NEW LARGE VALUE ESTIMATES FOR DIRICHLET POLYNOMIALS

LARRY GUTH AND JAMES MAYNARD

Abstract. We prove new bounds for how often Dirichlet polynomials can take large values. This gives improved estimates for a Dirichlet polynomial of length N taking values of size close to $N^{3/4}$, which is the critical situation for several estimates in analytic number theory connected to prime numbers and the Riemann zeta function. As a consequence, we deduce a zero density estimate $N(\sigma,T)\le T^{30(1-\sigma)/13+o(1)}$ and asymptotics for primes in short intervals of length $x^{17/30+o(1)}$.

1. Introduction

In this paper we prove new bounds for the frequency of large values of Dirichlet polynomials. This gives improved estimates for a Dirichlet polynomial of length N taking values of size close to $N^{3/4}$. Our main result is the following.

Theorem 1.1 (Large values estimate). *Suppose (b_n) is a sequence of complex numbers with $|b_n|\le 1$, and $(t_r)_{r\le R}$ is a sequence of 1-separated points in $[0,T]$ such that*

$$\Big|\sum_{n=N}^{2N} b_n n^{it_r}\Big|\ge V$$

for all $r\le R$. Then we have

$$R\le T^{o(1)}\Big(N^2V^{-2}+N^{18/5}V^{-4}+TN^{12/5}V^{-4}\Big).$$

The bound of Theorem 1.1 can be compared with the bound

$$R\le T^{o(1)}\Big(N^2V^{-2}+T\min(NV^{-2},N^4V^{-6})\Big) \tag{1.1}$$

coming from combining the classical Mean Value Theorem for Dirichlet polynomials and the Montgomery-Halasz-Huxley large values estimate. Together these previous results give stronger bounds when V is smaller than $N^{7/10-\epsilon}$ or bigger than $N^{8/10+\epsilon}$, but Theorem 1.1 gives a stronger bound when $N^{7/10+\epsilon}<V<N^{8/10-\epsilon}$ and $N\le T^{5/6-\epsilon}$. There are various improvements of the large values estimate which will supersede ours when V is a bit larger than $N^{3/4}$ (see [Iv, Chapter 11], for example), but Theorem 1.1 represents the first substantive improvement on the bound (1.1) coming from the Mean Value Theorem when $V\le N^{3/4}$ or on bounds when V is close to $N^{3/4}$.

The key interest in this result is that for many applications in analytic number theory, it is the case when $V \approx N^{3/4}$ which is the critical limiting scenario. In this situation the Mean Value Theorem and Large Values Estimate both give a bound of roughly $N^{1/2} + TN^{-1/2}$ whereas Theorem 1.1 gives roughly $N^{3/5} + TN^{-3/5}$. Therefore we obtain an improvement for N smaller than $T^{10/11}$, and correspondingly we expect Theorem 1.1 to lead to a quantitative improvement to any result where this covers the limiting situation.

One well-studied situation where the limiting case is improved by Theorem 1.1 is zero-density estimates for the Riemann Zeta function $\zeta(s)$. Let $N(\sigma, T)$ be the number of zeroes of $\zeta(s)$ in the rectangle $\Re(s) \geq \sigma$ and $|\Im(s)| \leq T$. After early work by Carlson [Ca], Ingham [In] proved the bound

$$N(\sigma, T) \leq T^{\frac{3(1-\sigma)}{2-\sigma}+o(1)}, \tag{1.2}$$

ultimately relying on the Mean Value Theorem for Dirichlet polynomials. Huxley [Hu], building on work of Montgomery [M3] and Halasz [Ha] and ultimately relying on the Montgomery-Halasz large values estimate, proved the bound

$$N(\sigma, T) \leq T^{\frac{3(1-\sigma)}{3\sigma-1}+o(1)}. \tag{1.3}$$

This improves on (1.2) for $\sigma > 3/4$ (corresponding to when the term N^4V^{-6} is smaller than NV^{-2} for $V = N^\sigma$ in $\min(NV^{-2}, N^4V^{-6})$ in (1.1)).

When $\sigma \approx 3/4$ the bounds (1.2) and (1.3) coincide, and the critical situation turns out to be related to values of size $V = N^{3/4}$ of Dirichlet polynomials of length $N = T^{4/5}$, where both estimates give $R \leq T^{3/5+o(1)}$. In this situation Theorem 1.1 gives an improved estimate of $R \leq T^{13/25+o(1)}$. Incorporating Theorem 1.1 into the zero density machinery gives the following result.

Theorem 1.2 (Zero density estimate)**.** *Let $N(\sigma, T)$ denote the number of zeros ρ of $\zeta(s)$ with $\Re(\rho) \geq \sigma$ and $|\Im(\rho)| \leq T$. Then we have*

$$N(\sigma, T) \ll T^{15(1-\sigma)/(3+5\sigma)+o(1)}.$$

Combining this with Ingham's estimate when $\sigma \leq 7/10$, we obtain

$$N(\sigma, T) \ll T^{30(1-\sigma)/13+o(1)}. \tag{1.4}$$

The exponent $30/13$ improves on the previous exponent of $12/5$ due to Huxley [Hu]. This has the following corollaries for the distribution of primes in short intervals.

Corollary 1.3 (Count of primes in short intervals)**.** *Let $y \in [x^{17/30+\epsilon}, x^{0.99}]$. Then we have*

$$\pi(x+y) - \pi(x) = \frac{y}{\log x} + O_\epsilon\Big(y \exp(-\sqrt[4]{\log x})\Big).$$

The exponent $\frac{17}{30}$ improves on the previous exponent $\frac{7}{12}$ due to Huxley [Hu].

Corollary 1.4 (Count of primes in 'almost-all' short intervals)**.** *Let $y \in [X^{2/15+\epsilon}, X^{0.99}]$. Then for all but $O(X\exp(-\sqrt[4]{\log x}))$ choices of $x \in [X, 2X]$ we have*

$$\pi(x+y) - \pi(x) = \frac{y}{\log x} + O_\epsilon\Big(y\exp(-\sqrt[4]{\log x})\Big).$$

The exponent $\frac{2}{15}$ improves on the previous exponent $\frac{1}{6}$ due to Huxley [Hu].

We expect there to be various further applications of Theorem 1.1 (and the underlying ideas) to improving quantitative estimates related to the primes and similar objects.

1.1. **Background.** Given an integer $N \in \mathbb{N}$ and a sequence $(b_n)_{n\in[N,2N]}$, a Dirichlet polynomial is a trigonometric sum of the form

$$D(t) = \sum_{n=N}^{2N} b_n e^{it\log n}. \tag{1.5}$$

We say such a polynomial has length N, and we study the question how often a Dirichlet polynomial can be large in the interval $[0, T]$. The precise question can be formulated as follows.

Main Question. *Suppose that $D(t)$ is a Dirichlet polynomial of the form* (1.5) *with $|b_n| \le 1$ for all n. Suppose that $W \subset [0, T]$ is a 1-separated set of points t where $|D(t)| > N^\sigma$. What is the largest possible cardinality of W, in terms of N, T, and σ?*

There are many examples where $|W| \gtrsim N^{2-2\sigma}$. Montgomery's large value conjecture [M2, page 142, Conjecture 2] predicts that under some natural conditions this lower bound is essentially tight.

Conjecture 1.5 (Montgomery's large value conjecture)**.** *Let $\sigma > 1/2$ and $D(t) = \sum_{n=N}^{2N} b_n e^{it\log n}$ with $|b_n| \le 1$. Suppose $W \subset [0, T]$ is a 1-separated set such that $|D(t)| > N^\sigma$ for $t \in W$. Then*

$$|W| \le C(\sigma)T^{o(1)}N^{2-2\sigma}.$$

Using a simple orthogonality argument, it has long been known that for any $T \ge N$,

$$|W| \lesssim TN^{1-2\sigma}, \tag{1.6}$$

which corresponds to the term NV^{-2} in $\min(NV^{-2}, N^4V^{-6})$ in (1.1).

This basic estimate has been improved in several regimes. An integer power of a Dirichlet polynomial is a Dirichlet polynomial, and applying (1.6) to powers of D

gives improved bounds when N is fairly small compared with T. For $\sigma \leq 3/4$, the best previous bounds on our main question came from this approach. In particular, if $\sigma \leq 3/4$ and if $N \in [T^{2/3}, T]$, then (1.6) represented the best known bound.

In the late 60s, Montgomery [M], building on ideas of Halasz [Ha] and Turán [HT], showed that Conjecture 1.5 is true if σ is sufficiently large. Indeed, (1.1) gives Conjecture 1.5 when $N^\sigma > N^{1/2}T^{1/4}$. The large value method gives very strong information for large σ, but it gives no information if $\sigma \leq 3/4$ as Montgomery explains in [M2, page 141]. For σ closer to 1, there have been several refinements of the underlying ideas (see, for example, [Bo, HB3, Ju]).

If one knows some more structure about the set of large values of a Dirichlet polynomial, then one can hope to have improved bounds. For example, another important result is Heath-Brown's work about the behavior of Dirichlet polynomials on difference sets:

Theorem 1.6 (Heath-Brown, [HB])**.** *Suppose that $\mathcal{T}$ is a 1-separated set of points in an interval of length T. Let $|a_n| \lessapprox 1$ be a complex sequence. Then*

$$\sum_{t_1, t_2 \in \mathcal{T}} \Big| \sum_{n=N}^{2N} a_n n^{i(t_1 - t_2)} \Big|^2 \lessapprox |\mathcal{T}|^2 N + |\mathcal{T}| N^2 + |\mathcal{T}|^{5/4} T^{1/2} N.$$

In the range $N \in [T^{2/3}, T]$, the last term can be ignored and the right-hand side is essentially sharp. Theorem 1.6 gives strong information about our main question if the set W has a lot of additive structure, such as arithmetic progressions.

We will make this precise using the idea of additive energy, which we define as follows. For a finite set W, let

$$E(W) := \#\{w_1, w_2, w_3, w_4 \in W : |w_1 + w_2 - w_3 - w_4| < 1\}. \tag{1.7}$$

A simple consequence of Theorem 1.6 is the following (we give a proof, as well as refined bounds, in Section 11).

Lemma 1.7. *Let $N \in [T^{2/3}, T]$, $\sigma > 1/2$ and $D(t) = \sum_{n=N}^{2N} b_n n^{it}$ with $|b_n| \leq 1$. Suppose $W \subset [0, T]$ is a 1-separated set such that $|D(t)| > N^\sigma$ for $t \in W$. Then*

$$E(W) \leq |W|^3 N^{1-2\sigma+o(1)} + |W|^2 N^{2-2\sigma+o(1)}.$$

If $E(W)$ is very large, say $E(W) > |W|^3 T^{-o(1)}$, then $E(W)$ is larger than the first term on the right hand side, so it must be bounded by the second term. This implies Conjecture 1.5 for Dirichlet polynomials D_N with $E(W) \gtrapprox |W|^3$. Thus we can obtain improved bounds for Dirichlet polynomials whose large value set has a lot of additive structure.

We introduce a new method that gives good bounds for Dirichlet polynomials on sets of small energy. We outline our method in the next section.

1.2. **Notations and conventions.** We write $A \lesssim B$ to mean that $A \leq CB$ for an absolute constant C, and $A \lesssim_z B$ to denote that the constant C may depend on the parameter z. We write $A \asymp B$ to mean $A \lesssim B$ and $B \lesssim A$ both hold, and $A \sim B$ to mean $B < A \leq 2B$. Similarly, we write $A \lessapprox B$ to mean that for any $\epsilon > 0$, there is a constant $C(\epsilon) > 0$ depending only on ϵ such that $A \leq C(\epsilon)T^\epsilon B$ for all large T, and $A \lessapprox_z B$ to denote a dependency on a parameter z. Asymptotic quantities such as $o(1)$ are interpreted as $T \to \infty$.

We use $e(x) := e^{2\pi i x}$ to denote the complex exponential. Summations will run over integers unless specified otherwise.

1.3. **Acknowledgements.** LG would like to thank Yuqiu Fu for many interesting discussions about this problem. LG is supported by a Simons Investigator Award. JM is supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 851318).

2. Sketch outline

For a heuristic description of our argument, let us consider a critical case of previous zero density estimates, where $\sigma = 3/4$ and we wish to improve upon the bound $|W| \lessapprox TN^{-1/2}$ for any set W of separated points in an interval of length T such that $|D_N(t)| > N^{3/4}$ when $t \in W$. We focus on the situation when $T = N^{1+\delta}$ for some small constant $\delta > 0$ (a corresponding improvement for larger values of T then follows by subdivision), and suppress many technical details (such as the presence of smoothing in most summations) for exposition.

Let M be the $|W| \times N$ matrix with entries

$$M_{t,n} = n^{it}$$

for $t \in W$ and $n \sim N$. We see that if $\mathbf{b} = (b_n)_{n\sim N}$ then

$$D_N(t) = \sum_n b_n n^{it} = (M\mathbf{b})_t,$$

and so we have (recalling $|b_n| \leq 1$)

$$N^{3/2}|W| = N^{2\sigma}|W| < \sum_{t\in W} |D_{\mathbf{b}}(t)|^2 = \|M\mathbf{b}\|_2^2 \leq \|M\|^2\|\mathbf{b}\|_2^2 \leq N\|M\|^2,$$

where $\|M\|$ is the matrix (operator) norm of M. Thus we wish to improve upon the bound $\|M\| \lessapprox T^{1/2}$ which follows from the Mean Value Theorem (or the Large Values Estimate), and we might hope that $\|M\| \approx |N|^{1/2}$. We note that

$$\|M\| = s_1(M)$$

where $s_1(M)$ is the largest singular value of M. This is also the square-root of the largest eigenvalue of M^*M. A simple bound for $s_1(M)$ is then given by the trace of powers of M^*M; for any $r \in \mathbb{N}$

$$s_1(M) \le \operatorname{tr}((M^*M)^r)^{1/2r}.$$

We might hope $\operatorname{tr}((M^*M)^r) \approx |W|N^r$, which would imply $s_1(M) \le N^{1/2}|W|^{1/2r}$. Unfortunately it appears to be very difficult to estimate these traces accurately for large r, and we would only improve upon the bound $|W| \lessapprox TN^{-1/2}$ if $(TN^{-1/2})^{1/r} \le TN^{-1}$, which requires r to be large. Nevertheless, a variant of this bound for $r = 3$ (which reduces the contribution from some of the trivial terms) allows one to show a bound similar to

$$s_1(M)^6 \lessapprox \Big| \sum_{\substack{t_1,t_2,t_3 \in W \\ |t_j - t_k| > T^\epsilon \,\forall j,k}} \sum_{n_1,n_2,n_3 \sim N} n_1^{i(t_1-t_2)} n_2^{i(t_2-t_3)} n_3^{i(t_3-t_1)} \Big|.$$

The right hand side would be $\operatorname{tr}((M^*M)^3)$ if we didn't have the lower bounds on $|t_i - t_j|$ (and so we avoid the $t_1 = t_2 = t_3$ terms). To improve upon the bound $|W| \lessapprox TN^{-1/2}$ we wish to show the right hand side is a bit smaller than T^3.

The natural approach to estimating the right hand side would be to apply Poisson summation and stationary phase to each of the inner sums, which would yield a bound of roughly

$$\frac{N^3}{T^{3/2}} \Big| \sum_{t_1,t_2,t_3 \in W} \sum_{m_1,m_2,m_3 \sim T/N} m_1^{i(t_1-t_2)} m_2^{i(t_2-t_3)} m_3^{i(t_3-t_1)} c_{t_1-t_2} c_{t_2-t_3} c_{t_3-t_1} \Big|$$

for some coefficients c_t of size 1. Unfortunately the $c_{t_i - t_j}$ coefficients link the variables t_1, t_2 and t_3, and this makes it difficult to show any cancellation over these summations. Without any cancellation in the t_1, t_2, t_3 sums we would be limited to a bound of $|W|^3 N^{3/2} \approx T^3$ even if we obtained square-root cancellation in the sums over m_1, m_2, m_3, and so this would fail to give the desired improvement on T^3 (but could potentially give results if $\sigma > 3/4$).

A key observation is that it can be beneficial to avoid the use of stationary phase. Applying Poisson summation without simplifying the Fourier integrals yields a bound of roughly

$$\begin{aligned} &N^3 \sum_{t_1,t_2,t_3 \in W} \sum_{|m_1|,|m_2|,|m_3| \sim T/N} \int_{[1,2]^3} e^{-2\pi i N \mathbf{m}\cdot\mathbf{u}} \Big(\frac{u_1}{u_3}\Big)^{it_1} \Big(\frac{u_2}{u_1}\Big)^{it_2} \Big(\frac{u_3}{u_1}\Big)^{it_3} d\mathbf{u} \\ &\le N^3 \sum_{|\mathbf{m}| \sim T/N} \Big| \int_{[1,2]^3} e^{-2\pi i N \mathbf{m}\cdot\mathbf{u}} R\Big(\frac{u_1}{u_3}\Big) R\Big(\frac{u_2}{u_1}\Big) R\Big(\frac{u_3}{u_2}\Big) d\mathbf{u} \Big|, \end{aligned}$$

where $R(x) := \sum_{x\in W} x^{it}$. By not simplifying the integrals over $\mathbf{u}$ directly we have given up a factor of $T^{3/2}$, but now the variables t_1, t_2, t_3 are nicely separated from one another, and we can hope to show cancellation in these sums by showing the R function exhibits cancellation. However, even square-root cancellation in $R(x)$ would only win back a factor of $|W|^{3/2}$ which is less than the $T^{3/2}$ factor we gave up by not applying stationary phase. Therefore we need to exploit simultaneous cancellation in the integral over $\mathbf{u}$ and the R functions.

The next important observation is that the arguments of the R functions only lie in a two dimensional subvariety $z_1z_2z_3 = 1$. If we change variables to $v_1 = u_1/u_3$ and $v_2 = u_2/u_3$ then the integral above is roughly

$$\int_{[1/2,2]^2}\Big(\int_{[1,2]} e^{-2\pi i N(m_1v_1+m_2v_2+m_3)u_3}du_3\Big)R(v_1)R\Big(\frac{v_2}{v_1}\Big)R\Big(\frac{1}{v_2}\Big)dv_1dv_2.$$

There is a large amount of cancellation in the inner integral unless $m_1v_1 + m_2v_2 + m_3 \approx 0$, and this allows us to win back a factor of T. We are then reduced to bounding expressions of the form

$$\frac{N^3}{T}\sum_{|\mathbf{m}|\sim T/N}\Big|\int_{\substack{(v_1,v_2)\in[1/2,2]^2\\ m_1v_1+m_2v_2+m_3=0}} R(v_1)R\Big(\frac{v_2}{v_1}\Big)R\Big(\frac{1}{v_2}\Big)dv_1\Big|. \tag{2.1}$$

If we had uniform square-root cancellation in the R functions, we would now get a bound of $T^2|W|^{3/2}$, which comfortably beats the desired bound of T^3 when $|W| \approx TN^{-1/2}$, and so we would get a corresponding improvement to the large value estimates and zero density results.

Of course, we cannot expect to prove uniform square-root cancellation in the R function for arbitrary sets W. Nevertheless, one can show that R does exhibit cancellation on average; $\|R(v)\|^2_{L^2([1/2,2])} \lessapprox |W|$ and $\|R(v)\|^4_{L^4([1/2,2])} \lessapprox E(W)$, where $E(W)$ from (1.7) counts approximate additive quadruples in W. Treating the $\mathbf{m}$ summation trivially, this would give a bound

$$T^2|W|^{1/2}E(W)^{1/2}, \tag{2.2}$$

which would beat the target of T^3 if $E(W)$ is a bit smaller than $T^2/|W| \approx N^2|W|^3/T^2$. Thus we obtain good bounds whenever $E(W)$ is small. In particular, when $T \approx N^{1+\delta}$ with δ small, this means that we can handle any set W except those for which the additive energy is close to the maximal possible value of $|W|^3$.

We would like to complement this with an argument for when $E(W)$ is large. As mentioned in the introduction, Heath-Brown's result becomes useful in this situation. Lemma 1.7 can handle the case when $E(W)$ is larger than $|W|^2N^{1/2} \approx |W|^3N/T$, which isn't quite strong enough to cover all ranges. A refinement of this lemma can handle the situation when $E(W)$ is larger than $|W|^3N^2/T^2$ (at least when δ is small and $|W| \approx TN^{-1/2}$), which then allows us to get a small improvement in all cases except when $E(W) \approx |W|^3N^2/T^2$.

To overcome this final obstacle when $E(W) \approx |W|^3N^2/T^2$, we go back to (2.1), and exploit the averaging over $\mathbf{m}$. After a change of variables we need to consider

$$\frac{N^3}{T}\sum_{|m_1|,|m_2|,|m_3|\sim T/N}\int_{v_1\in[1/2,2]}\Big|R(v_1)R\Big(\frac{m_1v_1+m_3}{m_2v_1}\Big)R\Big(\frac{m_1v_1+m_3}{m_2}\Big)\Big|dv_1.$$

The only way the bound above could be tight is if there is a sparse set U on which $R(u)$ is large, and such that for most $u\in U$ and most $|m_1|,|m_2|,|m_3|\sim T/N$ we have $(m_1u+m_3)/m_2\in U$ and $(m_1u+m_3)/(m_2u)\in U$. We show that there cannot be a small set U which has this property of being approximately closed under such affine transformations, and this ultimately leads to an improvement of (2.2) to roughly

$$TN|W|^{1/2}E(W)^{1/2}.$$

This now gives an improvement on the desired bound of T^3 whenever $E(W)$ is smaller than $|W|^3$, and so the bounds obtained on $E(W)$ stemming from Heath-Brown's result above are now sufficient to give an improvement for any size of $E(W)$. Therefore we obtain an improvement on the bound $|W| \lessapprox TN^{-1/2}$ for arbitrary W.

3. Reduction to Main Proposition

We fix a smooth function $w:\mathbb{R}\to\mathbb{R}_{\geq 0}$ supported on $[1,2]$ with $\|w^{(j)}\|_\infty = O_j(1)$ for all $j\in\mathbb{Z}_{\geq 0}$ and with $w(t)=1$ for $t\in[6/5,9/5]$.

Proposition 3.1. *Let $\sigma\in[7/10,8/10]$ and $\epsilon>0$. If b_n is a sequence of complex numbers with $|b_n|\leq 1$ and $\mathcal{W}$ is a set of T^ϵ-separated points in an interval of length $T=N^{6/5}$ such that*

$$\Big|\sum_n w\Big(\frac{n}{N}\Big)b_n n^{it}\Big| \geq N^\sigma$$

for all $t\in\mathcal{W}$. Then we have

$$|\mathcal{W}| \leq TN^{(12-20\sigma)/5+o_\epsilon(1)}.$$

Proof of Theorem 1.1 assuming Proposition 3.1. As mentioned in the introduction, the result follows from (1.1) if $V\leq N^{7/10+o(1)}$ or if $V\geq N^{8/10-o(1)}$, so we may assume $V\in[4N^{7/10},N^{8/10}]$, in which case $N^2V^{-2}\leq N^{18/5}V^{-2}$. By splitting D_N into 3 separate pieces (and using the triangle bound), we see that it suffices to show the result when $b_n=0$ unless $n\in[6N/5,9N/5]$. Since the function w is 1 on $[6/5,9/5]$, we then have that $b_n=b_nw(n/N)$, so we may insert the weight $w(n/N)$. Finally, by now relaxing the vanishing of the coefficients and letting $V=N^\sigma$, it suffices to show that whenever $\sigma \in [7/10, 8/10]$, (a_n) is a 1-bounded complex sequence and W is a set of 1-separated points such that

$$\Big|\sum_n w\Big(\frac{n}{M}\Big)a_n n^{it}\Big| \geq M^\sigma$$

for each $t \in W$, we have

$$|W| \leq T^{o(1)}(M^{18/5-4\sigma} + TM^{12/5-4\sigma}).$$

We now fix $\epsilon > 0$ and choose $W' \subset W$ so that W' is T^ϵ-separated and $|W'| \geq |W|/T^\epsilon$ (this can be achieved by picking the smallest element of W and then repeatedly choosing the next smallest element which is at least T^ϵ away from all picked elements). If $T \leq N^{6/5}$, then we apply Proposition 3.1 to bound W' directly, which implies that

$$|W| \leq T^\epsilon |W'| \leq N^{(18-20\sigma)/5+2\epsilon+o_\epsilon(1)}.$$

Letting $\epsilon \to 0$ sufficiently slowly then gives the result in this case. If instead $T > N^{6/5}$, then we divide W' into $\lceil T/N^{6/5}\rceil$ subsets W'_j each supported on an interval of length $N^{6/5}$, and we apply Proposition 3.1 to bound each W'_j separately. This gives

$$|W| \leq T^\epsilon \sum_{j \leq \lceil T/N^{6/5}\rceil} |W'_j| \leq T^{1+\epsilon} N^{(12-20\sigma)/5+\epsilon+o_\epsilon(1)}.$$

Letting $\epsilon \to 0$ sufficiently slowly then gives the result in this case too. □

4. The matrix M_W and its singular values

Now we begin to work on the proof of Proposition 3.1. Recall that D_N is a Dirichlet polynomial of the form

$$D_N(t) = \sum_n w\Big(\frac{n}{N}\Big) b_n n^{it},$$

where w is a smooth bump supported on $[1, 2]$, as defined in Section 3.

Given a set $W \subseteq \mathbb{R}$, let M_W be the $|W| \times N$ matrix with entries

$$M_{t,n} = w(n/N)n^{it}, \tag{4.1}$$

where $t \in W$ and $n \sim N$.

Lemma 4.1 (Large values of Dirichlet polynomials controlled by singular values)**.** *Let M_W be the matrix defined in* (4.1)*, and $s_1(M_W)$ its largest singular value. If $|D_N(t)| \geq N^\sigma$ on W and if $|b_n| \leq 1$, then we have*

$$|W| \lesssim N^{1-2\sigma} s_1(M_W)^2.$$

Proof. Let $\mathbf{b}$ be the vector with components b_n. Then note that for each t in W,

$$D_N(t) = \sum_n w(n/N) b_n n^{it} = (M_W \mathbf{b})_t.$$

Therefore we can relate the behavior of D_N on W (for arbitrary $\mathbf{b}$) to properties of the matrix M_W, in particular its singular values. We write $s_j(M_W)$ for the j^{th} singular value of M_W, with the convention that $s_1(M_W) \geq s_2(M_W) \geq \dots \geq s_k(M_W)$. Let M_W have singular value decomposition $M_W = U\Sigma V$, so that Σ is a rectangular matrix with $\Sigma_{ii} = s_i(M_W)$ and $\Sigma_{ij} = 0$ if $i \neq j$, and U, V are unitary matrices.

If $|D_N(t)| \geq N^\sigma$ on W, then we see

$$\begin{aligned} |W| N^{2\sigma} \leq \sum_{t \in W} |D_N(t)|^2 = (M_W \mathbf{b})^* M_W \mathbf{b} = (V\mathbf{b})^* \Sigma^2 V \mathbf{b} \\ \leq s_1(M_W)^2 \|V\mathbf{b}\|_{\ell^2}^2 \\ = s_1(M_W)^2 \|\mathbf{b}\|_{\ell^2}^2. \end{aligned}$$

Finally, if $|b_n| \leq 1$ then $\|\mathbf{b}\|_{\ell^2}^2 \lesssim N$. Substituting this into the expression above and rearranging now gives the result. □

Now $s_1(M_W)$ is equal to the square root of the largest eigenvalue value of the $W \times W$ matrix $M_W M_W^*$, with entries

$$(M_W M_W^*)_{t_1,t_2} = \sum_n w\left(\frac{n}{N}\right)^2 n^{i(t_1-t_2)}.$$

A simple bound for $s_1(M_W)$ is therefore to use the trace: for any integer $r \geq 1$ we have

$$s_1(M_W)^2 = s_1(M_W M_W^*) \leq \Big(\sum_{j=1}^k s_j(M_W M_W^*)^r\Big)^{1/r} = \operatorname{tr}((M_W M_W^*)^r)^{1/r}.$$

One might guess that $s_j(M_W) \lessapprox N^{1/2}$ for all j, in which case we would have $\operatorname{tr}((M_W M_W^*)^r)^{1/r} \lessapprox |W|^{1/r} N$. If one could establish such a sharp bound on $\operatorname{tr}((M_W^* M_W)^r)$ for large r, this would give Conjecture 1.5. Unfortunately we do not know how to obtain good bounds when $r \geq 4$, so we work with $r = 3$. In this case, even a sharp bound $\operatorname{tr}((M_W M_W^*)^3)^{1/3} \lessapprox |W|^{1/3} N$ would only yield $|W| \lessapprox N^{3-3\sigma}$, which is worse than the bounds established by previous works. To get around this issue, we note that if we are in the extreme scenario when

$$\frac{1}{k}\sum_{i=1}^{k} s_i(M_W)^6 = \Big(\frac{1}{k}\sum_{i=1}^{k} s_i(M_W)^2\Big)^3$$

then in fact we must have that all the singular values are the same, and so $s_1(M_W) = \mathrm{tr}((M_W M_W^*)^3)^{1/6}k^{-1/6}$, a significant improvement on the bound $\mathrm{tr}((M_W M_W^*)^3)^{1/6}$ we had before. Similarly, we would expect that if $\mathrm{tr}((M_W M_W^*)^3)$ is close to $\mathrm{tr}(M_W M_W^*)^3/k^2$, then we would also get an improved bound on $s_1(M_W)$ since most of the contribution to $\mathrm{tr}((M_W M_W^*)^3)$ would be coming from the many other singular values. The following lemma makes this precise, stating that we can essentially replace $\mathrm{tr}((M_W M_W^*)^3)$ with the difference $\mathrm{tr}((M_W M_W^*)^3) - \mathrm{tr}(M_W M_W^*)^3/k^2$ for the purposes of bounding $s_1(M_W)$.

Lemma 4.2 (Bound for singular values in terms of traces)**.** *Let A be an $m \times n$ complex matrix. Then we have*

$$s_1(A) \le 2\Big(\mathrm{tr}((AA^*)^3) - \frac{\mathrm{tr}(AA^*)^3}{m^2}\Big)^{1/6} + 2\Big(\frac{\mathrm{tr}(AA^*)}{m}\Big)^{1/2}.$$

Proof. Recall that $\mathrm{tr}((AA^*)^j) = \sum_{i=1}^m \lambda_i^j$ where $\lambda_1, \dots, \lambda_m$ are the eigenvalues of the $m \times m$ matrix AA^*, which are real and non-negative, and that $s_1(A) = \max_i \lambda_i^{1/2}$. We see that it is sufficient to show for any non-negative reals $x_1, \dots x_k$

$$x_1 \le 2\Big(\sum_{i=1}^{k} x_i^6 - \frac{(\sum_{i=1}^k x_i^2)^3}{k^2}\Big)^{1/6} + 2\Big(\frac{\sum_{i=1}^k x_i^2}{k}\Big)^{1/2}. \tag{4.2}$$

By Hölder's inequality, we have that $\sum_{i=2}^k x_i^6 \ge (\sum_{i=2}^k x_i^2)^3/(k-1)^2 \ge (\sum_{i=2}^k x_i^2)^3/k^2$. Thus

$$\begin{aligned} x_1^6 = \sum_{i=1}^{k} x_i^6 - \sum_{i=2}^{k} x_i^6 &\le \sum_{i=1}^{k} x_i^6 - \frac{(\sum_{i=2}^k x_i^2)^3}{k^2} \\ &\le \Big(\sum_{i=1}^{k} x_i^6 - \frac{(\sum_{i=1}^k x_i^2)^3}{k^2}\Big) + 3x_1^2\frac{(\sum_{i=1}^k x_i^2)^2}{k^2}. \end{aligned}$$

This in turn implies (4.2) (if (4.2) failed then x_1^6 would be larger than the right hand side above.) □

Thus we wish to estimate $\mathrm{tr}(M_W M_W^*)$ and $\mathrm{tr}((M_W M_W^*)^3)$. In both cases we expand the trace and use Poisson summation as a first step. In anticipation of this, we introduce the function

$$h_t(u) := w(u)^2 u^{it}, \tag{4.3}$$

which appears in the sums defining the coefficients of $M_W M_W^*$. We first record a basic tail estimate for the Fourier transform $\hat{h}_t$.

Lemma 4.3 (Non-stationary phase)**.** *Let $h_t(u) = w(u)^2 u^{it}$. Then we have*

(1) For any integer $j \geq 0$ we have
$$\hat{h}_t(\xi) \lesssim_j (1+|t|)^j/|\xi|^j.$$
(2) For any integer $j \geq 0$ we have
$$\hat{h}_t(\xi) \lesssim_j (1+|\xi|)^j/|t|^j.$$

Proof. Since $\|w^{(j)}\|_\infty \lesssim_j 1$ for all $j \geq 0$, we have that $\|h_t^{(j)}\|_\infty \lesssim_j 1+|t|^j$ for all $j \geq 0$. Thus, by integration by parts (and using that w is compactly supported), we have that

$$\hat{h}_t(\xi) = \int e(-\xi u) h_t(u) du = \frac{1}{\xi^j}\int e(-\xi u) h_t^{(j)}(u) du \lesssim_j \frac{1+|t|^j}{|\xi|^j}.$$

Similarly, if $g_\xi(u) = e(-\xi u) w(u)^2$ then $\|g_\xi^{(j)}\|_\infty \lesssim_j 1+|\xi|^j$, so integration by parts gives

$$\hat{h}_t(\xi) = \int g_\xi(u) u^{it} du = \frac{1}{(it+1)\cdots(it+j)}\int g_\xi^{(j)}(u) u^{it+j} du \lesssim_j \frac{1+|\xi|^j}{|t|^j}. \qquad \square$$

Lemma 4.4 (Hibert-Schmidt Norm estimate)**.** *If $W \subset \mathbb{R}$ is a finite set with $|W| \leq N^{O(1)}$, then*
$$\mathrm{tr}(M_W M_W^*) = N|W|\,\|w\|_{L^2}^2 + O(N^{-100}).$$

Proof. Expanding the trace, we see that

$$\mathrm{tr}(M_W M_W^*) = \sum_{t \in W}\sum_n w(n/N)^2 = |W|\sum_{n\in\mathbb{Z}} h_0(n/N).$$

Because h_0 is a smooth compactly supported function, the sum $\sum_n h_0(n/N)$ is very close to the integral $N\int_{\mathbb{R}} h_0(u) d\xi = N\|w\|_{L^2}^2$. We can get a precise estimate using Poisson summation, which gives (separating the term $m = 0$)

$$\sum_n h_0(n/N) = N\sum_m \widehat{h_0}(Nm) = N\widehat{h_0}(0) + O\Big(\sum_{m\neq 0} |\widehat{h_0}(Nm)|\Big).$$

The first term on the right hand side is $N\|w\|_{L^2}^2$ and the second term is $O(N^{-100})$ by Lemma 4.3. $\square$

Lemma 4.5 (Expansion of the cubic trace)**.** *Let W be T^ϵ-separated. Then we have*

$$\mathrm{tr}((M_W M_W^*)^3) = N^3|W|\|w\|_{L^2}^6 + \sum_{m \in \mathbb{Z}^3\setminus\{0\}} I_m + O_\epsilon(T^{-100}),$$

where

$$I_m := N^3 \sum_{t_1,t_2,t_3\in W} \hat{h}_{t_1-t_2}(m_1N)\hat{h}_{t_2-t_3}(m_2N)\hat{h}_{t_3-t_1}(m_3N).$$

Proof. First we expand $\mathrm{tr}((M_W M_W^*)^3)$ as the sum S, given by

$$\begin{aligned} S &= \sum_{n_1,n_2,n_3\in\mathbb{Z}} \sum_{t_1,t_2,t_3\in W} w\Big(\frac{n_1}{N}\Big)^2 w\Big(\frac{n_2}{N}\Big)^2 w\Big(\frac{n_3}{N}\Big)^2 n_1^{i(t_1-t_2)} n_2^{i(t_2-t_3)} n_3^{i(t_3-t_1)} \\ &= \sum_{t_1,t_2,t_3\in W} \sum_{n_1,n_2,n_3\in\mathbb{Z}} h_{t_1-t_2}\Big(\frac{n_1}{N}\Big) h_{t_2-t_3}\Big(\frac{n_2}{N}\Big) h_{t_3-t_1}\Big(\frac{n_3}{N}\Big), \end{aligned}$$

where, as in (4.3), we have $h_t(u) = w(u)^2 u^{it}$. We now perform Poisson summation in n_1, n_2, n_3, which gives

$$S = N^3 \sum_{m_1,m_2,m_3\in\mathbb{Z}} \sum_{t_1,t_2,t_3\in W} \hat{h}_{t_1-t_2}(m_1N)\hat{h}_{t_2-t_3}(m_2N)\hat{h}_{t_3-t_1}(m_3N) = \sum_{m\in\mathbb{Z}^3} I_m.$$

Finally, we separate the term $m_1 = m_2 = m_3 = 0$, which contributes

$$I_0 = N^3 \sum_{t_1,t_2,t_3\in W} \hat{h}_{t_1-t_2}(0)\hat{h}_{t_2-t_3}(0)\hat{h}_{t_3-t_1}(0).$$

Since W is T^ϵ-separated, $\hat{h}_{t_1-t_2}(0) \lesssim_\epsilon T^{-200}$ if $t_1 \neq t_2$ by Lemma 4.3. Thus the terms in the I_0 above are negligible unless $t_1 = t_2 = t_3$, and so

$$I_0 = N^3 \sum_{t\in W} \hat{h}_0(0)^3 + O_\epsilon(T^{-100}) = |W| \|w\|_{L^2}^6 + O_\epsilon(T^{-100}).$$

Putting this together gives the result. □

Putting together Lemmas 4.1-Lemma 4.5, and noting that the $N^3|W|\|w\|_{L^2}^3$ term cancels with $\mathrm{tr}(M_W M_W^*)^3/|W|^2$, gives the following.

Proposition 4.6. *Let W be T^ϵ-separated, and let $|b_n| \le 1$ be such that $|D_N(t)| > N^\sigma$ for all $t \in W$. Then we have*

$$|W| \lesssim_\epsilon N^{2-2\sigma} + N^{1-2\sigma}\Big(\sum_{m\in\mathbb{Z}^3\setminus\{0\}} I_m\Big)^{1/3},$$

where I_m is the quantity defined in Lemma 4.5.

The first term above corresponds to the best possible estimate is $|W| \lesssim N^{2-2\sigma}$ of Conjecture 1.5, and so our task is reduced to getting a good bound for the sum of I_m.

5. The pieces of the sum S

Recall from Proposition 4.6, we have

$$|W|^3 \lesssim N^{6-6\sigma} + N^{3-6\sigma} \sum_{m\in\mathbb{Z}^3\setminus\{0\}} I_m,$$

where

$$I_m = N^3 \sum_{t_1,t_2,t_3\in W} \hat{h}_{t_1-t_2}(m_1 N)\hat{h}_{t_2-t_3}(m_2 N)\hat{h}_{t_3-t_1}(m_3 N).$$

To get started, we note a few cases when $|\hat{h}_{t_1-t_2}(mN)|$ is easy to understand via Lemma 4.3. Since W is T^ϵ-separated, we see that if $t_1 \neq t_2$, by Lemma 4.3 we have

$$|\hat{h}_{t_1-t_2}(0)| \lesssim_\epsilon T^{-100}. \tag{5.1}$$

On the other hand, if $t_1 = t_2$, then we have

$$\hat{h}_{t_1-t_2}(0) = \hat{h}_0(0) = \int w^2(u)du \asymp 1. \tag{5.2}$$

If $t_1 = t_2$ but $m \neq 0$, then by Lemma 4.3 we have

$$|\hat{h}_{t_1-t_2}(mN)| \lesssim m^{-100}N^{-100}. \tag{5.3}$$

Finally, if $m > T^{1+\epsilon}/N$, then since W is contained in an interval of length T, we have by Lemma 4.3 and taking $j = \lceil 200/\epsilon\rceil + 100$

$$|\hat{h}_{t_1-t_2}(mN)| \lesssim_\epsilon \frac{T^{100}}{(mN)^{100}}\Big(\frac{T}{T^{1+\epsilon}}\Big)^{200/\epsilon} \lesssim T^{-100}m^{-100}. \tag{5.4}$$

With this in mind, we divide the sum into pieces

$$\sum_{m\in\mathbb{Z}^3\setminus\{0\}} I_m = S_1 + S_2 + S_3, \tag{5.5}$$

where S_1 contains the terms where exactly one m_i is non-zero, S_2 contains the terms where exactly two m_i are non-zero, and S_3 contains the terms where all three m_i are non-zero.

We will see that S_1 is negligible. In the next section, we will bound S_2 using Heath-Brown's theorem, Theorem 1.6. The main part of the paper is concerned with studying S_3, which contains most of the terms and is most difficult.

Proposition 5.1 (S_1 bound)**.** *We have*

$$S_1 = O_\epsilon(T^{-10}).$$

Proof. By symmetry, we see that

$$S_1 \leq 3N^3 \sum_{t_1,t_2,t_3\in W} \sum_{m_3\neq 0} |\hat{h}_{t_1-t_2}(0)\hat{h}_{t_2-t_3}(0)\hat{h}_{t_3-t_1}(m_3N)|.$$

By (5.4) (using the trivial bound $|\hat{h}_t(\xi)| \lesssim 1$ for the other factors), terms with $|m_3| > T^{1+\epsilon}/N$ contribute

$$\lesssim_\epsilon N^3|W|^3 \sum_{m>T^{1+\epsilon}/N} T^{-100}m^{-100} \lesssim T^{-10}.$$

Thus we may restrict attention to terms with $|m_3| < T^{1+\epsilon}/N$. Next we consider terms with $t_1 \neq t_2$. Using (5.1) to bound $|\hat{h}_{t_1-t_2}(0)|$ (and the trivial bound $\hat{h}_t \lesssim 1$ for the remaining factors), we see the terms with $t_1 \neq t_2$ and $|m_3| < T^{1+\epsilon}/N$ contribute

$$\lesssim_\epsilon N^3|W|^3\frac{T^{1+\epsilon}}{N}T^{-100} \lessapprox T^{-10}.$$

Similarly, the terms with $t_2 \neq t_3$ contribute $O_\epsilon(T^{-10})$. The remaining terms have $t_1 = t_2 = t_3$. For these terms we apply (5.3) to bound $|\hat{h}_{t_3-t_1}(m_3N)|$, which shows that the terms with $t_1 = t_2 = t_3$ also contribute $O_\epsilon(T^{-10})$. This gives the result. □

6. The contribution of S_2

The aim of this section is to establish the following bound for the sum S_2, which ultimately relies on Heath-Brown's estimate 1.6.

Proposition 6.1 (S_2 bound)**.** *For any choice of* $k \in \mathbb{N}$

$$S_2 \lessapprox N^2|W|^2 + TN|W|^{2-1/k} + N^2|W|^2\Big(\frac{T^{1/2}}{|W|^{3/4}}\Big)^{1/k}.$$

The proof of this proposition relies on the following consequence of stationary phase, which is part of the well-known 'reflection principle' for Dirichlet polynomials, or the approximate functional equation (values of a Dirichlet polynomial of length N at $t \in [T, 2T]$ are determined by values of a Dirichlet polynomial of length T/N).

Lemma 6.2 (Approximate functional equation)**.** *For every* t *with* $|t| \sim T_0 \geq T^\epsilon$, *we have*

$$\Big|\sum_{m\neq 0} \hat{h}_t(mN)\Big| \lesssim \frac{1}{T_0^{1/2}} \int_{u\lessapprox 1} \Big|\sum_{m\lessapprox T_0/N} m^{-i(t+u)}\Big| du + O(T^{-100}).$$

Although somewhat standard, we will give a detailed proof of Lemma 6.2 below. Let us first use it to bound S_2.

Proof of Proposition 6.1 assuming Lemma 6.2. Recall that S_2 is the sum of those I_m where exactly two m_i are non-zero. By symmetry, we have

$$S_2 = 3N^3 \sum_{m_1,m_2\neq 0}\sum_{t_1,t_2,t_3\in W} \hat{h}_{t_1-t_2}(m_1N)\hat{h}_{t_2-t_3}(m_2N)\hat{h}_{t_3-t_1}(0).$$

If $t_1 \neq t_3$, then (5.1) shows that the last factor $\hat{h}_{t_3-t_1}(0)$ is $O_\epsilon(T^{-100})$, and so using the bound $\hat{h}_t(u) \lesssim (1+|t|^2)/|u|^2$ from Lemma 4.3 for the remaining factors, these terms contribute $O_\epsilon(T^{-10})$ in total. Therefore we have

$$S_2 = 3N^3\hat{h}_0(0) \sum_{m_1,m_2\neq 0}\sum_{t_1,t_2\in W} \hat{h}_{t_1-t_2}(m_1N)\hat{h}_{t_2-t_1}(m_2N) + O_\epsilon(T^{-10}).$$

Since $h_t(u) = w(u)^2u^{it}$, we have $h_{-t}(u) = \overline{h_t(u)}$, and so $\hat{h}_{-t}(\xi) = \overline{\hat{h}_t(-\xi)}$. In particular,

$$\hat{h}_{t_2-t_1}(m_2N) = \overline{\hat{h}_{t_1-t_2}(-m_2N)}.$$

Therefore, we can simplify the last equation to get

$$S_2 = 3N^3\hat{h}_0(0)\sum_{t_1,t_2\in W}\Big|\sum_{m\neq 0}\hat{h}_{t_1-t_2}(mN)\Big|^2 + O_\epsilon(T^{-10}).$$

Remark. *Heath-Brown's theorem, Theorem 1.6, gives a good estimate for the sum* $\sum_{t_1,t_2\in W}|\sum_{m\in\mathbb{Z}}\hat{h}_{t_1-t_2}(mN)|^2$. *However we cannot apply it immediately because we need to be careful to leave out the term with* $m=0$. *This term is related to* I_0 *which we handled carefully in the previous section.*

If $t_1 = t_2$, then $\sum_{m\neq 0}\hat{h}_{t_1-t_2}(mN)$ is negligible by (5.3) and (5.4). So, splitting the sum dyadically according to the size of t_1-t_2, we find

$$S_2 \lessapprox N^3 \sup_{\substack{M=2^j\\ T^\epsilon/N<M<2T/N}} \sum_{\substack{t_1\neq t_2\in W\\ |t_1-t_2|\sim MN}}\Big|\sum_{m\neq 0}\hat{h}_{t_1-t_2}(mN)\Big|^2 + O_\epsilon(T^{-10}).$$

If $|t_1-t_2|\sim MN$, then $\sum_{m\neq 0}\hat{h}_{t_1-t_2}(mN)$ can be approximated by a Dirichlet polynomial of length M. Indeed, by Lemma 6.2, for such t_1,t_2 we have

$$\Big|\sum_{m\neq 0}\hat{h}_{t_1-t_2}(mN)\Big| \lesssim \frac{1}{M^{1/2}N^{1/2}}\int_{|u|\lessapprox 1}\Big|\sum_{1\leq m\lessapprox M} m^{-i(t-u)}\Big|du + O(T^{-100}).$$

Squaring and summing over $t_1,t_2\in W$ with $|t_1-t_2|\sim NM$ gives

$$S_2 \lessapprox \sup_{\substack{M\le 2T/N\\ |u|\lessapprox 1}} \frac{N^2}{M} \sum_{\substack{t_1\neq t_2\in W\\ |t_1-t_2|\sim MN}} \Big|\sum_{1\le m\lessapprox M} m^{i(t-u)}\Big|^2 + O(T^{-10}).$$

We can now drop the condition $|t_1-t_2|\sim MN$ for an upper bound, and split the summation range $m\lessapprox M$ into dyadic intervals. Noting that choosing all a_m to be zero apart from one gives a contribution which dominates the error term, we find

$$S_2 \lessapprox \frac{N^2}{M}\sum_{t_1,t_2\in W}\Big|\sum_{m\sim M} a_m m^{i(t_1-t_2)}\Big|^2 \tag{6.1}$$

for some choice of $M\lessapprox T/N$ and some coefficients $|a_m|\le 1$.

We apply Hölder's inequality to this sum, and rewrite the $2k^{th}$ power of the Dirichlet polynomial as the 2^{nd} power of a bigger Dirichlet polynomial. For any choice of positive integer k, we find that

$$\sum_{t_1,t_2\in W}\Big|\sum_{m\sim M} a_m m^{i(t_1-t_2)}\Big|^2 \le |W|^{2-2/k}\Big(\sum_{t_1,t_2\in W}\Big|\sum_{m\asymp M^k} b_m m^{i(t_1-t_2)}\Big|^2\Big)^{1/k} \tag{6.2}$$

for some coefficients $b_m\le M^{o_k(1)}$ (by the divisor bound). Theorem 1.6 bounds sums of this type. We recall the statement.

Theorem (Heath-Brown)**.** *Let $\mathcal{T}$ be a 1-separated set of reals, contained in an interval of length T. Let $|a_n|\lessapprox 1$ be a complex sequence. Then*

$$\sum_{t_1,t_2\in\mathcal{T}}\Big|\sum_{n=N}^{2N} a_n n^{i(t_1-t_2)}\Big|^2 \lessapprox |\mathcal{T}|^2N+|\mathcal{T}|N^2+|\mathcal{T}|^{5/4}T^{1/2}N.$$

This result implies that

$$\sum_{t_1,t_2\in W}\Big|\sum_{m\asymp M^k} b_m m^{i(t_1-t_2)}\Big|^2 \lessapprox_k |W|M^{2k}+|W|^2M^k+|W|^{5/4}T^{1/2}M^k. \tag{6.3}$$

Substituting (6.2) and (6.3) back into (6.1), we see that

$$\begin{aligned} S_2 &\lessapprox_k \frac{N^2}{M}(|W|^2M+M^2|W|^{2-1/k}+|W|^2MT^{1/2k}|W|^{-3/4k})\\ &\lessapprox_k N^2|W|^2+TN|W|^{2-1/k}+N^2|W|^2\Big(\frac{T^{1/2}}{|W|^{3/4}}\Big)^{1/k}. \qquad (6.4)\end{aligned}$$

This gives the result. □

Now we return to the proof of Lemma 6.2, which roughly says $\hat{h}_t(mN)$ can be thought of as a smoothed version of $t^{-1/2}m^{it}$ supported on $m \asymp t/N$.

Proof of Lemma 6.2. By Lemma 4.3, when $|t| \geq T^{\epsilon}$ we have $\hat{h}_t(0) = O(T^{-100})$ so we can include the term $m = 0$ at the cost of an $O_\epsilon(T^{-100})$ error term. Let $W(s) := \int w(u)^2 u^{s-1} du$ be the Mellin transform of w^2, which is entire and has $|W(s)| \lesssim_j |s|^{-j}$ for all $j \in \mathbb{N}$. Then by Poisson summation and Mellin inversion $(w(u)^2 = (2\pi i)^{-1} \int_{2-i\infty}^{2+i\infty} W(s) u^{-s} ds)$, we have that

$$N^{1+it} \sum_m \hat{h}_t(mN) = \sum_n w(n/N)^2 n^{it} = \frac{1}{2\pi i} \int_{2-i\infty}^{2+i\infty} W(s) N^s \zeta(s-it) ds. \tag{6.5}$$

We move the line of integration to $\Re(s) = -1$, picking up a pole at $s = 1 + it$, and then on the remaining contour we use the functional equation of the zeta function (see [Da, Chapter 8]), which states that $\zeta(s) = G(s)\zeta(1-s)$, where

$$G(s) := \pi^{-1/2+s} \frac{\Gamma\left(\frac{1-s}{2}\right)}{\Gamma\left(\frac{s}{2}\right)}.$$

Thus we have (using the rapid decay of W to bound the contribution from the pole)

$$\frac{1}{2\pi i} \int_{2-i\infty}^{2+i\infty} W(s) N^s \zeta(s-it) ds = N^{1+it} W(1+it) + \frac{1}{2\pi i} \int_{-1-i\infty}^{-1+i\infty} W(s) N^s \zeta(s-it) ds$$

$$= O(T^{-100}) + \frac{1}{2\pi i} \int_{-1-i\infty}^{-1+i\infty} W(s) N^s G(s-it) \zeta(1-s+it) ds. \tag{6.6}$$

We fix a small constant $\eta > 0$ and split the sum defining $\zeta(1-s+it)$ at $M := T^\eta T_0/N$. More specificall, write $\zeta(1-s+it) = Z_1(1-s+it) + Z_2(1-s+it)$, where

$$Z_1(s) := \sum_{1 \leq m \leq M} m^{-s}, \qquad Z_2(s) := \sum_{m > M} m^{-s},$$

noting that $Z_2(s)$ converges absolutely when $\Re(s) > 1$. We substitute this into the integral above, and then move the integral involving Z_1 to $\Re(s) = 1$ and the integral involving Z_2 to $\Re(s) = -2k$ for a large $k \in \mathbb{N}$. This gives

$$\frac{1}{2\pi i} \int_{-1-i\infty}^{-1+i\infty} W(s) N^s G(s-it) \zeta(1-s+it) ds = I_1 + I_2, \tag{6.7}$$

where (making a change of variables $s = 1 + iu$ in I_1 and $s = -2k + iv$ in I_2)

$$I_1 := \frac{1}{2\pi}\int_{-\infty}^{\infty} W(1+iu)N^{1+iu}G(1+i(u-t))\Big(\sum_{1\le m\le M} m^{i(u-t)}\Big)du,$$
$$I_2 := \frac{1}{2\pi}\sum_{m>M}\int_{-\infty}^{\infty} W(-2k+iv)N^{-2k+iv}G(-2k+i(v-t))m^{-2k-1+i(v-t)}dv.$$

From the rapid decay of $W(s)$, we can truncate these integrals to $|u|,|v| \lessapprox 1$ at the cost of an $O(T^{-100})$ error term. It follows from well-known properties of the Gamma function (Stirling's formula and the functional equation; see [Da, Chapter 10]) that for $|u|,|v| \lessapprox 1$ and $|t| > N$ we have

$$\Big|\frac{\Gamma\Big(\frac{1+i(u-t)}{2}\Big)}{\Gamma\Big(\frac{i(t-u)}{2}\Big)}\Big| \lesssim \frac{1}{|t|^{1/2}},$$
$$\Big|\frac{\Gamma\Big(\frac{2k+1-i(v-t)}{2}\Big)}{\Gamma\Big(\frac{-2k+i(v-t)}{2}\Big)}\Big| = \Big|\frac{\Gamma\Big(\frac{2k+1-i(v-t)}{2}\Big)}{\Gamma\Big(\frac{2k+i(v-t)}{2}\Big)}\Big|\prod_{j=-k}^{k-1}|-j+i(v-t)/2| \lesssim_k |t|^{2k+1/2}.$$

Recalling that $|t| \sim T_0$, we have $G(1+i(u-t)) \lesssim T_0^{-1/2}$ and $G(-2k+i(v-t)) \lesssim_k T_0^{2k+1/2}$. Substituting these bounds in (and recalling $M = T^\eta T_0/N$), we find that

$$I_1 \lesssim \frac{N}{T_0^{1/2}}\int_{|u|\lessapprox 1}\Big|\sum_{1\le m\lesssim T^\eta M} m^{i(u-t)}\Big|du + O(T^{-100}),$$
$$I_2 \lessapprox_k \frac{T_0^{2k+1/2}N^{-2k}}{(T^\eta T_0/N)^{2k-1}}\sum_{m>M}\frac{1}{m^2}.$$

If we choose k sufficiently large in terms of η, we then see that $I_2 = O_\eta(T^{-200})$. Substituting these bounds into (6.7), and combining this with (6.6) and (6.5) gives the result on letting $\eta \to 0$. □

7. The contribution of S_3: a key cancellation

Now we begin to study S_3, which is the most difficult term. Recall that

$$S_3 = \sum_{m_1,m_2,m_3\neq 0} I_m,$$

where

$$I_m = N^3\sum_{t_1,t_2,t_3\in W}\hat{h}_{t_1-t_2}(m_1N)\hat{h}_{t_2-t_3}(m_2N)\hat{h}_{t_3-t_1}(m_3N).$$

By Lemma 4.3, $|\hat{h}_t(\xi)| \lessapprox_j (|\xi|+1)^j/|t|^j$ for any $j \in \mathbb{N}$, and so $\hat{h}_t$ is rapidly decaying when $|\xi|$ is much bigger than $|t|$, and hence I_m is negligible unless $|m| \lessapprox T/N$. Thus

$$S_3 = \sum_{0<|m_1|,|m_2|,|m_3| \lessapprox T/N} I_m + O(T^{-100}). \tag{7.1}$$

The first step in our argument is an estimate for $|I_m|$. We introduce the function

$$R(v) := \sum_{t\in W} v^{it} = \hat{W}(\log v) \tag{7.2}$$

which will play an important role in our analysis of S_3.

Proposition 7.1 (Cancellation within the I_m integrals)**.** *We have*

$$|I_m| \lesssim N^3 \int_{\substack{|m_1v_1+m_2v_2+m_3|\lessapprox \frac{1}{N}\\ v_1\asymp v_2\asymp 1}} \Big|R(v_1)R\Big(\frac{v_2}{v_1}\Big)R(v_2)\Big| dv_1dv_2 + O(T^{-200}).$$

Moreover, if $|m_1|\le|m_2|\le|m_3|$, then $|I_m| = O(T^{-200})$ unless $|m_2|\asymp|m_3|$.

Proof. Expanding the definition of $\hat{h}_t$ as an integral and swapping the order of summation and integration, we have

$$\begin{aligned} I_m &= N^3 \sum_{t_1,t_2,t_3\in W}\int_{\mathbb{R}^3} e(-N\mathbf{m}\cdot\mathbf{u})w_1(\mathbf{u})u_1^{i(t_1-t_2)}u_2^{i(t_2-t_3)}u_3^{i(t_3-t_1)}d\mathbf{u}\\ &= N^3\int_{\mathbb{R}^3} e(-N\mathbf{m}\cdot\mathbf{u})w_1(\mathbf{u})R\Big(\frac{u_1}{u_3}\Big)R\Big(\frac{u_2}{u_1}\Big)R\Big(\frac{u_3}{u_2}\Big)d\mathbf{u}, \qquad (7.3)\end{aligned}$$

where

$$w_1(\mathbf{u}) := w(u_1)^2w(u_2)^2w(u_3)^2, \tag{7.4}$$

In (7.3), the R functions depend on u_1/u_3, u_2/u_1 and u_3/u_2. We therefore rewrite the integral using these variables. We define v_1 and v_2 by

$$v_1 := \frac{u_1}{u_3}, \qquad v_2 := \frac{u_2}{u_3}.$$

We rewrite the integral I_m in terms of the variables v_1, v_2, u_3. When we change variables, the R factors depend on v_1, v_2 but not on u_3.

$$R\Big(\frac{u_1}{u_3}\Big)R\Big(\frac{u_2}{u_1}\Big)R\Big(\frac{u_3}{u_2}\Big) = R(v_1)R\Big(\frac{v_2}{v_1}\Big)R\Big(\frac{1}{v_2}\Big)$$

The exponential factor also works out in a nice way in the new variables:

$$e^{-N\mathbf{m}\cdot\mathbf{u}} = e^{-N(m_1u_1+m_2u_2+m_3u_3)} = e^{-N(m_1v_1+m_2v_2+m_3)u_3}.$$

And a Jacobian computation shows that

$$du_1du_2du_3 = u_3^2dv_1dv_2du_3.$$

So in the new variables, our integral I_m becomes

$$N^3\int_{\mathbb{R}^3} e(-N(m_1v_1+m_2v_2+m_3)u_3)w_1(\mathbf{u})R(v_1)R\Big(\frac{v_2}{v_1}\Big)R\Big(\frac{1}{v_2}\Big)u_3^2dv_1dv_2du_3.$$

Since the R factors don't involve u_3 we rewrite our formula to do the u_3 integral first:

$$N^3\int_{\mathbb{R}^2}\left(\int_{\mathbb{R}} e(-N(m_1v_1+m_2v_2+m_3)u_3)w_1(\mathbf{u})u_3^2du_3\right)R(v_1)R\Big(\frac{v_2}{v_1}\Big)R\Big(\frac{1}{v_2}\Big)dv_1dv_2.$$

A key observation in our proof is that we can analyze the norm of this inner integral very accurately using non-stationary phase. Recalling the definition (7.4) of w_1, we see that for any $j\in\mathbb{N}$, $w_1(\mathbf{u})u_3^2$ has j^{th} derivative with respect to u_3 bounded by $O_j(1)$ (since w is supported on $[1,2]$ with $\|w^{(\ell)}\|_\infty \lesssim_\ell 1$ for all $\ell\in\mathbb{N}$). Thus for any $\eta>0$, the inner integral is $O_\eta(T^{-300})$ unless $|m_1v_1+m_2v_2+m_3|\le T^\eta/N$ by repeated integration by parts. In general, the inner integral has size $\lesssim 1$. In addition, $w_1(\mathbf{u})$ vanishes unless $v_1,v_2\in[1/2,2]$, because

$$w_1(\mathbf{u}) = w(u_1)^2w(u_2)^2w(u_3)^2 = w\big(u_3v_1\big)^2w\big(u_3v_2\big)^2w(u_3)^2,$$

and $w(u)$ is supported on $u\in[1,2]$. Therefore, the inner integral vanishes unless $v_1\in[1/2,2]$ and $v_2\in[1/2,2]$. Using these bounds for the inner integral and the triangle inequality, we see that since $\eta>0$ was arbitrary

$$|I_m|\lessapprox N^3\int_{\substack{|m_1v_1+m_2v_2+m_3|\lessapprox\frac{1}{N}\\ v_1,v_2\in[1/2,2]}}\Big|R(v_1)R\Big(\frac{v_2}{v_1}\Big)R\Big(\frac{1}{v_2}\Big)\Big|dv_1dv_2+O(T^{-200}).$$

Since $|R(v)|=|R(1/v)|$, we can replace $R(1/v_2)$ by $R(v_2)$. (This is not really important, but it makes later computations cleaner.)

Finally, this integral vanishes unless we can find $v_1\asymp 1$ and $v_2\asymp 1$ so that $m_1v_1+m_2v_2+m_3$ is almost zero. If $|m_1|\le|m_2|\le|m_3|$, this can only happen if $|m_2|\asymp|m_3|$. This gives the last claim in the proposition. □

Remark. *The cancellation in the inner integral when* $|m_1v_1 + m_2v_2 + m_3|$ *is not* $\lessapprox 1/N$ *is one of the key observations in our proof. This cancellation is specific to Dirichlet polynomials as opposed to more general trigonometric polynomials. For instance, one may consider a 'generalized' Dirichlet polynomial of the form* $\widetilde{D}(t) = \sum_{n\sim N} b_n e^{it\phi(n)}$, *where the function* $\phi(n)$ *has smoothness and convexity properties similar to those of* $\log n$. *One can follow the argument above, but the inner integral will have the form* $\int e(Ng_{m,v_1,v_2}(u_3))du_3$ *for some function* $g_{m,v_1,v_2}(u_3)$. *In general, the function* $g_{m,v_1,v_2}(u_3)$ *will not be linear in* u_3. *So one would have to use stationary phase as opposed to non-stationary phase, and the bounds would not work out as they do here.*

Because of the last claim in Proposition 7.1, we can restrict attention to m with $0 < |m_1| \le |m_2| \asymp |m_3|$. The domain of integration can be rewritten in the form

$$\Big|v_2 - \frac{m_1v_1 + m_3}{-m_2}\Big| \lessapprox \frac{1}{|m_2|N} \asymp \frac{1}{|m_3|N}.$$

So the domain of integration is essentially the $\frac{1}{N|m_3|}$-neighborhood of the curve $v_2 = \frac{m_1v_1+m_3}{-m_2}$. Therefore, $|I_m|$ is morally bounded by

$$\frac{N^3}{N|m_3|}\int_{v_1\asymp 1}\Big|R(v_1)R\Big(\frac{m_1v_1+m_3}{-m_2v_1}\Big)R\Big(\frac{m_1v_1+m_3}{-m_2}\Big)\Big|dv_1.$$

We can make this rigorous by using a smoothed version of R. Suppose that $\widetilde{\psi}(x)$ is a smooth bump which is $\asymp 1$ for $|x| \lessapprox 1$ and supported in $|x| \lessapprox 1$ (i.e. satisfying $\|\widetilde{\psi}^{(j)}\|_\infty \lessapprox_j 1$ for all $j \in \mathbb{N}$). Define a smoothed version of $|R(u)|$ by

$$\widetilde{R}_M(u) := \Big(\int NM\widetilde{\psi}(NM(u-u'))\widetilde{\psi}(e^u)|R(u')|^2du'\Big)^{1/2}. \tag{7.5}$$

The following proposition gives an expansion of S_3 in terms of such integrals.

Proposition 7.2 (Expansion of S_3)**.** *There is a choice of* $0 < M_1 \le M \lessapprox T/N$ *such that*

$$S_3 \lessapprox \frac{N^2}{M}\sum_{|m_1|\sim M_1,|m_2|,|m_3|\sim M}\widetilde{I}_m + O(T^{-100}),$$

where

$$\widetilde{I}_m := \int_{v_1\asymp 1}\Big|R(v_1)\widetilde{R}_M\Big(\frac{m_1v_1+m_3}{m_2v_1}\Big)\widetilde{R}_M\Big(\frac{m_1v_1+m_3}{m_2}\Big)\Big|dv_1.$$

Proof. Recall from (7.1) that S_3 is bounded by

$$S_3 \leq \sum_{0<|m_1|,|m_2|,|m_3| \lessapprox T/N} |I_m| + O(T^{-100}).$$

From (7.3), we have

$$I_m = N^3 \int_{\mathbb{R}^3} e(-N\mathbf{m} \cdot \mathbf{u}) w_1(\mathbf{u}) R\Big(\frac{u_1}{u_3}\Big) R\Big(\frac{u_2}{u_1}\Big) R\Big(\frac{u_3}{u_2}\Big) d\mathbf{u}.$$

From the definition (7.2) for $R(v)$, we see that $R(1/v) = \overline{R}(v)$. Therefore we see that $I_{(m_1,m_2,m_3)} = \overline{I}_{(m_2,m_1,m_3)}$, and similarly for any other transposition of (m_1, m_2, m_3). Thus $|I_m|$ is invariant under any permutation of (m_1, m_2, m_3), and so we can reduce to the case $|m_1| \leq |m_2| \leq |m_3|$ at the cost of a factor of 6. By Proposition 7.1, such terms are negligible unless $|m_2| \asymp |m_3|$. Thus, by choosing dyadic scales to maximize the right hand side, we find that there is an $M_1 \leq M \lessapprox T/N$ such that

$$S_3 \lessapprox \sum_{\substack{|m_1|\sim M_1 \\ |m_2|\sim M \\ |m_3|\sim M}} |I_m| + O(T^{-100}). \tag{7.6}$$

By Proposition 7.1, we have for $m_2 \asymp M$

$$|I_m| \lesssim N^3 \int_{v_1 \asymp 1} |R(v_1)| \Bigg(\int_{\left|v_2 - \frac{m_1 v_1 + m_3}{-m_2}\right| \lessapprox \frac{1}{MN}} \Big| R\Big(\frac{v_2}{v_1}\Big) R(v_2) \Big| dv_2 \Bigg) dv_1 + O(T^{-200}).$$

Using Cauchy-Schwarz, we bound the inner integral by

$$\Bigg(\int_{\left|v_2 - \frac{m_1 v_1 + m_3}{-m_2}\right| \lessapprox \frac{1}{MN}} \Big| R\Big(\frac{v_2}{v_1}\Big) \Big|^2 dv_2 \Bigg)^{1/2} \Bigg(\int_{\left|v_2 - \frac{m_1 v_1 + m_3}{-m_2}\right| \lessapprox \frac{1}{MN}} \Big| R(v_2) \Big|^2 dv_2 \Bigg)^{1/2}$$

$$\lesssim \frac{1}{MN} \widetilde{R}_M \left(\frac{m_1 v_1 + m_3}{-m_2 v_1} \right) \widetilde{R}_M \Big(\frac{m_1 v_1 + m_3}{-m_2} \Big).$$

Thus we find that

$$|I_m| \lesssim \frac{N^2}{M} \widetilde{I}_m + O(T^{-100}).$$

Finally, since we are summing over m_2 with $|m_2| \sim M$, we can replace $-m_2$ with m_2 without changing the overall sum. Substituting this into our expression (7.6) for S_3 above then gives the result. □

8. Basic estimate for the low energy case

In this section, we begin to estimate S_3 using Proposition 7.2. Recall that $R(v) = \sum_{t\in W} v^{it}$. The best bound for $|R(v)|$ we can hope for is square root cancellation: $|R(v)| \lesssim |W|^{1/2}$. If indeed $|R(v)| \approx |W|^{1/2}$ for all $v \asymp 1$, then we get $S_3 \lessapprox N^2M^2|W|^{3/2} \lessapprox T^2|W|^{3/2}$. We will see more generally that this bound holds whenever the energy of W is very small.

Proposition 8.1 (S_3 controlled by energy)**.** *If W is a T^ϵ-separated set contained in an interval of length T, then*

$$S_3 \lessapprox T^2|W|^{1/2}E(W)^{1/2}.$$

Remark. *If we look at the critical case when $T = N^{5/4}$ and $\sigma = 3/4$, then this estimate (together with our bounds for S_2) gives an improvement to the basic orthogonality estimate (1.6) when $E(W)$ is significantly below $|W|^{7/3}$. In the special case when $E(W) \approx |W|^2$, this Proposition is enough to prove our main theorem, Theorem 1.1. On the other hand, when $E(W)$ is very large, then we will get good estimates using Theorem 1.6. However there is an intermediate range of energy that is not yet covered. In the next two sections, we will develop a strengthening of this proposition that gives new estimates for a wider range of energies.*

We begin with some basic lemmas about the moments of R.

Lemma 8.2 (L^2 bound)**.** *Let W be a T^ϵ-separated set contained in an interval of length T. Then*

$$\int_{v\asymp 1} |R(v)|^2 dv \lesssim_\epsilon |W|.$$

Proof. Let $\psi_1(v)$ be a smooth bump function which majorizes the range of integration of the integral in the lemma and is supported on $v \asymp 1$ (so satisfies $\|\psi_1^{(j)}\|_\infty \lesssim_j 1$ for all $j \in \mathbb{N}$). Then we have

$$\int_{v\asymp 1} |R(v)|^2 dv \le \int \psi_1(v)|R(v)|^2 dv.$$

Recall from (7.2) that

$$R(v) = \sum_{t\in W} v^{it} = \hat{W}(\log v),$$

and let $\psi_2(\tau) := e^\tau \psi_1(e^\tau)$. Then, making a change of variables $v = e^\tau$

$$\int \psi_1(v)|R(v)|^2 dv = \int \psi_2(\tau)|\hat{W}(\tau)|^2 d\tau = \sum_{t_1,t_2\in W} \hat{\psi}_2(t_1 - t_2).$$

Note that ψ_2 is a smooth bump around the origin with $\|\psi_2^{(j)}\|_\infty \lesssim_j 1$ for all $j \in \mathbb{N}$, so $|\hat{\psi}_2(\xi)| \lesssim_j |\xi|^{-j}$ for any $j \in \mathbb{N}$. Thus the terms with $t_1 = t_2$ contribute $\lesssim |W|$

to the sum above. Since W is T^ϵ-separated, if $t_1 \neq t_2$ we have that $\hat{\psi}_2(t_1 - t_2) \lesssim_\epsilon T^{-100}$, so the terms with $t_1 \neq t_2$ are negligible. Thus the total sum is $O_\epsilon(|W|)$, as required. □

Lemma 8.3 (L^4 bound)**.** *For any M we have*

$$\int_{v \asymp 1} |\widetilde{R}_M(v)|^4 dv \lessapprox E(W) \qquad \textit{and} \qquad \int_{v \asymp 1} |R(v)|^4 dv \lessapprox E(W).$$

Proof. From the definition (7.5) of $\widetilde{R}$ and Cauchy-Schwarz, we have

$$\int_{v \asymp 1} |\widetilde{R}_M(v)|^4 dv \lessapprox N^2 M^2 \int\limits_{\substack{v \asymp 1 \\ |u-v| \lessapprox 1/NM \\ |u'-v| \lessapprox 1/NM}} |R(u')|^4 du' du dv \lessapprox \int_{u' \asymp 1} |R(u')|^4 du'.$$

Therefore it suffices to prove the result for R. We recall that $R(v) = \hat{W}(\log v)$, so by a change of variables $\tau = e^v$ we see that it suffices to show

$$\int_{\tau \asymp 1} |\hat{W}(\tau)|^4 d\tau \lessapprox E(W).$$

Let $\eta > 0$ and let ψ be a smooth bump supported on $\tau \lesssim 1$ such that $\psi(\tau/T^\eta)$ majorizes the range of integration. Then we see that

$$\begin{aligned}
\int_{\tau \asymp 1} |\hat{W}(\tau)|^4 d\tau &\leq \int \psi\Big(\frac{\tau}{T^\eta}\Big) |\hat{W}(\tau)|^4 d\tau \\
&= \sum_{t_1, t_2, t_3, t_4 \in W} \int \psi\Big(\frac{\tau}{T^\eta}\Big) e(\tau(t_1 + t_2 - t_3 - t_4)) d\tau \\
&= T^\eta \sum_{t_1, t_2, t_3, t_4 \in W} \hat{\psi}\Big(T^\eta (t_3 + t_4 - t_1 - t_2)\Big).
\end{aligned}$$

Since $\hat{\psi}$ decays rapidly, we may restrict the summation to $|t_1 + t_2 - t_3 - t_4| \leq 1$ at the cost of an $O_\eta(T^{-100})$ error term. The remaining terms contribute $\lesssim T^\eta E(W)$. Thus, letting $\eta \to 0$ we obtain

$$\int_{\tau \asymp 1} |\hat{W}(\tau)|^4 d\tau \lessapprox E(W).$$ □

We can now quickly prove Proposition 8.1.

Proof of Proposition 8.1. Starting with Proposition 7.2, we have for some $M_1 \leq M \lessapprox T/N$

$$S_3 \lessapprox \sum_{\substack{|m_1|\sim M_1\\ |m_2|,|m_3|\sim M}} \frac{N^2}{M}\int_{v_1\asymp 1}\Big|R(v_1)\widetilde{R}_M\Big(\frac{m_1v_1+m_3}{m_2v_1}\Big)\widetilde{R}_M\Big(\frac{m_1v_1+m_3}{m_2}\Big)\Big|dv_1.$$

Using Hölder's inequality, we find that the integral over v_1 is bounded by

$$\Big(\int_{v_1\asymp 1}|R(v_1)|^2dv_1\Big)^{1/2}\Big(\int_{v_1\asymp 1}\Big|\widetilde{R}_M\Big(\frac{m_1v_1+m_3}{m_2v_1}\Big)\Big|^4dv_1\Big)^{1/4} \times\Big(\int_{v_1\asymp 1}\Big|\widetilde{R}_M\Big(\frac{m_1v_1+m_3}{m_2}\Big)\Big|^4dv_1\Big)^{1/4}.$$

In the second integral we do a change of variables $u=\frac{m_1v_1+m_3}{m_2v_1}$ with Jacobian factor $\asymp 1$. In the third integral we do a change of variables $u=\frac{m_1v_1+m_3}{m_2}$ with a Jacobian factor of norm $\sim M/M_1$. Therefore we obtain

$$S_3 \lesssim N^2M^2\Big(\int_{v_1\asymp 1}|R(v_1)|^2dv_1\Big)^{1/2}\Big(\int_{u\asymp 1}\Big|\widetilde{R}_M(u)\Big|^4du\Big)^{1/2}$$

Using $M\lessapprox T/N$ and Lemmas 8.2 and 8.3, we find

$$S_3\lessapprox T^2|W|^{1/2}E(W)^{1/2}. \qquad \square$$

When $E(W)\approx |W|^2$, the bound from Proposition 8.1 is the best bound for S_3 we know how to prove. But for larger $E(W)$, we can improve the bound for S_3. Let us indicate the general direction here, and then we will develop the tool we need in the next section.

Ignoring some technical smoothing, we morally have

$$S_3 \lessapprox \sum_{|m_1|\sim M_1,|m_2|,|m_3|\sim M}\frac{N^2}{M}\int_{v_1\asymp 1}\Big|R(v_1)R\Big(\frac{m_1v_1+m_3}{m_2v_1}\Big)R\Big(\frac{m_1v_1+m_3}{m_2}\Big)\Big|dv_1.$$

We can split up the domain of integration into pieces where the first R factor has size $\sim A_1$, the second R factor has size $\sim A_2$, and the third R factor has size $\sim A_3$. For simplicity, suppose that $A_1=A_2=A_3=A$, which we expect to be the critical case, and focus on the value of A that dominates the integral. If $A\approx |W|^{1/2}$, then we get the bound corresponding to minimal energy. If A is larger, then $|R(v)|\sim A$ for only a small subset $U_A\subset\{v\asymp 1\}$ by Lemma 8.2. If the simple analysis in Proposition 8.1 was sharp, it would mean that for most $v\in U_A$ and most m_1,m_2,m_3, we have $\frac{m_1v+m_3}{m_2}\in U_A$. We will see that a small set U cannot be approximately invariant under this large set of affine transformations. In the next section, we will prove a precise estimate in this spirit, and then we will use it to give stronger bounds for S_3 when the energy is greater than $|W|^2$.

9. Summing over affine transformations

Given $M > 0$ and a compactly supported smooth function f, we define

$$J(f) := \sup_{0<M_1,M_2,M_3<M} \int \Big(\sum_{|m_1|\sim M_1, m_2\sim M_2, |m_3|\lesssim M_3} f\Big(\frac{m_1 u + m_3}{m_2}\Big)\Big)^2 du,$$

which is an average of sums of affine transformations of f. The aim of this section is to establish the following general bound for $J(f)$.

Proposition 9.1 (Equidistribution over affine transformations)**.** *Suppose that $f(u)$ is non-negative and supported on $u \asymp 1$ and that $|\hat{f}(\xi)| \lessapprox_j (|\xi|/T)^j$ for all $j \in \mathbb{N}$. Then*

$$J(f) \lessapprox M^6 \Big(\int f(u) du \Big)^2 + M^4 \int f(u)^2 du.$$

Remark. *By a simple Cauchy-Schwarz argument, we can bound the left hand side by $M^6 \int |f(u)|^2 du$. This bound is tight if $f(u)$ is a smooth bump on $u \sim 1$. But the Proposition improves on this bound when $f(u)$ is sparse. To get a sense of what it means, it's good to imagine the example that $f = \chi_U$ where U is a union of (smoothed) $1/T$-intervals. The proposition says that if $|U|$ is much smaller than 1, then the sets $\{\frac{m_1 U + m_3}{m_2}\}_{m_1,m_2,m_3\sim M}$ cannot overlap too much.*

There are two key examples which correspond to the two terms on the right-hand side. If U is a random subset, then the left-hand side is comparable to $M^6 \left(\int f(u)du\right)^2$. If U is the $1/T$ neighborhood of the set of rational numbers p/q with $p \sim q \sim B$ with the denominators having a divisor of size $\sim M$, then the left-hand side is comparable to $M^4 \int |f(u)|^2 du$ (the subsum where m_1 divides the denominator of the closest such rational to u gives roughly this contribution).

The following lemma is the main technical result used to prove Proposition 9.1, which is based on a fairly long Fourier analytic argument.

Lemma 9.2 (Iterative bound for $J(f)$)**.** *Let f be as in Proposition 9.1. Then there is a bump function $\psi(x)$ supported on $|x| \lessapprox 1$ such that*

$$J(f) \lessapprox M^6 \left(\int f(u) du \right)^2 + \left(M^4 \int f(u)^2 du \right)^{1/2} J(\widetilde{f})^{1/2}.$$

where $\widetilde{f}$ is defined in terms of $\psi(x)$ by

$$\widetilde{f}(u) := \int T\psi(T(u-u'))f(u')du'$$

Before we prove Lemma 9.2, we first show how to deduce Proposition 9.1 from it.

Proof of Proposition 9.1 assuming Lemma 9.2. We wish to show that for any $\epsilon > 0$ there is a $C(\epsilon) > 0$ such that

$$J(f) \le C(\epsilon)T^{\epsilon}\Big(M^6\Big(\int f(u)du\Big)^2 + M^4\int f(u)^2du\Big). \tag{9.1}$$

We wish to prove this by downwards induction on ϵ. As a base case, the result clearly holds for $\epsilon = 100$. By induction, we can assume that (9.1) holds with $3\epsilon/2$ in place of ϵ and seek to establish (9.1). We first apply Lemma 9.2 to give

$$J(f) \lessapprox M^6\Big(\int f(u)du\Big)^2 + \Big(M^4\int f(u)^2du\Big)^{1/2} J(\widetilde{f})^{1/2}.$$

Since $f(u)$ is supported on $u \asymp 1$ and $\psi(x)$ is supported on $x \lessapprox 1$, we see that $\widetilde{f}(u)$ is also supported on $u \asymp 1$. Thus we may apply the induction hypothesis to $J(\widetilde{f})$, which shows

$$J(\widetilde{f}) \lesssim_{\epsilon} T^{3\epsilon/2}\Big(M^6\Big(\int \widetilde{f}(u)du\Big)^2 + M^4\int \widetilde{f}(u)^2du\Big).$$

Since $\widetilde{f}$ is a smoothed version of f, we can bound $\int \widetilde{f}(u)du \lessapprox \int f(u)du$ and $\int \widetilde{f}(u)^2du \lessapprox \int f(u)^2du$. Substituting these bounds back into our bound for $J(f)$, we find

$$J(f) \lessapprox_{\epsilon} T^{3\epsilon/4}\left(M^6\Big(\int f(u)du\Big)^2 + M^4\int f(u)^2du\right).$$

Thus (9.1) holds for $C(\epsilon)$ sufficiently large, which completes the induction. □

We now return to the proof of Lemma 9.2.

Proof of Lemma 9.2. The most interesting situation is when $M_1 = M_2 = M_3 = M$. We encourage the reader to keep this case in mind on first reading.

We let $\psi_1(x)$ be a smooth bump supported on $|x| \lesssim 1$ so that $\psi_1(m_3/M_3)$ majorizes the summation condition $m_3 \lesssim M_3$. Thus we can bound the inner sum in $J(f)$ by

$$g(u) := \sum_{|m_1|\sim M_1, m_2\sim M_2}\sum_{m_3}\psi_1\Big(\frac{m_3}{M_3}\Big)f\Big(\frac{m_1u+m_3}{m_2}\Big).$$

Squaring and integrating over u, and then applying Plancherel gives (for the choice of M_1, M_2, M_3 achieving the supremum)

$$J(f) \leq \int |g(u)|^2 du = \int |\hat{g}(\xi)|^2 d\xi. \tag{9.2}$$

We wish to estimate $\hat{g}(\xi)$. We have

$$\hat{g}(\xi) = \sum_{|m_1|\sim M_1, m_2\sim M_2} \int \sum_{m_3} \psi_1\Big(\frac{m_3}{M_3}\Big) f\Big(\frac{m_1 u + m_3}{m_2}\Big) e(-\xi u) du.$$

We do a change of variables: $\widetilde{u} = u + \frac{m_3}{m_1}$, so that $f(\frac{m_1 u + m_3}{m_2}) = f(\frac{m_1 \widetilde{u}}{m_2})$. In the new variables, we get

$$\hat{g}(\xi) = \sum_{|m_1|\sim M_1, m_2\sim M_2} \left(\int f\Big(\frac{m_1 \widetilde{u}}{m_2}\Big) e(-\xi \widetilde{u}) d\widetilde{u} \right) \left(\sum_{m_3 \in \mathbb{Z}} \psi_1\Big(\frac{m_3}{M_3}\Big) e\Big(\frac{m_3}{m_1}\xi\Big) \right).$$

The integral in parentheses is $\frac{m_2}{m_1}\hat{f}(\frac{m_2}{m_1}\xi)$. The first key point in our analysis is that we can explicitly do the last sum by Poisson summation. It is equal to $M_3 \sum_\ell \hat{\psi}_1(M_3\ell - \frac{\xi}{m_1}))$. So all together we have

$$\hat{g}(\xi) = \sum_{|m_1|\sim M_1} M_3 \sum_{\ell} \hat{\psi}\Big(M_3\Big(\frac{\xi}{m_1} - \ell\Big)\Big) \sum_{m_2\sim M_2} \frac{m_2}{m_1} \hat{f}\Big(\frac{m_2}{m_1}\xi\Big).$$

Since $\hat{\psi}$ is rapidly decaying, $\hat{\psi}(M_3(\frac{\xi}{m_1} - \ell))$ is negligible unless $|\xi - \ell m_1| \lessapprox \frac{M_1}{M_3}$. Therefore, we have

$$|\hat{g}(\xi)| \leq \sum_{|m_1|\sim M_1} \sum_{\ell : |\xi - m_1 \ell| \lessapprox \frac{M_1}{M_3}} M_3 \Big| \sum_{m_2\sim M_2} \frac{m_2}{m_1} \hat{f}\Big(\frac{m_2}{m_1}\xi\Big)\Big| + O(T^{-100}). \tag{9.3}$$

We have to estimate $\int_{\mathbb{R}} |\hat{g}(\xi)|^2 d\xi$. For a small $\eta > 0$, we break up the domain of integration into the region $|\xi| \leq T^\eta M_1/M_3$, the region $T^\eta M_1/M_3 < |\xi| \leq T^2$ and the remainder:

$$\int_{\mathbb{R}} |\hat{g}(\xi)|^2 d\xi = \underbrace{\int_{|\xi|\leq T^\eta \frac{M_1}{M_3}} |\hat{g}(\xi)|^2 d\xi}_{I} + \underbrace{\int_{T^\eta \frac{M_1}{M_3} < |\xi| \leq T^2} |\hat{g}(\xi)|^2 d\xi}_{II} + \underbrace{\int_{|\xi|\geq T^2} |\hat{g}(\xi)|^2 d\xi}_{III}. \tag{9.4}$$

If $|\xi| > T^2$, then since $\hat{f}(\xi)$ is rapidly decaying for $|\xi| > T$, we see that $\hat{f}(m_2\xi/m_1)$ is negligible and $|\hat{g}(\xi)| \lesssim T^{-100}|\xi|^{-2}$. Thus

$$III = O(T^{-100}). \tag{9.5}$$

If $|\xi| \le T^\eta M_1/M_3$, then in the sum over ℓ in (9.3), the only terms that contribute have $|\ell| \lesssim T^\eta/M_3 \lesssim T^\eta$. Therefore, this sum is $O(T^\eta)$. Thus, by Cauchy-Schwarz we obtain

$$|\hat{g}(\xi)|^2 \lesssim T^{2\eta} M_1 \sum_{|m_1|\sim M_1} M_3^2 \Big| \sum_{m_2 \sim M_2} \frac{m_2}{m_1} \hat{f}\Big(\frac{m_2}{m_1}\xi\Big)\Big|^2 \le T^{2\eta} M_2^4 M_3^2 \sup_\xi |\hat{f}(\xi)|^2.$$

Therefore

$$\begin{aligned} I = \int_{|\xi|\le T^\eta M_1/M_3} |\hat{g}(\xi)|^2 d\xi &\lesssim T^{3\eta} M_1 M_2^4 M_3 \sup_\xi |\hat{f}(\xi)|^2 \\ &\lesssim T^{3\eta} M^6 \Big(\int f(u) du\Big)^2. \end{aligned} \tag{9.6}$$

Now suppose that $T^\eta M_1/M_3 < |\xi| \le T^2$. We return to (9.3) and consider the number of terms in the outer double sum. Let $s = m_1 \ell$. In this range, s must be a non-zero integer in the $\lessapprox M_1/M_3$ neighborhood of ξ as soon as T is sufficiently large in terms of η. The number of such integers s is $\lessapprox 1 + M_1/M_3$. Since $|\xi| \le T^2$ and s is non-zero, each such integer s has $\lessapprox 1$ factorizations as $s = m_1\ell$. All together the number of terms in the outer double sum is $\lessapprox 1 + M_1/M_3$. Therefore, we can use Cauchy-Schwarz to bound $|\hat{g}(\xi)|^2$ by

$$\lessapprox \Big(1 + \frac{M_1}{M_3}\Big) \sum_{|m_1|\sim M_1} \sum_{\substack{\ell \\ |\xi - \ell m_1| \lessapprox M_1/M_3}} M_3^2 \Big| \sum_{m_2\sim M_2} \frac{m_2}{m_1} \hat{f}\Big(\frac{m_2}{m_1}\xi\Big)\Big|^2 + O(T^{-200}).$$

Therefore the term II satisfies

$$II \lessapprox (M_1 M_3 + M_3^2) \sum_{|m_1|\sim M_1} \sum_\ell \int_{|\xi - \ell m_1| \lessapprox M_1/M_3} \Big| \sum_{m_2\sim M_2} \frac{m_2}{m_1} \hat{f}\Big(\frac{m_2}{m_1}\xi\Big)\Big|^2 d\xi.$$

Morally this integral does not depend on m_1, and we can make this precise by changing variables. For each m_1, ℓ, we write $\xi = \ell m_1 + \frac{m_1}{M_3}\tau$ and do a change of variables to get (extending the range of integration slightly for an upper bound so we have a range independent of m_1)

$$II \lesssim (M_1 + M_3) \sum_\ell \int_{|\tau|\lessapprox 1} \Big| \sum_{m_2\sim M_2} m_2 \hat{f}\Big(\ell m_2 + \frac{m_2}{M_3}\tau\Big)\Big|^2 d\tau.$$

Since $\hat{f}(\xi) = O(T^{-200})$ unless $|\xi| \lessapprox T$, we can restrict the sum over ℓ to the range $|\ell| \lessapprox T/M_2$ at the cost of a negligible error. We introduce a bump $\psi_2(x)$ supported on $|x| \lessapprox 1$ so that $\psi_2(M_2\ell/T)$ majorizes this summation condition, and bound the last sum by

$$\lesssim (M_1 + M_3)\Sigma_{II} + O(T^{-100}),$$

where

$$\Sigma_{II} := \sum_{\ell} \psi_2\Big(\frac{M_2\ell}{T}\Big) \int_{|\tau| \lessapprox 1} \Big| \sum_{m_2 \sim M_2} m_2 \hat{f}\Big(\ell m_2 + \frac{m_2}{M_1}\tau\Big)\Big|^2 d\tau.$$

We write out $\hat{f}$ as an integral, expand out the square and bring the summation over ℓ and integration over τ on the inside to get

$$\Sigma_{II} = \int\int \sum_{m_2, m_2' \sim M_2} m_2 m_2' f(u) f(u') \Sigma_1 \Sigma_2 du' du,$$

where

$$\Sigma_1 := \int_{|\tau| \lessapprox 1} e\Big(\tau\Big(\frac{m_2'}{M_1}u' - \frac{m_2}{M_1}u\Big)\Big) d\tau,$$
$$\Sigma_2 := \sum_{\ell} \psi_2\Big(\frac{M_2\ell}{T}\Big) e\Big(\ell\Big(m_2'u' - m_2 u\Big)\Big).$$

Trivially we have $|\Sigma_1| \lessapprox 1$. By Poisson summation, and the rapid decay of $\hat{\psi}_2$, we have

$$\begin{aligned}\Sigma_2 &= \frac{T}{M_2} \sum_j \hat{\psi}_2\Big(\frac{j - m_2'u' + m_2 u}{M_2/T}\Big) \\ &= \frac{T}{M_2} \sum_{\substack{j \\ |j - m_2'u' + m_2 u| \lessapprox M_2/T}} \hat{\psi}_2\Big(\frac{j - m_2'u' + m_2 u}{M_2/T}\Big) + O(T^{-100}).\end{aligned}$$

Substituting these back into our bound (9.7) for Σ_{II}, we find

$$\Sigma_{II} \lessapprox M_2 \int f(u) \sum_{m_2, m_2' \sim M_2} \sum_j T \int_{|u' - \frac{m_2 u + j}{m_2'}| \lessapprox \frac{1}{T}} f(u') du' du.$$

This gives

$$II \lessapprox M_2(M_1 + M_3) \int f(u) \sum_{m_2, m_2' \sim M_2} \sum_j T \int_{|u' - \frac{m_2 u + j}{m_2'}| \lessapprox \frac{1}{T}} f(u') du' du.$$

Since $\hat{f}(\xi)$ rapidly decays for $|\xi| > T$, f is morally almost constant on intervals of length $1/T$. We let $\psi(x)$ be a smooth bump function supported on $x \lessapprox 1$, and then define $\widetilde{f}$ to be a slightly smoothed version of f given by

$$\widetilde{f}(u) := \int T\psi(T(u-u'))f(u')du'$$

We can choose ψ such that the integral over u' in the bound for II above is bounded by $\widetilde{f}(\frac{m_2u+j}{m_2'})$. Thus we find (recalling that f is supported on $u \asymp 1$ and $M_1, M_2, M_3 \le M$)

$$II \lessapprox M^2 \int_{u\asymp 1} f(u) \sum_{m_2,m_2'\sim M_2} \sum_j \widetilde{f}\Big(\frac{m_2u+j}{m_2'}\Big)du.$$

Now we apply Cauchy-Schwarz to get

$$II \lessapprox \Big[M^4 \int f(u)^2 du\Big]^{1/2} \Big[\int_{u\asymp 1}\Big(\sum_{m_2,m_2'\sim M_2, j\in\mathbb{Z}} \widetilde{f}\Big(\frac{m_2u+j}{m_2'}\Big)\Big)^2 du\Big]^{1/2}. \tag{9.7}$$

Since $u \asymp 1$ and $f(x)$ is supported on $x \asymp 1$, we may restrict the summation over j to $j \lesssim M_2$. Thus we see that the second term in square brackets is bounded by $J(\widetilde{f})$. Putting together (9.2), (9.4), (9.5), (9.6) and (9.7) we find that for any $\eta > 0$, provided T is sufficiently large in terms of η, we have

$$J(f) \lessapprox T^{3\eta} M^6 \Big(\int f(u)du\Big)^2 + \Big(M^4 \int f(u)^2 du\Big)^{1/2} J(\widetilde{f})^{1/2}.$$

This now gives the result on letting $\eta \to 0$ sufficiently slowly with T. □

10. Further bounds for S_3

In this section, we use our bounds for sums over affine transformations to improve our bound for S_3. We will get the following estimate.

Proposition 10.1 (Refined S_3 bound)**.** *If W is a T^ϵ-separated set contained in an interval of length T, then*

$$S_3 \lessapprox T^2|W|^{3/2} + TN|W|^{1/2}E(W)^{1/2}. \tag{10.1}$$

Remark. *Comparing with Proposition 8.1, when $E(W) \approx |W|^2$ both propositions give $S_3 \lessapprox T^2|W|^{3/2}$. But when $E(W)$ is large, Proposition 10.1 is better by a factor of N/T. If we look at the critical case when $T = N^{5/4}$ and $\sigma = 3/4$, then this estimate (together with our bounds for S_2) gives an improvement to the basic orthogonality estimate (1.6) when $E(W)$ is significantly below $|W|^3$. In other words, this Proposition gives an improvement unless the energy is essentially maximal. But* *when $E(W)$ is very large, then we will get good estimates in the next section using Theorem 1.6.*

Proof. Recall from Proposition 7.2 that we have

$$S_3 \lessapprox \frac{N^2}{M} \int_{v_1 \asymp 1} |R(v_1)| \sum_{\substack{|m_1| \sim M_1 \\ |m_2|, |m_3| \sim M}} \left| \widetilde{R}_M\Big(\frac{m_1 v_1 + m_3}{m_2 v_1}\Big) \widetilde{R}_M\Big(\frac{m_1 v_1 + m_3}{m_2}\Big) \right| dv_1.$$

Thus by Cauchy-Schwarz we have

$$S_3 \lesssim \frac{N^2}{M} S_{3,1}^{1/2} S_{3,2}^{1/2},$$

where

$$\begin{aligned} S_{3,1} &:= \int_{v \asymp 1} |R(v)|^2 dv \lesssim W, \\ S_{3,2} &:= \int_{v \asymp 1} \Big(\sum_{|m_1| \sim M_1, |m_2|, |m_3| \sim M} \Big| \widetilde{R}\Big(\frac{m_3 + m_1 v}{m_2 v}\Big) \widetilde{R}\Big(\frac{m_3 + m_1 v}{m_2}\Big) \Big| \Big)^2 dv. \end{aligned}$$

By Cauchy-Schwarz again, we have that

$$S_{3,2} \lesssim S_{3,3}^{1/2} S_{3,4}^{1/2},$$

where

$$\begin{aligned} S_{3,3} &:= \int_{v \asymp 1} \Big(\sum_{|m_1| \sim M_1, |m_2|, |m_3| \sim M} \Big| \widetilde{R}\Big(\frac{m_3 + m_1 v}{m_2 v}\Big) \Big|^2 \Big)^2 dv, \\ S_{3,4} &:= \int_{v \asymp 1} \Big(\sum_{|m_1| \sim M_1, |m_2|, |m_3| \sim M} \Big| \widetilde{R}\Big(\frac{m_3 + m_1 v}{m_2}\Big) \Big|^2 \Big)^2 dv. \end{aligned}$$

We bound $S_{3,3}$ and $S_{3,4}$ using Proposition 9.1. To bound $S_{3,4}$, we use $f(v) = \psi(v) \widetilde{R}(v)^2$, where $\psi(v)$ is a smooth bump supported on $v \asymp 1$. To control $S_{3,3}$, we make a change of variables $u = 1/v$ and rewrite $S_{3,3}$ as

$$S_{3,3} \lesssim \int_{u \asymp 1} \Big(\sum_{|m_1| \sim M_1, |m_2|, |m_3| \sim M} \Big| \widetilde{R}\Big(\frac{m_3 u + m_1}{m_2}\Big) \Big|^2 \Big)^2 du.$$

Then we use Proposition 9.1 with $f(u) = \psi(u) |\widetilde{R}(u)|^2$. To apply Proposition 9.1, we have to check that $\hat{f}(\xi)$ is rapidly decaying for $|\xi| > T$. Both $R(v)$ and $\widetilde{R}(v)$

have the desired Fourier decay. Let's check for $\widetilde{R}(v)$. Recall that $\widetilde{\psi}(u)$ is a smooth bump supported on $|u| \lessapprox 1$ and that

$$\widetilde{R}(v)^2 = \int MN\widetilde{\psi}(MN(u-u'))|R(u)|^2 du.$$

Therefore, $\widetilde{R}(v)^2$ has Fourier transform rapidly decaying when $|\xi| > MN$, and $T \gtrapprox MN$. Then it follows that f has the desired Fourier decay. Proposition 9.1 then gives the bounds

$$S_{3,3}, S_{3,4} \lessapprox M^6\Big(\int_{v\asymp 1} |R(v)|^2 dv\Big)^2 + M^4 \int_{v\asymp 1} |R(v)|^4 dv.$$

Applying Lemmas 8.2 and 8.3, we get

$$S_{3,3}, S_{3,4} \lessapprox M^6|W|^2 + M^4 E(W).$$

The same bound holds for $S_{3,2}$ since $S_{3,2} \leq S_{3,3}^{1/2} S_{3,4}^{1/2}$. Then we get

$$\begin{aligned} S_3 \lessapprox \frac{N^2}{M} S_{3,1}^{1/2} S_{3,2}^{1/2} &\lessapprox \frac{N^2}{M} |W|^{1/2} (M^6|W|^2 + M^4 E(W))^{1/2} \\ &= N^2 M^2 |W|^{3/2} + N^2 M |W|^{1/2} E(W)^{1/2}. \end{aligned}$$

Since $M \lessapprox T/N$, we get

$$S_3 \lessapprox T^2|W|^{3/2} + TN|W|^{1/2}E(W)^{1/2}. \qquad \square$$

11. Energy Bound

In this section, we prove bounds related to the energy of W, which show that a Dirichlet polynomial cannot be too large on a set of large energy. These bounds ultimately rely on Heath-Brown's bound Theorem 1.6, and are closely related to (and refine) arguments in [HB2]. Recall that in equation (1.7), we defined the energy of a finite set $W \subset \mathbb{R}$ by

$$E(W) := \#\{t_1, t_2, t_3, t_4 \in W : |t_1 + t_2 - t_3 - t_4| < 1\}.$$

We will prove two bounds about the behavior of Dirichlet polynomials on sets of high energy. The first bound is Lemma 1.7. We recall the statement here.

Lemma. *Let $N \in [T^{2/3}, T]$, $\sigma > 1/2$ and $D(t) = \sum_{n=N}^{2N} b_n n^{it}$ with $|b_n| \leq 1$. Suppose $W \subset [0,T]$ is a 1-separated set such that $|D(t)| > N^\sigma$ for $t \in W$. Then*

$$E(W) \leq |W|^3 N^{1-2\sigma+o(1)} + |W|^2 N^{2-2\sigma+o(1)}.$$

Combining Lemma 1.7 with our previous results is enough to get an improvement on previous estimates in the key scenario $T = N^{4/5}$, $|W| = T^{3/5}$. The second bound is a little more complicated, but it leads to stronger estimates in our applications.

Proposition 11.1 (Bound for energy)**.** *Suppose that $D_N(t) = \sum_{n\sim N} b_n n^{it}$ with $|b_n| \leq 1$. Suppose that W is a 1-separated set contained in an interval of length T, and that $|D_N(t)| \geq N^\sigma$ for $t \in W$. If $T^{3/4} \leq N \leq T$, then*

$$E(W) \lessapprox |W|N^{4-4\sigma} + |W|^{21/8}T^{1/4}N^{1-2\sigma} + |W|^3N^{1-2\sigma}. \tag{11.1}$$

Since the algebra is a little messy, we take a moment to process the bounds. For one thing, if $E(W)$ is very large, we get very strong bounds on σ. For instance, if $E(W) \approx |W|^3$, and if $N^{2/3} \leq T \leq N$, then Lemma 1.7 gives a sharp estimate: either $N^\sigma \lessapprox N^{1/2}$ or $|W|N^{2\sigma} \lessapprox N^2$. More important for our application is when we get an improvement on the basic orthogonality bound $|W|N^{2\sigma} \lesssim TN$. The full equations are a little messy, but if we plug in the key scenario $N = T^{4/5}$ and $|W| = T^{3/5}$, then Lemma 1.7 gives an improvement when $E(W) \geq |W|^{\frac{8}{3}+\epsilon}$ and Proposition 11.1 gives an improvement when $E(W) \geq |W|^{\frac{7}{3}}$ (for some $\epsilon > 0$).

If $|W| \approx TN^{1-2\sigma}$, then the bound in Proposition 11.1 would be $(N/T)^2|W|^3 + (|W|^{5/8}T^{-6/8})|W|^3 + |W|^4/T$. The first term will be the most important for us, and generally sets the limitations on our bounds. The second term will be negligible in practice (and could be improved with a bit more effort). The final term corresponds to the additive energy of a random set in an interval of length T.

An immediate consequence of Proposition 11.1 is a good bound for the key term S_3 by substituting the bound of Proposition 11.1 into Proposition 10.1.

Proposition 11.2 (S_3 Bound)**.** *Let $N \geq T^{3/4}$. Then we have*

$$S_3 \lessapprox T^2|W|^{3/2} + T|W|N^{3-2\sigma} + T|W|^2N^{3/2-\sigma} + T^{9/8}|W|^{29/16}N^{3/2-\sigma}.$$

Now we turn to the proof of Proposition 11.1.

Suppose that $D_N(t) = \sum_{n\sim N} b_n n^{it}$ and $|D_N(t)| \geq N^\sigma$ on W. Morally, we also have $|D_N(t)| \gtrapprox N^\sigma$ if the distance from t to W is $\lesssim 1$. In particular, if $|t_1+t_2-t_3-t_4| \lessapprox 1$, then morally $|D_N(t_1+t_2-t_3)| \gtrapprox N^\sigma$. Therefore morally we should expect that

$$E(W) \lessapprox N^{-2\sigma} \sum_{t_1,t_2,t_3 \in W} \Big|D_N(t_1+t_2-t_3)\Big|^2 = N^{-2\sigma} \sum_{n_1,n_2\sim N} R\Big(\frac{n_1}{n_2}\Big)^2 R\Big(\frac{n_2}{n_1}\Big).$$

We can make this moral argument literally true by working with a slightly smoothed version of D_N.

Lemma 11.3 (Dirichlet polynomials do not vary too fast)**.** *We have*

$$|D_N(t)| \lessapprox \int_{|u-t|\lessapprox 1} |D_N(u)|du + O(T^{-100}).$$

Proof. Let $\psi(x)$ be a smooth bump which is supported on $|2\pi x - \log N| \lesssim 1$ and is equal to 1 on $[(2\pi)^{-1}\log N, (2\pi)^{-1}\log 2N]$. Then we have

$$D_N(t) = \sum_n w(n/N) b_n n^{it} = \sum_n w(n/N) b_n n^{it} \psi\Big(\frac{\log n}{2\pi}\Big) = \int \hat{\psi}(\xi) D_N(t-\xi) d\xi.$$

By the rapid decay of $\hat{\psi}$ we may restrict to $|\xi| \lessapprox 1$ at the cost of an error $O(T^{-100})$. □

Lemma 11.4 (Energy controlled by discrete 3^{rd} moment)**.** *We have that*

$$E(W) \lessapprox N^{-2\sigma} \sum_{n_1,n_2 \sim N} \Big| R\Big(\frac{n_1}{n_2}\Big)\Big|^3.$$

Proof. Since $|D_N(t)| > N^\sigma$ for $t \in W$, we have

$$E(W) = \sum_{\substack{t_1,t_2,t_3,t_4 \in W \\ |t_1+t_2-t_3-t_4| \le 1}} 1 \le N^{-2\sigma} \sum_{\substack{t_1,t_2,t_3,t_4 \in W \\ |t_1+t_2-t_3-t_4| \le 1}} |D_N(t_4)|^2.$$

By Lemma 11.3 and Cauchy-Schwarz, we have for $|t_1+t_2-t_3-t_4| \le 1$

$$|D_N(t_4)|^2 \lessapprox \int_{|u-t_4| \lessapprox 1} |D_N(u)|^2 du \lessapprox \int_{|u-t_1+t_2-t_3| \lessapprox 1} |D_N(u)|^2 du.$$

Since W is T^ϵ-separated, given t_1, t_2, t_3 there is at most 1 choice of $t_4 \in W$ such that $|t_1+t_2-t_3-t_4| \lessapprox 1$. Thus we see that

$$\begin{aligned} E(W) &\lessapprox N^{-2\sigma} \sum_{t_1,t_2,t_3 \in W} \int_{s \lessapprox 1} |D_N(t_1+t_2-t_3-s)|^2 ds \\ &= N^{-2\sigma} \sum_{n_1,n_2} w\Big(\frac{n_1}{N}\Big) w\Big(\frac{n_2}{N}\Big) \int_{s \lessapprox 1} \Big(\frac{n_2}{n_1}\Big)^{is} R\Big(\frac{n_1}{n_2}\Big)^2 R\Big(\frac{n_2}{n_1}\Big) ds \\ &\lessapprox N^{-2\sigma} \sum_{n_1,n_2 \sim N} \Big| R\Big(\frac{n_1}{n_2}\Big)\Big|^3. \end{aligned}$$

□

Next we note that Heath-Brown's theorem (Theorem 1.6) gives bounds for $\sum_{n_1,n_2\sim N} \Big|R\Big(\frac{n_1}{n_2}\Big)\Big|^2$ and also $\sum_{n_1,n_2\sim N} \Big|R\Big(\frac{n_1}{n_2}\Big)\Big|^4$.

Lemma 11.5 (Discrete second moment)**.** *For any $M \ge 1$,*

$$\sum_{n_1,n_2 \sim M} \Big| R\Big(\frac{n_1}{n_2}\Big)\Big|^2 \lessapprox |W| M^2 + |W|^2 M + |W|^{5/4} T^{1/2} M.$$

Proof. We have that

$$\sum_{n_1,n_2\sim M}\Big|R\Big(\frac{n_1}{n_2}\Big)\Big|^2=\sum_{t_1,t_2\in W}\Big|\sum_{n\sim M}n^{i(t_1-t_2)}\Big|^2,$$

so by Theorem 1.6 this is

$$\lessapprox |W|M^2+|W|^2M+|W|^{5/4}T^{1/2}M. \qquad \square$$

Lemma 1.7 is now a quick consequence of our arguments so far.

Proof of Lemma 1.7. By Lemma 11.4 and the trivial bound $|R(x)|\le|W|$ we have

$$E(W)\lessapprox N^{-2\sigma}\sum_{n_1,n_2\sim N}\Big|R\Big(\frac{n_1}{n_2}\Big)\Big|^3\le |W|N^{-2\sigma}\sum_{n_1,n_2\sim N}\Big|R\Big(\frac{n_1}{n_2}\Big)\Big|^2.$$

Lemma 11.5 (which is just Theorem 1.6) now shows for $N>T^{2/3}$ we have

$$E(W)\lessapprox |W|^3N^{1-2\sigma}+|W|^2N^{2-2\sigma}. \qquad \square$$

To do better we look at higher moments to avoid the potentially wasteful use of the trivial bound $|R(x)|\le|W|$.

Lemma 11.6 (Discrete fourth moment)**.** *For any $M\ge 1$,*

$$\sum_{n_1,n_2\sim M}\Big|R\Big(\frac{n_1}{n_2}\Big)\Big|^4\lessapprox M^2E(W)+|W|^4M+E(W)^{3/4}|W|T^{1/2}M.$$

Proof. We split the sum in the R function according to the number of representations of u as approximately t_1-t_2. Let $\lfloor x\rfloor$ denote the largest integer $\le x$, and define

$$U_B:=\Big\{u\in\mathbb{Z}:\ \#\{(t_1,t_2)\in W^2:\ \lfloor t_1-t_2\rfloor=u\}\sim B\Big\}.$$

Clearly U_B is empty if $B<1/2$ or if $B>|W|$. Thus, using Cauchy-Schwarz

$$\begin{aligned}|R(x)|^4=\Big|\sum_{t_1,t_2\in W}x^{i(t_1-t_2)}\Big|^2&=\Big|\sum_{B=2^j}\sum_{u\in U_B}\sum_{\substack{t_1,t_2\in W\\ \lfloor t_1-t_2\rfloor=u}}x^{i(t_1-t_2)}\Big|^2\\&\lessapprox\sum_{B=2^j\le|W|}\Big|\sum_{u\in U_B}\sum_{\substack{t_1,t_2\in W\\ \lfloor t_1-t_2\rfloor=u}}x^{i(t_1-t_2)}\Big|^2.\end{aligned}$$

Taking $x=n_1/n_2$ and summing over $n_1,n_2\sim M$ then gives

$$\begin{aligned}\sum_{n_1,n_2\sim M}\Big|R\Big(\frac{n_1}{n_2}\Big)\Big|^4 &\lessapprox \sup_{B\le |W|}\sum_{n_1,n_2\sim M}\Big|\sum_{u\in U_B}\sum_{\substack{t_1,t_2\in W\\ \lfloor t_1-t_2\rfloor=u}}\Big(\frac{n_1}{n_2}\Big)^{i(t_1-t_2)}\Big|^2\\ &\le \sup_{B\le|W|}\Big|\sum_{u_1,u_2\in U_B}\Big(\sum_{\substack{t_1,t_3\in W\\ \lfloor t_1-t_3\rfloor=u_1}}1\Big)\Big(\sum_{\substack{t_2,t_4\in W\\ \lfloor t_2-t_4\rfloor=u_2}}1\Big)\sup_{|s|\lesssim 1}\Big|\sum_{n\sim M}n^{i(u_1-u_2+s)}\Big|^2\\ &\lessapprox \sup_{B\le|W|}B^2\sum_{u_1,u_2\in U_B}\sup_{|s|\lesssim 1}\Big|\sum_{n\sim M}n^{i(u_1-u_2+s)}\Big|^2.\end{aligned}$$

By using Lemma 11.3 to replace the supremum with an integral, and then applying Theorem 1.6, we find

$$\begin{aligned}\sum_{n_1,n_2\sim M}\Big|R\Big(\frac{n_1}{n_2}\Big)\Big|^4 &\lessapprox \sup_{B\le|W|}B^2\int_{t\lessapprox 1}\sum_{u_1,u_2\in U_B}\Big|\sum_{n\sim M}n^{i(u_1-u_2+t)}\Big|^2dt\\ &\lessapprox \sup_{B\le|W|}B^2\Big(M^2|U_B|+|U_B|^2M+T^{1/2}|U_B|^{5/4}M\Big).\end{aligned}$$

We have that $B|U_B|\lessapprox |W|^2$ and $B^2|U_B|\lessapprox E(W)$, so this gives

$$\sum_{n_1,n_2\sim M}\Big|R\Big(\frac{n_1}{n_2}\Big)\Big|^4\lessapprox M^2E(W)+|W|^4M+E(W)^{3/4}|W|T^{1/2}M. \qquad \square$$

To bound $\sum_{n_1,n_2\sim N}|R\Big(\frac{n_1}{n_2}\Big)|^3$, we could use Hölder:

$$\sum_{n_1,n_2\sim N}\Big|R\Big(\frac{n_1}{n_2}\Big)\Big|^3\le\Bigg(\sum_{n_1,n_2\sim N}\Big|R\Big(\frac{n_1}{n_2}\Big)\Big|^2\Bigg)^{1/2}\Bigg(\sum_{n_1,n_2\sim N}\Big|R\Big(\frac{n_1}{n_2}\Big)\Big|^4\Bigg)^{1/2}$$

and then bound the two factors using Lemmas 11.5 and 11.6. However, this Hölder step is somewhat lossy. If n_1'/n_2' is a rational number of small height, then the sum $\sum_{n_1,n_2\sim N}\Big|R\Big(\frac{n_1}{n_2}\Big)\Big|^p$ counts $|R(n_1'/n_2')|^p$ many times – because there are many $n_1,n_2\sim N$ with $n_1/n_2=n_1'/n_2'$. The 4^{th} moment tends to be dominated by n_1,n_2 with large $\gcd(n_1,n_2)$. But the 2^{nd} moment tends to be dominated by n_1,n_2 with small $\gcd(n_1,n_2)$. Therefore, instead of doing Hölder immediately, we now split our argument according to the size of $\gcd(n_1,n_2)$.

Let $d=\gcd(n_1,n_2)$ and $n_1=n_1'd$, $n_2=n_2'd$ for some $n_1',n_2'\sim N/d$ with $\gcd(n_1',n_2')=1$. Thus we have for any choice of parameter D (dropping the coprimality constraint when d is large)

$$E(W)\le N^{-2\sigma}\sum_{d\le D}\sum_{\substack{n_1',n_2'\sim N/d\\ \gcd(n_1',n_2')=1}}\Big|R\Big(\frac{n_1'}{n_2'}\Big)\Big|^3+N^{-2\sigma}\sum_{d\ge D}\sum_{n_1',n_2'\sim N/d}\Big|R\Big(\frac{n_1'}{n_2'}\Big)\Big|^3. \tag{11.2}$$

First we consider small d. When d is small enough, the distinct fractions n_1'/n_2' are very well distributed and so it makes sense to compare our sum with $\int_{v\asymp 1}|R(v)|^3dv$.

We recall that W is contained in an interval of length T. Morally, $|\hat{W}(\tau)|$ is locally constant on intervals of length $1/T$. Since $R(v)=\hat{W}(\log v)$, we see that for $v\asymp 1$, $|R(v)|$ is morally locally constant at scale $1/T$. We make this precise in the following lemma:

Lemma 11.7. *For $v\asymp 1$,*

$$|R(v)|\lesssim T\int_{|v'-v|\lessapprox 1/T}|R(v')|dv'+O(T^{-100}).$$

Proof. Since $v\asymp 1$, we can do a change of variables, $\tau=\log v$, and it suffices to prove that

$$|\hat{W}(\tau)|\lesssim T\int_{|\tau'-\tau|\lessapprox 1/T}|\hat{W}(\tau')|d\tau'+O(T^{-100}).$$

We know that W is contained in an interval of length T; call this $[T_0,T_0+T]$. Let ψ be a smooth bump which is 1 on $[0,1]$. Then we have

$$\hat{W}(\tau)=\sum_{t\in W}e(-t\tau)=\sum_{t\in W}e(-t\tau)\psi\Big(\frac{t-T_0}{T}\Big)=\int\hat{\psi}(\xi)\hat{W}\Big(\tau-\frac{\xi}{T}\Big)e\Big(\frac{T_0\xi}{T}\Big)d\xi.$$

By the rapid decay of $\hat{\psi}$, we may restrict the integral to $\xi\lessapprox 1$ at the cost of an $O(T^{-100})$ error term. Since $\hat{\psi}\lesssim 1$ this then gives the result. □

Lemma 11.8 (Small GCD terms)**.** *We have*

$$\sum_{\substack{n_1,n_2\sim N\\ \gcd(n_1,n_2)\le D}}\Big|R\Big(\frac{n_1}{n_2}\Big)\Big|^3\lessapprox\Big(DT+N^2\Big)|W|^{1/2}E(W)^{1/2}.$$

Proof. Let $d=\gcd(n_1,n_2)$ and $n_1=n_1'd$, $n_2=n_2'd$ for some $n_1',n_2'\sim N/d$ with $\gcd(n_1',n_2')=1$. By Lemma 11.7, we have

$$\sum_{\substack{n_1',n_2'\sim N/d\\ \gcd(n_1',n_2')=1}}\Big|R\Big(\frac{n_1'}{n_2'}\Big)\Big|^3\lesssim T\int_{v\asymp 1}|R(v)|^3\Big(\sum_{\substack{n_1,n_2'\sim N/d\\ \gcd(n_1',n_2')=1\\ |v-n_1'/n_2'|\lessapprox 1/T}}1\Big)dv.$$

Since the fractions n_1'/n_2' are d^2/N^2-separated, we have that the inner sum over n_1',n_2' on the right hand side is $\lessapprox 1+N^2/(d^2T)$. Thus we find

$$\sum_{d\leq D}\sum_{\substack{n_1',n_2'\sim N/d\\ \gcd(n_1',n_2')=1}}\Big|R\Big(\frac{n_1'}{n_2'}\Big)\Big|^3 \lessapprox \sum_{d\leq D}\Big(T+\frac{N^2}{d^2}\Big)\int_{v\asymp 1}|R(v)|^3dv$$
$$\leq \sum_{d\leq D}\Big(T+\frac{N^2}{d^2}\Big)\Big(\int_{v\asymp 1}|R(v)|^2dv\Big)^{1/2}\Big(\int_{v\asymp 1}|R(v)|^4dv\Big)^{1/2}$$
$$\lesssim \Big(DT+N^2\Big)|W|^{1/2}E(W)^{1/2}. \qquad \square$$

We choose $D:=N^2/T$, so this gives

$$\sum_{\substack{n_1,n_2\sim N\\ \gcd(n_1,n_2)\leq D}}\Big|R\Big(\frac{n_1}{n_2}\Big)\Big|^3 \lessapprox N^2|W|^{1/2}E(W)^{1/2}. \tag{11.3}$$

Lemma 11.9 (Large GCD terms). *Let $D=N^2/T$ and $N\geq T^{3/4}$. Then we have*

$$\sum_{\substack{n_1,n_2\sim N\\ \gcd(n_1,n_2)\geq D}}\Big|R\Big(\frac{n_1}{n_2}\Big)\Big|^3 \lesssim N|W|^3+NT^{1/4}|W|^{21/8}+E(W)^{1/2}|W|^{1/2}N^2.$$

Proof. As in the previous lemma, we let $d=\gcd(n_1,n_2)$ and $n_1=n_1'd$, $n_2=n_2'd$. When d is large, we keep the discrete summation over n_1',n_2' and apply Cauchy-Schwarz directly, giving

$$\sum_{n_1',n_2'\sim N/d}\Big|R\Big(\frac{n_1'}{n_2'}\Big)\Big|^3 \lesssim \Big(\sum_{n_1',n_2'\sim N/d}\Big|R\Big(\frac{n_1'}{n_2'}\Big)\Big|^4\Big)^{1/2}\Big(\sum_{n_1',n_2'\sim N/d}\Big|R\Big(\frac{n_1'}{n_2'}\Big)\Big|^2\Big)^{1/2}.$$

Now we can bound the factors on the right-hand side by Lemmas 11.5 and 11.6, with $M=N/d$. This gives

$$\sum_{n_1',n_2'\sim N/d}\Big|R\Big(\frac{n_1'}{n_2'}\Big)\Big|^3 \lessapprox \Big(\frac{|W|N^2}{d^2}+\frac{|W|^2N}{d}+\frac{|W|^{5/4}T^{1/2}N}{d}\Big)^{1/2}$$
$$\times\Big(\frac{N^2E(W)}{d^2}+\frac{N|W|^4}{d}+\frac{E(W)^{3/4}|W|T^{1/2}N}{d}\Big)^{1/2}.$$

Summing over $d\geq D$, using Cauchy-Schwarz, and recalling that $D=N^2/T$ then gives

$$\sum_{\substack{n_1,n_2\sim N\\ \gcd(n_1,n_2)\geq D}}\Big|R\Big(\frac{n_1}{n_2}\Big)\Big|^3 \lessapprox \Big(\frac{|W|N^2}{D}+|W|^2N+|W|^{5/4}T^{1/2}N\Big)^{1/2}$$
$$\times\Big(\frac{N^2E(W)}{D}+N|W|^4+E(W)^{3/4}|W|T^{1/2}N\Big)^{1/2}.$$

Next we work on simplifying and organizing the algebra. Recall that we have $N > T^{3/4}$ and $D = N^2/T$. Therefore, we have $|W|^{5/4}T^{1/2}N > |W|T = |W|N^2/D$, and we can ignore the first term in the first factor. Thus the above expression is bounded by

$$\lessapprox \Big(|W|^2 N + |W|^{5/4}T^{1/2}N\Big)^{1/2}\Big(E(W)T + N|W|^4 + E(W)^{3/4}|W|T^{1/2}N\Big)^{1/2}. \tag{11.4}$$

There are two main cases, depending on whether $|W| > T^{2/3}$ or not. If $|W| > T^{2/3}$ then $|W|^2N > |W|^{5/4}T^{1/2}N$, and so the first factor is dominated by $|W|^2N$. We turn to the second factor. If $|W| > T^{2/3}$, then $N|W|^4 > N|W|^{13/4}T^{1/2} \geq E(W)^{3/4}|W|T^{1/2}N$. Also, since $N > T^{3/4} > T^{1/2}$, $N|W|^4 > |W|^3T \geq E(W)T$. So the second factor is dominated by $N|W|^4$. Therefore, if $|W| > T^{2/3}$, (11.4) simplifies to

$$\sum_{\substack{n_1,n_2\sim N\\ \gcd(n_1,n_2)\geq D}} \Big|R\Big(\frac{n_1}{n_2}\Big)\Big|^3 \lessapprox N|W|^3. \tag{11.5}$$

Now suppose $|W| < T^{2/3}$. Recalling that $N > T^{3/4} > T^{1/2}$, we see that $|W|^2N < |W|^{5/4}T^{1/2}N$, so the first factor is dominated by $|W|^{5/4}T^{1/2}N$. Turning to the second factor, we see that $E(W)^{3/4}|W|T^{1/2}N > E(W)T$. Thus, if $|W| < T^{2/3}$ we see that (11.4) simplifies to

$$\begin{aligned}\sum_{\substack{n_1,n_2\sim N\\ \gcd(n_1,n_2)\geq D}} \Big|R\Big(\frac{n_1}{n_2}\Big)\Big|^3 &\lessapprox \Big(|W|^{5/4}T^{1/2}N\Big)^{1/2}\Big(N|W|^4 + E(W)^{3/4}|W|T^{1/2}N\Big)^{1/2}\\ &\lessapprox NT^{1/4}|W|^{21/8} + E(W)^{1/2}|W|^{1/2}N^2\Big(\frac{T^{1/2}|W|^{5/8}}{E(W)^{1/8}N}\Big). \qquad (11.6)\end{aligned}$$

Since $E(W) > |W|^2$ and $|W| < T^{2/3}$ and $N > T^{3/4}$, we see that $T^{1/2}|W|^{5/8} < E(W)^{1/8}N$, so the final term in (11.6) is $O(|W|^{1/2}E(W)^{1/2}N^2)$. Thus, combining (11.5) and (11.6), we find that provided $N > T^{3/4}$, regardless of the size of W, we have

$$\sum_{\substack{n_1,n_2\sim N\\ \gcd(n_1,n_2)\geq D}} \Big|R\Big(\frac{n_1}{n_2}\Big)\Big|^3 \lessapprox N|W|^3 + NT^{1/4}|W|^{21/8} + E(W)^{1/2}|W|^{1/2}N^2. \qquad \square$$

Proof of Proposition 11.1. First we use Lemma 11.4 to give

$$E(W) \lessapprox N^{-2\sigma}\sum_{n_1,n_2\sim N}\Big|R\Big(\frac{n_1}{n_2}\Big)\Big|^3.$$

Splitting according to whether $\gcd(n_1, n_2) \le D = N^2/T$ or not, we find by Lemma 11.8 and Lemma 11.9 that

$$E(W) \le N^{-2\sigma}\Big(N|W|^3 + NT^{1/4}|W|^{21/8} + E(W)^{1/2}|W|^{1/2}N^2\Big).$$

This simplifies to give

$$E(W) \lessapprox |W|N^{4-4\sigma} + |W|^{21/8}T^{1/4}N^{1-2\sigma} + |W|^3N^{1-2\sigma}. \qquad \square$$

12. Proof of results on large values of Dirichlet polynomials

In this section we prove our main results on the large values of Dirichlet polynomials by assembling the tools in the previous sections.

Proof of Proposition 3.1. Suppose that $|D_N(t)| \ge N^\sigma$ on the set W contained in an interval of length $T = N^{6/5}$. By Proposition 4.6 and (5.5), we have

$$|W| \lesssim_\epsilon N^{2-2\sigma} + N^{1-2\sigma}\Big(\sum_{m\in\mathbb{Z}^3\setminus\{0\}} I_m\Big)^{1/3} = N^{2-2\sigma} + N^{1-2\sigma}\Big(S_1+S_2+S_3\Big)^{1/3}.$$

By Proposition 5.1, S_1 is negligible. We bound S_2 by Proposition 6.1, and S_3 by Proposition 11.2. Therefore we get for any choice of $k\in\mathbb{N}$

$$\begin{aligned}|W|^3N^{6\sigma-3} &\lessapprox N^3 + S_2 + S_3\\ &\lessapprox N^3 + |W|^2N^2 + TN|W|^{2-1/k} + N^2|W|^{2-3/4k}T^{1/2k} + T^2|W|^{3/2}\\ &\qquad + T|W|N^{3-2\sigma} + T|W|^2N^{3/2-\sigma} + T^{9/8}|W|^{29/16}N^{3/2-\sigma}.\end{aligned}$$

In this formula k comes from the bound for S_2. It is a positive integer that we can choose. The last inequality rearranges to give

$$\begin{aligned}|W| &\lessapprox N^{2-2\sigma} + N^{5-6\sigma} + T^{\frac{k}{k+1}}N^{(4-6\sigma)\frac{k}{k+1}} + N^{(5-6\sigma)\frac{4k}{4k+3}}T^{\frac{2}{4k+3}} + T^{4/3}N^{2-4\sigma}\\ &\qquad + T^{1/2}N^{3-4\sigma} + TN^{9/2-7\sigma} + T^{18/19}N^{72/19-112\sigma/19}.\end{aligned} \tag{12.1}$$

We choose $k=4$. We also simplify the formulas using $T = N^{6/5}$.

$$\begin{aligned}|W| &\lessapprox T\Big(N^{(4-10\sigma)/5} + N^{(19-30\sigma)/5} + N^{(74-120\sigma)/25} + N^{(298-480\sigma)/95}\\ &\qquad + N^{(12-20\sigma)/5} + N^{(9-14\sigma)/2} + N^{(354-560\sigma)/95}\Big)\\ &\lessapprox TN^{(4-10\sigma)/5} + TN^{(12-20\sigma)/5} + TN^{(9-14\sigma)/2}.\end{aligned}$$

If $\sigma\in[7/10, 8/10]$ the first and third terms can be dropped and we get

$$|W| \lessapprox TN^{(12-20\sigma)/5}. \qquad \square$$

When $N > T^{5/6}$ the process of going from Proposition 3.1 to Theorem 1.1 is somewhat wasteful since it actually bounds the number of large values in $[0, N^{6/5}]$. By using a variation of the above argument, the bound in Theorem 1.1 could be improved. We record one such improvement here.

Proposition 12.1 (Large values estimate for $N \geq T^{5/6}$)**.** *Suppose $(b_n)_{n\in[N,2N]}$, $(t_r)_{r\leq R}$ are as in Theorem 1.1, and that $T^{5/6} \leq N \leq T$ and $V = N^\sigma$ with $\sigma \geq 7/10$. Then we have*

$$R \lessapprox N^{2-2\sigma} + T^{1/2}N^{3-4\sigma} + \inf_{k\in\mathbb{N}}\Big(T^{\frac{k}{k+1}}N^{(4-6\sigma)\frac{k}{k+1}} + N^{(5-6\sigma)\frac{4k}{4k+3}}T^{\frac{2}{4k+3}}\Big).$$

In particular, we have

$$R \lessapprox N^{2-2\sigma} + T^{1/2}N^{3-4\sigma} + T^{(30\sigma-21)/5}N^{(46-60\sigma)/5}.$$

Proposition 12.1 implies that the $N^{18/5}V^{-4}$ term in Theorem 1.1 can be replaced by $T^{1/2}N^{3-4\sigma} + T^{(30\sigma-21)/5}N^{(46-60\sigma)/5}$, which is smaller. When $\sigma = 3/4$, Proposition 12.1 improves on (1.1) by a factor of $(T/N)^{1/2}$. Since we anticipate the main uses of Theorem 1.1 to be when $N < T^{5/6}$ we content ourselves to the simpler formulation of Theorem 1.1.

Proof. Jutila's large values estimate [Ju, Theorem (1.4)] with $k = 3$ gives

$$R \lessapprox N^{2-2\sigma} + TN^{(10-16\sigma)/3} + TN^{18-24\sigma},$$

which implies our bound for $\sigma \geq 39/50$ since the second and third terms above are then both smaller than $T^{1/2}N^{3-4\sigma}$. Thus we only need to consider $\sigma \in [7/10, 39/50]$. The bound now follows from (12.1), (1.1) and a little algebra. We find for $\sigma \in [7/10, 39/50]$ and $N \in [T^{5/6}, T]$ we have that

$$T^{1/2}N^{3-4\sigma} \gtrapprox T^{4/3}N^{2-4\sigma} + TN^{9/2-7\sigma} + T^{18/19}N^{72/19-112\sigma/19}.$$

Therefore the $T^{1/2}N^{3-4\sigma}$ term in (12.1) dominates the 5^{th}, 7^{th} and 8^{th} terms in (12.1). We also have

$$T^{1/2}N^{3-4\sigma} \gtrapprox \min(TN^{4-6\sigma}, N^{2-2\sigma}) + \min(TN^{1-2\sigma}, N^{5-6\sigma}).$$

Therefore, by combining (12.1) and (1.1) we find that

$$R \lessapprox N^{2-2\sigma} + T^{1/2}N^{3-4\sigma} + \inf_{k\in\mathbb{N}}\Big(T^{\frac{k}{k+1}}N^{(4-6\sigma)\frac{k}{k+1}} + N^{(5-6\sigma)\frac{4k}{4k+3}}T^{\frac{2}{4k+3}}\Big).$$

This gives the first bound. For the final bound, we note that if $N^{4-4\sigma} > T$ then $TN^{1-2\sigma} < T^{1/2}N^{3-4\sigma}$ so the bound follows from (1.1). Therefore we may assume that $N^{4-4\sigma} \le T$. In this case $N^{15-18\sigma} \le T^2$ and so in the expression in parentheses above, the first term is decreasing in k and the second term is increasing in k. Thus, for the expression to be less than $T^{(30\sigma-21)/5}N^{(46-60\sigma)/5}$ we require

$$\frac{-73+138n+90\sigma-180n\sigma}{12(1-n)(7-10\sigma)} \le k \le \frac{-21+46n+30\sigma-60n\sigma}{2(1-n)(13-15\sigma)}$$

where $n := \log N/\log T$. The upper bound of this interval is always at least 1 for $n \in [5/6, 1)$, $\sigma \in [7/10, 39/50]$ and the length of this interval is

$$1 + \frac{5(6n-5)(-41+123\sigma-90\sigma^2)}{12(1-n)(10\sigma-7)(13-15\sigma)}.$$

Thus the interval has length at least 1 whenever $\sigma \in [7/10, 39/50]$, and so there is a choice of $k \in \mathbb{N}$ giving the desired bound. □

13. Applications to Riemann zeta function and prime numbers

13.1. **Proof of Theorem 1.2.** As noted in the introduction, Theorem 1.2 follows from Ingham's result (1.2) if $\sigma \le 7/10$ and Huxley's result (1.3) if $\sigma \ge 8/10$, so we may assume that $\sigma \in [7/10, 8/10]$. Clearly it suffices to show the bound of Theorem 1.2 for zeros with imaginary part in $[T, 2T]$, since the result for $[0, T]$ then follows by considering $T/2, T/4, \dots$ in place of T.

We now briefly recall the classical zero-detecting methodology, referring the reader to [M3, Chapter 12] or [MP, Appendix C] for more complete details. Given a parameter N, we let

$$D_N(s) := \sum_{n\in[N,2N]} b_n n^{-s},$$
$$b_n := \Big(\sum_{\substack{d|n\\ d\le 2T^{1/\sqrt{\log\log T}}}} \mu(d)\Big) w_0\Big(\frac{n}{N}\Big)\exp\Big(-\frac{n}{T^{1/2}}\Big).$$

Here w_0 is a fixed smooth bump function supported on $[1, 2]$.

Then, a non-trivial zero $\rho = \beta + i\gamma$ of $\zeta(s)$ with $\beta \ge \sigma$ and $\gamma \in [T, 2T]$ is called a 'Type I zero' if there is a choice of $N \in [T^{1/\log\log T}, T^{1/2}(\log T)^2]$ such that $D_N(\rho) \ge 1/(3\log T)$. If it is not a Type I zero then it is a 'Type II zero', and the number of Type II zeros is $\le T^{2-2\sigma}(\log T)^{O(1)}$ by [MP, Lemma 33]. Thus it suffices to bound the number of Type I zeros. There are $O(\log T)$ choices of N so we focus on the value of N which gives the largest number of Type I zeros.

We now make a slight modification to D_N to remove the dependencies on the real parts. Let $\rho = \beta + i\gamma$ be a Type I zero with $\beta \ge \sigma$, and let $\psi(u)$ be a smooth function equal to $e^{u(\beta-\sigma)}$ on $[\log N, \log 2N]$ and supported on $[(\log N)/2, 2\log N]$ with $\|\psi^{(j)}(t)\|_\infty \lesssim_j t^{-j}$ for all $j \in \mathbb{N}$. We then note that by Fourier expansion

$$D_N(\rho) = \sum_{n\in[N,2N]} b_n n^{-\sigma-i\gamma}\psi(\log n) = \frac{1}{2\pi i}\int_\xi \hat{\psi}(\xi)\Big(D_N(\sigma+i(\gamma+2\pi\xi))\Big)d\xi.$$

Since $\hat{\psi}$ is rapidly decreasing, we may truncate the integral to $\xi \lessapprox 1$ at the cost of an $O(T^{-100})$ error term. Therefore we see that if ρ is a Type I zero, we have $|D_N(\sigma+i\gamma+i\xi)| \gtrapprox 1$ for some $\xi \lessapprox 1$. There are $O(\log T)$ non-trivial zeros $\rho = \beta+i\gamma$ with $\gamma \in [t, t+1]$ for any $t \in [T, 2T]$. Therefore we can find a 1-separated set of points $(s_r)_{r\le R}$ in $[T, 2T]$ with $|D_N(\sigma+is_r)| \gtrapprox 1$ and the number R of points satisfies $R \gtrapprox N(\sigma, 2T) - N(\sigma, T)$. Let

$$\widetilde{b}_n := \Big(\frac{N}{n}\Big)^\sigma b_n, \qquad \widetilde{D}_N(t) := \sum_{n\in[N,2N]} \widetilde{b}_n n^{it} = N^\sigma D_N(\sigma+it).$$

Thus it suffices to show that if $N < T^{1/2+o(1)}$ and W is a 1-separated set in $[T, 2T]$ such that $|\widetilde{D}_N(t)| \gtrapprox N^\sigma$, we have $|W| \lessapprox T^{15(1-\sigma)/(3+5\sigma)+o(1)}$.

If $T^{5/(3+5\sigma)} \le N^2 \le T^{75(1-\sigma)/(54+30\sigma-100\sigma^2)}$, then we use Theorem 1.1 applied to the Dirichlet polynomial $\widetilde{D}_N^2$ of length N^2, which shows that (noting that $\sigma \in [7/10, 8/10]$ implies that the N^2V^{-2} term is dominated by the $N^{18/5}V^{-4}$ term)

$$\begin{aligned}|W| &\lessapprox (T^{75(1-\sigma)/(54+30\sigma-100\sigma^2)})^{18/5-4\sigma} + T(T^{5/(3+5\sigma)})^{12/5-4\sigma}\\ &\lessapprox T^{15(1-\sigma)/(3+5\sigma)}.\end{aligned}$$

If instead N lies outside of these ranges, we can use classical estimates. If $T^{2/3} < N^2 < T^{5/(3+5\sigma)}$ then the Mean Value Theorem applied to $\widetilde{D}_N^3$ gives

$$|W| \lessapprox N^{6-6\sigma} \lessapprox T^{15(1-\sigma)/(3+5\sigma)}.$$

If $T \gtrapprox N^2 > T^{75(1-\sigma)/(54+30\sigma-100\sigma^2)}$ then the Mean Value Theorem applied to N^2 gives

$$|W| \lessapprox T(T^{75(1-\sigma)/(54+30\sigma-100\sigma^2)})^{1-2\sigma} = T^{(129-195\sigma+50\sigma^2)/(54+30\sigma-100\sigma^2)}.$$

A quick calcuation verifies $(129-195\sigma+50\sigma^2)/(54+30\sigma-100\sigma^2) < 15(1-\sigma)/(3+5\sigma)$. Finally, if $N < T^{1/3}$, we choose $k \ge 3$ such that $N^k \le T \le N^{k+1}$. The Mean Value Theorem applied to N^k and N^{k+1} gives

$$\begin{aligned}|W| &\lessapprox \min\Big(N^{(k+1)(2-2\sigma)}, TN^{k(1-2\sigma)}\Big)\\ &\lessapprox \Big(N^{(k+1)(2-2\sigma)}\Big)^{\frac{k(2\sigma-1)}{k+2-2\sigma}}\Big(TN^{k(1-2\sigma)}\Big)^{\frac{(k+1)(2-2\sigma)}{k+2-2\sigma}}\\ &= T^{(k(2-2\sigma)+2-2\sigma)/(k+2-2\sigma)}\\ &\lessapprox T^{(4-4\sigma)/(3-\sigma)}.\end{aligned}$$

Here we noted that the penultimate bound is decreasing in k and so maximized at $k = 3$. A quick calculation verifies that $(4-4\sigma)/(3-\sigma) < 15(1-\sigma)/(3+5\sigma)$, so this gives an acceptable bound too.

Remark. *For the purposes of proving a zero density estimate of the form $N(\sigma,T) \lessapprox T^{A(1-\sigma)}$ with A a fixed constant as small as possible, the critical case in our work is when $\sigma = 7/10$, $N = T^{10/13}$, we subdivide $[0,T]$ into intervals of length $T_1 = T^{12/13}$ and where the set W of large values on each subinterval has $|W| \approx T_1^{2/3}$ and $E(W) \approx |W|^{5/2} \approx |W|^4/T_1$. In this critical situation our bounds for S_1 and S_2 are both best possible, and so any further improvement would have to come from the S_3 term. Our bound for $E(W)$ is also likely to be difficult to improve since a random set W would have $E(W) \approx |W|^4/T_1$. The argument of Section 9 is also essentially tight if the R function is taking $T_1^{2/3}$ values of size $|W|/M \approx T_1^{1/2}$ and the set of these values is highly concentrated on rationals with numerator and denominator of size $T_1^{1/3}$.*

13.2. **Proof of Corollary 1.3 and Corollary 1.4.** These are well-known to follow quickly from (1.4), but for completeness we give a proof. By partial summation, it suffices to prove corresponding results for the Von Mangoldt function in place of the prime indicator function. By the explicit formula (see, for example [Da, Chapter 17]) we have for any choice of $T \geq 2$

$$\sum_{n\in[x,x+y]}\Lambda(n) = y - \sum_{|\rho|\leq T}\Big(\frac{(x+y)^\rho - x^\rho}{\rho}\Big) + O\Big(\frac{x(\log x)^3}{T}\Big).$$

We choose $T = xy^{-1}\exp(2\sqrt[4]{\log x})$ so the error term is $O(y\exp(-\sqrt[4]{\log x}))$, and note that the term in parentheses is $\int_x^{x+y} t^{\rho-1}dt \ll yx^{\Re(\rho)-1}$. Therefore, by considering $1/\log x$-separated values of σ, we find that

$$\sum_{n\in[x,x+y]}\Lambda(n) = y + O\Big(y(\log x)\sup_\sigma x^{\sigma-1}N(\sigma,T)\Big) + O(y\exp(-\sqrt[4]{\log x})).$$

By combining (1.4) with a slightly stronger result (such as [Ju] or [M3, Theorem 12.1]) that loses at most logarithmic factors for σ closer to 1, we have the bound

$$N(\sigma,T) \lesssim T^{(30/13+o(1))(1-\sigma)}(\log T)^{O(1)}. \tag{13.1}$$

Using this and the Vinogradov-Korobov zero-free bound $N(\sigma,T)=0$ for $\sigma\geq 1-c(\log T)^{-2/3}(\log\log T)^{-1/3}$ (for a suitable constant $c>0$; see [M3, Corollary 11.4], we find that

$$\sup_{\sigma} x^{\sigma-1}N(\sigma,T)\lesssim(\log T)^{O(1)}\sup_{\sigma\leq 1-c(\log T)^{-5/7}}\Big(\frac{T^{30/13+o(1)}}{x}\Big)^{1-\sigma}$$
$$\lesssim_\epsilon \exp(-\sqrt[4]{\log x})$$

provided $T<x^{13/30-\epsilon/2}$. Recalling that $T=xy^{-1}\exp(2\sqrt[4]{\log x})$ and $y\geq x^{17/30+\epsilon}$, this gives Corollary 1.3.

For Corollary 1.4, we first let $\delta=X^{-13/15+\epsilon/2}$. By splitting $[x,x+y]$ into intervals of length δx, we see that

$$\int_X^{2X}\Big(\sum_{n\in[x,x+y]}\Lambda(n)-y\Big)^2dx\lesssim\frac{y^2}{\delta^2X^2}\int_X^{2X}\Big(\sum_{n\in[x,x+\delta x]}\Lambda(n)-\delta x\Big)^2dx+O(\delta^2X^3).$$

If the corollary was false, the left hand side would be significantly larger than $y^2X\exp(-3\sqrt[4]{\log x})$, so it suffices to show that the integral on the right hand side is $\lesssim\delta^2X^3\exp(-3\sqrt[4]{\log x})$. Applying the Explicit Formula as above with $T=\delta^{-1}\exp(4\sqrt[4]{\log x})$, we see that it suffices to show that

$$\int_X^{2X}\Big|\sum_{|\rho|<T}x^\rho\Big(\frac{(1+\delta)^\rho-1}{\rho}\Big)\Big|^2dx\lesssim\delta^2X^3\exp(-3\sqrt[4]{\log x}).\tag{13.2}$$

Expanding the sum, and performing the integral over x, we obtain

$$\sum_{|\rho_1|,|\rho_2|\leq T}\Big(\frac{(1+\delta)_1^\rho-1}{\rho_1}\Big)\overline{\Big(\frac{(1+\delta)^{\rho_2}-1}{\rho_2}\Big)}\int_X^{2X}x^{\rho_1+\overline{\rho_2}}dx\lesssim\delta^2\sum_{|\rho_1|,|\rho_2|<T}\frac{x^{\Re(\rho_1)+\Re(\rho_2)+1}}{|\rho_1+\overline{\rho_2}+1|}.$$

Since $x^{\Re(\rho_1)+\Re(\rho_2)+1}\leq x^{2\Re(\rho_1)+1}+x^{2\Re(\rho_2)+1}$, and noting that (since there are $O(\log T)$ zeros in a horizontal strip of height 1) we have

$$\sum_{|\rho_2|<T}|1+z+\overline{\rho_2}|^{-1}\lesssim(\log T)^2$$

for any $|z|<T$ with $\Re(z)\geq 0$. Thus we find that

$$\int_X^{2X}\Big|\sum_{|\rho|<T}x^\rho\Big(\frac{(1+\delta)^\rho-1}{\rho}\Big)\Big|^2dx\lesssim(\log X)^2\delta^2\sup_\sigma x^{2\sigma+1}N(\sigma,T).$$

As above, applying (1.4) and the zero-free region, we have that

$$\sup_{\sigma} x^{2\sigma+1} N(\sigma,T) \lesssim x^3 \sup_{\sigma\leq 1-c(\log T)^{-5/7}} \Big(\frac{T^{30/13+o(1)}}{x^2}\Big)^{1-\sigma} \lesssim_\epsilon x^3 \exp(-10\sqrt[4]{\log x}),$$

on recalling that $T=\delta^{-1}\exp(4\sqrt[4]{\log x}) \lessapprox x^{13/15-\epsilon/3}$. Putting this together then gives (13.2), as required.

Remark. *By using a prime decomposition (such as the Heath-Brown Identity) and Mellin inversion, it is possible to relate the count of primes in short intervals directly to Dirichlet polynomials. The critical situation for both Corollary 1.3 and Corollary 1.4 is handling a product of six Dirichlet polynomials each of size roughly* $x^{1/6}$ *(this was the limiting case in the earlier work of Huxley* [Hu] *too). As in* [HB4]*, by bounding this contribution corresponding to six almost equal sized primes using a sieve method, one could obtain an asymptotic estimate in the slightly larger range* $y\in[x^{17/30-\epsilon},x^{0.99}]$ *and* $y\in[X^{2/15-\epsilon},X^{0.99}]$ *at the cost of a worse error term of size roughly* $O(\epsilon^5 y/\log x)$*.*

References

[Bo] J. Bourgain, On large values estimates for Dirichlet polynomials and the density hypothesis for the Riemann zeta function, Internat. Math. Res. Notices, (2000), no. 2, 133–146.

[Ca] F. Carlson. Uber die Nullstellen der Dirichletschen Reihen und der Riemannschen ζ-Funktion. Ark. Mat. Astron. Fys., 15(20):28 pp., 1921

[Da] H. Davenport, *Multiplicative number theory*, third ed.. Springer-Verlag, New York, 2000.

[Ha] G. Halasz, Uber die Mittelwerte multiplikativer zahlentheoretischer Funktionen. Acta Math. Acad. Sci. Hungaricae 19, 365-403 (1968).

[HT] G. Halasz, and P. Turan, On the distribution of roots of Riemann zeta and allied functions, I, Journal of Number Theory 1.1 (1969): 121-137.

[HB] D. R. Heath-Brown, A large values estimate for Dirichlet polynomials. Journal of the London Mathematical Society 2.1 (1979): 8-18.

[HB2] D. R. Heath-Brown, The Differences between Consecutive Primes, II. Journal of the London Mathematical Society, 2.19 (1979): 207-220.

[HB3] D. R. Heath-Brown, Zero Density Estimates for the Riemann Zeta-Function and Dirichlet L-Functions. Journal of the London Mathematical Society, 2.19 (1979): 221-232.

[HB4] D. R. Heath-Brown, Prime Numbers in Short Intervals and a Generalized Vaughan Identity. Canadian Journal of Mathematics, 34.6 (1982): 1365–1377.

[Hu] M. N. Huxley. On the difference between consecutive primes. Invent. Math., 15:164-170, 1972

[In] A. E. Ingham. On the estimation of $N(\sigma,T)$. Quart. J. Math. Oxford Ser., 11:291–292, 1940.

[Iv] A. Ivic. *The Riemann Zeta-Function*, Wiley, 1985 .

[IK] H. Iwaniec and E. Kowalski. *Analytic number theory*, Vol. 53. American Mathematical Soc., 2021.

[Ju] M. Jutila, Zero-density estimates for L-functions, Acta Arith., 32 (1977), 55–62.

[MP] J. Maynard and K. Pratt, Half-isolated zeros and zero-density estimates. arXiv preprint arXiv:2206.11729 (2022).

[M] H. Montgomery, Mean and large values of Dirichlet polynomials. Inventiones mathematicae 8.4 (1969): 334-345.

[M2] H. Montgomery, *Ten Lectures On The Interface Between Analytic Number Theory And Harmonic Analysis*, No. 84. American Mathematical Soc., 1994.

[M3] H. Montgomery. Topics in multiplicative number theory. Lecture Notes in Mathematics, Vol. 227. Springer-Verlag, Berlin-New York, 1971.