Proc. London Math. Soc. (3) 95 (2007) 653–686
doi:10.1112/plms/pdm021

HIGHER CORRELATIONS OF DIVISOR SUMS RELATED TO PRIMES III: SMALL GAPS BETWEEN PRIMES

D. A. GOLDSTON AND C. Y. YILDIRIM

Abstract

We use divisor sums to approximate prime tuples and moments for primes in short intervals. By connecting these results to classical moment problems we are able to prove that, for any $\eta > 0$, a positive proportion of consecutive primes are within $\frac{1}{4} + \eta$ times the average spacing between primes.

Authors' note. This paper was written in 2004, prior to the solution, in [**8**], of the problem considered here. In [**8**] it is shown that $\Delta = 0$. While the main result in Theorem 1 has now been superseded, we believe the method used here is both of interest and future utility in other applications. In particular, the work of Green and Tao [**12**] on arithmetic progressions of primes makes use of Proposition 1 of this paper.

1. *Introduction*

Finding mathematical proofs for easily observed properties of the distribution of prime numbers is a difficult and often humbling task, at least for the authors of this paper. The twin prime conjecture is a famous example of this, but we are concerned here with the much more modest problem of proving that there are arbitrarily large primes that are 'unusually close' together. Statistically this means that we seek consecutive primes whose distance apart is substantially less than the average distance between consecutive primes. Let p_n denote the nth prime; then by the prime number theorem the average gap size $p_{n+1} - p_n$ between consecutive primes is $\log p_n$. Thus we define

$$\Delta = \liminf_{n\to\infty} \left(\frac{p_{n+1} - p_n}{\log p_n} \right), \tag{1.1}$$

so that Δ is the smallest number for which there will be infinitely many gaps between consecutive primes of size less than $\Delta + \epsilon$ times the average size. It is empirically evident that

$$\Delta = 0, \tag{1.2}$$

but at the time of writing this has not been proved. Up to now three different unconditional methods have been invented which provide non-trivial estimates for Δ.

1. *The Hardy–Littlewood and Bombieri–Davenport method.* In the mid-1920s Hardy and Littlewood used the circle method to obtain a conditional result which in 1965 Bombieri and Davenport [**1**] both improved and made unconditional. This approach can be interpreted as a second moment method using a truncated divisor sum as an approximation of $\Lambda(n)$, the von Mangoldt function (see the introduction in [**9**]). The method proves that

$$\Delta \leqslant \tfrac{1}{2}. \tag{1.3}$$

Received 8 November 2005; revised 17 March 2007; published online 16 July 2007.

2000 *Mathematics Subject Classification* 11N05 (primary), 11P32 (secondary).

The research of Goldston was supported by NSF; that of Yıldırım was supported by TÜBİTAK

(In the unpublished paper *Partitio Numerorum VII*, Hardy and Littlewood proved, assuming the Generalized Riemann Hypothesis, that $\Delta \leqslant \frac{2}{3}$. In 1940 Rankin [20] refined the method of Hardy and Littlewood to show that $\Delta \leqslant (1+4\Theta)/5$, where Θ is the supremum of the real parts of all the zeros of all Dirichlet L-functions. In particular, if the Generalized Riemann Hypothesis ($\Theta = \frac{1}{2}$) is assumed, this gives $\Delta \leqslant \frac{3}{5}$.)

2. *The Erdős method.* By the prime number theorem we have $\Delta \leqslant 1$. Erdős [4] in 1940 was the first to prove unconditionally that $\Delta < 1$. He used the sieve upper bound for primes differing by an even number k,

$$\sum_{n \leqslant N} \Lambda(n)\Lambda(n+k) \leqslant (\mathcal{B}+\epsilon)\mathfrak{S}(k)N, \tag{1.4}$$

where $\mathfrak{S}(k)$ is the singular series and $\mathcal{B}$ is a constant. By this bound there can not be too many pairs of primes with the same difference, and therefore the distribution function for prime gaps must spread out from the average. This method gives the result

$$\Delta \leqslant 1 - \frac{1}{2\mathcal{B}}. \tag{1.5}$$

The value $\mathcal{B} = 4$ of Bombieri and Davenport [1] (see also [11] or [13]) or $\mathcal{B} = 3.5$ of Bombieri, Friedlander and Iwaniec [2], or even slightly smaller values may be used here. (The value $\mathcal{B} = 3.5$ only holds for k not too large as a function of N in (1.4), but this is acceptable for (1.5).)

3. *The Maier method.* In 1988 Maier [17] found certain (rather sparsely distributed) intervals where there are e^{γ} more primes than the expected number, and therefore within these intervals the average spacing is reduced by a factor of $e^{-\gamma}$. Hence

$$\Delta \leqslant e^{-\gamma} = 0.56145\ldots . \tag{1.6}$$

In contrast to the first two methods, this method does not produce a positive proportion of small prime gaps.

These three methods may be combined to obtain improved results. Huxley [15, 16] combined the first two methods making use of a weighted version of the first method to find that

$$\Delta \leqslant 0.44254\ldots \quad (\text{using } \mathcal{B}=4), \qquad \Delta \leqslant 0.43494\ldots \quad (\text{using } \mathcal{B}=3.5), \tag{1.7}$$

and Maier combined his method with Huxley's method with $\mathcal{B} = 4$ to obtain

$$\Delta \leqslant (0.44254\ldots)e^{-\gamma} = 0.24846\ldots . \tag{1.8}$$

This last result is the best result known up to now, and as we have seen uses all three of the previously known methods.

For several years we have been developing tools for dealing with higher correlations of short divisor sums which approximate primes. Our first results appeared in [9], and, with considerable help from other mathematicians, we have greatly simplified and improved on these results in [10]. In the former paper we had an application to small gaps between primes based on approximating a third moment. In particular, we recovered the result (1.3). The method is based on the same approximation that underlies the method of Bombieri and Davenport, but it detects small prime gaps in a different way. In this paper we extend that argument to all moments and obtain the limit of this method.

Let $\pi(N)$ denote the number of primes less than or equal to N.

Theorem 1. *Let r be any positive integer. For any fixed $\lambda > \left(\sqrt{r} - \frac{1}{2}\right)^2$ and N sufficiently large, we have*

$$\sum_{\substack{p_n \leqslant N \\ p_{n+r}-p_n \leqslant \lambda \log p_n}} 1 \gg_r \pi(N). \tag{1.9}$$

In particular, for any fixed $\eta > 0$ and all sufficiently large $N > N_0(\eta)$, a positive proportion of gaps $p_{n+1} - p_n$ for $p_n \leqslant N$ are less than $(\frac{1}{4} + \eta)\log N$, and

$$\Delta \leqslant \tfrac{1}{4}. \tag{1.10}$$

Our results depend on the level of distribution of primes in arithmetic progressions, and Theorem 1 makes use of the Bombieri–Vinogradov theorem. If for primes up to N the level of distribution in arithmetic progressions is assumed to be $N^{\vartheta-\epsilon}$ for any $\epsilon > 0$, then Theorem 1 holds with $\lambda > \left(\sqrt{r} - \sqrt{\vartheta/2}\right)^2$. Hence, assuming the Elliott–Halberstam conjecture that $\vartheta = 1$ holds, we obtain the improved result that

$$\liminf_{n\to\infty}\left(\frac{p_{n+r}-p_n}{\log p_n}\right) \leqslant \left(\sqrt{r} - \frac{1}{\sqrt{2}}\right)^2, \tag{1.11}$$

and, in particular,

$$\Delta \leqslant \left(\tfrac{3}{2} - \sqrt{2}\right) = 0.085786\ldots < \tfrac{1}{11}. \tag{1.12}$$

There are several improvements that can be made in our results. First, we can incorporate Maier's method into our method. This is a straightforward adaptation of the argument Maier used to combine his method with Huxley's result, although the result is complicated by the need to prove our propositions in the next section when they are summed over arithmetic progressions. Second, and more significantly, we have found in joint work with J. Pintz better approximations for prime tuples than those used in this paper, and these lead to significantly stronger results. These results will appear in future papers.

This paper is organized as follows. In Section 2 we present our method and state the two main propositions needed in the proof. In Section 3 we prove some lemmas which are used in the later sections. In Sections 4 and 5 we prove the propositions. In Section 6 we examine an optimization problem related to the Poisson distribution which is used in the proof of Theorem 1, and finally in Section 7 we prove Theorem 1.

Notation. We will take ϵ to be any sufficiently small positive number whose value can be changed from equation to equation, and similarly C, c, and c' will denote small fixed positive constants whose value may change from equation to equation. We will let A denote a large positive constant which may be taken as large as we wish, but is fixed throughout the paper. For a finite set $\mathcal{A}$ we let $|\mathcal{A}|$ denote the number of elements in $\mathcal{A}$. We will sometimes write $\mathcal{A} = \mathcal{A}_k$ if $|\mathcal{A}| = k$. For a vector $\mathbf{H}$ we denote the number of components by $|\mathbf{H}|$. A dash in a summation sign $\sum'$ indicates that all the summation variables are relatively prime to each other, and further any sum without a lower bound on the summation variables will have the variables start with the value 1. Empty sums will have the value zero, and empty products will have the value 1. We will make use of the Iverson notation that, for a statement P, $[P]$ is 1 if P is true and is 0 if P is false.

Acknowledgements. We are indebted to Andrew Granville and Kannan Soundararajan who have greatly clarified and simplified our method. We have made extensive use of their results here. We have also benefited from ideas of Enrico Bombieri, Brian Conrey, Percy Deift, David Farmer, John Friedlander, Roger Heath-Brown, Hugh Montgomery, Yoichi Motohashi, Michael Rubinstein, Peter Sarnak, and Jimena Sivak. We have used work of Dashiell Fryer, an undergraduate student at San Jose State University in the MARC program, who worked on properties of Laguerre polynomials needed in our method. The first-named author also thanks the American Institute of Mathematics where much of the collaboration mentioned above took place. In a recent preprint [12] Ben Green and Terence Tao proved a landmark result on arithmetic progressions of primes. One tool they used was the current Proposition 1 from an earlier (not widely distributed) preprint of this paper. They corrected an oversight in our original proof which we have incorporated into our Lemma 2 and the proof of Proposition 1.

2. *Approximating prime tuples*

Our approach for finding small gaps between primes is to compute approximations of the moments for the number of primes in short intervals, and this computation uses short divisor sums to approximate prime tuples. Given a positive integer h, let

$$\mathcal{H} = \{h_1, h_2, \ldots, h_k\}, \qquad \text{with } 0 \leqslant h_1, h_2, \ldots, h_k \leqslant h \text{ distinct integers,} \tag{2.1}$$

and let $\nu_p(\mathcal{H})$ denote the number of distinct residue classes modulo p that the elements of $\mathcal{H}$ occupy. We define the singular series

$$\mathfrak{S}(\mathcal{H}) = \prod_p \left(1 - \frac{1}{p}\right)^{-k} \left(1 - \frac{\nu_p(\mathcal{H})}{p}\right). \tag{2.2}$$

If $\mathfrak{S}(\mathcal{H}) \neq 0$ then $\mathcal{H}$ is said to be *admissible*. Thus $\mathcal{H}$ is admissible if and only if $\nu_p(\mathcal{H}) < p$ for all p.

Letting $\Lambda(n)$ denote the von Mangoldt function, define

$$\Lambda(n; \mathcal{H}) = \Lambda(n + h_1)\Lambda(n + h_2) \ldots \Lambda(n + h_k). \tag{2.3}$$

The Hardy–Littlewood prime tuple conjecture [14] states that for $\mathcal{H}$ admissible,

$$\sum_{n \leqslant N} \Lambda(n; \mathcal{H}) = N\big(\mathfrak{S}(\mathcal{H}) + o(1)\big), \qquad \text{as } N \to \infty. \tag{2.4}$$

(This is trivially true if $\mathcal{H}$ is not admissible.) We approximate $\Lambda(n)$ as in our earlier work [6] by using the truncated divisor sum

$$\Lambda_R(n) = \sum_{\substack{d \mid n \\ d \leqslant R}} \mu(d) \log \frac{R}{d}, \tag{2.5}$$

and then approximate $\Lambda(n; \mathcal{H})$ by

$$\Lambda_R(n; \mathcal{H}) = \Lambda_R(n + h_1)\Lambda_R(n + h_2) \ldots \Lambda_R(n + h_k). \tag{2.6}$$

For convenience we also define $\Lambda_R(n; \mathcal{H}) = 1$ if $\mathcal{H} = \emptyset$. Our method is founded on the following two propositions which allow us to obtain information about primes. Suppose $\mathcal{H}_1$ and $\mathcal{H}_2$ are both sets of distinct positive integers that are less than or equal to h, with $|\mathcal{H}_1| = k_1$ and $|\mathcal{H}_2| = k_2$, and let $k = k_1 + k_2$. We always assume $k \geqslant 1$.

PROPOSITION 1. *Let $\mathcal{H} = \mathcal{H}_1 \cup \mathcal{H}_2$ and $r = |\mathcal{H}_1 \cap \mathcal{H}_2|$. If $R = o(N^{1/k}(\log R)^{1-|\mathcal{H}|/k})$ and $h \leqslant R^A$ for any large constant $A > 0$, then we have for $R, N \to \infty$,*

$$\sum_{n \leqslant N} \Lambda_R(n; \mathcal{H}_1)\Lambda_R(n; \mathcal{H}_2) = N\big(\mathfrak{S}(\mathcal{H}) + o_k(1)\big)(\log R)^r. \tag{2.7}$$

PROPOSITION 2. *Let $\mathcal{H} = \mathcal{H}_1 \cup \mathcal{H}_2$, $r = |\mathcal{H}_1 \cap \mathcal{H}_2|$, and $1 \leqslant h_0 \leqslant h$. Let $\mathcal{H}_0 = \mathcal{H} \cup \{h_0\}$, and $r_0 = r$ if $h_0 \notin \mathcal{H}$ and $r_0 = r+1$ if $h_0 \in \mathcal{H}$. If $R \ll_k N^{1/(2k)}(\log N)^{-B(k)}$ for a sufficiently large positive constant $B(k)$, and $h \leqslant R^{1/(2k)}$, then we have for $R, N \to \infty$,*

$$\sum_{n \leqslant N} \Lambda_R(n;\mathcal{H}_1)\Lambda_R(n;\mathcal{H}_2)\Lambda(n+h_0) = N\big(\mathfrak{S}(\mathcal{H}_0) + o_k(1)\big)(\log R)^{r_0}. \tag{2.8}$$

If the Elliott–Halberstam conjecture is assumed, then equation (2.8) holds for $R \ll_k N^{1/k-\epsilon}$ with any $\epsilon > 0$.

The restriction on the size of R in Proposition 2 can be improved in the situation when $h_0 \in \mathcal{H}_1 \cup \mathcal{H}_2$. If we let $k^* = k - |\mathcal{H}_1 \cap \{h_0\}| - |\mathcal{H}_2 \cap \{h_0\}|$, then we see that the reduction of cases at the start of the proof of Proposition 2 implies that Proposition 2 holds in the range $R \ll_k N^{1/(2k^*)}(\log N)^{-B(k)}$ except in the trivial case when $k = 2$ and $k^* = 0$ where the result holds for $R \leqslant N$. In the case of the Elliott–Halberstam conjecture, k can also be replaced by k^* in the bound for R.

We actually prove both propositions with the error term $o_k(1)$ replaced by a series of lower order terms, which however are not needed in any of our applications.

If we take $\mathcal{H}_2 = \emptyset$ in Proposition 1, we have, for $R = o(N^{1/k})$ and $h \leqslant R^A$ for any given constant $A > 0$, that for $R, N \to \infty$,

$$\sum_{n \leqslant N} \Lambda_R(n;\mathcal{H}) = N\big(\mathfrak{S}(\mathcal{H}) + o_k(1)\big), \tag{2.9}$$

in agreement with the Hardy–Littlewood prime tuple conjecture (2.4).

In applying these propositions it is critical to have some form of positivity in the argument. For example, in the special case when $\mathcal{H}_2 = \emptyset$, Proposition 2 takes the form, for $R \leqslant N^{1/(2k)}(\log N)^{-B(k)}$,

$$\sum_{n \leqslant N} \Lambda_R(n;\mathcal{H})\Lambda(n+h_0) = \begin{cases} N\big(\mathfrak{S}(\mathcal{H}) + o_k(1)\big)\log R & \text{if } h_0 \in \mathcal{H}, \\ N\big(\mathfrak{S}(\mathcal{H}_0) + o_k(1)\big) & \text{if } h_0 \notin \mathcal{H}, \end{cases} \tag{2.10}$$

which would seem to exhibit that our approximation detects primes. However, since $\Lambda_R(n;\mathcal{H})$ is not non-negative, it is impossible to conclude anything about primes from (2.10) alone.

On the other hand, consider instead the special case of Proposition 2 where $\mathcal{H}_1 = \mathcal{H}_2 = \mathcal{H}$ which gives, on taking $|\mathcal{H}| = k$, for $R \leqslant N^{1/(4k)}(\log N)^{-B(k)}$,

$$\sum_{n \leqslant N} \Lambda_R(n;\mathcal{H})^2\Lambda(n+h_0) = \begin{cases} N\big(\mathfrak{S}(\mathcal{H}) + o_k(1)\big)(\log R)^{k+1} & \text{if } h_0 \in \mathcal{H}, \\ N\big(\mathfrak{S}(\mathcal{H}_0) + o_k(1)\big)(\log R)^{k} & \text{if } h_0 \notin \mathcal{H}. \end{cases} \tag{2.11}$$

The restriction on the size of R here makes it impossible to conclude from (2.11) that any given tuple $\mathcal{H}$ will contain two or more primes, but Granville and Soundararajan found a simple argument which uses the non-negativity of $\Lambda_R(n;\mathcal{H})^2$ to prove from (2.11) that

$$\Delta \leqslant \tfrac{3}{4}. \tag{2.12}$$

To prove their result, we need a formula of Gallagher that as $h \to \infty$,

$$\sum_{\substack{1 \leqslant h_1, h_2, \ldots, h_k \leqslant h \\ \text{distinct}}} \mathfrak{S}(\mathcal{H}) = h^k + O_k(h^{k-1/2+\epsilon}). \tag{2.13}$$

(Granville (unpublished) and Montgomery and Soundararajan [**19**] have recently proved more precise results, but these are not needed here.) We fix $k \geqslant 1$; the argument works equally well for any k, and we can take $k = 1$ if we wish. Suppose now that

$$R = N^{1/(4k)}(\log N)^{-B(k)}, \qquad h \ll \log N.$$

By differencing, equation (2.11) continues to hold when the sum on the left-hand side is over $N < n \leqslant 2N$, and therefore we have

$$\sum_{n=N+1}^{2N}\Bigg(\sum_{1\leqslant h_0\leqslant h}\Lambda(n+h_0)\Bigg)\Lambda_R(n;\mathcal{H})^2 \sim kN\mathfrak{S}(\mathcal{H})(\log R)^{k+1} + \sum_{\substack{1\leqslant h_0\leqslant h\\ h_0\neq h_i,\ 1\leqslant i\leqslant k}} N\mathfrak{S}(\mathcal{H}_0)(\log R)^k.$$

Also by Proposition 1,

$$\sum_{n=N+1}^{2N}\Lambda_R(n;\mathcal{H})^2 \sim N\mathfrak{S}(\mathcal{H})(\log R)^k,$$

and therefore we find, on summing over all distinct tuples $1 \leqslant h_1, h_2, \dots, h_k \leqslant h$ and applying (2.13), that, for ρ a fixed number and $h, N \to \infty$,

$$\begin{aligned}\sum_{n=N+1}^{2N}\Bigg(\sum_{1\leqslant h_0\leqslant h}\Lambda(n+h_0)-\rho\log N\Bigg)\Bigg(\sum_{\substack{1\leqslant h_1,h_2,\dots,h_k\leqslant h\\ \text{distinct}}}\Lambda_R(n;\mathcal{H})^2\Bigg)\\ \sim Nh^k(\log R)^k\big(k\log R + h - \rho\log N\big)\\ \sim Nh^k(\log R)^k\big(h - \big(\rho - \tfrac14\big)\log N\big).\end{aligned}$$

Since

$$\Lambda_R(n) \leqslant d(n)\log R \ll n^{\epsilon},$$

we see that the contribution in the sum above from terms where $n + h_0$ is a prime power is $\ll N^{1/2+\epsilon}$ which is negligible, and therefore we may restrict the sum over h_0 to terms where $n + h_0$ is prime. The right-hand side above is positive if $h > (\rho - \frac14)\log N$, which implies with this restriction on h that there is a value of n, with $N < n \leqslant 2N$, such that

$$\sum_{\substack{1\leqslant h_0\leqslant h\\ n+h_0\ \text{prime}}}\log(n+h_0) > \rho\log N.$$

If $\rho > 1$, this implies that for N sufficiently large there are at least two terms in this sum, and thus by taking $\rho \to 1^+$ we obtain (2.12).

The proof of Theorem 1 is a refinement of the above argument, where we detect primes by the square of the linear combination of tuple approximations

$$a_0 + \sum_{j=1}^{k} a_j\Bigg(\sum_{\substack{1\leqslant h_1,h_2,\dots,h_j\leqslant h\\ \text{distinct}}}\Lambda_R(n;\mathcal{H}_j)\Bigg). \tag{2.14}$$

Here the a_j are available to optimize the argument. While it is possible to use (2.14) directly, we have chosen in the proof of Theorem 1 to first approximate moments, which highlights the Poisson model which the prime numbers are thought to satisfy. This method also has the advantage of simplifying the combinatorics that occur in the problem. The moment method leads to an optimization problem which is familiar in the theory of orthogonal polynomials, the solution of which was provided to us by Enrico Bombieri and Percy Deift. The final result that we obtain depends on the asymptotics of the smallest zero of a certain sequence of Laguerre polynomials; these results are obtained by Sturm comparison type theorems and have appeared in the literature; Michael Rubinstein first pointed these out to us.

3. *Lemmas*

We recall the Riemann zeta-function defined for $\mathrm{Re}(s) > 1$ by

$$\zeta(s) = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1}. \tag{3.1}$$

The zeta-function is analytic everywhere except for a simple pole with residue 1 at $s = 1$, and therefore

$$\zeta(s) - \frac{1}{s-1} \tag{3.2}$$

is an entire function. We need to use a classical zero-free region result. By [**23**, Theorem 3.11 and (3.11.8)] there exists a small positive constant C such that $\zeta(\sigma + it) \neq 0$ in the region

$$\sigma \geqslant 1 - \frac{C}{\log(|t| + 2)} \tag{3.3}$$

for all t, and further

$$\zeta(\sigma + it) - \frac{1}{\sigma - 1 + it} \ll \log(|t| + 2), \qquad \frac{1}{\zeta(\sigma + it)} \ll \log(|t| + 2), \tag{3.4}$$

in this region. Let (c) denote the contour $s = c + it$, with $-\infty < t < \infty$, and let $\mathcal{L}$ denote the contour given by

$$s = -\frac{C}{\log(|t| + 2)} + it. \tag{3.5}$$

Lemma 1. *We have, for $R \geqslant 2$ and $c > 0$,*

$$\frac{1}{2\pi i} \int_{(c)} \frac{1}{\zeta(1+s)} \frac{R^s}{s^2}\, ds = 1 + O\big(e^{-c'\sqrt{\log R}}\big), \tag{3.6}$$

and, for any fixed constant B,

$$\int_{\mathcal{L}} \big(\log(|s| + 2)\big)^B \left| \frac{R^s\, ds}{s^2} \right| \ll e^{-c'\sqrt{\log R}}, \tag{3.7}$$

and

$$\int_{(1/\log R)} \big(\log(|s| + 2)\big)^B \left| \frac{R^s\, ds}{s^2} \right| \ll \log R. \tag{3.8}$$

Proof. We first prove (3.7). The integral to be bounded is, for any $w \geqslant 2$,

$$\begin{aligned}
&\ll \int_{-\infty}^{\infty} R^{-C/\log(|t|+2)} \frac{\big(\log(|t|+2)\big)^B}{(|t| + C)^2}\, dt \\
&\ll (\log w)^B \int_0^w R^{-C/\log(t+2)}\, dt + \int_w^{\infty} \frac{(\log t)^B}{t^2}\, dt \\
&\ll (w(\log w)^B) e^{-C\log R/\log w} + \frac{(\log w)^B}{w},
\end{aligned}$$

and on choosing $\log w = \frac{1}{2}\sqrt{C\log R}$ we see that this is

$$\ll (C\log R)^{B/2} e^{-\sqrt{C\log R}/2} \ll e^{-c'\sqrt{\log R}},$$

which proves (3.7).

To prove (3.6), we note that by the second bound in (3.4) the integrand in (3.6) vanishes as $|t| \to \infty$ in the region to the right of $\mathcal{L}$, and therefore we can move the contour (c) to the left to $\mathcal{L}$, pass the simple pole at $s = 0$ with residue 1, and obtain

$$\frac{1}{2\pi i}\int_{(c)} \frac{1}{\zeta(1+s)}\frac{R^s}{s^2}\,ds = 1 + \frac{1}{2\pi i}\int_{\mathcal{L}} \frac{1}{\zeta(1+s)}\frac{R^s}{s^2}\,ds.$$

On $\mathcal{L}$ we deduce from (3.4) that $1/\zeta(1+s) \ll \log(|t|+2)$, and therefore we may use the estimate (3.7) to obtain (3.6). Finally, the left-hand side of (3.8) is

$$\ll \int_{|t|\leqslant 1/\log R} (\log R)^2\,dt + \int_{|t|>1/\log R} \frac{\big(\log(|t|+2)\big)^B}{t^2}\,dt \ll \log R. \qquad \square$$

Lemma 2. *Let $f_R(s_1,s_2)$ be analytic in the strip $-B \leqslant \sigma_1,\sigma_2 \leqslant b$ for some positive constants B and b, and suppose also that, for any $\epsilon > 0$, $f_R(s_1,s_2) \ll e^{\epsilon\sqrt{\log R}}$ in this strip as $|t_1|,|t_2| \to \infty$. For $R \geqslant 2$ and $0 < c_1, c_2 \leqslant c$, let*

$$\mathcal{U}(R) = \frac{1}{(2\pi i)^2}\int_{(c_2)}\int_{(c_1)} f_R(s_1,s_2)\frac{\zeta(1+s_1+s_2)}{\zeta(1+s_1)\zeta(1+s_2)}\frac{R^{s_1+s_2}}{{s_1}^2{s_2}^2}\,ds_1\,ds_2. \tag{3.9}$$

Then

$$\mathcal{U}(R) = f_R(0,0)\log R + \mathcal{C}_R + O\big(e^{-c'\sqrt{\log R}}\big), \tag{3.10}$$

where

$$\mathcal{C}_R = \frac{\partial f_R}{\partial s_2}(0,0) + \frac{1}{2\pi i}\int_{L_2} f_R(-s,s)\frac{ds}{\zeta(1+s)\zeta(1-s)s^4}, \tag{3.11}$$

where L_2 is defined in (3.12) and (3.13) below.

Proof. One would expect to proceed by moving both contours to the left to $\mathcal{L}$. There is, however, a complication because the integrand now contains the function $\zeta(1+s_1+s_2)$ which necessitates that also s_1+s_2 be restricted to the region to the right of $\mathcal{L}$ if we wish to use the bounds in (3.4). (This was pointed out to us by J. Sivak and also Y. Motohashi. This problem was handled in similar ways in [**21**] and in [7]. We follow here our method in [**8**].)

Let

$$V = e^{\sqrt{\log R}} \tag{3.12}$$

and define the contours, for $j = 1$ or 2,

$$\begin{aligned}
L'_j &= \left(\frac{4^{-j}c}{\log V}\right) = \left\{\frac{4^{-j}c}{\log V} + it\colon -\infty < t < \infty\right\},\\
L_j &= \left\{\frac{4^{-j}c}{\log V} + it\colon |t| \leqslant 4^{-j}V\right\},\\
\mathcal{L}_j &= \left\{-\frac{4^{-j}c}{\log V} + it\colon |t| \leqslant 4^{-j}V\right\},\\
H_j &= \left\{\sigma_j \pm i4^{-j}V\colon |\sigma_j| \leqslant \frac{4^{-j}c}{\log V}\right\}.
\end{aligned} \tag{3.13}$$

Now, by the bound for f_R and (3.4), the integrand of $\mathcal{U}(R)$ is

$$\ll e^{\epsilon\sqrt{\log R}}\big(\log(|t_1|+2)\log(|t_2|+2)\big)^2 \max(1,|s_1+s_2|^{-1})R^{\sigma_1+\sigma_2}|s_1|^{-2}|s_2|^{-2}$$

provided s_1, s_2, and s_1+s_2 are on or to the right of $\mathcal{L}$. Thus if s_1 and s_2 are to the right of L'_2, this integrand vanishes as $|t_1| \to \infty$ or $|t_2| \to \infty$, and we may shift the contours c_1 and c_2 to L'_1 and L'_2, respectively, without changing the value of $\mathcal{U}(R)$. Next, we truncate these contours so that they may be replaced with L_1 and L_2; the error introduced by this is, since

$|s_1+s_2| \geqslant 5c/(16\log V)$ on these contours,

$$\ll (\log V)e^{\epsilon\sqrt{\log R}}R^{5c/(16\log V)}\left(\int_{-\infty}^{\infty}\frac{(\log|t|+2)^2}{|c/(16\log V)+it|^2}\,dt\right)\left(\int_{V/16}^{\infty}\frac{(\log t)^2}{t^2}\,dt\right)$$
$$\ll \frac{(\log V)^4}{V^{1-5c/16-\epsilon}} \ll e^{-c'\sqrt{\log R}}.$$

Hence

$$\mathcal{U}(R) = \frac{1}{(2\pi i)^2}\int_{L_2}\int_{L_1} f_R(s_1,s_2)\frac{\zeta(1+s_1+s_2)}{\zeta(1+s_1)\zeta(1+s_2)}\frac{R^{s_1+s_2}}{{s_1}^2{s_2}^2}\,ds_1\,ds_2 + O\big(e^{-c'\sqrt{\log R}}\big). \quad (3.14)$$

To replace the s_1-contour along L_1 with the contour along $\mathcal{L}_1$, we consider the rectangle formed by L_1, H_1 and $\mathcal{L}_1$ which contains poles of the integrand as a function of s_1 at $s_1=0$ and $s_1=-s_2$. Hence we see that

$$\begin{aligned}
\mathcal{U}(R) &= \frac{1}{2\pi i}\int_{L_2}\operatorname*{Res}_{s_1=0}\left(f_R(s_1,s_2)\frac{\zeta(1+s_1+s_2)}{\zeta(1+s_1)\zeta(1+s_2)}\frac{R^{s_1+s_2}}{{s_1}^2{s_2}^2}\right)ds_2\\
&\quad+\frac{1}{2\pi i}\int_{L_2}\operatorname*{Res}_{s_1=-s_2}\left(f_R(s_1,s_2)\frac{\zeta(1+s_1+s_2)}{\zeta(1+s_1)\zeta(1+s_2)}\frac{R^{s_1+s_2}}{{s_1}^2{s_2}^2}\right)ds_2\\
&\quad+\frac{1}{(2\pi i)^2}\int_{L_2}\int_{\mathcal{L}_1\cup H_1} f_R(s_1,s_2)\frac{\zeta(1+s_1+s_2)}{\zeta(1+s_1)\zeta(1+s_2)}\frac{R^{s_1+s_2}}{{s_1}^2{s_2}^2}\,ds_1\,ds_2 + O\big(e^{-c'\sqrt{\log R}}\big)\\
&= \frac{1}{2\pi i}\int_{L_2} f_R(0,s_2)\frac{R^{s_2}}{{s_2}^2}\,ds_2 + \frac{1}{2\pi i}\int_{L_2} f_R(-s_2,s_2)\frac{ds_2}{\zeta(1-s_2)\zeta(1+s_2){s_2}^4}\\
&\quad+\frac{1}{(2\pi i)^2}\int_{L_2}\int_{\mathcal{L}_1\cup H_1} f_R(s_1,s_2)\frac{\zeta(1+s_1+s_2)}{\zeta(1+s_1)\zeta(1+s_2)}\frac{R^{s_1+s_2}}{{s_1}^2{s_2}^2}\,ds_1\,ds_2 + O\big(e^{-c'\sqrt{\log R}}\big)\\
&= I_1+I_2+I_3+O\big(e^{-c'\sqrt{\log R}}\big).
\end{aligned} \quad (3.15)$$

Here the contours along $\mathcal{L}_1$ and H_1 are oriented clockwise. To evaluate I_1, we consider the rectangle formed by L_2, H_2 and $\mathcal{L}_2$ which contains a double pole at $s_2=0$, and obtain, by (3.7) of Lemma 1 and the bound for f_R and (3.4),

$$\begin{aligned}
I_1 &= f_R(0,0)\log R + \frac{\partial f_R}{\partial s_2}(0,0) + \frac{1}{2\pi i}\int_{\mathcal{L}_2\cup H_2} f_R(0,s_2)\frac{R^{s_2}}{{s_2}^2}\,ds_2\\
&= f_R(0,0)\log R + \frac{\partial f}{\partial s_2}(0,0) + O\big(e^{-c'\sqrt{\log R}}\big).
\end{aligned} \quad (3.16)$$

Since I_2 is included in $\mathcal{C}_R$, to complete the proof of Lemma 2 we only need to show that $I_3 \ll e^{-c'\sqrt{\log R}}$. This is done in the same way as we handled I_1; we consider the rectangle formed by L_2, H_2 and $\mathcal{L}_2$, which however in this case contains a simple pole at $s_2=0$, and obtain

$$\begin{aligned}
I_3 &= \frac{1}{2\pi i}\int_{\mathcal{L}_2\cup H_2} f_R(s_1,0)\frac{R^{s_1}}{{s_1}^2}\,ds_1\\
&\quad+\int_{\mathcal{L}_1\cup H_1}\int_{\mathcal{L}_2\cup H_2} f_R(s_1,s_2)\frac{\zeta(1+s_1+s_2)R^{s_1+s_2}}{\zeta(1+s_1)\zeta(1+s_2){s_1}^2{s_2}^2}\,ds_1\,ds_2.
\end{aligned} \quad (3.17)$$

By (3.7) of Lemma 1 and the bound for f_R and (3.4) both of these integrals are $\ll e^{-c'\sqrt{\log R}}$ which completes the proof. □

4. *Proof of Proposition* 1

Let

$$\kappa = |\mathcal{H}_1\cup\mathcal{H}_2|, \qquad r = |\mathcal{H}_1\cap\mathcal{H}_2|, \qquad k = |\mathcal{H}_1|+|\mathcal{H}_2| = k_1+k_2, \quad (4.1)$$

and therefore $0 \leqslant r, k_1, k_2 \leqslant \kappa$ and

$$\kappa = k - r. \tag{4.2}$$

Next, without loss of generality, we take

$$\begin{aligned} &\mathbf{H} = (h_1, h_2, \dots, h_k), \\ &\mathcal{H}_1 = \{h_1, h_2, \dots, h_{k_1}\}, \\ &\mathcal{H}_2 = \{h_{k_1+1}, h_{k_1+2}, \dots, h_k\}, \\ &\mathcal{H}_1 \cap \mathcal{H}_2 = \{h_1, h_2, \dots, h_r\}, \\ &h_1 = h_k, \quad h_2 = h_{k-1}, \quad \dots, \quad h_r = h_{k-r+1} = h_{\kappa+1}, \\ &\mathcal{H} := \mathcal{H}_1 \cup \mathcal{H}_2 = \{h_1, h_2, \dots, h_\kappa\}. \end{aligned} \tag{4.3}$$

Here $r = 0$ when $\mathcal{H}_1 \cap \mathcal{H}_2 = \emptyset$ and the fourth and fifth lines in (4.3) may be removed. With this notation we have

$$\begin{aligned} \mathcal{S}_k(\mathcal{H}_1, \mathcal{H}_2) &= \sum_{n \leqslant N} \Lambda_R(n; \mathcal{H}_1) \Lambda_R(n; \mathcal{H}_2) \\ &= \sum_{n \leqslant N} \sum_{\substack{d_1, d_2, \dots, d_k \leqslant R \\ d_i | n+h_i,\ 1 \leqslant i \leqslant k}} \prod_{i=1}^{k} \mu(d_i) \log \frac{R}{d_i} \\ &= \sum_{d_1, d_2, \dots, d_k \leqslant R} \left(\prod_{i=1}^{k} \mu(d_i) \log \frac{R}{d_i} \right) \sum_{\substack{n \leqslant N \\ d_i | n+h_i,\ 1 \leqslant i \leqslant k}} 1. \end{aligned} \tag{4.4}$$

Let

$$D_k = [d_1, d_2, \dots, d_k], \tag{4.5}$$

the least common multiple of $d_1, d_2, \dots, d_k$. The sum over n above is zero unless $(d_i, d_j) \mid h_j - h_i$, for $1 \leqslant i < j \leqslant k$, in which case the sum runs through a unique residue class modulo D_k, and we have

$$\sum_{\substack{n \leqslant N \\ d_j | n+h_j,\ 1 \leqslant j \leqslant k}} 1 = \frac{N}{D_k} + O(1). \tag{4.6}$$

Hence

$$\begin{aligned} \mathcal{S}_k(\mathcal{H}_1, \mathcal{H}_2) &= N \sum_{\substack{d_1, d_2, \dots, d_k \leqslant R \\ (d_i, d_j) | h_j - h_i,\ 1 \leqslant i < j \leqslant k}} \frac{1}{D_k} \prod_{j=1}^{k} \mu(d_j) \log \frac{R}{d_j} + O(R^k) \\ &= N \mathcal{T}_k(\mathcal{H}_1, \mathcal{H}_2) + O(R^k). \end{aligned} \tag{4.7}$$

We next decompose the d_i into relatively prime factors. Let $\mathcal{P}(k)$ be the set of all non-empty subsets of the set of k elements $\{1, 2, \dots, k\}$. (This is just the power set with the empty set removed.) For $\mathcal{B} \in \mathcal{P}(k)$, we let $\mathcal{P}_{\mathcal{B}}(k)$ denote the set of all members of $\mathcal{P}(k)$ for which $\mathcal{B}$ is a subset. Thus for example if $k = 4$ then

$$\mathcal{P}_{\{1,2\}}(4) = \{\{1,2\}, \{1,2,3\}, \{1,2,4\}, \{1,2,3,4\}\}.$$

Since the d_i are squarefree we can decompose them into the relatively prime factors

$$d_i = \prod_{\nu \in \mathcal{P}_{\{i\}}(k)} a_\nu, \qquad \text{for } 1 \leqslant i \leqslant k, \tag{4.8}$$

where a_ν is the product of all the primes that precisely divide all the d_i for which $i \in \nu$, and none of the other d_i. This decomposition is unique and the $2^k - 1$ factors a_ν are pairwise relatively prime to each other.

We next denote by $\mathcal{D}(\mathbf{H})$ the divisibility conditions

$$(d_i, d_j) = \prod_{\nu \in \mathcal{P}_{\{i,j\}}(k)} a_\nu \;\Big|\; h_j - h_i, \qquad \text{for } 1 \leqslant i < j \leqslant k, \tag{4.9}$$

and have

$$\mathcal{T}_k(\mathcal{H}_1, \mathcal{H}_2) = \sum_{\substack{d_1, d_2, \ldots, d_k \leqslant R \\ \mathcal{D}(\mathbf{H})}} \left(\prod_{\nu \in \mathcal{P}(k)} \frac{\mu(a_\nu)^{|\nu|}}{a_\nu} \right) \left(\prod_{j=1}^{k} \log \frac{R}{d_j} \right). \tag{4.10}$$

We now apply the formula, for $c > 0$,

$$\frac{1}{2\pi i} \int_{c - i\infty}^{c + i\infty} \frac{x^s}{s^2}\, ds = \begin{cases} 0 & \text{if } 0 < x \leqslant 1, \\ \log x & \text{if } x \geqslant 1, \end{cases} \tag{4.11}$$

and have

$$\mathcal{T}_k(\mathcal{H}_1, \mathcal{H}_2) = \frac{1}{(2\pi i)^k} \int_{(c_k)} \cdots \int_{(c_2)} \int_{(c_1)} F(s_1, s_2, \ldots, s_k) \prod_{j=1}^{k} \frac{R^{s_j}}{{s_j}^2}\, ds_j, \tag{4.12}$$

where

$$F(s_1, s_2, \ldots, s_k) = \sideset{}{'}\sum_{\substack{a_\nu,\, \nu \in \mathcal{P}(k) \\ \mathcal{D}(\mathbf{H})}} \prod_{\nu \in \mathcal{P}(k)} \frac{\mu(a_\nu)^{|\nu|}}{{a_\nu}^{1 + \tau_\nu}}, \tag{4.13}$$

and

$$\tau_\nu = \sum_{j \in \nu} s_j. \tag{4.14}$$

(Y. Motohashi has pointed out to us that we could also define

$$F(s_1, s_2, \ldots, s_k) = \prod_p \left(\sum_{\epsilon} \frac{\mu(p^{\epsilon_1}) \ldots \mu(p^{\epsilon_k})}{p^{\epsilon_1} \ldots p^{\epsilon_k} p^{\epsilon_1 s_1 + \ldots + \epsilon_k s_k}} \right)$$

where the subscript ϵ indicates that we sum over the values $0 \leqslant \epsilon_j \leqslant 1$, for $1 \leqslant j \leqslant k$, with the conditions $(p^{\epsilon_i}, p^{\epsilon_j}) \mid h_i - h_j$, for $1 \leqslant i < j \leqslant k$. Using this definition would simplify some of the calculations in this section.)

We next consider the divisibility conditions $\mathcal{D}(\mathbf{H})$. The variables a_ν indexed by the singleton sets $\nu = \{j\}$, with $1 \leqslant j \leqslant k$, are not constrained by these divisibility conditions, and therefore can contain any prime as a factor. Further, if $r \geqslant 1$, then $h_j - h_i = 0$ for $j = k - i + 1$ and $1 \leqslant i \leqslant r$. Thus these constraints drop out of $\mathcal{D}(\mathbf{H})$ and the unconstrained variables are both the singleton sets $\nu = \{j\}$, with $1 \leqslant j \leqslant k$, and also the doubleton sets $\nu = \{i, k - i + 1\}$, with $1 \leqslant i \leqslant r$. (If $r = 0$ there are none of these doubleton sets.) The remaining a_ν are constrained by at least one of the divisibility relations, and therefore must divide some $h_j - h_i$ so that they can only contain prime factors less than or equal to h. We therefore see that we can write

$F(s_1,\dots,s_k)$ as the Euler product, for $\sigma_j>0$, $1\leqslant j\leqslant k$,

$$F(s_1,\dots,s_k)=\prod_{p\leqslant h}\left(1-\sum_{j=1}^{k}\frac{1}{p^{1+s_j}}+f_{\mathbf{H}}(p;s_1,s_2,\dots,s_k)\right)\times\prod_{p>h}\left(1-\sum_{j=1}^{k}\frac{1}{p^{1+s_j}}+\sum_{j=1}^{r}\frac{1}{p^{1+s_j+s_{k-j+1}}}\right), \tag{4.15}$$

where

$$f_{\mathbf{H}}(p;s_1,s_2,\dots,s_k)=\sum_{\substack{\nu\in\mathcal{P}(k),|\nu|\geqslant 2\\ p\mid h_j-h_i \text{ for all } i,j\in\nu}}\frac{(-1)^{|\nu|}}{p^{1+\tau_\nu}}. \tag{4.16}$$

Factoring out the dominant zeta-factors we write

$$F(s_1,s_2,\dots,s_k)=G_{\mathbf{H}}(s_1,s_2,\dots,s_k)\prod_{j=1}^{r}\frac{\zeta(1+s_j+s_{k-j+1})}{\zeta(1+s_j)\zeta(1+s_{k-j+1})}\prod_{j=r+1}^{\kappa}\frac{1}{\zeta(1+s_j)}, \tag{4.17}$$

and proceed to analyze $G_{\mathbf{H}}$. Let

$$\Delta:=\prod_{1\leqslant i<j\leqslant\kappa}|h_j-h_i|\leqslant h^{k^2}, \tag{4.18}$$

so that this product is over all the non-zero differences of h_i and h_j for $1\leqslant i<j\leqslant k$. (Here of course Δ is not the same function as in the first section.) From the discussion above (4.15),

$$f_{\mathbf{H}}=\sum_{j=1}^{r}\frac{1}{p^{1+s_j+s_{k-j+1}}}$$

unless $p\mid\Delta$, and therefore

$$G_{\mathbf{H}}(s_1,s_2,\dots,s_k)=\prod_{p\mid\Delta}\left(\frac{1-\sum_{j=1}^{k}\frac{1}{p^{1+s_j}}+f_{\mathbf{H}}(p;s_1,s_2,\dots,s_k)}{\prod_{j=1}^{k}\left(1-\frac{1}{p^{1+s_j}}\right)\prod_{j=1}^{r}\left(1-\frac{1}{p^{1+s_j+s_{k-j+1}}}\right)^{-1}}\right)h(s_1,s_2,\dots,s_k), \tag{4.19}$$

where

$$h(s_1,s_2,\dots,s_k)=\prod_{p\nmid\Delta}\left(\frac{1-\sum_{j=1}^{k}\frac{1}{p^{1+s_j}}+\sum_{j=1}^{r}\frac{1}{p^{1+s_j+s_{k-j+1}}}}{\prod_{j=1}^{k}\left(1-\frac{1}{p^{1+s_j}}\right)\prod_{j=1}^{r}\left(1-\frac{1}{p^{1+s_j+s_{k-j+1}}}\right)^{-1}}\right). \tag{4.20}$$

Let

$$s^*=-\sum_{j=1}^{k}\min(\sigma_j,0). \tag{4.21}$$

Taking $\sigma_j\geqslant-\frac{1}{5}$, for $1\leqslant j\leqslant k$, we have

$$h(s_1,s_2,\dots s_k)\ll_k\prod_{p}\left(1+O_k\left(\frac{1}{p^{6/5}}\right)\right)\ll_k 1, \tag{4.22}$$

and thus in this region we have

$$\begin{aligned}G_{\mathbf{H}}(s_1, s_2, \ldots, s_k) &\ll_k \prod_{p\mid\Delta}\left(1 + O_k\left(\frac{1}{p^{1-s^*}}\right)\right)\\ &\ll_k \exp\left(a(k)\sum_{p\leqslant k^2\log(2h)}\frac{1}{p^{1-s^*}}\right)\\ &\ll_k \exp\left(a(k)(k^2\log(2h))^{s^*}\sum_{p\leqslant k^2\log(2h)}\frac{1}{p}\right)\\ &\ll_k \exp\big(b(k)(\log(2h))^{s^*}\log\log\log 2h\big),\end{aligned} \tag{4.23}$$

where the sum which was originally over $p \mid \Delta$ has been majorized by replacing these primes by the primes $2, 3, \ldots, p_m$ with $m = \nu(\Delta)$ and using the fact that $2 \cdot 3 \cdot \ldots \cdot p_m \leqslant \Delta$ and (4.18) to see by the prime number theorem that $p_m \leqslant k^2 \log(2h)$. By this bound and (4.17) we see that if $r \geqslant 1$ then F has simple poles at $s_i + s_{k-i+1} = 0$, for $1 \leqslant i \leqslant r$. By (3.4), for s_i with $1 \leqslant i \leqslant k$, and $s_i + s_{k-i+1}$ with $1 \leqslant i \leqslant r$, to the right of $\mathcal{L}$,

$$\begin{aligned}&F(s_1, s_2, \ldots, s_k)\\ &\quad\ll_k \exp\big(b(k)(\log(2h))^{s^*}\log\log\log 2h\big)\prod_{j=1}^{k}\log^2(|t_j|+2)\prod_{i=1}^{r}\max\left(1, \frac{1}{|s_i + s_{k-i+1}|}\right).\end{aligned} \tag{4.24}$$

We are now ready to begin the evaluation of $\mathcal{T}_k(\mathcal{H}_1, \mathcal{H}_2)$. By (4.24) we see that the integrand in (4.12) goes to zero as any one of the variables $|t_j| \to \infty$. We will first move successively the contours (c_j), for $r+1 \leqslant j \leqslant \kappa$, to $\mathcal{L}$; by (4.17) these correspond to the cases where the integrand has only a simple pole at $s_j = 0$. If $r = \kappa$ there are none of these terms and we skip ahead to (4.28). Thus, moving c_{r+1} to $\mathcal{L}$ and passing a simple pole at $s_{r+1} = 0$ we obtain

$$\begin{aligned}\mathcal{T}_k(\mathcal{H}_1, \mathcal{H}_2) &= \frac{1}{(2\pi i)^{k-1}}\left(\prod_{\substack{j=1\\ j\neq r+1}}^{k}\int_{(c_j)}\right) G_{\mathbf{H}}(s_1, s_2, \ldots, s_k)\Big|_{s_{r+1}=0}\left(\prod_{j=r+2}^{\kappa}\frac{R^{s_j}}{\zeta(1+s_j){s_j}^2}\,ds_j\right)\\ &\qquad\times\left(\prod_{j=1}^{r}\frac{\zeta(1+s_j+s_{k-j+1})R^{s_j+s_{k-j+1}}}{\zeta(1+s_j)\zeta(1+s_{k-j+1}){s_j}^2(s_{k-j+1})^2}\,ds_j\,ds_{k-j+1}\right)\\ &\quad+\frac{1}{(2\pi i)^{k}}\left(\prod_{\substack{j=1\\ j\neq r+1}}^{k}\int_{(c_j)}\right)\int_{\mathcal{L}} F(s_1, s_2, \ldots, s_k)\prod_{j=1}^{k}\frac{R^{s_j}}{{s_j}^2}\,ds_j.\end{aligned} \tag{4.25}$$

We bound the second multiple integral on the right by moving all the contours (c_j), with $j \neq r+1$, to $(1/\log R)$ which leaves the value of the integral unchanged. If s_j and s_{k-j+1} are on $(1/\log R)$, then

$$\frac{\zeta(1+s_j+s_{k-j+1})}{\zeta(1+s_j)\zeta(1+s_{k-j+1})} \ll (\log R)\log(2+|s_j+s_{k-j+1}|)\log(2+|s_j|)\log(2+|s_{k-j+1}|). \tag{4.26}$$

In the multiple integral $s^* = -\sigma_{r+1} \leqslant C/\log 2$ for σ_{r+1} on $\mathcal{L}$, and we take a fixed $C < \frac{1}{2}\log 2$. Then by (4.17), (4.24), (4.26) and Lemma 1, the second term in (4.25) is

$$\begin{aligned} &\ll_k \exp\big(b(k)(\log(2h))^{1/2}\big)\left(\int_{(1/\log R)} \log^2(|s|+2)\left|\frac{R^s\,ds}{s^2}\right|\right)^{\kappa-r-1} (\log R)^r \\ &\qquad \times \left(\int_{(1/\log R)} \log^2(|s|+2)\left|\frac{R^s\,ds}{s^2}\right|\right)^{2r} \times \int_{\mathcal{L}} \log^2(|s|+2)\left|\frac{R^s\,ds}{s^2}\right| \\ &\ll_k \exp\big(b(k)(\log(2h))^{1/2}\big)(\log R)^{\kappa+2r-1} e^{-c'\sqrt{\log R}} \\ &\ll_k e^{-c'_k\sqrt{\log R}}, \end{aligned} \tag{4.27}$$

where we used $\log 2h \ll \log R$ for the last line.

We continue this process, next moving (c_{r+2}) to $\mathcal{L}$ in the first term, and estimating the secondary term as above, and so on successively through the contours up to (c_κ). Hence we conclude that

$$\mathcal{T}_k(\mathcal{H}_1, \mathcal{H}_2) = \mathcal{U}_r + O_k\big(e^{-c'_k\sqrt{\log R}}\big), \tag{4.28}$$

where

$$\begin{aligned} \mathcal{U}_r = \frac{1}{(2\pi i)^{2r}} &\left(\prod_{j=1}^{r} \int_{(c_{k-j+1})} \int_{(c_j)}\right) G_1(s_1, s_2, \ldots, s_r, s_{\kappa+1}, s_{\kappa+2}, \ldots s_k) \\ &\times \prod_{j=1}^{r} \frac{\zeta(1+s_j+s_{k-j+1})R^{s_j+s_{k-j+1}}}{\zeta(1+s_j)\zeta(1+s_{k-j+1}){s_j}^2(s_{k-j+1})^2}\, ds_j\, ds_{k-j+1}, \end{aligned} \tag{4.29}$$

and

$$\begin{aligned} &G_1(s_1, s_2, \ldots, s_r, s_{\kappa+1}, s_{\kappa+2}, \ldots, s_k) \\ &\qquad = G_{\mathbf{H}}(s_1, s_2, \ldots, s_r, 0, 0, \ldots, 0, s_{\kappa+1}, s_{\kappa+2}, \ldots, s_k). \end{aligned} \tag{4.30}$$

We will now prove that

$$\mathcal{U}_r = G_{\mathbf{H}}(0,0,\ldots,0)(\log R)^r + \sum_{j=1}^{r} \mathcal{A}_j(\mathcal{H})(\log R)^{r-j} + O_k\big(e^{-c'_k\sqrt{\log R}}\big), \tag{4.31}$$

where the $\mathcal{A}_j(\mathcal{H})$ are explicitly computable arithmetic functions which for $1 \leqslant h \leqslant R^A$ with any $A > 0$ satisfy the bound

$$\mathcal{A}_j(\mathcal{H}) \ll_k (\log\log 2h)^{b(k)}. \tag{4.32}$$

We will also prove at the end of this section that

$$G_{\mathbf{H}}(0,0,\ldots,0) = \mathfrak{S}(\mathcal{H}). \tag{4.33}$$

From these results Proposition 1 follows.

The multiple integral in $\mathcal{U}_r$ would decouple into a product of double integrals evaluated in Lemma 2 if G_1 were a constant, but since this is not the case, we need to apply Lemma 2 inductively. To do this we need estimates for the partial derivatives of $G_{\mathbf{H}}$. Let

$$\mathbf{a} = (a_1, a_2, \ldots, a_k),$$

and define

$$D_{\mathbf{a}}G_{\mathbf{H}} = \frac{\partial^{a_1}}{\partial {s_1}^{a_1}} \frac{\partial^{a_2}}{\partial {s_2}^{a_2}} \cdots \frac{\partial^{a_k}}{\partial {s_k}^{a_k}} G_{\mathbf{H}}(s_1, s_2, \ldots, s_k). \tag{4.34}$$

We have, for $\sigma_j > -{c'}_k$, with $1 \leqslant j \leqslant k$,

$$D_{\mathbf{a}}G_{\mathbf{H}} \ll_k (\log\log 2h)^{b'(k)} \exp\big(b(k)(\log(2h))^{s^*}\log\log\log 2h\big). \tag{4.35}$$

To obtain these estimates, we logarithmically differentiate $G_{\mathbf{H}}$ to see that

$$\frac{\partial G_{\mathbf{H}}}{\partial s_1} \ll_k |G_{\mathbf{H}}| \left(\sum_{p \leqslant k^2 \log(2h)} \frac{\log p}{p^{1-s^*}} \right).$$

The sum above is bounded by

$$\ll (k^2 \log(2h))^{s^*} \sum_{p \leqslant k^2 \log(2h)} \frac{\log p}{p} \ll (k^2 \log(2h))^{s^*} \log(k^2 \log(2h)),$$

and (4.35) follows in this case by (4.23). By the product rule, further partial derivatives will satisfy the above bound with the sum having $\log p$ replaced by $(\log p)^{c(k)}$, which only changes the value of $b'(k)$ in (4.35).

We first consider the case $r = 1$ in (4.29). By Lemma 2 applied with $f_R = G_1$ we see by (4.23) that the conditions of the lemma are satisfied and therefore

$$\mathcal{U}_1 = G_{\mathbf{H}}(0, 0, \ldots, 0) \log R + \mathcal{A}_1 + O_k\big(e^{-c'_k \sqrt{\log R}}\big),$$

where

$$\mathcal{A}_1 = \frac{\partial G_1}{\partial s_k}(0, 0, \ldots, 0) + \frac{1}{2\pi i} \int_{L_2} G_1(-s_1, s_1) \frac{ds_1}{\zeta(1+s_1)\zeta(1-s_1){s_1}^4}.$$

It remains to prove that $\mathcal{A}_1$ satisfies (4.32). By (4.35) the first term in $\mathcal{A}_1$ satisfies this bound. In the integral term we move the contour L_2 to the imaginary axis with a semicircle of radius $\delta = 1/\log(k^2 \log(2h))$ to the right of the double pole at $s_1 = 0$. Using (3.4) and (4.23) we see that the part of the integral over the contour on the imaginary axis is bounded by

$$\ll_k (\log\log 2h)^{b(k)} \int_\delta^\infty \frac{(\log(t+2))^2}{t^4}\, dt \ll_k \frac{(\log\log 2h)^{b(k)}}{\delta^3} \ll_k (\log\log 2h)^{b'(k)},$$

and the integral over the contour on the semicircle is bounded by

$$\ll_k \pi\delta \times \frac{(\log\log 2h)^{3b(k)}}{\delta^2} \ll_k (\log\log 2h)^{b'(k)},$$

which completes the proof for $r = 1$.

For the general case of (4.29), we move all the contours to $(1/\log R)$ and apply Lemma 2 for the double integral over s_1 and s_k to obtain

$$\begin{aligned} \mathcal{U}_r = \frac{1}{(2\pi i)^{2r-2}} \int_{(1/\log R)} \cdots \int_{(1/\log R)} & \left(G_1\Big|_{\substack{s_1=0\\ s_k=0}} \log R + G_2 + O_r\big(e^{-c'_r\sqrt{\log R}}\big) \right) \\ & \times \prod_{j=2}^{r} \frac{\zeta(1+s_j+s_{k-j+1}) R^{s_j+s_{k-j+1}}}{\zeta(1+s_j)\zeta(1+s_{k-j+1})(s_{k-j+1})^2 {s_j}^2}\, ds_j\, ds_{k-j+1} \\ = \mathcal{U}_{r-1} \log R + \mathcal{U}'_{r-1} + O_r\big(e^{-c'_r\sqrt{\log R}}\big), & \end{aligned} \tag{4.36}$$

where the error term was estimated using (4.26) as in (4.27). Here

$$\begin{aligned} G_2(s_2, s_3, \ldots, s_r, s_{\kappa+1}, s_{\kappa+2}, \ldots, s_{k-1}) &= \frac{\partial G_1}{\partial s_k}(0, s_2, s_3, \ldots, s_r, s_{\kappa+1}, s_{\kappa+2}, \ldots, s_{k-1}, 0) \\ &\quad + \frac{1}{2\pi i} \int_{L_2} G_1(-s_1, s_2, \ldots, s_r, s_{\kappa+1}, \ldots, s_{k-1}, s_1) \frac{ds_1}{\zeta(1+s_1)\zeta(1-s_1){s_1}^4}, \end{aligned}$$

and therefore $\mathcal{U}'_{r-1}$ is of the same form as $\mathcal{U}_{r-1}$ with G_1 replaced by a partial derivative of G_1 or an absolutely convergent integral of G_1 with respect to the variable s_1 when $s_k = -s_1$. As we saw in the case $r = 1$, both of these terms satisfy (4.32) and (4.35). We now apply Lemma 2 for the double integral over s_2 and s_{k-1}, and continue this process until all the variables are exhausted. We thus arrive at (4.31) and the bound (4.32) follows by (4.35) and the argument used in the case $r = 1$.

It remains to prove (4.33). By (4.19) and (4.20) we have

$$G_{\mathbf{H}}(0,0,\ldots,0) = \prod_{p|\Delta}\left(1-\frac{k}{p}+f_{\mathbf{H}}(p;0,0,\ldots,0)\right)\left(1-\frac{1}{p}\right)^{-\kappa}\prod_{p\nmid\Delta}\left(1-\frac{\kappa}{p}\right)\left(1-\frac{1}{p}\right)^{-\kappa},$$

where, by (4.16),

$$f_{\mathbf{H}}(p;0,0,\ldots,0) = \frac{1}{p}\sum_{\substack{\nu\in\mathcal{P}(k),|\nu|\geqslant 2\\ p|h_j-h_i \text{ for all } i,j\in\nu}}(-1)^{|\nu|}.$$

Therefore by (2.2) we need to prove that

$$\sum_{\substack{\nu\in\mathcal{P}(k),|\nu|\geqslant 2\\ p|h_j-h_i \text{ for all } i,j\in\nu}}(-1)^{|\nu|} = k-\nu_p(\mathcal{H}). \tag{4.37}$$

If $\nu_p(\mathcal{H}) = q$, then $h_1, h_2, \ldots, h_k$ must fall into q distinct residue classes, say $r_1, r_2, \ldots, r_q$ (mod p). Let

$$\mathcal{M}_p(\ell) = \{m\colon h_m \equiv r_\ell \pmod p\}, \qquad \text{for } 1\leqslant \ell\leqslant q.$$

Thus given p, the sets $\mathcal{M}_p(\ell)$, with $1\leqslant\ell\leqslant q$, give a disjoint partition of the set $\{1,2,\ldots,k\}$, and therefore

$$\sum_{\ell=1}^{q}|\mathcal{M}_p(\ell)| = k. \tag{4.38}$$

The conditions $p \mid h_j - h_i$ hold if and only if h_i and h_j are in the same residue class modulo p and thus if and only if i and j are in $\mathcal{M}_p(\ell)$ for some ℓ. Hence the $\nu\in\mathcal{P}(k)$ which will satisfy $p \mid h_j - h_i$ for all $i,j\in\nu$ are precisely the subsets of $\mathcal{M}_p(\ell)$ with at least two elements

$$\tilde{\mathcal{M}}_p = \bigcup_{\ell=1}^{q}\{\nu\colon \nu\subset\mathcal{M}_p(\ell), |\nu|\geqslant 2\}.$$

We conclude, using (4.38), that

$$\begin{aligned}\sum_{\substack{\nu\in\mathcal{P}(k),|\nu|\geqslant 2\\ p|h_j-h_i \text{ for all } i,j\in\nu}}(-1)^{|\nu|} &= \sum_{\nu\in\tilde{\mathcal{M}}_p}(-1)^{|\nu|}\\ &= \sum_{\ell=1}^{q}\sum_{\nu\subset\mathcal{M}_p(\ell),|\nu|\geqslant 2}(-1)^{|\nu|}\\ &= \sum_{\ell=1}^{q}\sum_{j=2}^{|\mathcal{M}_p(\ell)|}(-1)^j\binom{|\mathcal{M}_p(\ell)|}{j}\\ &= \sum_{\ell=1}^{q}\big(-1+|\mathcal{M}_p(\ell)|\big)\\ &= \sum_{\ell=1}^{q}|\mathcal{M}_p(\ell)| - q\\ &= k-\nu_p(\mathcal{H}).\end{aligned}$$

5. *Proof of Proposition* 2

We first reduce the proof to the special case when $h_0 \notin \mathcal{H} = \mathcal{H}_1 \cup \mathcal{H}_2$. Let

$$\tilde{\mathcal{S}}(\mathcal{H}_1, \mathcal{H}_2, h_0) = \sum_{n \leqslant N} \Lambda_R(n; \mathcal{H}_1)\Lambda_R(n; \mathcal{H}_2)\Lambda(n + h_0). \tag{5.1}$$

Since trivially $|\Lambda_R(n)| \leqslant d(n) \log R$, we see that, for $i = 1, 2$,

$$\Lambda_R(n, \mathcal{H}_i) \leqslant (d(n) \log R)^{k_i}, \tag{5.2}$$

and since $d(n) \ll n^{\epsilon}$ and, in Proposition 2, $R \leqslant N^{1/2}$, we have

$$\begin{aligned}\tilde{\mathcal{S}}(\mathcal{H}_1, \mathcal{H}_2, h_0) &= \sum_{R<n\leqslant N} \Lambda_R(n; \mathcal{H}_1)\Lambda_R(n; \mathcal{H}_2)\Lambda(n + h_0) + O(R^{1+\epsilon}) \\ &= \sum_{\substack{R<n\leqslant N \\ n+h_0 \text{ prime}}} \Lambda_R(n; \mathcal{H}_1)\Lambda_R(n; \mathcal{H}_2)\Lambda(n + h_0) + O(N^{1/2+\epsilon}), \end{aligned}\tag{5.3}$$

where we have removed the prime powers in the last line. If $n + h_0$ is a prime greater than R then its only divisor less than or equal to R is $d = 1$, and therefore

$$\Lambda_R(n + h_0)\Lambda(n + h_0) = \Lambda(n + h_0) \log R.$$

Thus, if $h_0 \in \mathcal{H}_1 \cap \mathcal{H}_2$,

$$\tilde{\mathcal{S}}(\mathcal{H}_1, \mathcal{H}_2, h_0) = (\log R)^2 \sum_{\substack{R<n\leqslant N \\ n+h_0 \text{ prime}}} \Lambda_R(n; \mathcal{H}_1 - \{h_0\})\Lambda_R(n; \mathcal{H}_2 - \{h_0\}) \log(n + h_0) + O(N^{1/2+\epsilon});$$

if $h_0 \in \mathcal{H}_1 - \mathcal{H}_2$,

$$\tilde{\mathcal{S}}(\mathcal{H}_1, \mathcal{H}_2, h_0) = \log R \sum_{\substack{R<n\leqslant N \\ n+h_0 \text{ prime}}} \Lambda_R(n; \mathcal{H}_1 - \{h_0\})\Lambda_R(n; \mathcal{H}_2) \log(n + h_0) + O(N^{1/2+\epsilon});$$

and if $h_0 \notin \mathcal{H}_1 \cup \mathcal{H}_2$,

$$\tilde{\mathcal{S}}(\mathcal{H}_1, \mathcal{H}_2, h_0) = \sum_{\substack{R<n\leqslant N \\ n+h_0 \text{ prime}}} \Lambda_R(n; \mathcal{H}_1)\Lambda_R(n; \mathcal{H}_2) \log(n + h_0) + O(N^{1/2+\epsilon}).$$

In these sums we may once again include the terms that are less than or equal to R and the prime powers if we wish with the same error term, and therefore in each situation we have reduced the proof to the case when h_0 is distinct from the other h_i. Henceforth we can therefore take

$$h_0 \notin \mathcal{H}. \tag{5.4}$$

Proceeding to the proof, we have

$$\tilde{\mathcal{S}}(\mathcal{H}_1, \mathcal{H}_2, h_0) = \sum_{d_1,d_2,\ldots,d_k\leqslant R} \left(\prod_{i=1}^{k} \mu(d_i) \log \frac{R}{d_i}\right) \sum_{\substack{n\leqslant N \\ d_i|n+h_i,\ 1\leqslant i\leqslant k}} \Lambda(n + h_0). \tag{5.5}$$

By the Chinese Remainder theorem, the sum will run through an arithmetic progression modulo D_k provided $(d_i, d_j) \mid h_j - h_i$, for $1 \leqslant i < j \leqslant k$, and will be empty otherwise. As in (4.9) we denote these conditions by $\mathcal{D}(\mathcal{H})$. Using Iverson notation, we let

$$\psi(x; q, a) := \sum_{\substack{n\leqslant x \\ n\equiv a (\text{mod } q)}} \Lambda(n) = [(a, q) = 1]\frac{x}{\phi(q)} + E(x; q, a), \tag{5.6}$$

and have

$$\sum_{\substack{n\leqslant N\\ d_i|n+h_i,\ 1\leqslant i\leqslant k}} \Lambda(n+h_0) = [\mathcal{D}(\mathcal{H})]\big(\psi(N+h_0;D_k,a)-\psi(h_0,D_k,a)\big)$$
$$= [\mathcal{D}(\mathcal{H})]\psi(N;D_k,a)+O(h\log N), \tag{5.7}$$

where a is an integer satisfying the congruence relations $a\equiv h_0-h_j \pmod{d_j}$, for $1\leqslant j\leqslant k$. The term $\psi(N;D_k,a)$ has a non-zero main term if $(a,D_k)=1$, which is equivalent to

$$(d_j,h_j-h_0)=1, \qquad \text{for } 1\leqslant j\leqslant k, \tag{5.8}$$

and, if $(a,D_k)>1$ then $\psi(N;D_k,a)\ll(\log N)^2$; thus

$$\begin{aligned}\tilde{\mathcal{S}}(\mathcal{H}_1,\mathcal{H}_2,h_0) = N &\sum_{\substack{d_1,d_2,\ldots,d_k\leqslant R\\ \mathcal{D}(\mathcal{H})\\ (d_j,h_j-h_0)=1,\ 1\leqslant j\leqslant k}} \frac{1}{\phi(D_k)}\left(\prod_{i=1}^{k}\mu(d_i)\log\frac{R}{d_i}\right)\\ &+O\left(\sum_{d_1,d_2,\ldots,d_k\leqslant R}\left(\prod_{i=1}^{k}\mu^2(d_i)\log\frac{R}{d_i}\right)\max_{\substack{a(\text{mod } D_k)\\(a,D_k)=1}}|E(N;D_k,a)|\right)\\ &+O\big(R^kh(\log N)^2\big)\\ = N&\tilde{\mathcal{T}}_k(\mathcal{H}_1,\mathcal{H}_2,h_0)+O(\mathcal{E}_k)+O\big(R^kh(\log N)^2\big).\end{aligned} \tag{5.9}$$

We handle the error term $\mathcal{E}_k$ with the Bombieri–Vinogradov theorem. First, we have

$$\begin{aligned}\mathcal{E}_k &\ll (\log R)^k\sum_{d_1,d_2,\ldots,d_k\leqslant R}\mu^2(D_k)\max_{\substack{a(\text{mod } D_k)\\(a,D_k)=1}}|E(N;D_k,a)|\\ &\ll (\log R)^k\sum_{q\leqslant R^k}\mu^2(q)\max_{\substack{a(\text{mod } q)\\(a,q)=1}}|E(N;q,a)|\sum_{\substack{q=D_k\\ d_1,d_2,\ldots,d_k\leqslant R}}1.\end{aligned}$$

Given q, the number of ways to write $q=D_k$ (that is, write q as the least common multiple of k squarefree numbers) is bounded by $d(q)^k$, since each of the k numbers in the least common multiple must be a divisor of q. Applying Cauchy's inequality, we have

$$\begin{aligned}\mathcal{E}_k &\ll (\log R)^k\sum_{q\leqslant R^k}\mu^2(q)d(q)^k\max_{\substack{a(\text{mod } q)\\(a,q)=1}}|E(N;q,a)|\\ &\ll (\log R)^k\sqrt{\sum_{q\leqslant R^k}\frac{d(q)^{2k}}{q}}\sqrt{\sum_{q\leqslant R^k}q\max_{\substack{a(\text{mod } q)\\(a,q)=1}}|E(N;q,a)|^2}\,.\end{aligned}$$

We now use the estimate

$$\sum_{n\leqslant x}d(n)^k\ll_k x(\log x)^{2^k-1} \tag{5.10}$$

and the trivial estimate $|E(N;q,a)|\ll(N\log N)/q$ for $q\leqslant N$ to conclude that

$$\mathcal{E}_k\ll_k(\log R)^{4^k+k}\sqrt{N\log N}\sqrt{\sum_{q\leqslant R^k}\max_{\substack{a(\text{mod } q)\\(a,q)=1}}|E(N;q,a)|}\,.$$

By the Bombieri–Vinogradov theorem the last sum over q is $\ll N/(\log N)^A$ for any $A>0$ provided that

$$R^k\leqslant N^{1/2}(\log N)^{-B}, \tag{5.11}$$

where $B = B(A)$. We conclude under this condition that

$$\mathcal{E}_k \ll_k N(\log N)^{4^k+k+1/2-A/2} = o_k(N) \tag{5.12}$$

if $A > 2(4^k + k + \frac{1}{2})$. To complete the proof of the proposition we will now prove that, for $R^k \leqslant N$ and $h \leqslant R^{1/(2k)}$,

$$\tilde{\mathcal{T}}_k(\mathcal{H}_1, \mathcal{H}_2, h_0) = \mathcal{T}_{k+1}(\mathcal{H}_1 \cup \{h_0\}, \mathcal{H}_2) + O_k\big(e^{-c_k\sqrt{\log R}}\big), \tag{5.13}$$

which by (4.28), (4.32), and (4.33) completes the proof. To prove (5.13), we have

$$\begin{aligned}
&\mathcal{T}_{k+1}(\mathcal{H}_1 \cup \{h_0\}, \mathcal{H}_2)\\
&\quad = \sum_{\substack{d_0,d_1,\ldots,d_k \leqslant R\\ (d_i,d_j)|h_j-h_i,\ 0\leqslant i<j\leqslant k}} \frac{1}{[d_0, D_k]} \prod_{j=0}^{k} \mu(d_j) \log \frac{R}{d_j}\\
&\quad = \sum_{\substack{d_1,d_2,\ldots,d_k \leqslant R\\ (d_i,d_j)|h_j-h_i,\ 1\leqslant i<j\leqslant k}} \left(\prod_{j=1}^{k} \mu(d_j) \log \frac{R}{d_j}\right) \sum_{\substack{d_0 \leqslant R\\ (d_0,d_j)|h_j-h_0,\ 1\leqslant j\leqslant k}} \frac{\mu(d_0)}{[d_0, D_k]} \log \frac{R}{d_0}\\
&\quad = \sum_{\substack{d_1,d_2,\ldots,d_k \leqslant R\\ \mathcal{D}(\mathcal{H})}} \left(\prod_{j=1}^{k} \mu(d_j) \log \frac{R}{d_j}\right) T_1(\mathcal{H}_1 \cup \{h_0\}, \mathcal{H}_2).
\end{aligned} \tag{5.14}$$

On letting $g = (d_0, D_k)$ where $d_0 = gd'$, we see that $[d_0, D_k] = D_k d'$ and $(d', D_k) = 1$. Thus

$$\begin{aligned}
T_1(\mathcal{H}_1 \cup \{h_0\}, \mathcal{H}_2) &= \sum_{\substack{gd' \leqslant R\\ g|D_k\\ (g,d_j)|h_j-h_0,\ 1\leqslant j\leqslant k\\ (d',D_k)=1}} \frac{\mu(gd')}{d' D_k} \log \frac{R}{gd'}\\
&= \frac{1}{D_k} \sum_{\substack{g \leqslant R\\ g|D_k\\ (g,d_j)|h_j-h_0,\ 1\leqslant j\leqslant k}} \mu(g) \sum_{\substack{d' \leqslant R/g\\ (d',D_k)=1}} \frac{\mu(d')}{d'} \log \frac{R/g}{d'}.
\end{aligned}$$

For $\log m \ll \log R$ we have (by the prime number theorem or see [9, Lemma 2.1])

$$\sum_{\substack{d \leqslant R\\ (d,m)=1}} \frac{\mu(d)}{d} \log \frac{R}{d} = \frac{m}{\phi(m)} + O\big(e^{-c_1\sqrt{\log R}}\big). \tag{5.15}$$

Applying this and dropping the redundant condition $g \leqslant R$ since $g \leqslant \prod_{j=1}^{k}(g, d_j) \leqslant h^k \leqslant R$ when $h \leqslant R^{1/k}$, we see that

$$T_1(\mathcal{H}_1 \cup \{h_0\}, \mathcal{H}_2) = \frac{1}{\phi(D_k)} \sum_{\substack{g|D_k\\ (g,d_j)|h_j-h_0,\ 1\leqslant j\leqslant k}} \mu(g) + O\left(\frac{d(D_k)}{D_k} e^{-c_1\sqrt{\log(R/h^k)}}\right).$$

We now claim that, using Iverson notation,

$$\sum_{\substack{g|D_k\\ (g,d_j)|h_j-h_0,\ 1\leqslant j\leqslant k}} \mu(g) = [(d_j, h_j - h_0) = 1,\ 1 \leqslant j \leqslant k].$$

One way to see this is through the decomposition of the d_i into relative factors (4.8) from which we see that we can write $g = \prod_{\nu\in\mathcal{P}(k)} b_\nu$, where $b_\nu \mid a_\nu$ with the b_ν pairwise relatively prime to each other. Then the sum becomes

$$\prod_{j=1}^{k} \prod_{\nu \in \mathcal{P}_{\{j\}}(k)} \sum_{b_\nu | (d_j, h_j - h_0)} \mu(b_\nu) = [(d_j, h_j - h_0) = 1,\ 1 \leqslant j \leqslant k].$$

We conclude that

$$T_1(\mathcal{H}_1 \cup \{h_0\}, \mathcal{H}_2) = \frac{1}{\phi(D_k)}[(d_j, h_j - h_0) = 1,\ 1 \leqslant j \leqslant k] + O_k\left(\frac{d(D_k)}{D_K} e^{-c_1\sqrt{\log(R/h^k)}}\right),$$

and on substituting this result into (5.14) we have

$$\begin{aligned}\mathcal{T}_{k+1}(\mathcal{H}_1 \cup \{h_0\}, \mathcal{H}_2) = &\sum_{\substack{d_1, d_2, \ldots, d_k \leqslant R \\ \mathcal{D}(k) \\ (d_j, h_j - h_0) = 1,\ 1 \leqslant j \leqslant k}} \frac{1}{\phi(D_k)} \left(\prod_{j=1}^{k} \mu(d_j) \log \frac{R}{d_j}\right) \\ &+ O_k\left((\log R)^k e^{-c_1\sqrt{\log R/h^k}} \sum_{d_1, d_2, \ldots, d_k \leqslant R} \frac{d(D_k)}{D_k}\right).\end{aligned}$$

The first term is $\tilde{\mathcal{T}}_k(\mathcal{H}_1, \mathcal{H}_2, h_0)$ and, by (5.10),

$$\begin{aligned}\sum_{d_1, d_2, \ldots, d_k \leqslant R} \frac{d(D_k)}{D_k} &\ll \sum_{q \leqslant R^k} \frac{d(q)}{q} \sum_{q = D_k} 1 \\ &\ll \sum_{q \leqslant R^k} \frac{d(q)^{k+1}}{q} \\ &\ll_k (\log R)^{2^{k+1}}.\end{aligned}$$

Thus the error term is

$$\ll_k (\log R)^{2^{k+1}+k} e^{-c_1\sqrt{\log(R/h^k)}} \ll_k e^{-c_k\sqrt{\log R}},$$

which proves (5.13) if $h \leqslant R^{1/(2k)}$.

6. *Optimization of a quadratic form related to the Poisson distribution*

The content of this section and the proof given here was provided to us by E. Bombieri and P. Deift. The final tool we need for our proof of Theorem 1 is an optimization procedure related to the Poisson distribution. Let X be a Poisson random variable with expected value λ, defined by the discrete probability density function

$$p(j) = \text{Prob.}(X = j) = e^{-\lambda}\frac{\lambda^j}{j!}, \qquad \text{for } j = 0, 1, 2, \ldots\ . \tag{6.1}$$

We define an inner product with respect to this density function by

$$\begin{aligned}\langle f(x), g(x)\rangle &= \sum_{j=0}^{\infty} f(j) g(j) p(j) \\ &= e^{-\lambda} \sum_{j=0}^{\infty} f(j) g(j) \frac{\lambda^j}{j!}.\end{aligned} \tag{6.2}$$

The kth moment of the Poisson distribution is defined by

$$\begin{aligned}\mu_k(\lambda) = E(x^k) = \langle x^k, 1\rangle &= e^{-\lambda}\sum_{j=0}^{\infty}\frac{j^k\lambda^j}{j!}\\ &= e^{-\lambda}\left(\lambda\frac{d}{d\lambda}\right)^k\sum_{j=0}^{\infty}\frac{\lambda^j}{j!}\\ &= e^{-\lambda}\left(\lambda\frac{d}{d\lambda}\right)^k e^{\lambda}. \end{aligned} \tag{6.3}$$

More explicitly, we have

$$\mu_k(\lambda) = \sum_{\nu=1}^{k}\left\{ {k \atop \nu} \right\}\lambda^{\nu}, \tag{6.4}$$

where $\left\{ {k \atop \nu} \right\}$ denotes the Stirling numbers of the second type, defined to be the number of ways to partition a k-set (that is, a set with k elements) into ν non-empty subsets (not counting the order of the subsets). It is easy to see that

$$\left\{ {k \atop \nu} \right\} = \nu\left\{ {k-1 \atop \nu} \right\} + \left\{ {k-1 \atop \nu-1} \right\} \tag{6.5}$$

since the last element in our k-set either is put into its own singleton set or is put into one of the ν subsets which contain some of the earlier elements. To prove (6.4) we use the identity

$$\sum_{\nu=0}^{j}\nu!\left\{ {k \atop \nu} \right\}\binom{j}{\nu} = j^k. \tag{6.6}$$

This identity arises from counting the number of partitions of a k-set into at most j sets, where the order of these sets is counted. On one hand there are j choices for where to place each of the k elements, so this number is j^k, while on the other hand, if ν of these j sets are non-empty, then there are $\nu!\left\{ {k \atop \nu} \right\}$ such partitions and $\binom{j}{\nu}$ ways to choose the ν non-empty sets. Rewriting (6.6) in the form

$$\frac{j^k}{j!} = \sum_{\nu=0}^{j}\left\{ {k \atop \nu} \right\}\frac{1}{(j-\nu)!},$$

multiplying by $\lambda^j e^{-\lambda}$ and summing over j, we obtain, by (6.3),

$$\begin{aligned}\mu_k(\lambda) &= e^{-\lambda}\sum_{j=0}^{\infty}\frac{j^k\lambda^j}{j!}\\ &= e^{-\lambda}\sum_{j=0}^{\infty}\lambda^j\sum_{\nu=0}^{j}\left\{ {k \atop \nu} \right\}\frac{1}{(j-\nu)!}\\ &= \sum_{\nu=0}^{k}\left\{ {k \atop \nu} \right\}\lambda^{\nu},\end{aligned}$$

by interchanging the j and ν summations, which proves (6.4).

Our method for finding small gaps between primes leads us to define a second bilinear form given by

$$\begin{aligned}\langle f(x), g(x)\rangle_{\rho} &= \langle x-\rho, f(x)g(x)\rangle\\ &= \sum_{j=0}^{\infty}(j-\rho)f(j)g(j)p(j)\end{aligned} \tag{6.7}$$

where ρ is a real number. (This is not an inner product because it is not necessarily non-negative.) Letting $\mathbf{a} = (a_0, a_1, a_2, \ldots, a_k)$, consider

$$P_{\mathbf{a}}(x) = \sum_{i=0}^{k} a_i x^i, \tag{6.8}$$

and the associated quadratic form

$$\begin{aligned} Q = Q_{\mathbf{a}}(\lambda, \rho) &= \langle P_{\mathbf{a}}(x), P_{\mathbf{a}}(x)\rangle_\rho \\ &= \sum_{0\leqslant i,j\leqslant k} a_i a_j \langle x-\rho, x^{i+j}\rangle \\ &= \sum_{0\leqslant i,j\leqslant k} a_i a_j \left(\mu_{i+j+1}(\lambda) - \rho\mu_{i+j}(\lambda)\right) \\ &= \sum_{0\leqslant i,j\leqslant k} a_i a_j c_{i+j}, \end{aligned} \tag{6.9}$$

where we define

$$c_m = c_m(\lambda, \rho) = \mu_{m+1}(\lambda) - \rho\mu_m(\lambda). \tag{6.10}$$

The optimization problem we need to solve is to maximize Q over all vectors normalized by $a_k = 1$ when $\rho > 0$ is fixed. The solution involves the (generalized) Laguerre polynomials defined for $\alpha > -1$ by

$$L_n{}^{(\alpha)}(x) = \sum_{\nu=0}^{n} (-1)^\nu \binom{n+\alpha}{n-\nu} \frac{x^\nu}{\nu!}. \tag{6.11}$$

The zeros of the Laguerre polynomials are real, positive, and simple (see [**22**, Chapter 6]). We denote the smallest zero of $L_n{}^{(\alpha)}(x)$ by $x_1(n, \alpha)$. The solution of our problem is obtained in the following proposition.

PROPOSITION 3. *For each $k \geqslant 1$ and $\rho > k$ fixed, we have, for $0 < \lambda < x_1(k+1, \rho-k-1)$,*

$$\max_{a_k=1} Q_{\mathbf{a}}(\lambda, \rho) = -(k+1)!\lambda^k \frac{L_{k+1}{}^{(\rho-k-1)}(\lambda)}{L_k{}^{(\rho-k)}(\lambda)}. \tag{6.12}$$

Thus, for each $k \geqslant 1$ and $\rho > k$,

$$\inf\{\lambda > 0 \colon Q_{\mathbf{a}}(\lambda, \rho) > 0, a_k = 1\} = x_1(k+1, \rho-k-1). \tag{6.13}$$

The proof of this proposition will ultimately reduce to evaluating the determinant

$$D_k = \det \begin{vmatrix} c_0 & c_1 & c_2 & \cdots & c_k \\ c_1 & c_2 & c_3 & \cdots & c_{k+1} \\ c_2 & c_3 & c_4 & \cdots & c_{k+2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ c_k & c_{k+1} & c_{k+2} & \cdots & c_{2k} \end{vmatrix} = \det\left[c_{i+j}\right]_{\substack{i=0,1,2,\ldots,k \\ j=0,1,2,\ldots,k}}. \tag{6.14}$$

The solution of the optimization problem can be obtained by choosing $\mathbf{a}$ so that $P_{\mathbf{a}}(x)$ is orthogonal to all lower degree polynomials with respect to $\langle\ ,\ \rangle_\rho$. Thus we consider the k equations

$$\langle P_{\mathbf{a}}(x), x^i\rangle_\rho = 0, \qquad \text{for } i = 0, 1, 2, \ldots, k-1, \tag{6.15}$$

and prove the following lemma.

LEMMA 3. *If $D_{k-1} \neq 0$ for a given λ, then there is an (explicitly obtained) vector $\mathbf{a}$ with $a_k = 1$ which satisfies (6.15) and for which*

$$Q_{\mathbf{a}}(\lambda, \rho) = \frac{D_k}{D_{k-1}}. \tag{6.16}$$

Proof. We take $a_k = 1$. Equation (6.15) is equivalent to the equations

$$\sum_{j=0}^{k} a_j c_{i+j} = 0, \qquad \text{for } i = 0, 1, 2, \ldots, k-1. \tag{6.17}$$

If $\mathbf{a}$ satisfies these equations, then with δ_{ij} denoting the Kronecker delta, we have

$$\begin{aligned} Q &= \sum_{i=0}^{k} a_i \left(\sum_{j=0}^{k} a_j c_{i+j} \right) \\ &= \sum_{i=0}^{k} a_i \left(\delta_{ik} \sum_{j=0}^{k} a_j c_{i+j} \right) \\ &= \sum_{j=0}^{k} a_j c_{j+k}. \end{aligned} \tag{6.18}$$

On rewriting (6.17) in the form

$$\begin{aligned} c_0 a_0 + c_1 a_1 + c_2 a_2 + \ldots + c_{k-1} a_{k-1} &= -c_k, \\ c_1 a_0 + c_2 a_1 + c_3 a_2 + \ldots + c_k a_{k-1} &= -c_{k+1}, \\ &\vdots \\ c_{k-1} a_0 + c_k a_1 + c_{k+1} a_2 + \ldots + c_{2k-2} a_{k-1} &= -c_{2k-1}, \end{aligned} \tag{6.19}$$

we have by Cramer's rule (see [**24**]) that these equations have the solution

$$a_j = -\frac{D_{k-1}^{(j+1)}}{D_{k-1}}, \qquad \text{for } j = 0, 1, \ldots, k-1, \tag{6.20}$$

provided that $D_{k-1} \neq 0$, where $D_{k-1}^{(i)}$ is the determinant with the ith column of D_{k-1} replaced by the column $(c_k, c_{k+1}, \ldots, c_{2k-1})$. Thus (6.18) gives with this choice

$$Q = \frac{1}{D_{k-1}} \left(-\sum_{j=0}^{k-1} D_{k-1}^{(j+1)} c_{k+j} + D_{k-1} c_{2k} \right). \tag{6.21}$$

On the other hand, if we expand D_k into its cofactor expansion along the bottom row, we see that

$$D_k = \sum_{j=0}^{k} (-1)^{k+j} D_{k+1,j+1} c_{k+j},$$

where the minor $D_{i,j}$ is the determinant of the matrix where the ith row and jth column of D_k is removed. From (6.14) we see that

$$D_{k+1,j+1} = (-1)^{k-j-1} D_{k-1}^{(j+1)}, \quad \text{for } 1 \leqslant j \leqslant k-1, \qquad D_{k+1,k+1} = D_{k-1},$$

where the factor $(-1)^{k-j-1}$ results from shifting the last column of D_k by $k-j-1$ places to the left. Hence we conclude that

$$Q = \frac{D_k}{D_{k-1}}, \tag{6.22}$$

which is the required result. □

Our next lemma evaluates D_k.

LEMMA 4. *We have*

$$D_{k-1} = (-1)^k 1!\,2!\,3!\,\ldots\,k!\,\lambda^{k(k-1)/2} L_k{}^{(\rho-k)}(\lambda). \tag{6.23}$$

Proof. We first claim that

$$D_{k-1} = (-1)^k E_k \tag{6.24}$$

where

$$E_k = \det\begin{vmatrix} \mu_0 & \mu_1 & \mu_2 & \ldots & \mu_k \\ \mu_1 & \mu_2 & \mu_3 & \ldots & \mu_{k+1} \\ \mu_2 & \mu_3 & \mu_4 & \ldots & \mu_{k+2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mu_{k-1} & \mu_k & \mu_{k+1} & \ldots & \mu_{2k-1} \\ 1 & \rho & \rho^2 & \ldots & \rho^k \end{vmatrix} = \det\begin{vmatrix} \mu_{i+j} \\ \rho^j \end{vmatrix}_{\substack{i=0,1,2,\ldots,k-1 \\ j=0,1,2,\ldots,k}}, \tag{6.25}$$

for if in E_k we multiply the ℓth column by ρ and subtract this from the $(\ell+1)$th column for $\ell = 1, 2, \ldots, k$ we obtain

$$E_k = \det\begin{vmatrix} \mu_0 & c_0 & c_1 & \ldots & c_{k-1} \\ \mu_1 & c_1 & c_2 & \ldots & c_k \\ \mu_2 & c_2 & c_3 & \ldots & c_{k+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mu_{k-1} & c_{k-1} & c_k & \ldots & c_{2k-2} \\ 1 & 0 & 0 & \ldots & 0 \end{vmatrix},$$

and using the cofactor expansion along the bottom row gives $E_k = (-1)^k D_{k-1}$.

We now introduce the differential operators

$$D = \frac{d}{d\lambda}, \qquad \delta = \lambda D = \lambda\frac{d}{d\lambda}, \qquad \Delta = \delta + \lambda = \lambda\frac{d}{d\lambda} + \lambda. \tag{6.26}$$

Clearly we have the relations

$$\delta^k = \lambda^k D^k + \sum_{j=1}^{k-1} a_j(\lambda) D^j, \qquad \Delta^k = \delta^k + \sum_{j=0}^{k-1} b_j(\lambda)\delta^j \tag{6.27}$$

where $a_j(\lambda)$ and $b_j(\lambda)$ are polynomials of degree j in λ. Now by (6.4) and (6.5) we have $\mu_k = \Delta\mu_{k-1}$, and in general,

$$\mu_k = \Delta^i \mu_{k-i}, \quad \text{for } 0 \leqslant i \leqslant k, \qquad \mu_k = \Delta^k 1. \tag{6.28}$$

From this we see that

$$E_k = \det\begin{vmatrix} \mu_0 & \mu_1 & \mu_2 & \ldots & \mu_k \\ \Delta\mu_0 & \Delta\mu_1 & \Delta\mu_2 & \ldots & \Delta\mu_k \\ \Delta^2\mu_0 & \Delta^2\mu_1 & \Delta^2\mu_3 & \ldots & \Delta^2\mu_k \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \Delta^{k-1}\mu_0 & \Delta^{k-1}\mu_1 & \Delta^{k-1}\mu_2 & \ldots & \Delta^{k-1}\mu_k \\ 1 & \rho & \rho^2 & \ldots & \rho^k \end{vmatrix} = \det\begin{vmatrix} \Delta^i\mu_j \\ \rho^j \end{vmatrix}_{\substack{i=0,1,2,\ldots,k-1 \\ j=0,1,2,\ldots,k}}. \tag{6.29}$$

By the second relation in (6.27) we can replace Δ^i by δ^i and a linear combination of lower powers of δ, which can be eliminated by row operations. Thus we can replace Δ by δ in the above determinant without affecting its value, and then by the first relation in (6.27) and row operations we can replace δ^i by $\lambda^i D^i$ which on removing the factors of λ in each row gives

$$\begin{aligned} E_k &= \lambda^{k(k-1)/2} \det \begin{vmatrix} \mu_0 & \mu_1 & \mu_2 & \dots & \mu_k \\ D\mu_0 & D\mu_1 & D\mu_2 & \dots & D\mu_k \\ D^2\mu_0 & D^2\mu_1 & D^2\mu_3 & \dots & D^2\mu_k \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ D^{k-1}\mu_0 & D^{k-1}\mu_1 & D^{k-1}\mu_2 & \dots & D^{k-1}\mu_k \\ 1 & \rho & \rho^2 & \dots & \rho^k \end{vmatrix} \\ &= \lambda^{k(k-1)/2} \det \begin{vmatrix} D^i\mu_j \\ \rho^j \end{vmatrix}_{\substack{i=0,1,2,\dots,k-1 \\ j=0,1,2,\dots,k}} . \end{aligned} \tag{6.30}$$

We next need the relation

$$q(q-1)\dots(q-h+1) = h! \binom{q}{h} = \sum_{j=0}^{h} (-1)^{h-j} \begin{bmatrix} h \\ j \end{bmatrix} q^j, \tag{6.31}$$

where $\left[{h \atop j} \right]$ are the Stirling numbers of the first type, although we do not need to use any properties of these numbers. Then we have, by (6.3),

$$\begin{aligned} \lambda^h = e^{\lambda}\lambda^h e^{-\lambda} = \sum_{q=0}^{\infty} \frac{\lambda^{q+h}}{q!} e^{-\lambda} &= \sum_{q=0}^{\infty} q(q-1)\dots(q-h+1)\frac{\lambda^q}{q!} e^{-\lambda} \\ &= \sum_{q=0}^{\infty} \left(\sum_{j=0}^{h} (-1)^{h-j} \begin{bmatrix} h \\ j \end{bmatrix} q^j \right) \frac{\lambda^q}{q!} e^{-\lambda} \\ &= \sum_{j=0}^{h} (-1)^{h-j} \begin{bmatrix} h \\ j \end{bmatrix} \mu_j(\lambda). \end{aligned} \tag{6.32}$$

Thus, using column operations we have

$$\begin{aligned} E_k &= \lambda^{k(k-1)/2} \det \begin{vmatrix} 1 & \lambda & \lambda^2 & \dots & \lambda^k \\ D1 & D\lambda & D\lambda^2 & \dots & D\lambda^k \\ D^2 1 & D^2\lambda & D^2\lambda^2 & \dots & D^2\lambda^k \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ D^{k-1}1 & D^{k-1}\lambda & D^{k-1}\lambda^2 & \dots & D^{k-1}\lambda^k \\ 1 & 1!\binom{\rho}{1} & 2!\binom{\rho}{2} & \dots & k!\binom{\rho}{k} \end{vmatrix} \\ &= \lambda^{k(k-1)/2} \det \begin{vmatrix} D^i\lambda^j \\ j!\binom{\rho}{j} \end{vmatrix}_{\substack{i=0,1,2,\dots,k-1 \\ j=0,1,2,\dots,k}} . \end{aligned} \tag{6.33}$$

Expanding along the bottom row we see that

$$\det \begin{vmatrix} D^i\lambda^j \\ j!\binom{\rho}{j} \end{vmatrix}_{\substack{i=0,1,2,\dots,k-1 \\ j=0,1,2,\dots,k}} = \sum_{h=0}^{k} (-1)^{k-h} h! \binom{\rho}{h} \det[D^i\lambda^j]_{\substack{i=0,1,\dots,k-1 \\ j=0,1,\dots,k; j\neq h}} . \tag{6.34}$$

We will show below that

$$F_h = \det[D^i\lambda^j]_{\substack{i=0,1,\dots,k-1 \\ j=0,1,\dots,k; j\neq h}} = 1!\,2!\,\dots\,(k-1)! \binom{k}{h} \lambda^{k-h} \tag{6.35}$$

which then gives on retracing our steps

$$\begin{aligned}D_{k-1} &= (-1)^k \lambda^{k(k-1)/2} 1!\,2!\,\ldots\,(k-1)! \sum_{h=0}^{k} (-1)^{k-h} h! \binom{\rho}{h} \binom{k}{h} \lambda^{k-h} \\ &= (-1)^k \lambda^{k(k-1)/2} 1!\,2!\,\ldots\,k! L_k{}^{(\rho-k)}(\lambda),\end{aligned}$$

which proves Lemma 4.

We prove (6.35) by the following argument shown to us by Wasin So. We consider the complete upper triangular matrix

$$\begin{aligned}M = [D^i \lambda^j]_{\substack{i=0,1,\ldots,k\\ j=0,1,\ldots,k}} &= \begin{pmatrix} 1 & \lambda & \lambda^2 & \ldots & \lambda^k \\ D1 & D\lambda & D\lambda^2 & \ldots & D\lambda^k \\ D^2 1 & D^2\lambda & D^2\lambda^2 & \ldots & D^2\lambda^k \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ D^k 1 & D^k\lambda & D^k\lambda^2 & \ldots & D^k\lambda^k \end{pmatrix} \\ &= \left[i! \binom{j}{i} \lambda^{j-i} \right]_{\substack{i=0,1,\ldots,k\\ j=0,1,\ldots,k}}.\end{aligned}$$

Observe that $\det M = 1!\,2!\,\ldots k!$, and further that

$$M = TP, \qquad \text{where } T = [\delta_{ij} i!\,]_{\substack{i=0,1,\ldots,k\\ j=0,1,\ldots,k}} \text{ and } P = \left[\binom{j}{i} \lambda^{j-i} \right]_{\substack{i=0,1,\ldots,k\\ j=0,1,\ldots,k}}.$$

Now by the matrix inverse formula using minors we have

$$M^{-1} = \frac{1}{\det M} [(-1)^{i+j} D_{j,i}]_{\substack{i=0,1,\ldots,k\\ j=0,1,\ldots,k}},$$

where F_h occurs in this matrix as the minor $D_{k,h}$. Further,

$$M^{-1} = P^{-1} T^{-1},$$

where

$$T^{-1} = \left[\delta_{ij} \frac{1}{i!} \right]_{\substack{i=0,1,\ldots,k\\ j=0,1,\ldots,k}}, \qquad P^{-1} = \left[(-1)^{j-i} \binom{j}{i} \lambda^{j-i} \right]_{\substack{i=0,1,\ldots,k\\ j=0,1,\ldots,k}},$$

where we used the identity

$$\sum_{s=0}^{k} (-1)^s \binom{j}{s} \binom{s}{i} = (-1)^i \delta_{ij}.$$

From this last relation we have, letting $M^{-1} = [\bar{m}_{ij}]$,

$$F_h = (-1)^{k+h} (\det M) \bar{m}_{hk} = (-1)^{k+h} \frac{(-1)^{k-h}}{k!} 1!\,2!\,\ldots\,k! \binom{k}{h} \lambda^{k-h},$$

as desired. □

Proof of Proposition 3. Let $\mathbf{a}$ be the solution for (6.15) found in Lemma 3, which exists for any λ where $D_{k-1} \neq 0$, and let $\mathbf{b}$ be any other k-vector with $b_k = 1$. Then $P_{\mathbf{b}-\mathbf{a}}(x)$ is a polynomial of degree $k-1$ or less, and by the orthogonality property (6.15),

$$\begin{aligned}Q_{\mathbf{b}}(\lambda, \rho) &= \langle P_{\mathbf{a}}(x) + P_{\mathbf{b}-\mathbf{a}}(x), P_{\mathbf{a}}(x) + P_{\mathbf{b}-\mathbf{a}}(x) \rangle_\rho \\ &= \langle P_{\mathbf{a}}(x), P_{\mathbf{a}}(x) \rangle_\rho + \langle P_{\mathbf{b}-\mathbf{a}}(x), P_{\mathbf{b}-\mathbf{a}}(x) \rangle_\rho \\ &= Q_{\mathbf{a}}(\lambda, \rho) + Q_{\mathbf{b}-\mathbf{a}}(\lambda, \rho).\end{aligned}$$

In general by (6.7) and (6.9) for any $\mathbf{c} \neq \mathbf{0}$, assuming $\rho > 0$ is fixed, we have

$$\begin{aligned} Q_{\mathbf{c}}(\lambda, \rho) &= \sum_{j=0}^{\infty} (j-\rho)(P_{\mathbf{c}}(j))^2 p(j) \\ &= -\rho\langle (P_{\mathbf{c}}(x))^2, 1\rangle + O_{\mathbf{c}}(\lambda) \\ &< 0, \qquad \text{for } 0 < \lambda \leqslant \lambda_0(\mathbf{c}, \rho), \end{aligned}$$

where $\lambda_0(\mathbf{c}, \rho)$ is a small positive constant depending on $\mathbf{c}$. Thus

$$Q_{\mathbf{b}}(\lambda, \rho) \leqslant Q_{\mathbf{a}}(\lambda, \rho)$$

for $0 < \lambda < \lambda_0(\mathbf{c})$, proving that $Q_{\mathbf{a}}$ is maximal at least for small enough λ. This will continue to be true for larger λ as long as $Q_{\mathbf{c}} < 0$ for any $(k-1)$-vector $\mathbf{c}$, and therefore as long as the maximal Q for $(k-1)$-vectors is negative. By (5.1.14) of Szegö [**22**] we have

$$\frac{d}{dx} L_n{}^{(\alpha)}(x) = -L_{n-1}{}^{(\alpha+1)}(x),$$

and therefore we see that the sequence $\{L_k{}^{(\rho-k)}\}$ of Laguerre polynomials has the property that the negative of the derivative of a term is the previous term. (Thus the negative derivative of the Laguerre polynomial in the numerator in (6.12) is the Laguerre polynomial in the denominator.) Further, in this sequence of Laguerre polynomials the polynomials are all decreasing functions up to their first positive zero, and hence the sequence of smallest positive zeros $x_1(k, \rho-k)$ is a decreasing sequence. Starting with the trivial case when $k = 1$ we see successively that the Q_k with $\mathbf{a}$ satisfying (6.15) will be maximal for $0 < \lambda < x_1(k+1, \rho-k-1)$. This completes the proof of Proposition 3. □

Our next result evaluates the smallest positive zero $x_1(n, \alpha)$ asymptotically as $n \to \infty$.

LEMMA 5. *Let $L_n{}^{(\alpha)}(x)$, with $\alpha > -1$, denote the Laguerre polynomials. The zeros of $L_n{}^{(\alpha)}(x)$ are real, positive, and simple. Let $x_1(n, \alpha)$ denote the smallest zero of $L_n{}^{(\alpha)}(x)$. If $\alpha = \beta(n) - n$ and $\lim_{n\to\infty}(\beta(n)/n) = A > 0$, then*

$$\lim_{n\to\infty} \frac{x_1(n, \alpha)}{n} = \left(\sqrt{A} - 1\right)^2. \tag{6.36}$$

Proof. The properties of $L_n{}^{(\alpha)}(x)$ may be found in [**22**] by Szegö. Equation (6.36) is a special case of [**3**, Theorem 4.4]. A simple proof may be obtained by using the same argument as that found in [**18**] where a result corresponding to (6.36) for Jacobi polynomials is proved using Sturm comparison theory. By [**22**, (5.1.2)], the differential equation

$$u'' + \left(\frac{n + (\alpha+1)/2}{x} + \frac{1-\alpha^2}{4x^2} - \frac{1}{4} \right) u = 0 \tag{6.37}$$

has $u = e^{-x/2} x^{(\alpha+1)/2} L_n{}^{(\alpha)}(x)$ as a solution. Let

$$\begin{aligned} H_n(x) &:= \frac{n + (\alpha+1)/2}{x} + \frac{1-\alpha^2}{4x^2} - \frac{1}{4} \\ &= -\frac{x^2 - (4n + 2(\alpha+1))x + (\alpha^2 - 1)}{4x^2}, \end{aligned}$$

and denote the smaller root of the quadratic in the numerator by ${x_n}^-$. Then by the Sturm comparison argument in [**18**], and noting that $\lim_{n\to\infty}(\alpha/n) = A - 1$, we have

$$\begin{aligned}\lim_{n\to\infty}\frac{x_1(n,\alpha)}{n} &= \lim_{n\to\infty}\frac{{x_n}^-}{n}\\ &= \lim_{n\to\infty}\frac{2n+\alpha+1-\sqrt{4n^2+2\alpha+2+4n\alpha+4n}}{n}\\ &= 2+\lim_{n\to\infty}\frac{\alpha}{n}-\sqrt{4+4\lim_{n\to\infty}\frac{\alpha}{n}}\\ &= \big(\sqrt{A}-1\big)^2.\end{aligned}$$

□

7. *Gaps between primes*

In this section we prove Theorem 1. We want to examine statistically the number of primes in the interval $(n, n+h]$ for $N < n \leqslant 2N$ with $N \to \infty$. In this range the average distance between consecutive primes is $\log N$, and thus we will take h to be a multiple of this length. We therefore let

$$\psi(x) = \sum_{n\leqslant x}\Lambda(n), \tag{7.1}$$

$$\psi(n,h) = \psi(n+h) - \psi(n), \tag{7.2}$$

$$h = \lambda\log N, \tag{7.3}$$

and in this paper we assume that

$$\lambda \ll 1. \tag{7.4}$$

The model for our method is due to Gallagher [**5**], who proved that if the Hardy–Littlewood conjecture (2.4) holds uniformly for $h \ll \log N$ then one can asymptotically evaluate all the moments for the number of primes in intervals of length h. Thus assuming (2.4), Gallagher proved that

$$M_k(N,h,\psi) := \frac{1}{N(\log N)^k}\sum_{n=N+1}^{2N}(\psi(n,h))^k \sim \mu_k(\lambda), \tag{7.5}$$

as $N \to \infty$, where $\mu_k(\lambda)$ is the Poisson moment from (6.3) and (6.4).

In order to obtain unconditional results we make use of our approximation $\Lambda_R(n;\mathcal{H})$, where $\mathcal{H}$ is the set formed by the distinct numbers among $h_1, h_2, \ldots, h_k$. Taking $N < n \leqslant 2N$, we first need to approximate

$$\begin{aligned}\psi(n,h)^k &= \sum_{1\leqslant h_1,h_2,\ldots,h_k\leqslant h}\Lambda(n+h_1)\Lambda(n+h_2)\ldots\Lambda(n+h_k)\\ &= (1+o(1))\sum_{1\leqslant h_1,h_2,\ldots,h_k\leqslant h}(\log N)^{k-|\mathcal{H}|}\Lambda(n;\mathcal{H}).\end{aligned} \tag{7.6}$$

To define our approximation, we extend the definition of $\Lambda_R(n;\mathcal{H})$ in (2.6) to vectors (or lists) $\mathbf{H} = (h_1, h_2, \ldots, h_k)$ by letting

$$\Lambda_R(n;\mathbf{H}) := (\log R)^{k-|\mathcal{H}|}\Lambda_R(n;\mathcal{H}). \tag{7.7}$$

Thus our approximation of $\psi(n,h)^k$ is

$$\psi_R{}^{(k)}(n,h) := \sum_{1\leqslant h_1,h_2,\ldots,h_k\leqslant h}\Lambda_R(n;\mathbf{H}). \tag{7.8}$$

For convenience we also define ${\psi_R}^{(0)}(n,h) = 1$. We next define the approximate moments, letting $k = i + j$,

$$M_{ij}(R) = \frac{1}{N(\log R)^k} \sum_{n=N+1}^{2N} {\psi_R}^{(i)}(n,h){\psi_R}^{(j)}(n,h), \tag{7.9}$$

and note that $M_{00}(R) = 1$. We also need the mixed moments

$$\tilde{M}_{ij}(R) = \frac{1}{N(\log R)^{k+1}} \sum_{n=N+1}^{2N} {\psi_R}^{(i)}(n,h){\psi_R}^{(j)}(n,h)\psi(n,h), \tag{7.10}$$

for which we note by the prime number theorem that $\tilde{M}_{00}(R) \sim \lambda/\theta = \mu_1(\lambda/\theta)$, in accord with (7.11) and (7.13) below.

Using Propositions 1 and 2 we will prove asymptotically that these approximate moments are also Poisson moments with an increased expected value involving the truncation level R. Define θ by

$$R = N^\theta. \tag{7.11}$$

Proposition 4. *As $N \to \infty$ we have, for $k = i + j \geqslant 1$ and for any fixed $0 < \theta < 1/k$,*

$$M_{ij}(R) = (1 + o_k(1))\mu_k\left(\frac{\lambda}{\theta}\right), \tag{7.12}$$

and for any fixed $0 < \theta < 1/(2k)$,

$$\tilde{M}_{ij}(R) = (1 + o_k(1))\mu_{k+1}\left(\frac{\lambda}{\theta}\right). \tag{7.13}$$

Proof. By differencing, Propositions 1 and 2 continue to hold unchanged when we sum for $N < n \leqslant 2N$. We first extend Proposition 1 for vectors $\mathbf{H}_1$ and $\mathbf{H}_2$. Recalling the notation $|\mathbf{H}|$ which denotes the number of components of the vector $\mathbf{H}$, let $k = |\mathbf{H}_1| + |\mathbf{H}_2|$ and $\mathcal{H} = \mathcal{H}_1 \cup \mathcal{H}_2$, where $\mathcal{H}_i$ is the set of distinct components of $\mathbf{H}_i$. Then by (7.7) and Proposition 1 (note that the k in Propositions 1 and 2 is equal to $|\mathcal{H}_1| + |\mathcal{H}_2|$ here), we have for $R = o(N^{1/k})$,

$$\begin{aligned} \sum_{n=N+1}^{2N} \Lambda_R(n;\mathbf{H}_1)\Lambda_R(n;\mathbf{H}_2) &= (\log R)^{k-|\mathcal{H}_1|-|\mathcal{H}_2|} \sum_{n=N+1}^{2N} \Lambda_R(n;\mathcal{H}_1)\Lambda_R(n;\mathcal{H}_2) \\ &= N\big(\mathfrak{S}(\mathcal{H}) + o(1)\big)(\log R)^{k-|\mathcal{H}|}. \end{aligned} \tag{7.14}$$

Thus we see that this result depends on k and not the individual values of $|\mathbf{H}_1|$ and $|\mathbf{H}_2|$. Hence, letting $h_1, h_2, \ldots, h_k$ list the components of $\mathbf{H}_1$ and $\mathbf{H}_2$ (in any order), we have

$$\begin{aligned} M_{ij}(R) &= \frac{1}{N(\log R)^k} \sum_{1\leqslant h_1,h_2,\ldots,h_k\leqslant h} \sum_{n=N+1}^{2N} \Lambda_R(n;\mathbf{H}_1)\Lambda_R(n;\mathbf{H}_2) \\ &= \sum_{1\leqslant h_1,h_2,\ldots,h_k\leqslant h} \big(\mathfrak{S}(\mathcal{H}) + o_k(1)\big)(\log R)^{-|\mathcal{H}|}, \end{aligned}$$

provided $R = o(N^{1/k})$. We group terms in this sum according to the number of distinct values ν of $h_1, h_2, \ldots, h_k$, and denote these distinct values by ${h'}_1, {h'}_2, \ldots, {h'}_\nu$. There are $\left\{{k \atop \nu}\right\}$ ways to partition the k values h_i into these ν disjoint sets, and all of these will occur in the sum above.

Hence by (2.13) we have

$$\begin{aligned} M_{ij}(R) &= \sum_{\nu=1}^{k} \left\{ {k \atop \nu} \right\} h^{\nu}(1+o_k(1))(\log R)^{-\nu} \\ &= (1+o_k(1))\mu_k\left(\frac{\lambda}{\theta}\right) \end{aligned}$$

which proves the first part of Proposition 4. The second part is proved identically using Proposition 2. □

Now consider

$$\begin{aligned} \mathcal{S}_k &= \mathcal{S}_k(N,R,\lambda,\rho) \\ &= \frac{1}{N(\log R)^{2k+1}} \sum_{n=N+1}^{2N} (\psi(n,h) - \rho\log N)\big(P_k(\psi_R(n,h))\big)^2, \end{aligned} \tag{7.15}$$

where

$$P_k(\psi_R(n,h)) = \sum_{\ell=0}^{k} a_\ell {\psi_R}^{(\ell)}(n,h)(\log R)^{k-\ell}, \tag{7.16}$$

and the a_ℓ are arbitrary functions of N, R, k, λ, and ρ which are to be chosen to optimize the argument. On multiplying out we have

$$\begin{aligned} \mathcal{S}_k &= \frac{1}{N(\log R)^{2k+1}} \sum_{0\leqslant i,j\leqslant k} a_i a_j (\log R)^{2k-i-j} \\ &\quad \times \sum_{n=N+1}^{2N} \left(\psi(n,h) - \frac{\rho}{\theta}\log R\right) {\psi_R}^{(i)}(n,h){\psi_R}^{(j)}(n,h) \\ &= \sum_{0\leqslant i,j\leqslant k} a_i a_j \mathcal{M}_{ij}. \end{aligned} \tag{7.17}$$

Letting

$$\tilde{\lambda} = \frac{\lambda}{\theta}, \qquad \tilde{\rho} = \frac{\rho}{\theta}, \tag{7.18}$$

we have by Proposition 4, on taking $i+j=\kappa$ and assuming that $0<\theta<1/(2\kappa)$,

$$\begin{aligned} \mathcal{M}_{ij} &= \tilde{M}_{ij}(R) - \frac{\rho}{\theta}M_{ij}(R) \\ &= \mu_{\kappa+1}(\tilde{\lambda}) - \tilde{\rho}\mu_\kappa(\tilde{\lambda}) + o_\kappa\left(\frac{1}{\theta^{\kappa+1}}\right) \\ &= c_\kappa(\tilde{\lambda},\tilde{\rho}) + o_\kappa\left(\frac{1}{\theta^{\kappa+1}}\right), \end{aligned} \tag{7.19}$$

using the notation of (6.10) in the last line. To evaluate $\mathcal{S}_k$ we need to apply these results for $0\leqslant\kappa\leqslant 2k$, all of which will hold if we impose the condition

$$\frac{1}{4k+1} \leqslant \theta < \frac{1}{4k}. \tag{7.20}$$

Thus

$$\begin{aligned} \mathcal{S}_k &= \sum_{0\leqslant i,j\leqslant k} a_i a_j c_{i+j}(\tilde{\lambda},\tilde{\rho}) + o_k\left(\max_{1\leqslant\ell\leqslant k} |a_\ell|^2\right) \\ &= Q_{\mathbf{a}}(\tilde{\lambda},\tilde{\rho}) + o_k(1), \end{aligned} \tag{7.21}$$

since $\max_{1\leqslant \ell\leqslant k}|a_\ell|^2$ depends only on k for fixed λ and ρ. By Proposition 3 we obtain a sign change for $Q_{\mathbf{a}}(\tilde\lambda,\tilde\rho)$ at the smallest zero $x_1(k+1,\tilde\rho-k-1)$ of the Laguerre polynomial $L_{k+1}{}^{(\tilde\rho-k-1)}(\tilde\lambda)$, with $Q_{\mathbf{a}}(\tilde\lambda,\tilde\rho)$ negative for $0<\tilde\lambda<x_1(k+1,\tilde\rho-k-1)$ and positive for $x_1(k+1,\tilde\rho-k-1)<\tilde\lambda<x_1(k,\tilde\rho-k)$. Therefore by (7.21), $\mathcal{S}_k$ will also be positive for

$$x_1(k+1,\tilde\rho-k-1)+o_k(1)<\tilde\lambda<x_1(k,\tilde\rho-k)-o_k(1)\qquad \text{as } N\to\infty.$$

We apply Lemma 5 with $\beta(k)=\tilde\rho$; if we take sequences $\theta=\theta_k\to(1/(4k))^-$ and $\rho=\rho_k\to r^+$ as $k\to\infty$, then $A=4r$, and there exist constants $0<c_k<c'_k$, and $c_k,c'_k\to 0$, such that for

$$\left(\sqrt{r}-\tfrac{1}{2}\right)^2+c_k\leqslant\lambda\leqslant\left(\sqrt{r}-\tfrac{1}{2}\right)^2+c'_k$$

we have

$$\mathcal{S}_k\gg_k 1,\qquad \mathcal{S}_k>0. \tag{7.22}$$

Note that the Laguerre polynomials are well defined here since, by (7.20), $\tilde\rho-k>0$. The proof of Theorem 1 is now a standard deduction from (7.22); we follow our earlier proof in the last section of [9]. Define

$$Q_r^+(N,h)=\sum_{\substack{n=N+1\\ \pi(n+h)-\pi(n)>r}}^{2N} 1. \tag{7.23}$$

If n is an integer for which $\pi(n+h)-\pi(n)>r$ then there must be a j such that $n\leqslant p_j$ and $p_{j+r}\leqslant n+h$. Thus $p_{j+r}-p_j\leqslant h$ and $p_{j+r}-h\leqslant n\leqslant p_j<p_{j+r}$, so that there are at most h such values of n corresponding to each such gap. Therefore

$$Q_r^+(N,h)\ll_r h\sum_{\substack{N<p_n\leqslant 2N\\ p_{n+r}-p_n\leqslant h}} 1+O\big(Ne^{-c\sqrt{\log N}}\big), \tag{7.24}$$

where we have used the prime number theorem to remove the prime gaps overlapping the endpoints N and $2N$. (This can be done more explicitly as in [9].)

Next, we have, for N sufficiently large,

$$Q_r^+(N,h)=\sum_{\substack{n=N+1\\ \psi(n+h)-\psi(n)\geqslant\rho\log N}}^{2N} 1+O(N^{1/2}), \tag{7.25}$$

where ρ can be taken to be any number in the range $r<\rho<r+1$, and the error term comes from removing prime powers. By (7.25) and Cauchy's inequality we see that

$$\begin{aligned}
\mathcal{S}_k &\leqslant \frac{1}{N(\log R)^{2k+1}}\sum_{\substack{n=N+1\\ \psi(n+h)-\psi(n)\geqslant\rho\log N}}^{2N}\psi(n,h)\big(P_k(\psi_R(n,h))\big)^2\\
&\leqslant \frac{1}{N(\log R)^{2k+1}}\left(\sum_{\substack{n=N+1\\ \psi(n+h)-\psi(n)\geqslant\rho\log N}}^{2N}\big(P_k(\psi_R(n,h))\big)^2\right)^{1/2}\\
&\quad\times\left(\sum_{n=N+1}^{2N}\psi(n,h)^2\big(P_k(\psi_R(n,h))\big)^2\right)^{1/2}\\
&\leqslant \frac{\sqrt[4]{Q_r^+(N,h)+O(N^{1/2})}}{N(\log R)^{2k+1}}\left(\sum_{n=N+1}^{2N}\psi(n,h)^4\right)^{1/4}\left(\sum_{n=N+1}^{2N}\big(P_k(\psi_R(n,h))\big)^4\right)^{1/2}.
\end{aligned} \tag{7.26}$$

Hence, provided (7.22) holds we have

$$Q_r^+(N,h) + O(N^{1/2}) \gg_k \frac{\left(N(\log R)^{2k+1}\right)^4}{\left(\sum_{n=N+1}^{2N} \psi(n,h)^4\right)\left(\sum_{n=N+1}^{2N} \left(P_k(\psi_R(n,h))\right)^4\right)^2}. \tag{7.27}$$

We will prove below that subject to $h \ll \log N$ from (7.3) and (7.4) we have

$$\sum_{n=N+1}^{2N} \psi(n,h)^4 \ll N(\log N)^4 \tag{7.28}$$

and

$$\sum_{n=N+1}^{2N} \left(P_k(\psi_R(n,h))\right)^4 \ll N(\log N)^{4k}. \tag{7.29}$$

Therefore we conclude from (7.24)–(7.27) that, for $\lambda = \left(\sqrt{r} - \frac{1}{2}\right)^2 + c_k$,

$$\sum_{\substack{N<p_n\leqslant 2N\\ p_{n+r}-p_n\leqslant h}} 1 \gg_k \frac{N}{h} \gg_k \pi(N), \tag{7.30}$$

where $c_k \to 0^+$ as $k \to \infty$, which proves Theorem 1.

Before proceeding to the proofs of (7.28) and (7.29), we note that, for $N < n \leqslant 2N$, the trivial estimates $\psi(n,h) \ll h\log N$ and ${\psi_R}^{(k)}(n,h) \ll_k N^\epsilon$ immediately imply the bounds $\ll N^{1+\epsilon}$ in (7.28) and (7.29) from which (1.10) follows. To prove (7.28) we make use of the sieve bound

$$\sum_{n\leqslant N} \Lambda(n;\mathcal{H}_k) \leqslant (2^k k! + \epsilon)\mathfrak{S}(\mathcal{H}_k)N \tag{7.31}$$

(see [**13**, Theorem 5.7] or [**11**, Theorem 4 of 2.3.3]). Then by equations (7.6) and (2.13), we have

$$\begin{aligned}
\sum_{n=N+1}^{2N} \psi(n;h)^4 &= \sum_{n=N+1}^{2N} (1+o(1)) \sum_{1\leqslant h_1,h_2,h_3,h_4\leqslant h} (\log N)^{4-|\mathcal{H}|}\Lambda(n;\mathcal{H}) \\
&= (1+o(1)) \sum_{\nu=1}^{4} \left\{ {4 \atop \nu} \right\} (\log N)^{4-\nu} \sum_{\substack{1\leqslant h_1,\ldots,h_\nu\leqslant h\\ \text{distinct}}} \left(\sum_{n=N+1}^{2N} \Lambda(n;\mathcal{H}_\nu)\right) \\
&\leqslant (N+o(N)) \sum_{\nu=1}^{4} \left\{ {4 \atop \nu} \right\} 2^\nu \nu! (\log N)^{4-\nu} \left(\sum_{\substack{1\leqslant h_1,\ldots,h_\nu\leqslant h\\ \text{distinct}}} \mathfrak{S}(\mathcal{H}_\nu)\right) \\
&\leqslant \left(\sum_{\nu=1}^{4} \left\{ {4 \atop \nu} \right\} \nu!(2\lambda)^\nu + \epsilon \right) N\log^4 N,
\end{aligned} \tag{7.32}$$

which proves (7.28).

The proof of (7.29) is based on a generalization of Proposition 1 proved in [**10**] by the same method as that used in the proof of Proposition 1 in this paper. For $k \geqslant 1$, and $\mathcal{H} = \{h_1, h_2, \ldots, h_r\}$ with distinct integers h_i, and $\mathbf{a} = (a_1, a_2, \ldots, a_r)$, where $a_i \geqslant 1$ with $\sum_{i=1}^r a_i = k$, let

$$\mathcal{S}_k(N,\mathcal{H},\mathbf{a}) = \sum_{n=1}^{N} \Lambda_R(n+h_1)^{a_1}\Lambda_R(n+h_2)^{a_2}\ldots\Lambda_R(n+h_r)^{a_r}. \tag{7.33}$$

Then for $\max_i a_i \leqslant 4$, $\max_i |h_i| \leqslant R$ and $R \geqslant 2$ we have

$$\mathcal{S}_k(N, \mathcal{H}, \mathbf{a}) = \big(\mathcal{C}_k(\mathbf{a})\mathfrak{S}(\mathcal{H}) + o_k(1)\big)N(\log R)^{k-r} + O(R^k), \tag{7.34}$$

where the $\mathcal{C}_k(\mathbf{a})$ are constants that are computable rational numbers. On multiplying out the left-hand side of (7.29) we obtain a linear combination of $(k+1)^4$ terms of the form

$$\mathcal{T}(\ell_1, \ell_2, \ell_3, \ell_4) = (\log R)^{4k-\ell_1-\ell_2-\ell_3-\ell_4} \sum_{n=N+1}^{2N} \prod_{i=1}^{4} {\psi_R}^{(\ell_i)}(n, h),$$

for any $0 \leqslant \ell_1, \ell_2, \ell_3, \ell_4 \leqslant k$. Let $\ell = \ell_1 + \ell_2 + \ell_3 + \ell_4$; then $0 \leqslant \ell \leqslant 4k$ and we have

$$\mathcal{T}(\ell_1, \ell_2, \ell_3, \ell_4) = (\log R)^{4k-\ell} \sum_{1 \leqslant m_1, m_2, \ldots, m_\ell \leqslant h} \sum_{n=N+1}^{2N} \prod_{i=1}^{4} \Lambda_R(n, \mathbf{H}_i),$$

where $m_1, m_2, \ldots, m_\ell$ run through the components of the $\mathbf{H}_i$, for $1 \leqslant i \leqslant 4$. Letting $\mathcal{H}$ be the set of distinct components, we have by (7.7) and (7.34),

$$\sum_{n=N+1}^{2N} \prod_{i=1}^{4} \Lambda_R(n, \mathbf{H}_i) \ll_k N\mathfrak{S}(\mathcal{H})(\log R)^{\ell-|\mathcal{H}|},$$

and, by (2.13),

$$\begin{aligned} \mathcal{T}(\ell_1, \ell_2, \ell_3, \ell_4) &\ll_k (\log R)^{4k-\ell} \sum_{j=1}^{\ell} \sum_{\substack{1 \leqslant h_1, \ldots, h_j \leqslant h \\ \text{distinct}}} N\mathfrak{S}(\mathcal{H}_j)(\log R)^{\ell-j} \\ &\ll_k N(\log R)^{4k} \sum_{j=1}^{\ell} \frac{h^j}{(\log R)^j} \\ &\ll_k N(\log N)^{4k}, \end{aligned}$$

which proves (7.29).

References

1. E. Bombieri and H. Davenport, 'Small differences between prime numbers', *Proc. Roy. Soc. Ser. A* 293 (1966) 1–18.
2. E. Bombieri, J. B. Friedlander and H. Iwaniec, 'Primes in arithmetic progressions to large moduli', *Acta Math.* 156 (1986) 203–251.
3. H. Dette and W. J. Studden, 'Some new asymptotic properties for the zeros of Jacobi, Laguerre and Hermite polynomials', *Constr. Approx.* 11 (1995) 227–238.
4. P. Erdős, 'The difference of consecutive primes', *Duke Math. J.* 6 (1940) 438–441.
5. P. X. Gallagher, 'On the distribution of primes in short intervals', *Mathematika* 23 (1976) 4–9.
6. D. A. Goldston, 'On Bombieri and Davenport's theorem concerning small gaps between primes', *Mathematika* 39 (1992) 10–17.
7. D. A. Goldston, Y. Motohashi, J. Pintz and C. Y. Yildirim, 'Small gaps between primes exist', *Proc. Japan Acad.* 82A (2006) 61–65.
8. D. A. Goldston, J. Pintz and C. Y. Yildirim, 'Primes in tuples I', *Ann of Math.* to appear.
9. D. A. Goldston and C. Y. Yildirim, 'Higher correlations of divisor sums related to primes. I: triple correlations', *Integers* 3 (2003) A5, 66 pp. (electronic).
10. D. A. Goldston and C. Y. Yildirim, 'Higher correlations of divisor sums related to primes IV: k-correlations', AIM Preprint 2002-11, http://www.aimath.org/preprints.html.
11. G. Greaves, *Sieves in number theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete 43 (Springer, Berlin, 2001).
12. B. Green and T. Tao, 'The primes contain arbitrarily long arithmetic progressions', *Ann. of Math.* to appear.
13. H. Halberstam and H.-E. Richert, *Sieve methods* (Academic Press, New York, 1974).
14. G. H. Hardy and J. E. Littlewood, 'Some problems of "Partitio Numerorum": III on the expression of a number as a sum of primes', *Acta Math.* 44 (1923) 1–70.
15. M. N. Huxley, 'Small differences between consecutive primes II', *Mathematika* 24 (1977) 142–152.
16. M. N. Huxley, 'An application of the Fouvry–Iwaniec theorem', *Acta Arith.* XLIII (1984) 441–443.

17. H. Maier, 'Small differences between prime numbers', *Michigan Math. J.* 35 (1988) 323–344.
18. D. S. Moak, E. B. Saff and R. S. Varga, 'On the zeros of Jacobi polynomials ${P_n}^{(\alpha_n,\beta_n)}(x)$', *Trans. Amer. Math. Soc.* 249 (1979) 159–162.
19. H. L. Montgomery and K. Soundararajan, 'Primes in short intervals', *Comm. Math. Phys.* 252 (2004) 589–617.
20. R. A. Rankin, 'The difference between consecutive prime numbers. II', *Proc. Cambridge Philos. Soc.* 36 (1940) 255–266.
21. J. Sivak, 'Méthodes de crible appliquées aux sommes de Kloosterman et aux petits écarts entre nombres premiers', Thèse de Doctorat de l'Université Paris Sud (Paris XI), 2005.
22. G. Szegö, *Orthogonal polynomials*, American Mathematical Society Colloquium Publications 23 (AMS, Providence, RI, 1939).
23. E. C. Titchmarsh, *The theory of the Riemann zeta-function*, 2nd edn, edited and with a preface by D. R. Heath-Brown (The Clarendon Press, Oxford University Press, New York, 1986).
24. J. V. Uspensky, *Theory of equations* (McGraw-Hill, New York, 1948).

D. A. Goldston
Department of Mathematics
San Jose State University
San Jose, CA 95192
USA

goldston@math.sjsu.edu

C. Y. Yıldırım
Department of Mathematics
Boğaziçi University
Bebek
Istanbul, 34342
Turkey

and

Feza Gürsey Enstitüsü
Çengelköy
Istanbul, P.K. 6, 81220
Turkey

yalciny@boun.edu.tr