mean value theorem.

12. Dirichlet polynomials II: decompositions

In this lecture we will define the approximate decomposition of the Dirichlet polynomial $F(t) := \sum_{n\in[X,2X]} f(n)n^{it}$ that we will use to bound the integral

$$\int_{-\infty}^{\infty} |F(t)|^2 \min\Big(\frac{1}{X^2}, \frac{1}{H^2t^2}\Big)\, dt$$

that we considered in the last lecture. In fact we will do this twice; first to give the decomposition that we will actually use in our proof, and second to give the decomposition that is used in the original paper of Matomäki–Radziwiłł. The decomposition of Matomäki–Radziwiłł provides stronger quantitative bounds, but is a little more complicated.

Throughout we will assume that $f \in \mathcal{M}_0$ is real-valued and completely multiplicative. The central idea is to write $f(n)n^{it}$ as $f(p)p^{it}f(m)m^{it}$ and then to sum over p and m in various ranges. This is of course similar to what we did when proving Halasz's theorem. However, we need to be careful to weight $f(n)$ by an appropriate factor in order to take account of how many ways we can represent n as a product $n = pm$ with p and m in suitable ranges.

The central workhorse in our decomposition is the Turán–Kubilius inequality. I set this inequality as an exercise on the first examples sheet – and discussed it briefly in the examples class – but given its central rôle in what follows I think it is important that I lecture the proof as well.

Here is the result we will use.

Theorem 12.1 (Turán–Kubilius inequality)**.** *Let $g : \mathbb{N} \longrightarrow \mathbb{C}$ be an additive function, i.e. $g(mn) = g(m) + g(n)$ when $\gcd(m, n) = 1$. Let $\mathbb{E}_X(g) := \sum_{p^k\leqslant X} \frac{g(p^k)}{p^k}(1 - \frac{1}{p})$. Then:*

(1) $\frac{1}{X}\sum_{n\leqslant X} g(n) = \mathbb{E}_X(g) + O(1/\log X)$, if $|g(p^k)| \ll 1$ for all prime powers p^k.

(2) $\frac{1}{X}\sum_{n\leqslant X} |g(n) - \mathbb{E}_X(g)|^2 \ll \sum_{p^k\leqslant X} \frac{|g(p^k)|^2}{p^k}$.

One may think of this result as a bound for the 'mean' and 'variance' of the additive function g. In words, we have that 'additive functions are always close to their mean values'.

Consequence of the Turán–Kubilius inequality

The standard application of the Turán–Kubilius inequality is to establishing the normal order of $\omega(n)$, where $\omega(n) := \sum_{p|n} 1$. The function ω is additive, and

$$\mathbb{E}_X(\omega) = \sum_{p^k\leqslant X} \frac{\omega(p^k)}{p^k}\Big(1 - \frac{1}{p}\Big) = \sum_{p^k\leqslant X} \frac{1}{p^k}\Big(1 - \frac{1}{p}\Big) = \log\log X + O(1)$$

and

$$\sum_{p^k\leqslant X} \frac{|\omega(p^k)|^2}{p^k} = \sum_{p^k\leqslant X} \frac{1}{p^k} = \log\log X + O(1).$$

Therefore, by the Turán–Kubilius inequality we have

$$\frac{1}{X}\sum_{n\leqslant X} |\omega(n) - \log\log X|^2 \ll \log\log X.$$

We can use Chebyshev's inequality to deduce that for any function $\theta(X)$ such that $\theta(X) \to \infty$ as $X \to \infty$, we have

$$\frac{1}{X}|\{n \leqslant X : |\omega(n) - \log\log X| \geqslant \theta(X)\sqrt{\log\log X}\}| \ll \frac{\frac{1}{X}\sum_{n\leqslant X}|\omega(n) - \log\log X|^2}{\theta(X)^2 \log\log X}$$
$$\ll \frac{1}{\theta(X)^2} = o(1)$$

as $X \to \infty$. This certainly implies that for all $\varepsilon > 0$, almost all $n \leqslant X$ satisfy $(1 - \varepsilon)\log\log X \leqslant \omega(n) \leqslant (1 + \varepsilon)\log\log X$, i.e. almost all numbers have basically the same number of prime factors.

This result was first proved by Hardy–Ramanujan.It may seem (at least to me!) inherently surprising that such a result should be true. For one thing, we observe that

$$\frac{1}{X}\sum_{n\leqslant X} 2^{\omega(n)} \geqslant \frac{1}{X}\sum_{n\leqslant X}\mu^2(n)\tau(n) \asymp \log X$$

and yet

$$\frac{1}{X}\sum_{n\leqslant X} 2^{\log\log X} = \frac{1}{X}\sum_{n\leqslant X}(\log X)^{\log 2} \asymp (\log X)^{\log 2},$$

which is of a lower order of magnitude to $\log X$ since $\log 2 < 1$. Therefore we see that the main contribution to the average of $\sum_{n\leqslant X}\mu^2(n)\tau(n)$ is actually not from typical integers but rather from integers that have unusually many prime factors. (This is also true for the average $\sum_{n\leqslant X}\tau(n)$).

There is actually a rather simple proof of the normal order of $\omega(n)$ that doesn't use the full strength of the Turán–Kubilius inequality. This comes from just considering the primes $p \leqslant X^{1/100}$, say, for which the L^2 estimate on

$$\sum_{n\leqslant X}\Big(\sum_{\substack{p|n\\ p\leqslant X^{1/100}}} 1 - \log\log X\Big)^2$$

is easier to establish. In fact this approach can be used to approximate all the moments of $\omega(n) - \log\log X$, and establish an asymptotic normal distribution. This is the Erdős–Kac theorem, and will be on the example sheet.

The proof of the Turán–Kubilius inequality is not so deep, and indeed the result itself is not terribly strong, in the sense that it never saves more than a factor $\log\log X$ over a trivial bound. Indeed, in the case of $\omega(n)$ say, a trivial bound would lead to

$$\frac{1}{X}\sum_{n\leqslant X}(\omega(n) - \log\log X)^2 \ll (\log\log X)^2 + \frac{1}{X}\sum_{n\leqslant X}\omega(n)^2 \ll (\log\log X)^2,$$

which is only a factor of $\log\log X$ out from the Turán–Kubilius bound. Nonetheless, in the context of this section of the course, any saving is a good saving.

Proof of Turán–Kubilius. For part (1), we just calculate

$$\begin{aligned}\sum_{n\leqslant X} g(n) &= \sum_{n\leqslant X}\sum_{p^k\|n} g(p^k)\\ &= \sum_{p^k\leqslant X} g(p^k)\sum_{\substack{n\leqslant X\\ p^k\|n}} 1\\ &= \sum_{p^k\leqslant X} g(p^k)\Big(\Big\lfloor\frac{X}{p^k}\Big\rfloor - \Big\lfloor\frac{X}{p^{k+1}}\Big\rfloor\Big)\end{aligned}$$

$$= X \sum_{p^k \leqslant X} \frac{g(p^k)}{p^k}\Big(1 - \frac{1}{p}\Big) + O(\sum_{p^k \leqslant X} |g(p^k)|).$$

Since we assume that $|g(p^k)| \ll 1$ for all prime powers p^k we get the error term in the above is $O(\sum_{p^k \leqslant X} 1)$, which is

$$\ll \frac{X}{\log X} + \sum_{2 \leqslant k \leqslant \log X/\log 2} \sum_{p \leqslant X^{1/k}} 1 \ll \frac{X}{\log X} + X^{1/2} \log X \ll \frac{X}{\log X}.$$

This settles part (1) of the theorem.

For part (2) (which is the real meat), we expand out the square to get the sum of three terms, namely

$$\frac{\lfloor X \rfloor}{X} |\mathbb{E}_X(g)|^2, \quad -\frac{2}{X} \Re\Big(\overline{\mathbb{E}_X(g)} \sum_{n \leqslant X} g(n)\Big), \quad \text{and} \quad \frac{1}{X} \sum_{n \leqslant X} |g(n)|^2.$$

Each of these three terms will be approximately $|\mathbb{E}_X(g)|^2$, $-2|\mathbb{E}_X(g)|^2$, and $|\mathbb{E}_X(g)|^2$ respectively. So these main terms will cancel, and we will be left with the task of establishing that the error terms are at $\ll \sum_{p^k \leqslant X} |g(p^k)|^2/p^k$.

A useful preparatory observation is the following: by Cauchy–Schwarz we have

$$\begin{aligned} |\mathbb{E}_X(g)| &= \Big| \sum_{p^k \leqslant X} \frac{g(p^k)}{p^k}\Big(1 - \frac{1}{p}\Big)\Big| \\ &\leqslant \Big(\sum_{p^k \leqslant X} \frac{|g(p^k)|^2}{p^k}\Big)^{1/2} \Big(\sum_{p^k \leqslant X} \frac{1}{p^k}\Big)^{1/2} \\ &\ll \Big(\sum_{p^k \leqslant X} \frac{|g(p^k)|^2}{p^k}\Big)^{1/2} (\log\log X)^{1/2}. \end{aligned}$$

Then we have, firstly,

$$\begin{aligned} \frac{\lfloor X \rfloor}{X} |\mathbb{E}_X(g)|^2 &= |\mathbb{E}_X(g)|^2 + O\Big(\frac{1}{X} |\mathbb{E}_X(g)|^2\Big) \\ &= |\mathbb{E}_X(g)|^2 + O\Big(\frac{\log\log X}{X} \Big(\sum_{p^k \leqslant X} \frac{|g(p^k)|^2}{p^k}\Big)\Big). \end{aligned}$$

This error term is acceptable.

Regarding the second contributing term $-2\Re(\overline{\mathbb{E}_X(g)} \frac{1}{X} \sum_{n \leqslant X} g(n))$, by the argument for part (1) of the theorem we have

$$\begin{aligned} -2\Re\Big(\overline{\mathbb{E}_X(g)} \frac{1}{X} \sum_{n \leqslant X} g(n)\Big) &= -2\Re\Big(\overline{\mathbb{E}_X(g)} \Big(\mathbb{E}_X(g) + O\Big(\frac{1}{X} \sum_{p^k \leqslant X} |g(p^k)|\Big)\Big)\Big) \\ &= -2|\mathbb{E}_X(g)|^2 + O\Big(\frac{1}{X} |\mathbb{E}_X(g)| \sum_{p^k \leqslant X} |g(p^k)|\Big) \end{aligned}$$

Now by applying our bounds on $|\mathbb{E}_X(g)|$ from earlier, together with Cauchy–Schwarz on the $\sum_{p^k \leqslant X} |g(p^k)|$ term, we get

$$= -2|\mathbb{E}_X(g)|^2 + O\Big(\frac{1}{X} (\log\log X)^{1/2} \Big(\sum_{p^k \leqslant X} \frac{|g(p^k)|^2}{p^k}\Big)^{1/2} \Big(\sum_{p^k \leqslant X} \frac{|g(p^k)|^2}{p^k}\Big)^{1/2} \Big(\sum_{p^k \leqslant X} p^k\Big)^{1/2}\Big)$$

$$= -2|\mathbb{E}_X(g)|^2 + O\Big(\Big(\frac{\log\log X}{\log X}\Big)^{1/2}\Big(\sum_{p^k\leqslant X}\frac{|g(p^k)|^2}{p^k}\Big)\Big)$$

which is an acceptable error term.

The remaining issue is the third term $\frac{1}{X}\sum_{n\leqslant X}|g(n)|^2$, which will be the most complicated to control. Expanding the square we have

$$\begin{aligned}\frac{1}{X}\sum_{n\leqslant X}|g(n)|^2 &= \frac{1}{X}\sum_{n\leqslant X}\Big|\sum_{p^k\| n}g(p^k)\Big|^2\\ &= \frac{1}{X}\sum_{p^k,q^l\leqslant X}g(p^k)\overline{g(q^l)}\sum_{\substack{n\leqslant X\\ p^k\| n\\ q^l\| n}}1.\end{aligned}$$

There are now two cases to consider. If $p=q$, then we only get a contribution when $k=l$, yielding

$$\begin{aligned}\frac{1}{X}\sum_{p^k\leqslant X}|g(p^k)|^2\sum_{\substack{n\leqslant X\\ p^k\| n}}1 &= \frac{1}{X}\sum_{p^k\leqslant X}|g(p^k)|^2\Big(\Big\lfloor\frac{X}{p^k}\Big\rfloor-\Big\lfloor\frac{X}{p^{k+1}}\Big\rfloor\Big)\\ &\leqslant \sum_{p^k\leqslant X}\frac{|g(p^k)|^2}{p^k}\end{aligned}$$

which is an acceptable error term.

The remaining terms (when $p\neq q$) contribute

$$\frac{1}{X}\sum_{\substack{p^kq^l\leqslant X\\ p\neq q}}g(p^k)\overline{g(p^l)}\Big(\Big\lfloor\frac{X}{p^kq^l}\Big\rfloor-\Big\lfloor\frac{X}{p^{k+1}q^l}\Big\rfloor-\Big\lfloor\frac{X}{p^kq^{l+1}}\Big\rfloor+\Big\lfloor\frac{X}{p^{k+1}q^{l+1}}\Big\rfloor\Big)$$

by inclusion exclusion. This is

$$= \sum_{\substack{p^kq^l\leqslant X\\ p\neq q}}\frac{g(p^k)}{p^k}\frac{\overline{g(q^l)}}{q^l}\Big(1-\frac{1}{p}\Big)\Big(1-\frac{1}{q}\Big)+O\Big(\frac{1}{X}\sum_{\substack{p^kq^l\leqslant X\\ p\neq q}}|g(p^k)\overline{g(q^l)}|\Big)$$

$$= \sum_{p^kq^l\leqslant X}\frac{g(p^k)}{p^k}\frac{\overline{g(q^l)}}{q^l}\Big(1-\frac{1}{p}\Big)\Big(1-\frac{1}{q}\Big)+O\Big(\frac{1}{X}\sum_{p^kq^l\leqslant X}|g(p^k)\overline{g(q^l)}|\Big)+O\Big(\sum_{p,k,l:\,p^{k+l}\leqslant X}\frac{|g(p^k)\overline{g(p^l)}|}{p^{k+l}}\Big).$$

We have two error terms to contend with. By Cauchy–Schwarz, we have

$$\begin{aligned}\frac{1}{X}\sum_{p^kq^l\leqslant X}|g(p^k)\overline{g(q^l)}| &\ll \frac{1}{X}\Big(\sum_{p^kq^l\leqslant X}\frac{|g(p^k)|^2|g(q^l)|^2}{p^kq^l}\Big)^{1/2}\Big(\sum_{p^kq^l\leqslant X}p^kq^l\Big)^{1/2}\\ &\ll \frac{1}{X}\Big(\sum_{p^k\leqslant X}\frac{|g(p^k)|^2}{p^k}\Big)\Big(X\sum_{p^k\leqslant X}\sum_{q^l\leqslant X/p^k}1\Big)^{1/2}\end{aligned}$$

Considering just the second bracket here, we calculate

$$\sum_{p^k\leqslant X}\sum_{q^l\leqslant X/p^k}1\ll\sum_{p^k\leqslant X}\frac{X}{p^k\log(1+X/p^k)}\ll\frac{X\log\log X}{\log X}$$

(which can be seen by reducing the sum to a sum over primes p and then splitting into dyadic ranges, for examples). So, overall we have

$$\frac{1}{X}\sum_{p^kq^l\leqslant X}|g(p^k)\overline{g(q^l)}|\ll\Big(\frac{\log\log X}{\log X}\Big)^{1/2}\sum_{p^k\leqslant X}\frac{|g(p^k)|^2}{p^k},$$

which is an acceptable error.

We also have (by Cauchy again)

$$\sum_{p,k,l:\, p^{k+l}\leqslant X} \frac{|g(p^k)\overline{g(p^l)}|}{p^{k+l}} \ll \Big(\sum_{p,k,l:\, p^{k+l}\leqslant X} \frac{|g(p^k)|^2}{p^{k+l}}\Big)^{1/2}\Big(\sum_{p,k,l:\, p^{k+l}\leqslant X} \frac{|g(p^l)|^2}{p^{k+l}}\Big)^{1/2}$$
$$\ll \sum_{p^k\leqslant X} \frac{|g(p^k)|^2}{p^k} \sum_{\substack{l\\ p^l\leqslant X/p^k}} \frac{1}{p^l} \ll \sum_{p^k\leqslant X} \frac{|g(p^k)|^2}{p^k},$$

which is again an acceptable error.

It remains to consider the 'main term'

$$\sum_{p^kq^l\leqslant X} \frac{g(p^k)}{p^k}\frac{\overline{g(q^l)}}{q^l}\Big(1-\frac{1}{p}\Big)\Big(1-\frac{1}{q}\Big).$$

This is equal to

$$|\mathbb{E}_X(g)|^2 - \sum_{\substack{p^k\leqslant X\\ q^l\leqslant X\\ p^kq^l>X}} \frac{g(p^k)}{p^k}\frac{\overline{g(q^l)}}{q^l}\Big(1-\frac{1}{p}\Big)\Big(1-\frac{1}{q}\Big).$$

By Cauchy–Schwarz, the error term here is

$$\leqslant \Big(\sum_{\substack{p^k\leqslant X\\ q^l\leqslant X\\ p^kq^l>X}} \frac{|g(p^k)|^2}{p^k}\frac{|g(q^l)|^2}{q^l}\Big)^{1/2}\Big(\sum_{\substack{p^k\leqslant X\\ q^l\leqslant X\\ p^kq^l>X}} \frac{1}{p^kq^l}\Big)^{1/2}.$$

Here the first bracket is $\leqslant \sum_{p^k\leqslant X}|g(p^k)|^2/p^k$, and the second bracket (squared) is

$$\ll \sum_{p^k\leqslant X^{1/2}} \frac{1}{p^k}\sum_{X/p^k<q^l\leqslant X}\frac{1}{q^l} + \Big(\sum_{X^{1/2}<p^k\leqslant X}\frac{1}{p^k}\Big)^2 \ll \sum_{p^k\leqslant X^{1/2}}\frac{k\log p}{p^k\log X}+1 \ll 1,$$

since by partial summation we have

$$\sum_{X/p^k<q^l\leqslant X}\frac{1}{q^l} \ll \frac{(\log X)-\log(X/p^k)}{\log X}$$

for $p^k\leqslant X^{1/2}$. So the overall error is acceptable, and we conclude the theorem. □

We're now going to show a baby version of the Dirichlet series decomposition we are going to use in the next lecture. Let $[P,Q]\subset[1,X]$ be a certain interval of primes, and let

$$\omega_{[P,Q]}(n) := \sum_{\substack{p\in[P,Q]\\ p|n}} 1.$$

Then $\omega_{[P,Q]}$ is an additive function and, letting

$$W_{[P,Q]} := \sum_{p\in[P,Q]}\frac{1}{p} \approx \log\log Q - \log\log P$$

we have $W_{[P,Q]}\approx \mathbb{E}_X(\omega_{[P,Q]})$ and so

$$\sum_{n\leqslant X}(\omega_{[P,Q]}(n)-W_{[P,Q]})^2 \ll XW_{[P,Q]}$$

by the Turán–Kubilius inequality. (Once we pick precise thresholds P,Q in the next lecture, we can make all these approximations rigorous).

Let $f \in \mathcal{M}_0$ be real valued, completely multiplicative, and not 1-pretentious. By the triangle inequality we have

$$W_{[P,Q]}^2 \int_X^{2X} \Big| \sum_{x<n\leqslant x+H} f(n)\Big|^2 dx$$
$$\ll \int_X^{2X} \Big| \sum_{x<n\leqslant x+H} f(n) w_{[P,Q]}(n)\Big|^2 dx + \int_X^{2X} \Big| \sum_{x<n\leqslant x+H} f(n)(w_{[P,Q]}(n) - W_{[P,Q]})\Big|^2 dx.$$

The second term here is

$$\ll \int_X^{2X} \Big(\sum_{x<n\leqslant x+H} |f(n)|^2\Big)\Big(\sum_{x<n\leqslant x+H} (w_{[P,Q]}(n) - W_{[P,Q]})^2\Big) dx$$
$$\ll XH^2 W_{[P,Q]},$$

so we get

$$\int_X^{2X} \Big| \sum_{x<n\leqslant x+H} f(n)\Big|^2 dx = \frac{1}{W_{[P,Q]}^2} \int_X^{2X} \Big| \sum_{x<n\leqslant x+H} f(n) w_{[P,Q]}(n)\Big|^2 dx + O\Big(\frac{XH^2}{W_{[P,Q]}}\Big).$$

If $W_{[P,Q]} \to \infty$ as $X \to \infty$, the error term is negligible.

The utility of doing this comes from the fact that the arithmetic function $f(n)w_{[P,Q](n)}$ can be related to a Dirichlet series that admits a pleasant factorisation. Indeed, assuming that Q/P is a power of 2, we have (by the previous Lemma 11.2)

$$\frac{1}{X}\int_X^{2X} \Big|\frac{1}{H} \sum_{x<n\leqslant x+H} f(n) w_{[P,Q]}(n)\Big|^2 dx \ll \int_{-\infty}^{\infty} |A(t)|^2 \min\Big(\frac{1}{X^2}, \frac{1}{t^2H^2}\Big) dt + o(1)$$

where

$$A(t) = \sum_{j=1}^{O(\log(Q/P))} \sum_{2^{j-1}P\leqslant p<2^jP} f(p)p^{it} \sum_{\frac{X}{2^jP}\leqslant m<\frac{2X}{2^{j-1}P}} f(m)m^{it}.$$

Indeed, we have that the coefficient $a(n)$ of n^{it} in $A(t)$ is

$$\sum_{j=1}^{O(\log(Q/P))} \sum_{2^{j-1}P\leqslant p<2^jP} f(p) \sum_{\substack{\frac{X}{2^jP}\leqslant m<\frac{2X}{2^{j-1}P}\\ pm=n}} f(m)$$

and so we have $a(n) = 0$ unless $n \in [X/2, 4X]$, $|a(n)| \leqslant w_{[P,Q]}(n)$ for all n, and if $n \in [X, 2X)$ we have $a(n) = f(n)w_{[P,Q]}(n)$, by the complete multiplicativity of f. (This is the reason for choice of the range of m).

NB: technically we only proved Lemma 11.2 for functions f that were supported on $[X, 3X]$. Actually we'll need to consider functions supported on $[c_1X, c_2X]$ for more general c_1, c_2: we'll state the appropriate version next time.

Note that we have the factorisation (from Cauchy–Schwarz)

$$\int_{-\infty}^{\infty} |A(t)|^2 \min\Big(\frac{1}{X^2}, \frac{1}{t^2H}\Big) dt$$

$$\ll (\log(Q/P))^2 \max_{j\ll \log(Q/P)} \int_{-\infty}^{\infty} \Big|\sum_{2^{j-1}P\leqslant p<2^jP} f(p)p^{it}\Big|^2 \Big|\sum_{\frac{X}{2^jP}\leqslant m<\frac{2X}{2^{j-1}P}} f(m)m^{it}\Big|^2 \min\Big(\frac{1}{X^2},\frac{1}{t^2H^2}\Big)\,dt.$$

(When we do this in full detail next lecture we will deal with the dyadic scales slightly differently.) We will be playing off L^∞ bounds on $\sum_{2^{j-1}P\leqslant p<2^jP} f(p)p^{it}$ and L^2 bounds on $\sum_{X/2^jP\leqslant m<2X/2^{j-1}P} f(m)m^{it}$.

For the rest of today, I just want to spend 10-15 minutes or so discussing the different Dirichlet series decomposition that was used by Matomäki–Radziwiłłin their work. They centred their work around *Ramaré's identity*.

Lemma 12.2 (Ramaré's identity)**.** *For any $P<Q$ and for all $n\geqslant 1$ we have*

$$\sum_{\substack{p\in[P,Q]\\ p|n}} \frac{1}{\#\{q\in[P,Q]: q|\frac{n}{p}\}+1_{(p,n/p)=1}} = \begin{cases} 1 & \text{if } \exists p\in[P,Q] \text{ with } p|n\\ 0 & \text{otherwise.}\end{cases}$$

Proof. Let $A_{[P,Q],n}=\{p\in[P,Q]: p\|n\}$ and $B_{[P,Q],n}=\{p\in[P,Q]: p^2|n\}$. Let $a_{[P,Q],n}=|A_{[P,Q],n}|$ and $b_{[P,Q],n}=|B_{[P,Q],n}|$. Then

$$\begin{aligned}&\sum_{\substack{p\in[P,Q]\\ p|n}} \frac{1}{\#\{q\in[P,Q]: q|\frac{n}{p}\}+1_{(p,n/p)=1}}\\ &= \sum_{p\in A_{[P,Q],n}} \frac{1}{(a_{[P,Q],n}+b_{[P,Q],n}-1)+1} + \sum_{p\in B_{[P,Q],n}} \frac{1}{a_{[P,Q],n}+b_{[P,Q],n}}\\ &= 1\end{aligned}$$

provided the sum is non-empty. This settles the identity. □

Let

$$\theta_p(m) := \frac{1}{\#\{q\in[P,Q]: q|m\}+1_{(p,m)=1}}$$

be the weight function from the above identity and let $\mathcal{S}=\{n\in[X,2X]: \exists p\in[P,Q] \text{ with } p|n\}$. Then if $f\in\mathcal{M}_0$ is completely multiplicative we have

$$\sum_{\substack{n\in[X,2X]\\ n\in\mathcal{S}}} f(n)n^{it} = \sum_{p\in[P,Q]} f(p)p^{it} \sum_{\frac{X}{p}\leqslant m<\frac{2X}{p}} f(m)m^{it}\theta_p(m/p).$$

Now, we know from sieving (Lemma 4.1) that

$$\#\{n\in[X,2X]: n\notin\mathcal{S}\} = X\frac{\log P}{\log Q} = X(1+o(1))$$

for suitable choices of parameters Q and P. So one may approximate

$$\begin{aligned}&\frac{1}{X}\int_X^{2X} \Big|\frac{1}{H}\sum_{x<n\leqslant x+h} f(n)\Big|^2\,dx = \frac{1}{X}\int_X^{2X} \Big|\frac{1}{H}\sum_{\substack{x<n\leqslant x+h\\ n\in\mathcal{S}}} f(n)\Big|^2\,dx + o(1)\\ &\ll \int_{-\infty}^{\infty} \Big|\sum_{p\in[P,Q]} f(p)p^{it}\Big|^2 \Big|\sum_{\frac{X}{p}\leqslant m<\frac{2X}{p}} f(m)m^{it}\theta_p(m/p)\Big|^2 \min\Big(\frac{1}{X^2},\frac{1}{t^2H^2}\Big)\,dx + o(1).\end{aligned}$$

which is a similar decomposition to the one that we achieved earlier, except that we had an error of size $\approx (\log(\log Q/\log P))^{-1}$ rather than $(\log Q/\log P)^{-1}$ (which is a smaller error). So Ramaré's identity turns out to be a more efficient approach regarding the quantitative error terms, although the details are a bit more complicated to work through, owing to the fact that the weight $f(m)\theta_p(m/p)$ is not multiplicative.

This finishes our description of the general decomposition result, based on the Turán–Kubilius inequality that we will employ. Next time, we will replace the single scale $[P, Q]$ with multiple scales $[P_1, Q_1], \dots, [P_K, Q_K]$ and begin the full decomposition of the integral over $A(t)$, splitting according to the L^∞ size of the various Dirichlet polynomial factors.

13. Ladders of scales

In this rather technical lecture, after several lectures of preparation, we will attack the theorem of Matomäki–Radziwiłł head-on.

The time has come for me to decide exactly how strong a statement we will have time to prove in these lectures. Let $\log_k(X)$ denote the k^{th} iterated logarithm of X (i.e. $\log_2(X) = \log\log X$, $\log_3 X = \log\log\log X$ etc.). After some consideration, I've decided to present the full proof of the following result:

Theorem 13.1. *Let $f \in \mathcal{M}_0$ be completely multiplicative and real-valued, and assume that $\mathbb{D}(f, 1; \infty) = \infty$. Let X be an asymptotic parameter tending to infinity, and let $H = H(X)$ be a function with $\log_K(X) \leqslant H \leqslant X$ (for some fixed $K \in \mathbb{N}$). Then for all but $o_K(X)$ natural numbers $x \in [X, 2X)$ we have*

$$\Big|\frac{1}{H}\sum_{x<n\leqslant x+H} f(n)\Big| = o_K(1)$$

as $X \to \infty$. The error term can depend on $\mathbb{D}(f, 1; X)$, but is otherwise independent of f.

Our proof may be adapted to the proof of the full Matomäki–Radiwiłł theorem by taking care to note the dependency of our error terms on K, but we won't do this.

We have already shown that Theorem 13.1 is implied by the bound

$$\frac{1}{X}\int_X^{2X} \Big|\frac{1}{H}\sum_{x<n\leqslant x+H} f(n)\Big|^2 dx = o_K(1). \tag{8}$$

It will be a useful simplification to assume that $H = o(X)$ rather than just $H \leqslant X$. This is easy to do using Halasz's theorem. Indeed, if $H \geqslant \varepsilon X$ we have

$$\frac{1}{H}\sum_{x<n\leqslant x+H} f(n) = \frac{1}{H}((x+H)M_{x+H}(f) - xM_x(f)) \leqslant \varepsilon^{-1}o(1)$$

as $X \to \infty$, by Halasz's theorem (using the fact that $\mathbb{D}(f, 1; \infty) = \infty$). Choosing ε to be a slow-enough decreasing function of X, we conclude that

$$\frac{1}{X}\int_X^{2X} \Big|\frac{1}{H}\sum_{x<n\leqslant x+H} f(n)\Big|^2 dx = o(1)$$

when $H \geqslant \varepsilon X$. So we are left with the case $H < \varepsilon X$, which implies that $H = o(X)$.

We now make another simplifying observation. Suppose that $\log_K(X) \leqslant H^*(X) < H(X) = o(X)$, and suppose further that $H^*(X) = o(H(X))$. If (8) is known to hold for the smaller scale $H^*(X)$, then (8) also holds for the larger scale $H(X)$. Indeed, by splitting into smaller subintervals we observe that

$$\begin{aligned}\int_X^{2X} \Big|\sum_{x<n\leqslant x+H} f(n)\Big|^2 dx &= \int_X^{2X} \Big|\Big(\sum_{j=0}^{\lfloor H/H^*\rfloor} \sum_{x+jH^*<n\leqslant x+(j+1)H^*} f(n)\Big) + O(H^*)\Big|^2 dx \\ &\ll X(H^*)^2 + \frac{H}{H^*}\sum_{j=0}^{\lfloor H/H^*\rfloor} \int_{X-jH^*}^{2X-jH^*} \Big|\sum_{x<n\leqslant x+H^*} f(n)\Big|^2 dx \\ &\ll X(H^*)^2 + \frac{H}{H^*}\sum_{j=0}^{\lfloor H/H^*\rfloor} \Big(\int_X^{2X} \Big|\sum_{x<n\leqslant x+H^*} f(n)\Big|^2 dx + O(j(H^*)^3)\Big)\end{aligned}$$

$$\ll X(H^*)^2 + \frac{H}{H^*}\sum_{j=0}^{\lfloor H/H^*\rfloor}(o_K(X(H^*)^2) + O(j(H^*)^3))$$
$$\ll X(H^*)^2 + o_K(XH^2) + O(H^3) = o_K(XH^2)$$

as $X \to \infty$. Here we used Cauchy–Schwarz, the fact that the inner-integral and the assumption that (8) holds for the scale H^*.

So from now on I will assume without loss of generality that

$$H(X) = \exp((\log_{K+1} X)^3)$$

for some large fixed K. This turns out to be a more convenient function to work with than an iterated logarithm. Following on from the ideas of last time, it will be important to define some scales of primes $[P, Q]$ which we will use to factorise the relevant Dirichlet polynomials. Right at the start, let us fix some intervals $[P_1, Q_1']$, $[P_2, Q_2']$, $\dots$, $[P_K, Q_K']$, namely

$$P_1 = \exp((\log\log X)^2), \qquad Q_1' = \exp((\log\log X)^3)$$
$$P_2 = \exp((\log\log\log X)^2), \qquad Q_2' = \exp((\log\log\log X)^3)$$

and in general for $k = 1, \dots, K$ let

$$P_k = \exp((\log_{k+1}(X)^2), \qquad Q_k' = \exp((\log_{k+1}(X)^3).$$

For reference, we have

$$[P_K, Q_K'] = [\exp((\log_{K+1} X)^2), H].$$

In order to avoid certain technical issues regarding dyadic pigeonholing, it will be convenient for the ratio of the upper and lower bounds for our intervals are a power of 2. To that end, we let Q_k be the nearest integer to Q_k' for which Q_k of the form $\lceil 2^\ell P_k\rceil$ for some $\ell \in \mathbb{N}$. We have $Q_k/Q_k' \asymp 1$.

For ease of notation, we let

$$\omega_k(n) := \omega_{[P_k,Q_k]}(n) := \sum_{p|n : p\in[P_k,Q_k]} 1 \qquad \text{and } W_k := W_{[P_k,Q_k]} = \sum_{p\in[P_k,Q_k]}\frac{1}{p}.$$

We have

$$W_k = \log\log Q_k - \log\log P_k + O(1) = 3\log_{k+2}(X) - 2\log_{k+2}(X) + O(1) \gg \log_{k+2}(X),$$

which tends to infinity.

As we demonstrated last time, we can use the Turán–Kubilius inequality to introduce the weights ω_k into our L^2 averages. Indeed, by the triangle inequality, we have

$$(\prod_{k=1}^K W_k^2)\int_X^{2X}\Big|\sum_{x<n\leqslant x+H} f(n)\Big|^2 dx$$
$$\ll \int_X^{2X}\Big|\sum_{x<n\leqslant x+H} f(n)\prod_{k=1}^K\omega_k(n)\Big|^2 dx + \int_X^{2X}\Big|\sum_{x<n\leqslant x+H} f(n)\Big(\prod_{k=1}^K\omega_k(n) - \prod_{k=1}^K W_k\Big)\Big|^2 dx$$

Note that we have the telescoping identity

$$\prod_{k=1}^K\omega_k(n) - \prod_{k=1}^K W_k = \sum_{L=1}^K\Big(\prod_{k\leqslant L-1}\omega_k(n)\Big)(\omega_L(n) - W_L)\Big(\prod_{L+1\leqslant k\leqslant K} W_k\Big).$$

Furthermore we have the bound

$$\begin{aligned}\sum_{n\leqslant 3X}\prod_{k\leqslant L-1}\omega_k(n)^2(\omega_L(n)-W_L)^2 &= \sum_{\substack{p_1,q_1\in[P_1,Q_1]\\ \dots\\ p_{L_1},q_{L-1}\in[\dot{P}_{L-1},Q_{L-1}]}}\sum_{\substack{n\leqslant 3X\\ p_1,q_1|n\\ \dots\\ p_{L-1},q_{L-1}|n}}(\omega_L(n)-W_L)^2\\ &= \sum_{\substack{p_1,q_1\in[P_1,Q_1]\\ \dots\\ p_{L_1},q_{L-1}\in[\dot{P}_{L-1},Q_{L-1}]}}\sum_{n\leqslant 3X/\operatorname{lcm}(p_1,\dots,q_{L-1})}(\omega_L(n)-W_L)^2\\ &\ll XW_L\sum_{\substack{p_1,q_1\in[P_1,Q_1]\\ \dots\\ p_{L_1},q_{L-1}\in[\dot{P}_{L-1},Q_{L-1}]}}\frac{1}{\operatorname{lcm}(p_1,\dots,q_{L-1})}\\ &\ll_K XW_L\prod_{k\leqslant L-1}W_k^2.\end{aligned}$$

This follows since the value of $\omega_L(n)$ is unaffected by primes in the other ranges, and by the Turán–Kubilius inequality. For the final inequality we split into cases according to which of the primes are equal or not.The bound

$$\sum_{n\leqslant X}\prod_{k\leqslant L-1}\omega_k(n)^2\ll_K X\prod_{k\leqslant L-1}W_k^2$$

may be proved in the same way.

Therefore, by Cauchy–Schwarz, we have

$$\begin{aligned}&\int_X^{2X}\Big|\sum_{x<n\leqslant x+H}f(n)\Big(\prod_{k=1}^K\omega_k(n)-\prod_{k=1}^K W_k\Big)\Big|^2dx\\ &\ll_K\sum_{L\leqslant K}\Big(\prod_{k=L+1}^K W_k^2\Big)\int_X^{2X}\sum_{\substack{x<n\leqslant x+H\\ x<m\leqslant x+H}}\Big(\prod_{k\leqslant L-1}\omega_k(n)\Big)|\omega_L(n)-W_L|\Big(\prod_{k\leqslant L-1}\omega_k(m)\Big)|\omega_L(m)-W_L|\,dx\\ &\ll_K H\sum_{L\leqslant K}\Big(\prod_{k=L+1}^K W_k^2\Big)\sum_{\substack{n,m\leqslant 3X\\ |n-m|\leqslant H}}\Big(\prod_{k\leqslant L-1}\omega_k(n)\Big)|\omega_L(n)-W_L|\Big(\prod_{k\leqslant L-1}\omega_k(m)\Big)|\omega_L(m)-W_L|\,dx\\ &\ll H^2\sum_{L\leqslant K}\Big(\prod_{k=L+1}^K W_k^2\Big)\sum_{n\leqslant 3X}\Big(\prod_{k\leqslant L-1}\omega_k(n)^2\Big)(\omega_L(n)-W_L)^2\\ &\ll XH^2\sum_{L\leqslant K}W_L\prod_{\substack{k\leqslant K\\ k\neq L}}W_k^2.\end{aligned}$$

So,

$$\int_X^{2X}\Big|\sum_{x<n\leqslant x+H}f(n)\Big|^2dx\ll\Big(\prod_{k=1}^K W_k\Big)^{-2}\int_X^{2X}\Big|\sum_{x<n\leqslant x+H}f(n)\prod_{k=1}^K\omega_k(n)\Big|^2dx+O_K\Big(XH^2\Big(\sum_{L\leqslant K}W_L^{-1}\Big)\Big).$$

From our previous bounds we get

$$\sum_{L\leqslant K}W_L^{-1}\ll\sum_{L\leqslant K}\frac{1}{\log_{L+2}(X)}=o_K(1).$$

So the error term is acceptable, and our gaze moves to bounding the term

$$\int_X^{2X} \Big| \sum_{x<n\leqslant x+H} f(n) \prod_{k=1}^{K} \omega_k(n) \Big|^2 dx.$$

We are going to define a Dirichlet polynomial $A(t) = \sum_n a(n) n^{it}$ for which $a(n) = f(n) \prod_{k\leqslant K} \omega_k(n)$ for $n \in (X, 2X]$, but where $A(t)$ admits a useful factorisation. To do this, let us introduce some dyadic scales into proceedings. For $k \leqslant K$ and for j in the range $1 \leqslant j \leqslant \frac{\log(Q_k/P_k)}{\log 2}$, we let $P_k^{(j)} := 2^{j-1} P_j$, and $\mathcal{P}_k^{(j)}$ denote the primes in the closed interval $[P_k^{(j)}, 2P_k^{(j)}]$. For simplicity, we write $J_k := \frac{\log(Q_k/P_k)}{\log 2} \in \mathbb{N}$.

Then, let

$$A(t) = \sum_{\substack{j_1\leqslant J_1 \\ \dots \\ j_K\leqslant J_K}} \Big(\prod_{k=1}^{K} \sum_{p\in \mathcal{P}_k^{(j_k)}} f(p) p^{it} \Big) \Big(\sum_{m\in I_{j_1,\dots,j_K}} f(m) m^{it} \Big),$$

where $m \in I_{j_1,\dots,j_K}$ if

$$\frac{X}{2^K \prod_{k\leqslant K} P_k^{(j_k)}} \leqslant m \leqslant \frac{4X}{\prod_{k\leqslant K} P_k^{(j_k)}}.$$

Then $A(t) = \sum_n a(n) n^{it}$, where $a(n)$ is supported on $[X/2^K, X2^{K+2}]$, as promised we have $a(n) = f(n) \prod_{k\leqslant K} \omega_k(n)$ for all $n \in [X, 4X)$, and furthermore $|a(n)| \leqslant \prod_{k\leqslant K} \omega_k(n)$ for all n. So, by an adaptation of Lemma 11.2 (adapted to functions supported on $[X/2^K, X2^{K+2}]$), we have

$$\frac{1}{X} \int_X^{2X} \Big| \frac{1}{H} \sum_{x<n\leqslant x+H} f(n) \prod_{k=1}^{K} \omega_k(n) \Big|^2 dx \ll_K \int_{-\infty}^{\infty} |A(t)|^2 \min\Big(\frac{1}{X^2}, \frac{1}{t^2 H^2} \Big) dt + o_K(1).$$

(NB the reason we arrange for $a(n) = f(n) \prod_{k\leqslant K} \omega_k(n)$ for $n \in [X, 4X)$ rather than $n \in [X, 2X)$ is simply to avoid a slight technicality regarding the fact that n can be as large as $2X + H$ on the left-hand side.)

First we dispose of the portion of the integral with large t. Indeed, let $\theta = \theta(X) \to \infty$ as $X \to \infty$. Then from Montgomery's mean value theorem (after splitting into dyadic ranges) we get

$$\begin{aligned} \int_{|t|\geqslant X\theta/H} \frac{|A(t)|^2}{H^2 t^2} dt &\ll_K \frac{1}{H^2} \sum_{l\geqslant 0} \frac{H^2}{2^{2l} X^2 \theta^2} (2^l \frac{X\theta}{H} + X) \sum_n |a(n)|^2 \\ &\ll \sum_{l\geqslant 0} \frac{1}{2^{2l} X^2 \theta^2} (2^l \frac{X\theta}{H} + X) \sum_{n\leqslant 2^{K+2} X} \prod_{k\leqslant K} \omega_k(n)^2 \\ &\ll_K \Big(\frac{1}{\theta X H} + \frac{1}{X\theta^2} \Big) X \prod_{k\leqslant K} W_k^2 \end{aligned}$$

This error term is acceptable. So, we are left with showing that

$$\Big(\prod_{k\leqslant K} W_k^{-2} \Big) \cdot \int_{-X\theta/H}^{X\theta/H} |A(t)|^2 dt = o_K(X^2).$$

For notation convenience we write

$$D_{k,j_k}(t) := \sum_{p\in\mathcal{P}_k^{(j_k)}} f(p)p^{it}.$$

If $|D_{k,j_k}(t)|^2$ exhibits a strong cancellation over a particular set $t\in\mathcal{T}$, then we will be able to satisfactorily bound the contribution to the integral coming from those $t\in\mathcal{T}$. Of course for a general $f\in\mathcal{M}_0$ we know very little about the domain of cancellation $\mathcal{T}$, but, by an outrageously cunning argument of Matomäki–Radziwiłł, it turns out that we don't need to.

Let us define a sequence of exponents $\eta_L = (K-L+1)/(10K)$ for $L\leqslant K$. Then let

$$\mathcal{T}_L := \Big\{ t\in[-X\theta/H, X\theta/H] : |D_{L,j_L}(t)| \leqslant (P_L^{(j_L)})^{1-\eta_L} \text{ for all } j_L\leqslant J_L \\ \text{and, for all } l\geqslant L+1,\ |D_{l,j_l}(t)| > (P_l^{(j_l)})^{1-\eta_l} \text{ for some } j_l\leqslant J_l \Big\}$$

and let

$$\mathcal{E} := [-X\theta/H, X\theta/H] \setminus \bigcup_{L\leqslant K} \mathcal{T}_L.$$

We have set up the scales so that $|\mathcal{E}|\ll X^{3/8}$, and so we have a chance of using the Halas–Montgomery bound for the integral over $\mathcal{E}$ (which you recall gives non-trivial information all the way up to $X^{1/2-\varepsilon}$). It turns out that when $f=\lambda$ is the Liouville function the Halasz–Montgomery bound is enough, although for general f we will need to use a variation of the Halasz–Montgomery bound that is adapted to Dirichlet polynomials supported on the primes. That will be the topic of next lecture.

The rest of this lecture will be devoted to controlling the contribution from $t\in\mathcal{T}_L$, starting with some manipulations to reduce matters to a single dyadic scale. Indeed,

$$\begin{aligned}
&|A(t)|^2 \\
&= \Big| \sum_{\substack{j_1\leqslant J_1\\ \dots\\ j_K\leqslant J_K}} \Big(\prod_{k=L}^K \frac{1}{\log P_k^{(j_k)}}\Big)\Big(\prod_{k=L}^K \log P_k^{(j_k)}\Big)\Big(\prod_{k=1}^K D_{k,j_k}(t)\Big)\Big(\sum_{m\in I_{j_1,\dots,j_K}} f(m)m^{it}\Big)\Big|^2 \\
&\ll_K \Big(\sum_{\substack{j_L\leqslant J_L\\ \dots\\ j_K\leqslant J_K}} \prod_{k=L}^K \frac{1}{\log P_k^{(j_k)}}\Big) \sum_{\substack{j_L\leqslant J_L\\ \dots\\ j_K\leqslant J_K}} \Big(\prod_{k=L}^K \log P_k^{(j_k)}\Big)\Big| \sum_{\substack{j_1\leqslant J_1\\ \dots\\ j_{L-1}\leqslant J_{L-1}}} \Big(\prod_{k=1}^K D_{k,j_k}(t)\Big) \sum_{m\in I_{j_1,\dots,j_K}} f(m)m^{it}\Big|^2
\end{aligned}$$

Now, because of our choices of parameters, we always have

$$\sum_{j_k\leqslant J_k} \frac{1}{\log P_k^{(j_k)}} = \sum_{j_k\leqslant \frac{\log(Q_k/P_k)}{\log 2}} \frac{1}{j_k\log 2+\log P_k} \ll \log\log Q_k - \log\log P_k \ll W_k.$$

Using these decompositions on the range $t\in\mathcal{T}_L$, we get

$$\begin{aligned}
\Big(\prod_{k\leqslant K} W_k^{-2}\Big)\cdot \int\limits_{t\in\mathcal{T}_L} |A(t)|^2\,dt &\ll_K \Big(\prod_{k=1}^{L-1} W_k^{-2}\Big)\Big(\prod_{k=L}^K W_k^{-1}\Big) \sum_{\substack{j_L\leqslant J_L\\ \dots\\ j_K\leqslant J_K}} \Big(\prod_{k=L}^K \frac{1}{\log P_k^{(j_k)}}\Big) C_{j_L,\dots,j_K} \\
&\ll_K \Big(\prod_{k=1}^{L-1} W_k^{-2}\Big) \max_{j_L,\dots,j_K} C_{j_L,\dots,j_K}
\end{aligned}$$

where

$$C_{j_L,\dots,j_K} = \Big(\prod_{k=L}^{K}(\log P_k^{(j_k)})^2\Big)\int\limits_{t\in\mathcal{T}_L}\Big(\prod_{k=L}^{K}|D_{k,j_k}(t)|^2\Big)\Big|\sum_{\substack{j_1\leqslant J_1\\ \dots\\ j_{L-1}\leqslant J_{L-1}}}\Big(\prod_{k=1}^{L-1}D_{k,j_k}(t)\Big)\sum_{m\in I_{j_1,\dots,j_K}}f(m)m^{it}\Big|^2\,dt.$$

We have to show that

$$\Big(\prod_{k=1}^{L-1}W_k^{-2}\Big)C_{j_L,\dots,j_K} = o_K(X^2)$$

(for all choices of $j_L,\dots,j_K$). Just for a quick sanity check, note that the length of the Dirichlet polynomial

$$\Big(\prod_{k=L}^{K}D_{k,j_k}(t)\Big)\sum_{\substack{j_1\leqslant J_1\\ \dots\\ j_{L-1}\leqslant J_{L-1}}}\Big(\prod_{k=1}^{L-1}D_{k,j_k}(t)\Big)\sum_{m\in I_{j_1,\dots,j_K}}f(m)m^{it}$$

is $\asymp_K X$.

The case $L = K$

The contribution from $\mathcal{T}_K$ can be bounded by a simple direct argument. Indeed,

$$\begin{aligned}C_{j_K} &= (\log P_K^{(j_K)})^2\int\limits_{t\in\mathcal{T}_K}|D_{K,j_K}(t)|^2\Big|\sum_{\substack{j_1\leqslant J_1\\ \dots\\ j_{K-1}\leqslant J_{K-1}}}\Big(\prod_{k=1}^{K-1}D_{k,j_k}(t)\Big)\sum_{m\in I_{j_1,\dots,j_K}}f(m)m^{it}\Big|^2\,dt\\ &\ll (\log P_K^{(j_K)})^2(P_K^{(j_K)})^{2-2\eta_K}\int\limits_{-X\theta/H}^{X\theta/H}\Big|\sum_{\substack{j_1\leqslant J_1\\ \dots\\ j_{K-1}\leqslant J_{K-1}}}\Big(\prod_{k=1}^{K-1}D_{k,j_k}(t)\Big)\sum_{m\in I_{j_1,\dots,j_K}}f(m)m^{it}\Big|^2\,dt\end{aligned}$$

The integral is the L^2 mean value of a Dirichlet polynomial of length $\asymp_K X/P_K^{(j_K)}$ and with coefficients $c(n)$ satisfying $|c(n)|\leqslant\prod_{1\leqslant k\leqslant K-1}\omega_k(n)$. So by Montgomery's mean value theorem we have an overall bound of

$$\begin{aligned}&\ll_K (\log P_K^{(j_K)})^2(P_K^{(j_K)})^{2-2\eta_K}\Big(\frac{X\theta}{H}+\frac{X}{P_K^{(j_K)}}\Big)\sum_{n\ll_K X/P_K^{(j_K)}}\prod_{k=1}^{K-1}\omega_k(n)^2\\ &\ll_K \Big(\prod_{k=1}^{K-1}W_k^2\Big)(P_K^{(j_K)})^{2-2\eta_K+o(1)}\Big(\frac{X\theta}{H}+\frac{X}{P_K^{(j_K)}}\Big)\frac{X}{P_K^{(j_K)}}\\ &\ll_K \Big(\prod_{k=1}^{K-1}W_k^2\Big)X^2\Big(\theta H^{-1}(P_K^{(j_K)})^{1-\frac{1}{5K}+o(1)}+(P_K^{(j_K)})^{-\frac{1}{5K}+o(1)}\Big)\end{aligned}$$

which is acceptable if θ grows slowly enough, since $P_K^{(j_K)}\leqslant Q_K\ll H$. Note that we have some. but not too much, flexibility in how to choose this lowest scale, in the sense that it is very important that Q_K is not too much larger than H.

The case $1\leqslant L\leqslant K-1$

So how on earth do we deal with the larger scales $P_L^{(j_L)}$ (i.e. the scales with smaller L)? The issue is that the Dirichlet polynomial

$$\sum_{\substack{j_1 \leqslant J_1 \\ \vdots \\ j_{L-1} \leqslant J_{L-1}}} \Big(\prod_{k=1}^{L-1} D_{k,j_k}(t) \Big) \sum_{m \in I_{j_1,\dots,j_K}} f(m) m^{it} \tag{9}$$

has length $\asymp_K X / \prod_{k=L}^{K} P_k^{(j_k)}$, and this is too small for the Montgomery mean value theorem to give us an adequate bound, in the manner that it did above. The trick is to use the fact that we know that for all k in the range $L+1 \leqslant k \leqslant K$ there exists some scale j'_k for which $|D_{k,j'_k}(t)|$ is large. This will enable us to replace the Dirichlet polynomial with a longer Dirichlet polynomial, on which Montgomery's mean value theorem will be effective. Actually, it will be enough just to use this fact for $k = L+1$.

Indeed, from trivial bounds we have

$$\begin{aligned} &C_{j_L,\dots,j_K} = \\ &\Big(\prod_{k=L}^{K} (\log P_k^{(j_k)})^2 \Big) \int\limits_{t \in \mathcal{T}_L} \Big(\prod_{k=L}^{K} |D_{k,j_k}(t)|^2 \Big) \Big| \sum_{\substack{j_1 \leqslant J_1 \\ \vdots \\ j_{L-1} \leqslant J_{L-1}}} \Big(\prod_{k=1}^{L-1} D_{k,j_k}(t) \Big) \sum_{m \in I_{j_1,\dots,j_K}} f(m) m^{it} \Big|^2 dt \\ &\ll (P_L^{(j_L)})^{2-2\eta_L+o(1)} \int\limits_{t \in \mathcal{T}_L} \Big| \sum_{\substack{j_1 \leqslant J_1 \\ \vdots \\ j_{L-1} \leqslant J_{L-1}}} \Big(\prod_{k=1}^{L-1} D_{k,j_k}(t) \Big) \sum_{m \in I_{j_1,\dots,j_K}} f(m) m^{it} \Big|^2 dt. \end{aligned}$$

We know that for all $t \in \mathcal{T}_L$ there exists a scale j'_{L+1} for which $|D_{L+1,j'_{L+1}}(t)| \geqslant (P_{L+1}^{(j'_{L+1})})^{1-\eta_{L+1}}$. Decomposing $\mathcal{T}_L$ according to which scale j'_{L+1} occurs, and using the fact that $J_{L+1} \ll (P_L^{(j_L)})^{o(1)}$, we may bound $C_{j_L,\dots,j_K}$ above by

$$\ll_K (P_L^{(j_L)})^{2-2\eta_L+o(1)} \int\limits_{t \in \mathcal{T}_{L,\prime}} \Big| \sum_{\substack{j_1 \leqslant J_1 \\ \vdots \\ j_{L-1} \leqslant J_{L-1}}} \Big(\prod_{k=1}^{L-1} D_{k,j_k}(t) \Big) \sum_{m \in I_{j_1,\dots,j_K}} f(m) m^{it} \Big|^2 dt$$

where $\mathcal{T}_{L,\prime} \subset \mathcal{T}_L$ is the subset corresponding to a particular choice of scale j'_{L+1}.

By assumption, we therefore have the upper bound of

$$\begin{aligned} \ll_K (P_L^{(j_L)})^{2-2\eta_L+o(1)} &((P_{L+1}^{(j'_{L+1})})^{-2+2\eta_{L+1}})^d \\ &\int\limits_{-X\theta/H}^{X\theta/H} |D_{L+1,j'_{L+1}}(t)|^{2d} \Big| \sum_{\substack{j_1 \leqslant J_1 \\ \vdots \\ j_{L-1} \leqslant J_{L-1}}} \Big(\prod_{k=1}^{L-1} D_{k,j_k}(t) \Big) \sum_{m \in I_{j_1,\dots,j_K}} f(m) m^{it} \Big|^2 dt \end{aligned}$$

for any $d \in \mathbb{N}$. Choose

$$d = \Big\lceil \frac{\log P_L^{(j_L)}}{\log P_{L+1}^{(j'_{L+1})}} \Big\rceil,$$

in order to make $D_{L+1,j'_{L+1}}(t)^d$ have length $\approx P_L^{(j_L)}$, and therefore the entire Dirichlet polynomial to have length $\approx X$.

What remains is some sweat to estimate all of these various terms. This is a common theme in complicated analytic number theory proofs; one proceeds via heuristic arguments based on intuition and some simpler cases (i.e. roughly what you want the lengths of the

relevant Dirichlet polynomials to be), but then in the end there is no substitute for crunching the precise estimates.

We know that the coefficients $c(n)$ of the Dirichlet polynomial

$$\sum_{\substack{j_1\leqslant J_1\\ \dots\\ j_{L-1}\leqslant J_{L-1}}} \Big(\prod_{k=1}^{L-1} D_{k,j_k}(t)\Big) \sum_{m\in I_{j_1,\dots,j_K}} f(m)m^{it}$$

satisfy the bound $|c(n)|\leqslant \prod_{k\leqslant L-1}\omega_k(n)$, and are supported on a range $n\ll_K X/\prod_{L\leqslant k\leqslant K} P_k^{(j_k)}$. So, by Montgomery's mean value theorem, we get that

$$\int_{-X\theta/H}^{X\theta/H} |D_{L+1,j'_{L+1}}(t)|^{2d}\Big|\sum_{\substack{j_1\leqslant J_1\\ \dots\\ j_{L-1}\leqslant J_{L-1}}} \Big(\prod_{k=1}^{L-1} D_{k,j_k}(t)\Big) \sum_{m\in I_{j_1,\dots,j_K}} f(m)m^{it}\Big|^2 dt$$

is

$$\ll_K \Big(\frac{(P_{L+1}^{(j'_{L+1})})^d X}{\prod_{L\leqslant k\leqslant K} P_k^{(j_k)}} + \frac{X\theta}{H}\Big)\sum_{n\geqslant 1}\Big|\sideset{}{^*}\sum_{p_1\dots p_d m=n}\prod_{k\leqslant L-1}\omega_k(m)\Big|^2$$

where the sum $\sum^*$ is over each $p_i\in \mathcal{P}_{L+1}^{(j'_{L+1})}$ and m in some range

$$\frac{X}{\prod_{L\leqslant k\leqslant K} P_k^{(j_k)}} \ll_K m \ll_K \frac{X}{\prod_{L\leqslant k\leqslant K} P_k^{(j_k)}}.$$

(Note how I've relabelled the summation variable m.)

This summation is

$$\ll_K \sum_{\substack{p_1,\dots,p_d\in\mathcal{P}_{L+1}^{(j'_{L+1})}\\ q_1,\dots,q_d\in\mathcal{P}_{L+1}^{(j'_{L+1})}}} \sum_{\substack{n\\ p_1\dots p_d|n\\ q_1\dots q_d|n}} \prod_{k\leqslant L-1}\omega_k(n/(p_1\dots p_d))\omega_k(n/(q_1\dots q_d)),$$

where the range of n summation is

$$n\ll_K \frac{(P_{L+1}^{(j'_{L+1})})^d X}{\prod_{L\leqslant k\leqslant K} P_k^{(j_k)}}.$$

Because the primes in $\mathcal{P}_{L-1}^{(j'_{L-1})}$ have no affect on the value of $\omega_k(n)$ for $k\neq L-1$, the above is

$$\ll_K \sum_{\substack{p_1,\dots,p_d\in\mathcal{P}_{L-1}^{(j'_{L-1})}\\ q_1,\dots,q_d\in\mathcal{P}_{L-1}^{(j'_{L-1})}}} \sum_{\substack{n\\ p_1\dots p_d|n\\ q_1\dots q_d|n}} \prod_{k\leqslant L-1}\omega_k(n)^2.$$

The contribution when $\{p_1,\dots,p_d\}\cap\{q_1,\dots,q_d\}=\emptyset$ is

$$\ll_K \Big(\prod_{k\leqslant L-1} W_k^2\Big)\frac{(P_{L+1}^{(j'_{L+1})})^d X}{\prod_{L\leqslant k\leqslant K} P_k^{(j_k)}}\Big(\sum_{p\in\mathcal{P}_{L+1}^{(j'_{L+1})}}\frac{1}{p}\Big)^{2d} \ll \Big(\prod_{k\leqslant L-1} W_k^2\Big)\frac{(P_{L+1}^{(j'_{L+1})})^d X}{\prod_{L\leqslant k\leqslant K} P_k^{(j_k)}}$$

since

$$\sum_{p\in\mathcal{P}_{L-1}^{(j'_{L-1})}}\frac{1}{p}\leqslant 1$$

as $\mathcal{P}_{L-1}^{(j'_{L-1})}$ is a dyadic range. Bounding the contributions when $\{p_1,\dots,p_d\}\cap\{q_1,\dots,q_d\}\neq\emptyset$ similarly, we are left with a bound of

$$\ll_K d^{O(d)}\Big(\prod_{k\leqslant L-1}W_k^2\Big)\frac{(P_{L+1}^{(j'_{L+1})})^dX}{\prod_{L\leqslant k\leqslant K}P_k^{(j_k)}}.$$

Bringing everything together, we get that the mean value of the Dirichlet polynomial under consideration is

$$\ll_K d^{O(d)}\Big(\prod_{k\leqslant L-1}W_k^2\Big)\Big(\frac{(P_{L+1}^{(j'_{L+1})})^dX}{\prod_{L\leqslant k\leqslant K}P_k^{(j_k)}}+\frac{X\theta}{H}\Big)\frac{(P_{L+1}^{(j'_{L+1})})^dX}{\prod_{L\leqslant k\leqslant K}P_k^{(j_k)}}$$

so overall we have

$$\Big(\prod_{k=1}^{L-1}W_k^{-2}\Big)C_{j_L,\dots,j_K}$$

$$\ll_K d^{O(d)}(P_L^{(j_L)})^{2-2\eta_L+o(1)}((P_{L+1}^{(j'_{L+1})})^{-2+2\eta_{L+1}})^d\Big(\frac{(P_{L+1}^{(j'_{L+1})})^dX}{\prod_{L\leqslant k\leqslant K}P_k^{(j_k)}}+\frac{X\theta}{H}\Big)\frac{(P_{L+1}^{(j'_{L+1})})^dX}{\prod_{L\leqslant k\leqslant K}P_k^{(j_k)}}.$$

By the choice of d, we may upper bound this by X^2 times

$$\begin{aligned}&\ll_K d^{O(d)}(P_L^{(j_L)})^{2-2\eta_L+o(1)}((P_{L+1}^{(j'_{L+1})})^{-2+2\eta_{L+1}})^d(P_{L+1}^{(j'_{L+1})})^2\\&\ll_K d^{O(d)}(P_L^{(j_L)})^{-2\eta_L+2\eta_{L+1}+o(1)}(P_{L+1}^{(j'_{L+1})})^{O(1)}\\&\ll_K d^{O(d)}(P_L^{(j_L)})^{-\frac{1}{100K}}Q_{L+1}^{O(1)}.\end{aligned}$$

The right-hand side needs to be $o_K(1)$.

Finally (!) we get to use the explicit form of the scales $[P_k,Q_k]$ that we defined towards the start of the lecture. This means that

$$d\ll\frac{\log P_L^{(j_L)}}{\log P_{L+1}}\ll\frac{\log P_L^{(j_L)}}{\log_{L+2}(X)^2}$$

and

$$\log d\ll\log\log Q_L-\log\log P_{L+1}\ll\log_{L+2}(X),$$

and so

$$d^{O(d)}\ll\exp(O(d\log d))\ll\exp\Big(O\Big(\frac{\log P_L^{(j_L)}}{\log_{L+2}(X)}\Big)\Big)\ll(P_L^{(j_L)})^{o_K(1)}.$$

Since $Q_{L+1}^{O(1)}=(P_L^{(j_L)})^{o_K(1)}$ as well, we are left with an overall bound of $O_K((P_L^{(j_L)})^{-\frac{1}{200K}})$, which is $o_K(1)$ and is therefore acceptable. Phew!

So all that remains is to control the integral over the small exceptional set $\mathcal{E}$. That will be the topic of next lecture.

14. Endgame

Today we will will finish the proof of the Matomäki–Radziwiłł theorem (well, at least when $H \geqslant \log_K(X)$ for some fixed K). Actually, preparing the detailed notes for this lecture, I've realised that to give a full proof it will actually be necessary to restrict matters to the case in which the multiplicative function in question is the Liouville function λ. At the end I will indicate the ideas that go into generalising matters to arbitrary multiplicative functions, in some (but not total) detail.

There is comparatively little left to do, given the complications that have come our way thus far. However, we will nonetheless need to rely on some further deep estimates, given below. Here, $\pi(y) := \sum_{p \leqslant y} 1$ (which is a standard notation that somehow I have managed to avoid introducing up until now).

Theorem 14.1 (Vinogradov–Korobov bounds)**.** *For any $\delta > 0$, and for all $y \geqslant 2$ and $t \in \mathbb{R}$, we have*

$$\sum_{n \leqslant y} \lambda(n) n^{it} \ll_\delta y \exp\Big(- \frac{\log y}{(\log(y + |t|))^{\frac{2}{3}+\delta}}\Big),$$

$$\sum_{p \leqslant y} p^{it} \ll_\delta \frac{\pi(y)}{1+|t|} + y \exp\Big(- \frac{\log y}{(\log(y + |t|))^{\frac{2}{3}+\delta}}\Big),$$

and

$$\sum_{p \leqslant y} \Big(1 - \frac{p}{y}\Big) p^{it} \ll_\delta \frac{\pi(y)}{1+|t|^2} + y \exp\Big(- \frac{\log y}{(\log(y + |t|))^{\frac{2}{3}+\delta}}\Big).$$

I'm afraid that I won't have time to prove these bounds. In fact we're at least three lectures short on time for this, and the techniques would take us dramatically far afield from our central task. So, in lieu of a proof, I can give you:

- some contextual remarks, to help you appreciate the strength of these estimates in comparison to what we have proved before this point;
- a sketch proof using complex analytic methods, assuming a suitable zeros free region.

First the contextual remarks. The bounds are non-trivial for $|t|$ as large as $\exp((\log y)^{3/2-\delta})$, which is significantly larger than y. The prime number theorem with classical error term (the state of the art at the start of the 20th century) does not yield anything non-trivial for $|t| \gg y$.

Here's another perspective. Recall that back in Lecture 3 we showed that

$$|\zeta(1 + \frac{1}{\log y} + it)| \ll \log(|t| + 2),$$

and thereby demonstrated the lower bound

$$\mathbb{D}(1, n^{it}; y)^2 \geqslant \log\log y - \log\log(|t| + 2) - O(1)$$

when $|t| \geqslant 1$. Well this lemma also becomes trivial once $|t| \geqslant y$. However, a deep result of Vinogradov–Korobov from the middle of the last century (discovered independently in 1958, but building on previous work of Vinogradov), improves the zeta bound to

$$|\zeta(\sigma + it)| \ll (\log|t|)^{2/3} (\log\log|t|)^{1/3}$$

if $\sigma \geqslant 1 - c(\log|t|)^{-2/3}(\log\log|t|)^{-1/3}$ for some absolute $c > 0$. This is still the best known exponent, and leads to the bound

$$\mathbb{D}(1, n^{it}; y)^2 \gg \log\log y, \qquad |t| \leqslant \exp((\log y)^{1.4})),$$

say.

Now, the Vinogradov–Korobov bound is based on an exponential sum estimate on the short interval sum $\sum_{N\leqslant n\leqslant N+M} n^{it}$. After some Taylor expansion and bilinear sum techniques, one can prove highly non-trivial bounds on this exponential sum by bounding the number of solutions to the system of equations

$$\begin{aligned} x_1+\dots x_s &= x_{s+1}+\dots+x_{2s}\\ x_1^2+\dots x_s^2 &= x_{s+1}^2+\dots+x_{2s}^2\\ &\dots\dots\\ x_1^k+\dots x_s^k &= x_{s+1}^k+\dots+x_{2s}^k \end{aligned}$$

with $1\leqslant x_i\leqslant X$ for all $i\leqslant k$, for certain parameters k and s. Non-trivial bounds to the number of such solutions are known as 'Vinogradov's mean value theorem', and are a whole world in and of themselves. If you're interested, have a look at the second half of Chapter 8 of Iwaniec–Kowalski 'Analytic Number Theory'. Or if you're *really* interested, then you can look at the spectacular recent literature by Wooley and by Bourgain–Demeter–Guth (ask me for specific references if you want them).

Coming back down to earth, let me sketch how to use the zero-free region to establish some of the bounds in Theorem 14.1. [This will only make any sort of sense if you have attended a first course in analytic number theory.] Let me stress that it is not necessary to use Cauchy's Theorem to prove bounds of this strength. However, formulating such a proof is not easy, and wasn't worked out until Dimitris Koukoulopoulos did it in 2013.

Sketch proof. By Perron's formula one has

$$\sum_{n\leqslant y}\lambda(n)n^{it}=\frac{1}{2\pi i}\int\limits_{1+1/\log y-iy}^{1+1/\log y+iy}\frac{\zeta(2w-2it)}{\zeta(w-it)}\frac{y^w}{w}\,dw+O_\varepsilon(y^\varepsilon).$$

Now move the line of integration to $\Re(w)=1-(\log(y+|t|))^{-2/3-\delta}$, picking up no contributions from any zeros (by the Vinogradov–Korobov zero free region), and use known upper bounds on ζ and ζ^{-1} in this region.

For the other sums, apply Perron to the function F in order to obtain

$$\sum_{n\leqslant y}F(n/y)\Lambda(n)n^{it}=\frac{-1}{2\pi i}\int\limits_{1+1/\log y-iy}^{1+1/\log y+iy}\frac{\zeta'(w-it)}{\zeta(w-it)}y^w\widetilde{F}(w)dw+O_\varepsilon(y^\varepsilon),$$

where $\widetilde{F}(w)$ is the Mellin transform of F, and do the same thing. (If $F(x)=1_{[0,1]}(x)$ we get $\widetilde{F}(w)=w^{-1}$, whereas if $F(x)=(1-|x|)$ when $|x|\leqslant 1$ and $F(x)=0$ otherwise we get $\widetilde{F}(w)=(w(w+1))^{-1}$.) This time we pick up a pole at $w=1+it$. □

The third bound in this theorem is clearly closely related to the bound we spent a long time considering in Lecture 11, namely

$$\sum_{n\leqslant 2N}\Big(2-\frac{n}{N}\Big)n^{it}\ll\frac{N}{1+|t|^2}+(1+|t|)^{1/2}\log(2+|t|).$$

Let's remind ourselves where we are in our general effort. We were proving Matomäki–Radziwiłł for the function $H=\exp((\log_{K+1}(X)^3)$, where K was large and fixed. Last time, we reduced matters to the bound

$$\Big(\prod_{k\leqslant K}W_k^{-2}\Big)\int\limits_{-X\theta/H}^{X\theta/H}|A(t)|^2\,dt=o_K(X^2),$$

where $\theta = \theta(X)$ was any function tending to infinity with X, and $A(t)$ was a complicated Dirichlet polynomial of length $\approx X$ built out of a factorisation according to the scales $[P_1, Q_1], \dots, [P_K, Q_K]$ and a further dyadic decomposition, and $W_k \approx \log\log Q_k - \log\log P_k$. In detail,

$$A(t) = \sum_{\substack{j_1 \leqslant J_1 \\ \dots \\ j_K \leqslant J_K}} \Big(\prod_{k=1}^{K} \sum_{p \in \mathcal{P}_k^{(j_k)}} f(p)p^{it}\Big)\Big(\sum_{m \in I_{j_1,\dots,j_K}} f(m)m^{it}\Big),$$

where $m \in I_{j_1,\dots,j_K}$ if

$$\frac{X}{2^K \prod_{k\leqslant K} P_k^{(j_k)}} \leqslant m \leqslant \frac{4X}{\prod_{k\leqslant K} P_k^{(j_k)}},$$

and $J_k = \frac{\log Q_k}{(\log 2)(\log P_k)}$, $P_k^{(j_k)} := 2^{j_k - 1}P_k$ and $\mathcal{P}_k^{(j_k)}$ is the set of all primes in the interval $[P_k^{(j_k)}, 2P_k^{(j_k)})$.

We split the range of integration $[-X\theta/H, X\theta/H]$ into different sets $\mathcal{T}_1, \dots, \mathcal{T}_K$ and $\mathcal{E}$, according to whether the Dirichlet polynomials

$$D_{k,j_k}(t) := \sum_{p \in \mathcal{P}_k^{(j_k)}} f(p)p^{it}$$

were small or large. Using Montgomery's mean value theorem, together with an ingenious 'Dirichlet polynomial lengthening' device, we controlled the contribution from $\mathcal{T}_1, \dots, \mathcal{T}_K$.

What remains is to control the contribution over $\mathcal{E}$. An important first step will be to determine an upper-bound on $|\mathcal{E}|$. It turns out that Montgomery's mean value theorem will be adequate.

Lemma 14.2. *Let $2 \leqslant P \leqslant T$. Let $c(p)$ be any sequence of complex numbers, defined on primes p, with $|c(p)| \leqslant 1$. Let $V \geqslant 3$ be a real number and let $\mathcal{T}$ denote the set of values $|t| \leqslant T$ such that $|\sum_{p\leqslant P} c(p)p^{it}| \geqslant \pi(P)/V$. Then*

$$|\mathcal{T}| \ll (V^2 \log T)^{1+(\log T)/(\log P)}.$$

Proof. Let $k = \lceil (\log T)/(\log P) \rceil$ so that $P^k \geqslant T$. Write

$$\Big(\sum_{p\leqslant P} c(p)p^{it}\Big)^k = \sum_{n \leqslant P^k} c_k(n)n^{it}.$$

Note that $|c_k(n)| \leqslant k!$ and that

$$\sum_{n\leqslant P^k} |c_k(n)| \leqslant \Big(\sum_{p\leqslant P} |c(p)|\Big)^k \leqslant \pi(P)^k.$$

Therefore, from Montgomery's mean value theorem, we have

$$|\mathcal{T}|\Big(\frac{\pi(P)}{V}\Big)^{2k} \leqslant \int_{-T}^{T} \Big|\sum_{p\leqslant P} c(p)p^{it}\Big|^{2k} dt \ll (T + P^k) \sum_{n\leqslant P^k} |c_k(n)|^2 \ll k!P^k\pi(P)^k.$$

What remains is some explicit estimation of the various terms. Observe that the claimed bound in the lemma is trivial if $T = O(1)$, and similarly is also trivial if $P = O(1)$ (as the claimed bound is weaker than the trivial $|\mathcal{T}| \ll T$). Therefore we may assume that $P/\pi(P) \leqslant 2\log P$. From Stirling's Theorem (or otherwise) we get $k! \ll ((k-1)/2)^k$. Therefore

$$|\mathcal{T}| \ll V^{2k}k!(2\log P)^k \ll V^{2k}\Big(\frac{\log T}{2\log P}\Big)^k (2\log P)^k \ll (V^2 \log T)^k,$$

and this gives the lemma. □

The point about the choice of k is that it makes the length of the Dirichlet polynomial essentially equal to the length of integration, which is range where Montgomery's estimate is most powerful.

Let us apply this lemma to our exceptional set $\mathcal{E}$ from our main estimation. If $t \in \mathcal{E}$ we know that there is some $j_1 \leqslant J_1$ for which $|D_{1,j_1}(t)| \geqslant (P_1^{(j_1)})^{1-\eta_1} = (P_1^{(j_1)})^{9/10}$. By the previous lemma, the measure of such t is

$$\begin{aligned}
&\ll ((P_1^{(j_1)})^{\frac{2}{10}} \log(X\theta/H))^{1+(\log(X\theta/H))/(\log P_1^{(j_1)})} \\
&\ll \exp(\frac{2}{10}\log X + \frac{2}{10}\log Q_1 + \log X \frac{\log\log X}{\log P_1}) \\
&\ll \exp(\frac{2}{10}\log X + \frac{2}{10}(\log\log X)^3 + \frac{\log X}{\log\log X}) \\
&\ll X^{\frac{2}{10}+o(1)}.
\end{aligned}$$

Multiplying by J_1, which is $\ll \log\log X$, we we obtain

$$|\mathcal{E}| \ll X^{\frac{1}{4}},$$

say. Notice how we have relied greatly on the fact that $(\log P_1)/(\log\log X) \to \infty$ as $X \to \infty$, so we could not have applied such an argument to P_k for any $k \geqslant 2$.

Now, the Halasz–Montgomery mean value theorem gives

$$\int_{t\in\mathcal{E}} |A(t)|^2\,dt \ll_K (X + |\mathcal{E}|X^{1/2}\log X)\sum_n |a(n)|^2 \ll_K (X + X^{3/4+o(1)})X\prod_{k\leqslant K} W_k^2$$

since $|a(n)| \leqslant \prod_{k\leqslant K}\omega_k(n)$. Of course this doesn't actually give us the necessary bound of $o_K(X^2\prod_{k\leqslant K}W_k^2)$. But we're nearly there, and the term coming from 'average behaviour' over t, namely the second term, *is* acceptable. This is good news.

In order to improve the 'L^∞ part' of the mean value estimate, like before we are going to employ a prime decomposition. Define the interval $[P_0, Q_0']$ by

$$P_0 = \exp((\log X)^{0.9}), \qquad Q_0' = \exp((\log X)^{0.95})$$

and then as before we let Q_0 be equal to $2^{J_0+1}P_0$ for some $J_0 \in \mathbb{N}$ such that $Q_0/Q_0' \asymp 1$. The interval $[P_0, Q_0]$ is still contained within $[1, X]$, but is substantially larger than any of the previous intervals $[P_k, Q_k]$.

Now I am going to be a little naughty: I am going to 'redefine' the Dirichlet polynomial $A(t)$. This is because the cleanest way to use this final scale in the proof is actually to introduce it right at the beginning, and carrying through the entirety of the arguments last time regarding $D_1, \dots, D_k$ without alteration. I felt that introducing $[P_0, Q_0]$ last time (which has a different role to $[P_k, Q_k]$ for $k \geqslant 1$) would have been exceptionally confusing, which is why I didn't do it. But of course you might legitimately complain that arranging matters this way is in itself confusing!

We define

$$B(t) = \sum_{\substack{j_0\leqslant J_0\\ \dots \\ j_K\leqslant J_K}} \Big(\prod_{k=0}^{K}\sum_{p\in\mathcal{P}_k^{(j_k)}} f(p)p^{it}\Big)\Big(\sum_{m\in I'_{j_0,\dots,j_K}} f(m)m^{it}\Big),$$

where $m \in I'_{j_0,\dots,j_K}$ if

$$\frac{X}{2^{K+1}\prod_{0\leqslant k\leqslant K}P_k^{(j_k)}} \leqslant m \leqslant \frac{4X}{\prod_{0\leqslant k\leqslant K}P_k^{(j_k)}},$$

and we will work with $B(t)$ in lieu of $A(t)$.

By the same argument as last time one derives

$$\frac{1}{X}\int_X^{2X}\Big|\frac{1}{H}\sum_{x\leqslant n<x+H}f(n)\Big|^2\,dx\ll\Big(\prod_{k=0}^K W_k^{-2}\Big)\frac{1}{X^2}\int_{-X\theta/H}^{X\theta/H}|B(t)|^2\,dt+o_K(1).$$

One can define the sets $\mathcal{T}_1,\dots,\mathcal{T}_K$ identically as before, and bound their contribution. Our task is reduced to showing that

$$\Big(\prod_{k=0}^K W_k^{-2}\Big)\frac{1}{X^2}\int_{\mathcal{E}}|B(t)|^2\,dt=o_K(1),\tag{10}$$

where $\mathcal{E}$ is an arbitrary measurable set of measure $|\mathcal{E}|\ll X^{1/4}$.

In a by-now familiar manoeuvre (in the sense as it's the same as we did last time), we arrange a dyadic decomposition on scale $j_0\leqslant J_0$ to obtain

$$\int_{\mathcal{E}}|B(t)|^2\,dt\leqslant W_0^2\max_{j_0\leqslant J_0}\mathcal{C}_{j_0},.$$

Here

$$\mathcal{C}_{j_0}:=(\log P_0^{(j_0)})^2\int_{\mathcal{E}}\Big|\sum_{p\in\mathcal{P}_0^{(j_0)}}f(p)p^{it}\Big|^2|M(t)|^2\,dt,$$

where

$$M(t):=\sum_{\substack{j_1\leqslant J_1\\ \dots\\ j_K\leqslant J_K}}\Big(\prod_{k=1}^K\sum_{p\in\mathcal{P}_k^{(j_k)}}f(p)p^{it}\Big)\Big(\sum_{m\in I'_{j_0,\dots,j_K}}f(m)m^{it}\Big)$$

is a Dirichlet polynomial of length $\asymp_K X/P_0^{(j_0)}$ and whose coefficients $m(n)$ satisfy $|m(n)|\leqslant\prod_{1\leqslant k\leqslant K}\omega_k(n)$.

Using the tools that we have introduced so far, we may prove our main theorem in the case where $f=\lambda$ is the Liouville function. In this instance, the second bound of Theorem 14.1 implies that

$$\begin{aligned}\sum_{p\in\mathcal{P}_0^{(j_0)}}p^{it}&\ll\frac{P_0^{(j_0)}}{\log P_0^{(j_0)}}\frac{1}{1+|t|}+P_0^{(j_0)}\exp\Big(-\frac{\log P_0^{(j_0)}}{(\log(P_0^{(j_0)}+|t|))^{2/3+\delta}}\Big)\\&\ll\frac{P_0^{(j_0)}}{\log P_0^{(j_0)}}\Big(\frac{1}{1+|t|}+\exp(-(\log X)^{0.1})\Big)\end{aligned}$$

when $|t|\leqslant X$. (Note that it is important that $P_0\geqslant\exp((\log X)^{2/3})$). So, from applying the Halasz–Montgomery bound to the Dirichlet polynomial $M(t)$ we obtain

$$\begin{aligned}&\int_{\substack{t\in\mathcal{E}\\ |t|\geqslant(\log X)}}\Big|\sum_{p\in P_0^{(j_0)}}p^{it}\Big|^2|M(t)|^2\,dt\\&\ll_K\frac{(P_0^{(j_0)})^2}{(\log P_0^{(j_0)})^2(\log X)^2}\Big(\frac{X}{P_0^{(j_0)}}+|\mathcal{E}|X^{1/2}(\log X)\Big)\sum_n|m(n)|^2\\&\ll_K\frac{(P_0^{(j_0)})^2}{(\log P_0^{(j_0)})^2(\log X)^2}\Big(\frac{X}{P_0^{(j_0)}}+X^{3/4+o(1)}\Big)\Big(\frac{X}{P_0^{(j_0)}}\prod_{1\leqslant k\leqslant K}W_k^2\Big)\end{aligned}$$

$$\ll_K \frac{X^2}{(\log P_0^{(j_0)})^2(\log X)^2}\prod_{1\leqslant k\leqslant K} W_k^2.$$

So the contribution to $\mathcal{C}_{j_0}$ from $|t|\geqslant \log X$ is at most $\frac{X^2}{(\log X)^2}\prod_{1\leqslant k\leqslant K} W_k^2$, which is acceptable as $\log X\to\infty$ as $X\to\infty$.

For $|t|\leqslant \log X$, we can use the first part of Theorem 14.1 to show that for all $j_0\leqslant J_0,\dots,j_K\leqslant J_K$ we have

$$\Big|\sum_{m\in I'_{j_0,\dots,j_K}}\lambda(m)m^{it}\Big|\ll_K \frac{X}{\prod_{0\leqslant k\leqslant K}P_k^{(j_k)}}\exp(-(\log X)^{0.1}),$$

since $\prod_{0\leqslant k\leqslant K}P_k^{(j_k)}=X^{o(1)}$. Summing over the dyadic scales, we obtain

$$|M(t)|\ll_K \frac{X}{P_0^{(j_0)}}\exp(-(\log X)^{0.05})$$

in the same range. Thus

$$\begin{aligned}(\log P_0^{(j_0)})^2&\int_{|t|\leqslant\log X}\Big|\sum_{p\in\mathcal{P}_0^{(j_0)}}f(p)p^{it}\Big|^2|M(t)|^2\,dt\\ &\ll (P_0^{(j_0)})^2\int_{|t|\leqslant\log X}|M(t)|^2\,dt\\ &\ll (P_0^{(j_0)})^2(\log X)\Big(\frac{X}{P_0^{(j_0)}}\exp(-(\log X)^{0.1})\Big)^2\\ &\ll o(X^2\prod_{1\leqslant k\leqslant K}W_k^2).\end{aligned}$$

So $\mathcal{C}_{j_0}$ is controlled, and we are entirely done in the case of the Liouville function $\square$.

Addendum

But what if f is not the Liouville function? Then we really don't know anything about where exactly $\sum_p f(p)p^{it}$ is small, and in particular we can't assume that this Dirichlet polynomial is always small for large t. This was an important part of our argument above.

What to do? Well, if

$$\Big|\sum_{p\in\mathcal{P}_0^{(j_0)}}f(p)p^{it}\Big|\leqslant\frac{P_0^{(j_0)}}{(\log P_0^{(j_0)})^2},$$

say, then we really can conclude as in the above argument, using the Halasz-Montgomery bound for the Dirichlet polynomial $M(t)$. So the remaining case is when

$$\Big|\sum_{p\in\mathcal{P}_0^{(j_0)}}f(p)p^{it}\Big|\geqslant\frac{P_0^{(j_0)}}{(\log P_0^{(j_0)})^2},$$

which we know from Lemma 14.2 occurs extremely rarely. In fact, if $\mathcal{T}$ is the set of such t, then we know that

$$|\mathcal{T}|\ll((\log P_0^{(j_0)})^2\log X)^{1+(\log X)/(\log P_0^{(j_0)})}\ll\exp((\log X)^{0.11}),$$

say.

Now, assume for the rest of today that we have some good uniformity bound regarding how non-pretentious f is, say we know that

$$\mathbb{D}(f,n^{it};X)^2\gg\log\log X$$

for all $|t| \leqslant X$. We then appeal to a quantitative version of Halasz's Theorem, which we mentioned at the end of Lecture 5 (and which actually does follow easily from our Lecture 5 methods too). This would give us

$$\Big| \sum_{m \in I'_{j_0,\dots,j_K}} f(m) m^{it} \Big| \ll_K \frac{X}{\prod_{k=0}^{K} P_k^{(j_k)}} \frac{1}{(\log X)^{\delta}}$$

for some absolute constant $\delta > 0$, provided $|t| \leqslant X$. One can reduce to this strong non-pretentious case by removing the contribution from small $|t|$ by another complicated device, but we will not cover this. (One should look to Lemma 4 of the original paper of Matomäki–Radziwiłł).

Summing over all dyadic scales, and using the fact that $\prod_{1 \leqslant k \leqslant K} J_k \ll (\log X)^{o(1)}$, we obtain the upper bound

$$\mathcal{C}_{j_0} \ll_K \frac{X^2 (\log P_0^{(j_0)})^2}{P_0^{(j_0)} (\log X)^{\delta - o(1)}} \int\limits_{t \in \mathcal{T}} \Big| \sum_{p \in \mathcal{P}_0^{(j_0)}} f(p) p^{it} \Big|^2 dt.$$

A trivial bound here would give $\mathcal{C}_{j_0} \ll X^2 (\log X)^{-\delta + o(1)} |\mathcal{T}|$, but unfortunately it is not necessarily true that $|\mathcal{T}| \ll (\log X)^{\delta}$. We have hit a problem that we mentioned way back in Lecture 5, namely that even an optimal version of a general Halasz's theorem doesn't actually give us very strong cancellation.

An application of Halasz–Montgomery will also fail us, as this would give a bound of

$$\frac{X^2 (\log P_0^{(j_0)})^2}{P_0^{(j_0)} (\log X)^{\delta - o(1)}} \Big(P_0^{(j_0)} + |\mathcal{T}| X^{1/2} \log X \Big) \frac{P_0^{(j_0)}}{\log P_0^{(j_0)}},$$

which is too weak by a factor of $\log P_0^{(j_0)}$.

The problem was that the general Halasz–Montgomery mean value theorem wasn't sensitive to the fact that the Dirichlet polynomial in question here is supported on primes, which are sparse. However, Matomäki–Radziwiłł were able to prove a version of this mean value theorem which regains this sparseness factor in the specific case that the coefficients are supported on primes.

Theorem 14.3 (Halasz–Montgomery for primes)**.** *Let $T \geqslant 2$, let $\mathcal{T}$ be a measurable subset of $[-T, T]$. Then for any complex numbers $c(p)$ and any $\varepsilon > 0$,*

$$\int\limits_{\mathcal{T}} \Big| \sum_{p \leqslant P} c(p) p^{it} \Big|^2 dt \ll \Big(\frac{P}{\log P} + |\mathcal{T}| P \exp\Big(- \frac{\log P}{(\log(T+P))^{2/3+\varepsilon}} \Big) \Big) \sum_{p \leqslant P} |c(p)|^2.$$

Proof. I'll leave this to you, but you already have all the pre-requisite material to do this easily yourself. Follow the proof of the Halasz–Montgomery bound, but use the bound from Theorem 14.1 on

$$\sum_{p \leqslant P} \Big(1 - \frac{p}{P} \Big) p^{it}$$

in place of the bound on

$$\sum_{n \leqslant N} \Big(1 - \frac{n}{N} \Big) n^{it}.$$

□

15. Correlations of multiplicative functions

In the final two lectures of the course, we will discuss Tao's proof of the following theorem:

Theorem 15.1. *For all $\varepsilon > 0$, for all w large enough in terms of ε, and for all X with $X/\log X \geqslant w$, we have*

$$\Big| \sum_{X/w < n \leqslant X} \frac{\lambda(n)\lambda(n+1)}{n} \Big| \leqslant \varepsilon \log w. \tag{11}$$

One shouldn't read too much into the parameter w; at certain points in the argument, it will simply be convenient that n is not very small. Of course, Tao also proved a more general result for all non-pretentious $f \in \mathcal{M}_0$ (in which it is necessary to preclude cases in which $\mathbb{D}(f, \chi(n)n^{it}; X)$ is small, for all Dirichlet characters χ of conductor up to a certain size), but we won't attack the general case in these notes.

Note that the trivial bound on the left-hand side of equation (11) is $\log w$. Letting ε tend to 0 sufficiently slowly as $X \to \infty$, and $w = X/\log X$, one may conclude that

$$\sum_{n \leqslant X} \frac{\lambda(n)\lambda(n+1)}{n} = o(\log X)$$

as $X \to \infty$.

Unlike earlier in the course, we will not be focussed on providing absolutely all of the details of the proof of Theorem 15.1. Our focus will rather be in providing an overall flavour of the argument (although actually we will be able to furnish almost all of the details in the time available). Thus our exposition will fall somewhere in between a seminar-style sketch and an undergraduate-style lecture.

Why is the twin prime conjecture is hard?

Right back in Lecture 1, I noted that there was some heuristic analogy between the correlations $\lambda(n)\lambda(n+1)$ and $\Lambda(n)\Lambda(n+2)$. We seem to be a long way off proving that $\sum_{n \leqslant X} \Lambda(n)\Lambda(n+2) \to \infty$, and I think it is worth taking 10 minutes or so to discuss why this is so – even if only to make the achievement of understanding $\sum_{n \leqslant X} \lambda(n)\lambda(n+1)/n$ appear all the more impressive.

There are two tools that one might try to bring to bear on the sum $\sum_{n \leqslant X} \Lambda(n)\Lambda(n+2)$. The first is the circle method, namely by noting that

$$\sum_{n \leqslant X} \Lambda(n)\Lambda(n+2) = \int_0^1 \Big| \sum_{n \leqslant X} \Lambda(n) e(n\theta) \Big|^2 e(2\theta)\, d\theta.$$

One could split the interval into a union $\mathfrak{M}$ of major arcs, when θ is close to a rational with small denominator, and a minor arc $\mathfrak{m} = [0,1) \setminus \mathfrak{M}$, expecting that the main contribution to the integral will come from the major arcs. [A standard major arc would be something like

$$\mathfrak{M} := \bigcup_{q \leqslant \log^A X} \bigcup_{\substack{a \leqslant q \\ (a,q)=1}} \Big\{ \theta \in [0,1) : \Big| \theta - \frac{a}{q} \Big| \leqslant \frac{\log^B X}{Xq} \Big\},$$

for certain values of A and B, but there are others.] So,

$$\begin{aligned} \int_0^1 \Big| \sum_{n \leqslant X} \Lambda(n) e(n\theta) \Big|^2 e(2\theta)\, d\theta &= \int_{\theta \in \mathfrak{M}} \Big| \sum_{n \leqslant X} \Lambda(n) e(n\theta) \Big|^2 e(2\theta)\, d\theta + \int_{\theta \in \mathfrak{m}} \Big| \sum_{n \leqslant X} \Lambda(n) e(n\theta) \Big|^2 e(2\theta)\, d\theta \\ &\geqslant \int_{\theta \in \mathfrak{M}} \Big| \sum_{n \leqslant X} \Lambda(n) e(n\theta) \Big|^2 e(2\theta)\, d\theta - \int_{\theta \in \mathfrak{m}} \Big| \sum_{n \leqslant X} \Lambda(n) e(n\theta) \Big|^2 d\theta \end{aligned}$$

and actually one can compute the integral over $\theta \in \mathfrak{M}$ to give

$$2\prod_{p\geqslant 3}\Big(1-\frac{1}{(p-1)^2}\Big)X + o(X).$$

(Exercise.) One might hope that the integral of the absolute value over the minor arcs is $o(X)$. But it isn't! One may show that by removing the $e(2\theta)$ term one ends up with a minor arc term of size $\gg X$. So unless one can take advantage of the oscillations given by the $e(2\theta)$ term – which seems almost impossible – a naïve circle-method approach is doomed.

Another way of trying to detect primes is by a sieve. However, as a general rule, sieves have a hard time (a) proving lower bounds on the number of primes in a certain set (they work better for providing upper bounds), and (b) distinguishing between numbers with an even number of prime factors and numbers with an odd number of prime factors.

This second phenomenon is known as 'the parity problem'. This term is overused in the field – plenty of very clever people have (Friedlander, Iwaniec, Heath–Brown, Harman, Goldston–Pintz–Yildrim, Green, Maynard, Tao etc.) have found ways to use sieves to detect primes themselves, in certain situations – but there is one specific case in which one has a more precise articulation of this parity obstruction: the linear sieve.

Consider the set $\mathcal{A} = \{p+2 : p \in [X/2, X]\}$. We want to detect primes in $\mathcal{A}$. For all (odd) square-free $d \leqslant X^u$ one has

$$\sum_{\substack{n\in\mathcal{A}\\ d|n}} 1 \approx \frac{1}{d}\sum_{n\in\mathcal{A}} 1.$$

The linear sieve is a general tool for taking any sequence $\mathcal{A}$ that obeys this property, and finding bounds for the number of 'rough' integers in $\mathcal{A}$. Approximately speaking, it gives us two absolute functions f and F for which

$$f(u)\frac{|\mathcal{A}|}{\log(X^u)} <\approx \sum_{\substack{n\in\mathcal{A}\\ p|n\Rightarrow p>X^u}} 1 <\approx F(u)\frac{|\mathcal{A}|}{\log(X^u)}.$$

The functions f and F are defined in a somewhat complicated way, but all you need to know is that $f(u) = 0$ if $u \geqslant 1/2$. In particular, this lower bound does not allow us to establish the twin prime conjecture by sieving all the way up to $X^{1/2}$. So the twin prime conjecture cannot be proved by this general tool.

But the terrible thing (at least as far as proving the twin prime conjecture is concerned!) is that these functions f and F are optimal in this generality, in the sense that if one lets

$$\mathcal{B} = \{n : n \in [X/2, X],\, \lambda(n) = -1\},$$

then one can show using the prime number theorem that

$$\sum_{\substack{n\in\mathcal{B}\\ d|n}} 1 \approx \frac{1}{d}\sum_{n\in\mathcal{B}} 1$$

and

$$f(u)\frac{|\mathcal{B}|}{\log(X^u)} \approx \sum_{\substack{n\in\mathcal{B}\\ p|n\Rightarrow p>X^u}} 1.$$

So there is no way of improving the function $f(u)$ in this generality.

Note that the bad case $\mathcal{B}$ consisted of numbers with an odd number of prime factors. Bombieri showed back in the sixties (the Bombieri sieve) that this parity issue is, in some precise sense, the 'only' obstruction here. But to describe that work would take us too far

from our main theme.

Two preliminary results

Before starting the proof proper of Tao's result, I need to furnish you with two preliminary results. One is a consequence of the Matomäki–Radziwiłł theorem, and the other is a standard large-deviation inequality from probability theory.

Theorem 15.2 (Matomäki–Radziwiłł–Tao)**.** *If $H \to \infty$ as $X \to \infty$, with $H \leqslant X$, then*

$$\sup_{\alpha\in[0,1]} \int_X^{2X} \Big| \sum_{x<n\leqslant x+H} \lambda(n)e(n\alpha)\Big|\, dx = o(HX).$$

How is this proved? Well, when $\alpha = 0$ one sees that this is exactly the manner of average that we considered in the Matomäki–Radizwiłł theorem, and so cancellation follows. A similar argument works when α is in a 'major arc'. Indeed, observe that

$$\begin{aligned}\int_X^{2X} \Big| \sum_{\substack{x<n\leqslant x+H\\(n,r)=1}} \lambda(n)e(na/r)\Big|\, dx &= \int_X^{2X} \Big| \sum_{\substack{x<n\leqslant x+H\\(n,r)=1}} \lambda(n) \sum_{b\leqslant r} e\Big(\frac{nb}{r}\Big)\frac{1}{\varphi(r)} \sum_{\chi \bmod r} \chi(a^{-1}b)\Big|\, dx\\ &= \int_X^{2X} \Big|\frac{1}{\varphi(r)} \sum_{\chi \bmod r} \sum_{\substack{x<n\leqslant x+H\\(n,r)=1}} \lambda(n)\overline{\chi(a)} \sum_{b\leqslant r} e\Big(\frac{nb}{r}\Big)\chi(b)\Big|\, dx\\ &= \sum_{\chi \bmod r} \frac{|\tau(\chi)|}{\varphi(r)} \int_X^{2X} \Big| \sum_{\substack{x<n\leqslant x+H\\(n,r)=1}} \lambda(n)\overline{\chi}(n)\Big|\, dx,\end{aligned}$$

which (if r grows slowly enough in terms of H) can be controlled as $o(HX)$ using Matomäki–Radziwiłł on the function $\lambda\overline{\chi}$. (But note that we will need the general non-pretentious version for complex-valued multiplicative functions in $\mathcal{M}_0$, not just a version for real-valued functions).

The minor arc case is dealt with by very similar bounds to the ones we used back in Lecture 9 to deal with minor arcs in that case (albeit one needs to be rather more precise than we were being – these techniques are due to Katai, and Bourgain–Sarnak–Ziegler).

The long-average version of this bound, namely

$$\Big| \sum_{n\leqslant X} \lambda(n)e(n\theta)\Big| = O_A\Big(\frac{X}{\log^A X}\Big)$$

is due to Davenport, and is on the examples sheet.

The next pre-requisite concerns a large deviation inequality.

Theorem 15.3 (Hoeffding's inequality)**.** *Let $Z_1, \dots, Z_n$ be independent random variables with $Z_i \in [a, b]$ for all i. Then*

$$\mathbb{P}\Big(\Big| \sum_{i=1}^n Z_i - \sum_{i=1}^n \mathbb{E}(Z_i)\Big| \geqslant nt\Big) \leqslant 2\exp\Big(\frac{-2nt^2}{(b-a)^2}\Big).$$

In words, this says that the sum of a sequence of independent samples is exponentially unlikely to be far from its mean.

One may put some more precise heuristics in here. Suppose further that the Z_i were i.i.d with mean 0 and variance 1. Then by the central limit theorem we have

$$\mathbb{P}\Big(\Big|\sum_{i=1}^n Z_i\Big| \geqslant \theta\sqrt{n}\Big) \to \frac{1}{\sqrt{2\pi}}\int_\theta^\infty \exp\Big(-\frac{x^2}{2}\Big)\,dx$$

as $n \to \infty$, for every fixed θ. So, playing somewhat fast and lose with the uniformity in θ, we might expect that

$$\mathbb{P}\Big(\Big|\sum_{i=1}^n Z_i\Big| \geqslant tn\Big) \approx \int_{t\sqrt{n}}^\infty \exp\Big(-\frac{x^2}{2}\Big)\,dx \approx \exp\Big(-\frac{nt^2}{2}\Big).$$

So Hoeffding's inequality is saying that, in a rather general scenario, this quality of upper bound on the tail can be recovered.

The proof uses Markov's inequality, as applied to the distribution of the moment generating function. The independence assumption means that the moment generating function $\mathbb{E}\exp(t\sum_i Z_i)$ factorises as $\prod_i \mathbb{E}\exp(tZ_i)$, which is ultimately where the power of the bound comes from.

Introducing an extra summation variable

Now let us embark upon the proof of Theorem 15.1. Let us suppose for contradiction that for some values of the parameters we have

$$\Big|\sum_{X/w<n\leqslant X}\frac{\lambda(n)\lambda(n+1)}{n}\Big| \geqslant \varepsilon\log w,$$

where we may assume that w and X are large in terms of ε.

In a manner that we have seen countless times before, our first manoeuvre will be to use multiplicativity to introduce a further summation variable over primes for free. We will also use a translation trick to introduce a further summation over j ranging in a short interval. (This idea goes back at least to van der Corput in the 1920s).

Let $H_- \leqslant H_+$ be two scales, with H_- large enough in terms of ε (but H_+ much smaller than both w and X).

Lemma 15.4. *For any $H \in [H_-, H+]$, let $\mathcal{P}_H$ denote the set of primes between $\frac{\varepsilon^2}{2}H$ and $\varepsilon^2 H$. Then*

$$\Big|\sum_{X/w<n\leqslant X}\sum_{p\in\mathcal{P}_H}\sum_{\substack{j\\ j,j+p\in[1,H]}}\frac{\lambda(n+j)\lambda(n+j+p)}{n}1_{p|n+j}\Big| \gg \varepsilon\frac{H}{\log H}\log w.$$

Proof. For all $p \in \mathcal{P}_H$ we have

$$\begin{aligned}\sum_{X/w<n\leqslant X}\frac{\lambda(n)\lambda(n+1)}{n} &= \sum_{X/w<n\leqslant X}\frac{\lambda(p)\lambda(n)\lambda(p)\lambda(n+1)}{n}\\ &= \sum_{X/w<n\leqslant X}\frac{\lambda(pn)\lambda(pn+p)}{n}\\ &= O(\log p) + \sum_{X/pw<n\leqslant X/p}\frac{\lambda(pn)\lambda(pn+p)}{n}\\ &= O(\log p) + p\sum_{X/w<m\leqslant X}\frac{\lambda(m)\lambda(m+p)}{m}1_{p|m}.\end{aligned}$$

Note the crucial use of the logarithmic averaging here, to replace the range $X/w < n \leqslant X$ with $X/pw < n \leqslant X/p$ with limited loss.

Thus,

$$\sum_{X/w<m\leqslant X} \frac{\lambda(m)\lambda(m+p)}{m} 1_{p|m} = \frac{C}{p} - O\Big(\frac{\log p}{p}\Big),$$

where $|C| \geqslant \varepsilon \log w$. Now, for any j such that $j \in [1, H]$ and $j + p \in [1, H]$, we have

$$\begin{aligned}\sum_{X/w<m\leqslant X} \frac{\lambda(m)\lambda(m+p)}{m} 1_{p|m} &= \sum_{X/w<m+j\leqslant X} \frac{\lambda(m+j)\lambda(m+j+p)}{m+j} 1_{p|m+j} \\ &= \sum_{X/w<m\leqslant X} \frac{\lambda(m+j)\lambda(m+j+p)}{m} 1_{p|m+j} + O\Big(\frac{Hw}{X}\Big).\end{aligned}$$

Hence

$$\begin{aligned}&\sum_{X/w<n\leqslant X} \sum_{p\in\mathcal{P}_H} \sum_{\substack{j\\ j,j+p\in[1,H]}} \frac{\lambda(n+j)\lambda(n+j+p)}{n} 1_{p|n+j} \\ &\qquad = H(1+O(\varepsilon^2)) \sum_{p\in\mathcal{P}_H} \frac{C}{p} - O\Big(H \sum_{p\in\mathcal{P}_H} \frac{\log p}{p}\Big) - o_{X\to\infty}(1),\end{aligned}$$

which implies the lemma since

$$\sum_{p\in\mathcal{P}_H} \frac{1}{p} \gg \frac{1}{\log H}.$$

□

Consider

$$\sum_{X/w<n\leqslant X} \sum_{p\in\mathcal{P}_H} \sum_{\substack{j\\ j,j+p\in[1,H]}} \frac{\lambda(n+j)\lambda(n+j+p)}{n} 1_{p|n+j}$$

from the statement of the previous lemma. What would happen if we were to replace the worryingly singular expression $1_{p|n+j}$ with its 'average value' over n, namely $1/p$? It turns out that this expression we can handled by our knowledge of λ in almost-all short intervals.

Lemma 15.5.

$$\sum_{X/w<n\leqslant X} \sum_{p\in\mathcal{P}_H} \sum_{\substack{j\\ j,j+p\in[1,H]}} \frac{\lambda(n+j)\lambda(n+j+p)}{pn} \ll \varepsilon^2 \frac{H}{\log H} \log w.$$

Proof. One may verify the identity

$$\begin{aligned}H \sum_{r\leqslant H} e\Big(-\frac{rp}{H}\Big)\Big|\frac{1}{H}\sum_{j\leqslant H} \lambda(n+j)e\Big(-\frac{jr}{H}\Big)\Big|^2 &= \sum_{j_1,j_2\leqslant H} \lambda(n+j_1)\lambda(n+j_2)1_{j_2\equiv j_1+p \bmod H} \\ &= \sum_{\substack{j\\ j,j+p\in[1,H]}} \lambda(n+j)\lambda(n+j+p) + O(\varepsilon^2 H).\end{aligned}$$

This error will be acceptable, and so we are reduced to showing that

$$(\log H) \sum_{X/w<n\leqslant X} \frac{1}{n} \sum_{p\in\mathcal{P}_H} \frac{1}{p} \sum_{r\leqslant H} e\Big(-\frac{rp}{H}\Big)\Big|\frac{1}{H}\sum_{j\leqslant H} \lambda(n+j)e\Big(-\frac{jr}{H}\Big)\Big|^2 \ll \varepsilon^2 \log w.$$

Swapping orders of summation, we seek to bound

$$(\log H) \sum_{r\leqslant H} \Big|\sum_{p\in\mathcal{P}_H} \frac{1}{p} e\Big(-\frac{rp}{H}\Big)\Big| \sum_{X/w<n\leqslant X} \frac{1}{n}\Big|\frac{1}{H}\sum_{j\leqslant H} \lambda(n+j)e\Big(-\frac{jr}{H}\Big)\Big|^2.$$

Let $\mathcal{R}_H$ be the set of $r \leqslant H$ for which

$$\Big|\sum_{p\in\mathcal{P}_H}\frac{1}{p}e\Big(-\frac{rp}{H}\Big)\Big| \geqslant \frac{\varepsilon^2}{\log H}.$$

We conclude that the contribution from $r \notin \mathcal{R}_H$ is

$$\leqslant \varepsilon^2 \sum_{X/w<n\leqslant X}\frac{1}{n}\sum_{r\leqslant H}\Big|\frac{1}{H}\sum_{j\leqslant H}\lambda(n+j)e\Big(-\frac{jr}{H}\Big)\Big|^2 \leqslant \varepsilon^2\log w$$

by Parseval. The contribution from $r \in \mathcal{R}_H$ is at most

$$\ll \sum_{r\in\mathcal{R}_H}\sum_{X/w<n\leqslant X}\frac{1}{n}\Big|\frac{1}{H}\sum_{j\leqslant H}\lambda(n+j)e\Big(-\frac{jr}{H}\Big)\Big|^2 = o(|\mathcal{R}_H|\log w)$$

by the result of Matomäki–Radziwiłł–Tao mentioned earlier (splitting m into dyadic ranges and using the phase $\alpha = r/H$). This is where it is convenient that we have assumed that n is never small.

It suffices to show that $|\mathcal{R}_H| \ll_\varepsilon 1$. There are various ways to show this, but the most direct and low-tech is to note that

$$\sum_{r\leqslant H}\Big|\sum_{p\in\mathcal{P}_H}\frac{1}{p}e\Big(-\frac{rp}{H}\Big)\Big|^4 = H\sum_{p_1,p_2,p_3,p_4\in\mathcal{P}_H}\frac{1}{p_1p_2p_3p_4}1_{H|(p_1+p_2-p_3-p_4)}.$$

Because $p_i \leqslant \varepsilon^2 H$, and ε is small, we note that $p_1+p_2 = p_3+p_4$. Using the sieve or the circle method one can upper bound the number of solutions to give

$$|\mathcal{R}_H|\Big(\frac{\varepsilon^2}{\log H}\Big)^4 \leqslant \sum_{r\leqslant H}\Big|\sum_{p\in\mathcal{P}_H}\frac{1}{p}e\Big(-\frac{rp}{H}\Big)\Big|^4 \ll \frac{1}{\varepsilon^2(\log H)^4}.$$

So

$$|\mathcal{R}_H| \ll_\varepsilon 1$$

as required. The lemma follows. □

16. Lecture 16: The 'entropy decrement argument'

To summarise where we are currently at in the proof, we have parameters

$$\varepsilon^{-1} \lll H_- \leqslant H \leqslant H_+ \lll w \leqslant \frac{X}{\log X},$$

and our aim is to find some nontrivial upper bound on

$$\sum_{X/w<n\leqslant X} \sum_{p\in\mathcal{P}_H} \sum_{\substack{j\\ j,j+p\in[1,H]}} \frac{\lambda(n+j)\lambda(n+j+p)}{n} 1_{p|n+j},$$

where the trivial upper bound is $\frac{H}{\log H}\log w$. (Here $\mathcal{P}_H$ is the set of primes between $(\varepsilon^2/2)H$ and $\varepsilon^2 H$.) We have shown that

$$\sum_{X/w<n\leqslant X} \sum_{p\in\mathcal{P}_H} \sum_{\substack{j\\ j,j+p\in[1,H]}} \frac{\lambda(n+j)\lambda(n+j+p)}{pn} \ll \varepsilon^2 \frac{H}{\log H}\log w,$$

so it suffices to prove a bound on

$$\sum_{X/w<n\leqslant X} \sum_{p\in\mathcal{P}_H} \sum_{\substack{j\\ j,j+p\in[1,H]}} \frac{\lambda(n+j)\lambda(n+j+p)}{n}\Big(1_{p|n+j}-\frac{1}{p}\Big).$$

At this point, we are going to find it useful to introduce probabilistic language and notation. All the manipulations to follow could be done by writing out the explicit sums instead, but the resulting expressions would be exceedingly complicated to write down.

We treat the variable n as a random variable $\mathbf{n}$, distributed according to

$$\mathbb{P}(\mathbf{n}=n) = \begin{cases} \frac{1}{n}\Big(\sum_{X/w<m\leqslant X}\frac{1}{m}\Big)^{-1} & \text{if } n\in(X/w,X]\\ 0 & \text{otherwise.}\end{cases}$$

The task is then to prove a bound such as

$$\Big|\mathbb{E}\sum_{p\in\mathcal{P}_H}\sum_{\substack{j\\ j,j+p\in[1,H]}} \lambda(\mathbf{n}+j)\lambda(\mathbf{n}+j+p)\Big(1_{p|\mathbf{n}+j}-\frac{1}{p}\Big)\Big| \leqslant \varepsilon^2\frac{H}{\log H}.$$

We will not be able to do this for all H, but via another outrageously cunning argument ('entropy decrement') we will be able to find *some* suitable scale $H\in[H_-,H_+]$ for which we can prove such a bound.

Hoeffding's inequality rescues something for us here straight away. Recalling that $P_H=\prod_{p\in\mathcal{P}_H}p$, note first that for all j and p we have

$$\frac{1}{p} = \frac{1}{P_H}\sum_{y'\leqslant P_H} 1_{p|y'+j}.$$

We now fix the values of $\lambda(\mathbf{n}+j)$ for all $j=1,\dots,H$, calling these values $x=(x_1,\dots,x_H)$ (which is a vector with ±1 entries). Having done this, let $\mathcal{E}_x$ denote the set of exceptional residue classes y modulo P_H for which

$$\Big|\sum_{p\in\mathcal{P}_H}\sum_{\substack{j\\ j,j+p\in[1,H]}} x_jx_{j+p}1_{p|y+j} - \frac{1}{P_H}\sum_{y'\leqslant P_H}\sum_{p\in\mathcal{P}_H}\sum_{\substack{j\\ j,j+p\in[1,H]}} x_jx_{j+p}1_{p|y'+j}\Big| \geqslant \varepsilon^2\frac{H}{\log H}.$$

We were hoping to prove $\mathcal{E}_x = \emptyset$ for all x. This we can't do, but we can show that $|\mathcal{E}_x|$ must be exceedingly small. Precisely, we claim that

$$|\mathcal{E}_x| \leqslant P_H \exp\Big(- \varepsilon^{O(1)} \frac{H}{\log H}\Big).$$

(To give you a sense of the relative scales involved, note that $P_H \approx \exp(\varepsilon^2 H/\log H)$ by the prime number theorem.)

Why is this true? Consider the random variable $\mathbf{y} \in \mathbb{Z}/P_H\mathbb{Z}$ uniformly distributed, and let Z_p denote the random variable

$$Z_p := \sum_{\substack{j \\ j, j+p \in [1,H]}} x_j x_{j+p} 1_{p|\mathbf{y}+j}.$$

The family $(Z_p)_{p\in\mathcal{P}_H}$ is independent by construction, and each random variable Z_p is bounded by $O(\varepsilon^{-2})$ (since for all fixed values $\mathbf{y} = y$ there are at most $O(\varepsilon^{-2})$ values of j that can possibly contribute to the sum). Then the inequality we have claimed is just Hoeffding's inequality for suitable parameters, namely bounding

$$\mathbb{P}_{\mathbf{y}\sim \mathrm{Unif}(\mathbb{Z}/P_H\mathbb{Z})}\Big(\Big|\sum_{p\in\mathcal{P}_H} Z_p - \sum_{p\in\mathcal{P}_H} \mathbb{E}Z_p\Big| \geqslant |\mathcal{P}_H|\Big).$$

Let's begin to think about how we might use this result to control the overall average. Letting $\mathbf{X}_H$ be the random variable $(\lambda(\mathbf{n}+j))_{j=1}^H$, and $\mathbf{Y}_H$ be the random variable $\mathbf{n} \bmod P_H$, by the law of total expectation we have

$$\begin{aligned}
&\mathbb{E}\sum_{p\in\mathcal{P}_H} \sum_{\substack{j \\ j,j+p\in[1,H]}} \lambda(\mathbf{n}+j)\lambda(\mathbf{n}+j+p)1_{p|\mathbf{n}+j} \\
&= \sum_{x\in(-1,+1)^{[H]}} \mathbb{E}\Big(\sum_{p\in\mathcal{P}_H} \sum_{\substack{j \\ j,j+p\in[1,H]}} \lambda(\mathbf{n}+j)\lambda(\mathbf{n}+j+p)1_{p|\mathbf{n}+j} \,\Big|\, \mathbf{X}_H = x\Big)\mathbb{P}(\mathbf{X}_H = x) \\
&= \sum_{x\in(-1,+1)^{[H]}} \sum_{y \bmod P_H} \Big(\sum_{p\in\mathcal{P}_H} \sum_{\substack{j \\ j,j+p\in[1,H]}} x_j x_{j+p} 1_{p|y+j}\Big)\mathbb{P}(\mathbf{X}_H = x)\mathbb{P}(\mathbf{Y}_H = y \,|\, \mathbf{X}_H = x).
\end{aligned}$$

We have two different types of contributions: $y \notin \mathcal{E}_x$ and $y \in \mathcal{E}_x$. Suppose one knew the following lemma:

Lemma 16.1. *We let $x \in (-1,+1)^{[H]}$ be* good *if, for all subsets $S \subset \mathbb{Z}/P_H\mathbb{Z}$ of size $|S| \leqslant P_H \exp(-\varepsilon^{10} H/\log H)$, one has*

$$\mathbb{P}(\mathbf{Y}_H = S \,|\, \mathbf{X}_H = x) = o_{H_-\to\infty}(1).$$

Then there exists a scale $H \in [H_-, H_+]$ for which

$$\sum_{\substack{x\in(-1,+1)^{[H]} \\ x \text{ good}}} \mathbb{P}(\mathbf{X}_H = x) = 1 - o_{H_-\to\infty}(1).$$

Then the contribution from x not good is

$$\ll \sum_{\substack{x\in(-1,+1)^{[H]} \\ x \text{ not good}}} \frac{H}{\log H}\mathbb{P}(\mathbf{X}_H = x) \sum_{y \bmod P_H} \mathbb{P}(\mathbf{Y}_H = y \,|\, \mathbf{X}_H = x),$$

as at most $O(\varepsilon^{-2})$ values of j contribute to the sum, and this is

$$\ll \frac{H}{\log H} \sum_{\substack{x\in(-1,+1)^{[H]} \\ x \text{ not good}}} \mathbb{P}(\mathbf{X}_H = x) = o_{H_-\to\infty}(H/\log H),$$

which is acceptable.

The contribution from good x is

$$\sum_{\substack{x\in(-1,+1)^{[H]}\\ x \text{ is } good}} \sum_{y \bmod P_H} \Big(\sum_{p\in\mathcal{P}_H} \sum_{\substack{j\\ j,j+p\in[1,H]}} x_j x_{j+p} 1_{p|y+j}\Big)\mathbb{P}(\mathbf{X}_H = x)\mathbb{P}(\mathbf{Y}_H = y \,|\, \mathbf{X}_H = x)$$

$$= \sum_{\substack{x\in(-1,+1)^{[H]}\\ x \text{ good}}} \sum_{\substack{y \bmod P_H\\ y\notin\mathcal{E}_x}} \Big(\sum_{p\in\mathcal{P}_H} \sum_{\substack{j\\ j,j+p\in[1,H]}} \frac{x_j x_{j+p}}{p}\Big)\mathbb{P}(\mathbf{X}_H = x)\mathbb{P}(\mathbf{Y}_H = y \,|\, \mathbf{X}_H = x) + A_1 + A_2,$$

where the error term A_1 comes from the definition of $\mathcal{E}_x$ and satisfies

$$|A_1| \leqslant \sum_{x\in(-1,+1)^{[H]}} \sum_{y \bmod P_H} \Big(\varepsilon^2 \frac{H}{\log H}\Big)\mathbb{P}(\mathbf{X}_H = x)\mathbb{P}(\mathbf{Y}_H = y \,|\, \mathbf{X}_H = x) \leqslant \varepsilon^2 \frac{H}{\log H},$$

which is acceptable, and A_2 comes from the $y \in \mathcal{E}_x$ contribution and is bounded by

$$\begin{aligned} |A_2| &\ll \frac{H}{\log H} \sum_{\substack{x\in(-1,+1)^{[H]}\\ x \text{ is good}}} \mathbb{P}(\mathbf{X}_H = x)\mathbb{P}(\mathbf{Y}_H \in \mathcal{E}_x \,|\, \mathbf{X}_H = x) \\ &\ll o_{H_-\to\infty}(H/\log H). \end{aligned}$$

Regarding the main term, namely

$$\sum_{\substack{x\in(-1,+1)^{[H]}\\ x \text{ good}}} \sum_{\substack{y \bmod P_H\\ y\notin\mathcal{E}_x}} \Big(\sum_{p\in\mathcal{P}_H} \sum_{\substack{j\\ j,j+p\in[1,H]}} \frac{x_j x_{j+p}}{p}\Big)\mathbb{P}(\mathbf{X}_H = x)\mathbb{P}(\mathbf{Y}_H = y \,|\, \mathbf{X}_H = x),$$

one can use the same estimations as before to extend the ranges of summation to

$$\sum_{x\in(-1,+1)^{[H]}} \sum_{y \bmod P_H} \Big(\sum_{p\in\mathcal{P}_H} \sum_{\substack{j\\ j,j+p\in[1,H]}} \frac{x_j x_{j+p}}{p}\Big)\mathbb{P}(\mathbf{X}_H = x)\mathbb{P}(\mathbf{Y}_H = y \,|\, \mathbf{X}_H = x)$$

up to an acceptable error. This expression in turn is equal to

$$\sum_{x\in(-1,+1)^{[H]}} \Big(\sum_{p\in\mathcal{P}_H} \sum_{\substack{j\\ j,j+p\in[1,H]}} \frac{x_j x_{j+p}}{p}\Big)\mathbb{P}(\mathbf{X}_H = x).$$

This is the expression we estimated in Lemma 15.5 last lecture, using the Matomäki–Radziwiłł theorem. We derived a bound of $\ll \varepsilon^2 \frac{H}{\log H}$, which is acceptable.

So it remains to prove the lemma, i.e. Lemma 16.1. In words, we've reduced the whole theorem to trying to show that the weight of the sum

$$\sum_{X/w<n\leqslant X} \frac{1}{n} \sum_{p\in\mathcal{P}_H} \sum_{\substack{j\\ j,j+p\in[1,H]}} \lambda(n+j)\lambda(n+j+p)$$

is not concentrated on some very small collection of residue classes n modulo P_H. This is essentially equivalent to proving some weak independence property of the random variables $\mathbf{X}_H$ and $\mathbf{Y}_H$.

Tao's method for proving Lemma 16.1 was highly original, using concepts from information theory to locate the appropriate scale H. There were precedents for such an argument in the additive combinatorics and ergodic theory literature, but this was the first time that such ideas had appeared in analytic number theory in quite this fashion (though see the Green–Tao theorem for another application of of an increment argument in analytic number theory). Again, we stress that all the probabilities here may be written down explicitly in

terms of sums, but one misses out on substantial intuition behind the manipulation.

In order to get ready for the proof, let's make a short observation about what happens to our random variables under small translations.

Lemma 16.2. *Let r be a fixed integer with $|r| \leqslant H_+$. Then for any complex valued random variable $X(\mathbf{n})$, depending on $\mathbf{n}$ and bounded in magnitude by $O(1)$, one has*

$$\mathbb{E}(X(\mathbf{n})) = \mathbb{E}(X(\mathbf{n}+r)) + o_{w\to\infty}(1).$$

Proof. This is exactly the same observation has we made when we introduced the summation over j in the last lecture. The point is that

$$\begin{aligned}\frac{1}{\log w}\sum_{X/w<n\leqslant X}\frac{X(n+r)}{n} &= \frac{1}{\log w}\sum_{X/w+r<n\leqslant X+r}\frac{X(n)}{n-r}\\ &= \frac{1}{\log w}\sum_{X/w+r<n\leqslant X+r}\Big(\frac{X(n)}{n}+O\Big(\frac{H_+}{n(n+r)}\Big)\Big)\\ &= \frac{1}{\log w}\sum_{X/w<n\leqslant X}\frac{X(n)}{n}+O\Big(\frac{H_+}{\log w}\Big).\end{aligned}$$

□

A brief primer on entropy

For a discrete random variable $\mathbf{X}$ taking finitely many values, we will say that the *Shannon entropy* $\mathbb{H}(\mathbf{X})$ is defined by

$$\mathbb{H}(\mathbf{X}) := \sum_x \mathbb{P}(\mathbf{X}=x)\log\frac{1}{\mathbb{P}(\mathbf{X}=x)},$$

with the convention that $0\log 1/0 = 0$. One can think of $\mathbb{H}(\mathbf{X})$ as the amount of 'information', or the amount of 'randomness', that is encoded in the random variable $\mathbf{X}$. Entropy is always non-negative, and $\mathbb{H}(X)$ is maximised among random variables with the same range when $\mathbf{X}$ is uniformly distributed (this can be shown by Jensen's inequality, and in fact the uniform distribution is the unique maximiser). In this case $\mathbb{H}(\mathbf{X}) = \log N$, where N is the size of the range of $\mathbf{X}$.

Given two random variables $\mathbf{X}$ and $\mathbf{Y}$ taking finitely many values we define the *mutual information*

$$\mathbb{I}(\mathbf{X},\mathbf{Y}) := \mathbb{H}(\mathbf{X}) + \mathbb{H}(\mathbf{Y}) - \mathbb{H}(\mathbf{X},\mathbf{Y}).$$

One can think of $\mathbb{I}(\mathbf{X},\mathbf{Y})$ as the extra randomness that is present in both $\mathbf{X}$ and $\mathbf{Y}$ taken separately, over and above what is contained in the joint distribution of $(\mathbf{X},\mathbf{Y})$. One may show that $\mathbb{I}(\mathbf{X},\mathbf{Y}) \geqslant 0$, and that $\mathbb{I}(\mathbf{X},\mathbf{Y}) = 0$ if and only if $\mathbf{X}$ and $\mathbf{Y}$ are independent. To prove the first of these assertions, consider the conditional entropy

$$\begin{aligned}\mathbb{H}(\mathbf{X}|\mathbf{Y}) &:= \sum_y \mathbb{P}(\mathbf{Y}=y)\mathbb{H}(\mathbf{X}|\mathbf{Y}=y)\\ &= \sum_y \mathbb{P}(\mathbf{Y}=y)\sum_x \mathbb{P}(\mathbf{X}=x\,|\,\mathbf{Y}=y)\log\frac{1}{\mathbb{P}(\mathbf{X}=x\,|\,\mathbf{Y}=y)}.\end{aligned}$$

By using Jensen's inequality as applied to the concave function $u\mapsto u\log 1/u$, one concludes that

$$\mathbb{H}(\mathbf{X}|\mathbf{Y}) = \sum_x\sum_y \mathbb{P}(\mathbf{Y}=y)\mathbb{P}(\mathbf{X}=x\,|\,\mathbf{Y}=y)\log\frac{1}{\mathbb{P}(\mathbf{X}=x\,|\,\mathbf{Y}=y)}$$

$$\leqslant \sum_x \Big(\sum_y \mathbb{P}(\mathbf{Y}=y)\mathbb{P}(\mathbf{X}=x \,|\, \mathbf{Y}=y)\Big) \log \Big(\Big(\sum_y \mathbb{P}(\mathbf{Y}=y)\mathbb{P}(\mathbf{X}=x \,|\, \mathbf{Y}=y)\Big)^{-1}\Big)$$
$$= \sum_x \mathbb{P}(\mathbf{X}=x) \log \frac{1}{\mathbb{P}(\mathbf{X}=x)}$$
$$= \mathbb{H}(\mathbf{X}).$$

(In words, $\mathbf{X}$ cannot contain more randomness when certain auxiliary information is fixed than it can with no auxiliary information.) One then has the identity (easily proved) that

$$\mathbb{H}(\mathbf{X},\mathbf{Y}) = \mathbb{H}(\mathbf{X}|\mathbf{Y}) + \mathbb{H}(\mathbf{Y}) = \mathbb{H}(\mathbf{Y}|\mathbf{X}) + \mathbb{H}(\mathbf{X}),$$

and so

$$\mathbb{I}(\mathbf{X},\mathbf{Y}) = \mathbb{H}(\mathbf{X}) - \mathbb{H}(\mathbf{X}|\mathbf{Y}) \geqslant 0.$$

We also conclude the subadditivity of entropy, namely

$$\mathbb{H}(\mathbf{X},\mathbf{Y}) \leqslant \mathbb{H}(\mathbf{X}) + \mathbb{H}(\mathbf{Y}).$$

Now, let us specialise to the case in hand. We know that the random variable $\mathbf{X}_H$ takes at most 2^H values, and so $\mathbb{H}(\mathbf{X}_H) \ll H$. We also know that for each $y \bmod P_H$ we have $\mathbb{P}(\mathbf{Y}_H = y) = P_H^{-1} + o_{w\to\infty}(1)$, so $\mathbb{H}(\mathbf{Y}_H) = \log(P_H) + o_{w\to\infty}(1) \ll H$ for large enough parameters (crudely estimating P_H as $exp(O(H))$). So the trivial upper bound is $\mathbb{I}(\mathbf{X},\mathbf{Y}) \ll H$.

Our first lemma shows that a big enough improvement on this trivial bound will finish the matter

Lemma 16.3. *If $\mathbb{I}(\mathbf{X}_H,\mathbf{Y}_H) = o_{H_-\to\infty}(H/\log H)$ then Lemma 16.1 follows with this value of H (and so the whole theorem follows too).*

We're not going to go through all of the proof of this lemma in lectures: not because the proof is very difficult, but because the details are a little fiddly and tedious.

Start of a proof. We have

$$\sum_x \mathbb{P}(\mathbf{X}_H = x)(\mathbb{H}(\mathbf{Y}_H) - \mathbb{H}(\mathbf{Y}_H \,|\, \mathbf{X}_H = x)) = \mathbb{I}(\mathbf{X}_H,\mathbf{Y}_H) = o_{H_-\to\infty}\Big(\frac{H}{\log H}\Big).$$

Since

$$\mathbb{H}(\mathbf{Y}_H) - \mathbb{H}(\mathbf{Y}_H \,|\, \mathbf{X}_H = x) \geqslant \log P_H - o_{w\to\infty}(1) - \log P_H = -o_{w\to\infty}(1),$$

we conclude from Markov's inequality that the probability in x that

$$\mathbb{H}(\mathbf{Y}_H) - \mathbb{H}(\mathbf{Y}_H \,|\, \mathbf{X}_H = x) = o_{H_-\to\infty}\Big(\frac{H}{\log H}\Big)$$

is $1 - o_{H_-\to\infty}(1)$. Call such an x *very good*. It suffices to show that if x is very good than x is good, in the sense of Lemma 16.1, i.e. that for all $S \subset \mathbb{Z}/P_H\mathbb{Z}$ with $|S| \leqslant P_H \exp(-\varepsilon^{O(1)}H/\log H)$ one has $\mathbb{P}(\mathbf{Y}_H \in S \,|\, \mathbf{X}_H = x) = o_{H_-\to\infty}(1)$.

We leave the rest as an exercise. See Lemma 3.3 of Tao's original paper if you get stuck. □

So we have to get control on the mutual information somehow. This seems hard to do directly, but there is a critical observation that enables one to make progress; namely, if the mutual information is large at scale H, i.e. if $\mathbf{Y}_H$ very nearly determines $\mathbf{X}_H$, then $\mathbf{Y}_H$ also very nearly determines $\mathbf{X}_{kH}$ (ultimately due to the translation invariance of the distribution of $\mathbf{Y}_H$). But this is a stronger statement for larger k, as one has a random variable that potentially takes 2^{kH} values being controlled by a different random variable that only takes $P_H \approx \exp(\varepsilon^2 H/\log H)$ values. So $\mathbf{X}_{kH}$ is less random than it might otherwise be, in the sense that the entropy ratio $\mathbb{H}(\mathbf{X}_{kH})/kH$ is smaller than the entropy ratio $\mathbb{H}(\mathbf{X}_H)/H$.

This is the detailed lemma.

Lemma 16.4 (Large mutual information implies entropy decrement)**.** *Provided w is large enough, for all $H \in [H_-, H_+]$ one has*

$$\frac{\mathbb{H}(\mathbf{X}_{kH})}{kH} \leqslant \frac{\mathbb{H}(\mathbf{X}_H)}{H} - \frac{\mathbb{I}(\mathbf{X}_H, \mathbf{Y}_H)}{H} + O\Big(\frac{1}{k}\Big).$$

Proof. Fixed $H, H_1, H_2 \in [H_-, H_+]$. We temporarily define the random variable

$$\mathbf{X}_{H_1, H_1+H_2} := (\lambda(\mathbf{n}+j))_{j=H_1+1}^{H_1+H_2}.$$

By the approximate translation invariance of the sum over $\mathbf{n}$, one may show that

$$\mathbb{H}(\mathbf{X}_{H_1,H_1+H_2}) = \mathbb{H}(\mathbf{X}_{H_2}) + o_{w\to\infty}(1).$$

So, by subadditivity of entropy, we have

$$\mathbb{H}(\mathbf{X}_{H_1+H_2}) \leqslant \mathbb{H}(\mathbf{X}_{H_1}) + \mathbb{H}(\mathbf{X}_{H_1,H_1+H_2}) \leqslant \mathbb{H}(\mathbf{X}_{H_1}) + \mathbb{H}(\mathbf{X}_{H_2}) + o_{w\to\infty}(1).$$

But by conditioning on $\mathbf{Y}_H$ one can get an improved inequality. Indeed, by subadditivity again one has

$$\begin{aligned}\mathbb{H}(\mathbf{X}_{H_1+H_2}|\mathbf{Y}_H) &\leqslant \mathbb{H}(\mathbf{X}_{H_1}|\mathbf{Y}_H) + \mathbb{H}(\mathbf{X}_{H_1,H_1+H_2}|\mathbf{Y}_H)\\ &= \mathbb{H}(\mathbf{X}_{H_1}|\mathbf{Y}_H) + \mathbb{H}(\mathbf{X}_{H_1,H_1+H_2}|\mathbf{Y}_H + H_1 \bmod P_H),\end{aligned}$$

just by translating the sum over y in the definition of the conditional entropy in the second term. By translating $\mathbf{n} \mapsto \mathbf{n} - H_1$ and using the approximate translation invariance of the measure, the above is in turn equal to

$$\mathbb{H}(\mathbf{X}_{H_1}|\mathbf{Y}_H) + \mathbb{H}(\mathbf{X}_{H_2}|\mathbf{Y}_H) + o_{w\to\infty}(1).$$

So all in all we have

$$\mathbb{H}(\mathbf{X}_{H_1+H_2}|\mathbf{Y}_H) \leqslant \mathbb{H}(\mathbf{X}_{H_1}|\mathbf{Y}_H) + \mathbb{H}(\mathbf{X}_{H_2}|\mathbf{Y}_H) + o_{w\to\infty}(1),$$

so in particular we have

$$\mathbb{H}(\mathbf{X}_{kH}|\mathbf{Y}_H) \leqslant k\mathbb{H}(\mathbf{X}_H|\mathbf{Y}_H) + o_{X\to\infty}(1)$$

as long as $H_- \leqslant H \leqslant kH \leqslant H_+$. Overall, we end up with

$$\begin{aligned}\mathbb{H}(\mathbf{X}_{kH}) &= \mathbb{H}(\mathbf{X}_{kH}|\mathbf{Y}_H) + \mathbb{H}(\mathbf{Y}_H) - \mathbb{H}(\mathbf{Y}_H|\mathbf{X}_{kH})\\ &\leqslant \mathbb{H}(\mathbf{X}_{kH}|\mathbf{Y}_H) + \mathbb{H}(\mathbf{Y}_H)\\ &\leqslant k\mathbb{H}(\mathbf{X}_H|\mathbf{Y}_H) + \mathbb{H}(\mathbf{Y}_H) + o_{w\to\infty}(1)\\ &= k\mathbb{H}(\mathbf{X}_H) - k\mathbb{I}(\mathbf{X}_H, \mathbf{Y}_H) + \mathbb{H}(\mathbf{Y}_H) + o_{w\to\infty}(1).\end{aligned}$$

The lemma now follows from dividing through by kH. □

Such an entropy decrement cannot continue indefinitely, as $\mathbb{H}(\mathbf{X}_{kH})/kH \geqslant 0$. This puts a limit on how large the mutual information $\mathbb{I}(\mathbf{X}_H, \mathbf{Y}_H)$ can actually be. Making this precise, one can locate an appropriate scale.

Lemma 16.5 (An appropriate scale)**.** *There exists a scale $H \in [H_-, H_+]$ for which*

$$\mathbb{I}(\mathbf{X}_H, \mathbf{Y}_H) < \frac{H}{\log H \log\log\log H} = o\Big(\frac{H}{\log H}\Big).$$

Proof. Suppose for contradiction that $\mathbb{I}(\mathbf{X}_H, \mathbf{Y}_H) > \frac{H}{\log H \log\log\log H}$ for all $H_- \leqslant H \leqslant H_+$. For some C and J to be picked later, let us recursively define the natural numbers

$$H_- \leqslant H_1 \leqslant H_2 \leqslant \cdots \leqslant H_J$$

by setting $H_1 := H_-$ and

$$H_{j+1} := H_j \lfloor C \log H_j \log\log\log H_j \rfloor$$

for all $1 \leqslant j \leqslant J-1$. If H_+ is sufficiently large, then $H_J \leqslant H_+$, and if C is large enough then (from our previous observations)

$$\frac{\mathbb{H}(\mathbf{X}_{H_{j+1}})}{H_{j+1}} \leqslant \frac{\mathbb{H}(\mathbf{X}_{H_j})}{H_j} - \frac{1}{2\log H_j \log\log\log H_j}.$$

On the other hand, one can show that there is some large B (depending on C and H_-) for which

$$H_j \leqslant \exp(Bj\log j)$$

for all $2 \leqslant j \leqslant J$, which gives us

$$\frac{\mathbb{H}(\mathbf{X}_{H_{j+1}})}{H_{j+1}} \leqslant \frac{\mathbb{H}(\mathbf{X}_{H_j})}{H_j} - \frac{1}{2Bj\log j\log\log(Bj\log j)},$$

for all $2 \leqslant j \leqslant J$.

Telescoping, we have

$$0 \leqslant \frac{\mathbb{H}(\mathbf{X}_{H_J})}{H_J} \leqslant \frac{\mathbb{H}(\mathbf{X}_{H_2})}{H_2} - \sum_{j=2}^{J} \frac{1}{2Bj\log j\log\log(Bj\log j)} \leqslant O(1) - \sum_{j=2}^{J} \frac{1}{2Bj\log j\log\log(Bj\log j)}.$$

So

$$\sum_{j=2}^{J} \frac{1}{2Bj\log j\log\log(Bj\log j)} \ll 1.$$

But this sum diverges as $J \to \infty$, so this leads to a contradiction if J is sufficiently large. □

So we've found an appropriate scale H, and, after our large series of deductions, we've finally proved the main theorem of this section!

Postlude

We've covered a lot of ground in this course! Not just in terms of the main theorems we tackled, but also in terms of the number of different techniques from analytic number theory that we had to learn in order to attack these problems. You have seen:

- Dirichlet convolution identities
- major arc/minor decompositions
- Vinogradov's Type II bound
- expansion of congruence conditions in terms of multiplicative characters
- Van der corput B process (truncated poisson summation)
- Smoothing of sums
- L^2-L^∞ bounding technique (in the proof of Halasz's theorem)
- Gauss sums and primitive characters
- Dirichlet's approximation theorem
- Turan–Kubilius inequality
- Polya–Vinogradov inequality
- Ramaré's identity
- Montgomery's mean value theorem for Dirichlet polynomials
- Halasz–Montgomery mean value theorem for Dirichlet polynomials
- Halasz–Montgomery mean value theorem for Dirichlet polynomials supported on the primes
- Telescoping identities
- Dyadic decompositions
- Adding extra averaging variables, by multiplicative decompositions and by additive shifts

- Properties of Granville–Soundararajan distance
- Probabilistic interpretations and entropy decrement
- Hoeffding's inequality
- ...

not to mention any of the main theorems we have actually proved.

I hope you will find this to be a solid grounding for the rest of your careers in number theory.