

## RESEARCH

# Variants of the Selberg sieve, and bounded intervals containing many primes

DHJ Polymath

Correspondence:

[tao@math.ucla.edu](mailto:tao@math.ucla.edu)[michaelnielsen.org/polymath1/index.php?title=Bounded\\_gaps\\_between\\_primes](https://michaelnielsen.org/polymath1/index.php?title=Bounded_gaps_between_primes),

Full list of author information is available at the end of the article

## Abstract

For any  $m \geq 1$ , let  $H_m$  denote the quantity  $\liminf_{n \rightarrow \infty} (p_{n+m} - p_n)$ , where  $p_n$  is the  $n^{\text{th}}$  prime. A celebrated recent result of Zhang showed the finiteness of  $H_1$ , with the explicit bound  $H_1 \leq 70000000$ . This was then improved by us (the Polymath8 project) to  $H_1 \leq 4680$ , and then by Maynard to  $H_1 \leq 600$ , who also established for the first time a finiteness result for  $H_m$  for  $m \geq 2$ , and specifically that  $H_m \ll m^3 e^{4m}$ . If one also assumes the Elliott-Halberstam conjecture, Maynard obtained the bound  $H_1 \leq 12$ , improving upon the previous bound  $H_1 \leq 16$  of Goldston, Pintz, and Yıldırım, as well as the bound  $H_m \ll m^3 e^{2m}$ .

In this paper, we extend the methods of Maynard by generalizing the Selberg sieve further, and by performing more extensive numerical calculations. As a consequence, we can obtain the bound  $H_1 \leq 246$  unconditionally, and  $H_1 \leq 6$  under the assumption of the generalized Elliott-Halberstam conjecture. Indeed, under the latter conjecture we show the stronger statement that for any admissible triple  $(h_1, h_2, h_3)$ , there are infinitely many  $n$  for which at least two of  $n + h_1, n + h_2, n + h_3$  are prime, and also obtain a related disjunction asserting that either the twin prime conjecture holds, or the even Goldbach conjecture is asymptotically true if one allows an additive error of at most 2, or both. We also modify the “parity problem” argument of Selberg to show that the  $H_1 \leq 6$  bound is the best possible that one can obtain from purely sieve-theoretic considerations. For larger  $m$ , we use the distributional results obtained previously by our project to obtain the unconditional asymptotic bound  $H_m \ll m e^{(4 - \frac{28}{157})m}$ , or  $H_m \ll m e^{2m}$  under the assumption of the Elliott-Halberstam conjecture. We also obtain explicit upper bounds for  $H_m$  when  $m = 2, 3, 4, 5$ .

**Keywords:** Selberg sieve; Elliott-Halberstam conjecture; Prime gaps

## 1 Introduction

For any natural number  $m$ , let  $H_m$  denote the quantity

$$H_m := \liminf_{n \rightarrow \infty} (p_{n+m} - p_n),$$

where  $p_n$  denotes the  $n^{\text{th}}$  prime. The twin prime conjecture asserts that  $H_1 = 2$ ; more generally, the Hardy-Littlewood prime tuples conjecture [30] implies that  $H_m = H(m+1)$  for all  $m \geq 1$ , where  $H(k)$  is the diameter of the narrowest admissible  $k$ -tuple (see Section 3 for a definition of this term). Asymptotically, one has the bounds

$$\left(\frac{1}{2} + o(1)\right)k \log k \leq H(k) \leq (1 + o(1))k \log k$$

as  $k \rightarrow \infty$  (see Theorem 3.3 below); thus the prime tuples conjecture implies that  $H_m$  is comparable to  $m \log m$  as  $m \rightarrow \infty$ .

Until very recently, it was not known if any of the  $H_m$  were finite, even in the easiest case  $m = 1$ . In the breakthrough work of Goldston, Pintz, and Yıldırım [25], several results in this direction were established, including the following conditional result assuming the Elliott-Halberstam conjecture  $\text{EH}[\vartheta]$  (see Claim 2.2 below) concerning the distribution of the prime numbers in arithmetic progressions:

**Theorem 1.1** (GPY theorem) *Assume the Elliott-Halberstam conjecture  $\text{EH}[\vartheta]$  for all  $0 < \vartheta < 1$ . Then  $H_1 \leq 16$ .*

Furthermore, it was shown in [25] that any result of the form  $\text{EH}[\frac{1}{2} + 2\varpi]$  for some fixed  $0 < \varpi < 1/4$  would imply an explicit finite upper bound on  $H_1$  (with this bound equal to 16 for  $\varpi > 0.229855$ ). Unfortunately, the only results of the type  $\text{EH}[\vartheta]$  that are known come from the Bombieri-Vinogradov theorem (Theorem 2.3), which only establishes  $\text{EH}[\vartheta]$  for  $0 < \vartheta < 1/2$ .

The first unconditional bound on  $H_1$  was established in a breakthrough work of Zhang [65]:

**Theorem 1.2** (Zhang’s theorem)  $H_1 \leq 70\,000\,000$ .

Zhang’s argument followed the general strategy from [25] on finding small gaps between primes, with the major new ingredient being a proof of a weaker version of  $\text{EH}[\frac{1}{2} + 2\varpi]$ , which we call  $\text{MPZ}[\varpi, \delta]$ ; see Claim 2.4 below. It was quickly realized that Zhang’s numerical bound on  $H_1$  could be improved. By optimizing many of the components in Zhang’s argument, we were able [52, 53] to improve Zhang’s bound to

$$H_1 \leq 4680.$$

Very shortly afterwards, a further breakthrough was obtained by Maynard [38] (with related work obtained independently in unpublished work of Tao), who developed a more flexible “multidimensional” version of the Selberg sieve to obtain stronger bounds on  $H_m$ . This argument worked without using any equidistribution results on primes beyond the Bombieri-Vinogradov theorem, and amongst other things was able to establish finiteness of  $H_m$  for all  $m$ , not just for  $m = 1$ . More precisely, Maynard established the following results.

**Theorem 1.3** (Maynard’s theorem) *Unconditionally, we have the following bounds:*

(i)  $H_1 \leq 600$ .

(ii)  $H_m \leq Cm^3e^{4m}$  for all  $m \geq 1$  and an absolute (and effective) constant  $C$ .

*Assuming the Elliott-Halberstam conjecture  $\text{EH}[\vartheta]$  for all  $0 < \vartheta < 1$ , we have the following improvements:*

(iii)  $H_1 \leq 12$ .

(iv)  $H_2 \leq 600$ .

(v)  $H_m \leq Cm^3e^{2m}$  for all  $m \geq 1$  and an absolute (and effective) constant  $C$ .

For a survey of these recent developments, see [29].

In this paper, we refine Maynard’s methods to obtain the following further improvements.

**Theorem 1.4** *Unconditionally, we have the following bounds:*

(i)  $H_1 \leq 246$ .

(ii)  $H_2 \leq 398\,130$ .

(iii)  $H_3 \leq 24\,797\,814$ .

(iv)  $H_4 \leq 1\,431\,556\,072$ .

(v)  $H_5 \leq 80\,550\,202\,480$ .

(vi)  $H_m \leq Cm \exp((4 - \frac{28}{157})m)$  for all  $m \geq 1$  and an absolute (and effective) constant  $C$ .

*Assume the Elliott-Halberstam conjecture  $\text{EH}[\vartheta]$  for all  $0 < \vartheta < 1$ . Then we have the following improvements:*

(vii)  $H_2 \leq 270$ .

(viii)  $H_3 \leq 52\,116$ .

$$(ix) \ H_4 \leq 474\,266.$$

$$(x) \ H_5 \leq 4\,137\,854.$$

$$(xi) \ H_m \leq Cme^{2m} \text{ for all } m \geq 1 \text{ and an absolute (and effective) constant } C.$$

Finally, assume the generalized Elliott-Halberstam conjecture  $\text{GEH}[\vartheta]$  (see Claim 2.6 below) for all  $0 < \vartheta < 1$ . Then

$$(xii) \ H_1 \leq 6.$$

$$(xiii) \ H_2 \leq 252.$$

In Section 3 we will describe the key propositions that will be combined together to prove the various components of Theorem 1.4. As with Theorem 1.1, the results in (vii)-(xiii) do not require  $\text{EH}[\vartheta]$  or  $\text{GEH}[\vartheta]$  for all  $0 < \vartheta < 1$ , but only for a single explicitly computable  $\vartheta$  that is sufficiently close to 1.

Of these results, the bound in (xii) is perhaps the most interesting, as the parity problem [57] prohibits one from achieving any better bound on  $H_1$  than 6 from purely sieve-theoretic methods; we review this obstruction in Section 8. If one only assumes the Elliott-Halberstam conjecture  $\text{EH}[\vartheta]$  instead of its generalization  $\text{GEH}[\vartheta]$ , we were unable to improve upon Maynard's bound  $H_1 \leq 12$ ; however the parity obstruction does not exclude the possibility that one could achieve (xii) just assuming  $\text{EH}[\vartheta]$  rather than  $\text{GEH}[\vartheta]$ , by some further refinement of the sieve-theoretic arguments (e.g. by finding a way to establish Theorem 3.6(ii) below using only  $\text{EH}[\vartheta]$  instead of  $\text{GEH}[\vartheta]$ ).

The bounds (ii)-(vi) rely on the equidistribution results on primes established in our previous paper [52]. However, the bound (i) uses only the Bombieri-Vinogradov theorem, and the remaining bounds (vii)-(xiii) of course use either the Elliott-Halberstam conjecture or a generalization thereof.

A variant of the proof of Theorem 1.4(xii), which we give in Section 9, also gives the following conditional “near miss” to (a disjunction of) the twin prime conjecture and the even Goldbach conjecture:

**Theorem 1.5 (Disjunction)** *Assume the generalized Elliott-Halberstam conjecture  $\text{GEH}[\vartheta]$  for all  $0 < \vartheta < 1$ . Then at least one of the following statements is true:*

$$(a) \text{ (Twin prime conjecture) } H_1 = 2.$$

$$(b) \text{ (near-miss to even Goldbach conjecture) If } n \text{ is a sufficiently large multiple of six, then at least one of } n \text{ and } n-2 \text{ is expressible as the sum of two primes. Similarly with } n-2 \text{ replaced by } n+2. \text{ (In particular, every sufficiently large even number lies within 2 of the sum of two primes.)}$$

We remark that a disjunction in a similar spirit was obtained in [45], which established (prior to the appearance of Theorem 1.2) that either  $H_1$  was finite, or that every interval  $[x, x + x^\varepsilon]$  contained the sum of two primes if  $x$  was sufficiently large depending on  $\varepsilon > 0$ .

There are two main technical innovations in this paper. The first is a further generalization of the multidimensional Selberg sieve introduced by Maynard and Tao, in which the support of a certain cutoff function  $F$  is permitted to extend into a larger domain than was previously permitted (particularly under the assumption of the generalized Elliott-Halberstam conjecture). As in [38], this largely reduces the task of bounding  $H_m$  to that of efficiently solving a certain multidimensional variational problem involving the cutoff function  $F$ . Our second main technical innovation is to obtain efficient numerical methods for solving this variational problem for small values of the dimension  $k$ , as well as sharpened asymptotics in the case of large values of  $k$ .

The methods of Maynard and Tao have been used in a number of subsequent applications [18], [3], [60], [4], [35], [8], [50], [2], [39], [51], [48], [9], [49]. The techniques in this paper should be able to be used to obtain slight numerical improvements to such results, although we did not pursue these matters here.

### 1.1 Organization of the paper

The paper is organized as follows. After some notational preliminaries, we recall in Section 2 the known (or conjectured) distributional estimates on primes in arithmetic progressions that we will need to prove Theorem 1.4. Then, in Section 3, we give the key propositions that will be combined together to establish this theorem. One of these propositions, Lemma 3.4, is an easy application of the pigeonhole principle. Two further propositions, Theorem 3.5 and Theorem 3.6, use the prime distribution results from Section 2 to give asymptotics for certain sums involving sieve weights and the von Mangoldt function; they are established in Section 4. Theorems 3.8, 3.10, 3.12, 3.14 use the asymptotics established in Theorems 3.5, 3.6, in combination with Lemma 3.4, to give various criteria for bounding  $H_m$ , which all involve finding sufficiently strong candidates for a variety of multidimensional variational problems; these theorems are proven in Section 5. These variational problems are analysed in the asymptotic regime of large  $k$  in Section 6, and for small and medium  $k$  in Section 7, with the results collected in Theorems 3.9, 3.11, 3.13, 3.15. Combining these results with the previous propositions gives Theorem 3.2, which, when combined with the bounds on narrow admissible tuples in Theorem 3.3 that are established in Section 10, will give Theorem 1.4. (See also Table 1 for some more details of the logical dependencies between the key propositions.)

Finally, in Section 8 we modify an argument of Selberg to show that the bound  $H_1 \leq 6$  may not be improved using purely sieve-theoretic methods, and in Section 9 we establish Theorem 1.5 and make some miscellaneous remarks.

### 1.2 Notation

The notation used here closely follows the notation in our previous paper [52].

We use  $|E|$  to denote the cardinality of a finite set  $E$ , and  $\mathbf{1}_E$  to denote the indicator function of a set  $E$ , thus  $\mathbf{1}_E(n) = 1$  when  $n \in E$  and  $\mathbf{1}_E(n) = 0$  otherwise. In a similar spirit, if  $E$  is a statement, we write  $\mathbf{1}_E = 1$  when  $E$  is true and  $\mathbf{1}_E = 0$  otherwise.

All sums and products will be over the natural numbers  $\mathbb{N} := \{1, 2, 3, \dots\}$  unless otherwise specified, with the exceptions of sums and products over the variable  $p$ , which will be understood to be over primes.

The following important asymptotic notation will be in use throughout the paper:

**Definition 1.6** (Asymptotic notation) We use  $x$  to denote a large real parameter, which one should think of as going off to infinity; in particular, we will implicitly assume that it is larger than any specified fixed constant. Some mathematical objects will be independent of  $x$  and referred to as *fixed*; but unless otherwise specified we allow all mathematical objects under consideration to depend on  $x$  (or to vary within a range that depends on  $x$ , e.g. the summation parameter  $n$  in the sum  $\sum_{x \leq n \leq 2x} f(n)$ ). If  $X$  and  $Y$  are two quantities depending on  $x$ , we say that  $X = O(Y)$  or  $X \ll Y$  if one has  $|X| \leq CY$  for some fixed  $C$  (which we refer to as the *implied constant*), and  $X = o(Y)$  if one has  $|X| \leq c(x)Y$  for some function  $c(x)$  of  $x$  (and of any fixed parameters present) that goes to zero as  $x \rightarrow \infty$  (for each choice of fixed parameters). We use  $X \ll Y$  to denote the estimate  $|X| \leq x^{o(1)}Y$ ,  $X \asymp Y$  to denote the estimate  $Y \ll X \ll Y$ , and  $X \approx Y$  to denote the estimate  $Y \ll X \ll Y$ . Finally, we say that a quantity  $n$  is of *polynomial size* if one has  $n = O(x^{O(1)})$ .

If asymptotic notation such as  $O()$  or  $\ll$  appears on the left-hand side of a statement, this means that the assertion holds true for any specific interpretation of that notation. For instance, the assertion  $\sum_{n=O(N)} |\alpha(n)| \ll N$  means that for each fixed constant  $C > 0$ , one has  $\sum_{|n| \leq CN} |\alpha(n)| \ll N$ .

If  $q$  and  $a$  are integers, we write  $a|q$  if  $a$  divides  $q$ . If  $q$  is a natural number and  $a \in \mathbb{Z}$ , we use  $a \pmod{q}$  to denote the residue class

$$a \pmod{q} := \{a + nq : n \in \mathbb{Z}\}$$

and let  $\mathbb{Z}/q\mathbb{Z}$  denote the ring of all such residue classes  $a \pmod{q}$ . The notation  $b = a \pmod{q}$  is synonymous to  $b \in a \pmod{q}$ . We use  $(a, q)$  to denote the greatest common divisor of  $a$  and  $q$ , and  $[a, q]$  to denote the least common multiple.<sup>[1]</sup> We also let

$$(\mathbb{Z}/q\mathbb{Z})^\times := \{a \pmod{q} : (a, q) = 1\}$$

denote the primitive residue classes of  $\mathbb{Z}/q\mathbb{Z}$ .

We use the following standard arithmetic functions:

- (i)  $\varphi(q) := |(\mathbb{Z}/q\mathbb{Z})^\times|$  denotes the Euler totient function of  $q$ .
- (ii)  $\tau(q) := \sum_{d|q} 1$  denotes the divisor function of  $q$ .
- (iii)  $\Lambda(q)$  denotes the von Mangoldt function of  $q$ , thus  $\Lambda(q) = \log p$  if  $q$  is a power of a prime  $p$ , and  $\Lambda(q) = 0$  otherwise.
- (iv)  $\theta(q)$  is defined to equal  $\log q$  when  $q$  is a prime, and  $\theta(q) = 0$  otherwise.
- (v)  $\mu(q)$  denotes the Möbius function of  $q$ , thus  $\mu(q) = (-1)^k$  if  $q$  is the product of  $k$  distinct primes for some  $k \geq 0$ , and  $\mu(q) = 0$  otherwise.
- (vi)  $\Omega(q)$  denotes the number of prime factors of  $q$  (counting multiplicity).

We recall the elementary *divisor bound*

$$\tau(n) \ll 1 \tag{1}$$

whenever  $n \ll x^{O(1)}$ , as well as the related estimate

$$\sum_{n \leq x} \frac{\tau(n)^C}{n} \ll \log^{O(1)} x \tag{2}$$

for any fixed  $C > 0$ ; this follows for instance from [52, Lemma 1.3].

The *Dirichlet convolution*  $\alpha \star \beta: \mathbb{N} \rightarrow \mathbb{C}$  of two arithmetic functions  $\alpha, \beta: \mathbb{N} \rightarrow \mathbb{C}$  is defined in the usual fashion as

$$\alpha \star \beta(n) := \sum_{d|n} \alpha(d) \beta\left(\frac{n}{d}\right) = \sum_{ab=n} \alpha(a) \beta(b).$$

## 2 Distribution estimates on arithmetic functions

As mentioned in the introduction, a key ingredient in the Goldston-Pintz-Yıldırım approach to small gaps between primes comes from distributional estimates on the primes, or more precisely on the von Mangoldt function  $\Lambda$ , which serves as a proxy for the primes. In this work, we will also need to consider distributional estimates on more general arithmetic functions, although we will not prove any new such estimates in this paper, relying instead on estimates that are already in the literature.

More precisely, we will need averaged information on the following quantity:

**Definition 2.1** (Discrepancy) For any function  $\alpha: \mathbb{N} \rightarrow \mathbb{C}$  with finite support (that is,  $\alpha$  is non-zero only on a finite set) and any primitive residue class  $a \pmod{q}$ , we define the (signed) *discrepancy*  $\Delta(\alpha; a \pmod{q})$  to be the quantity

$$\Delta(\alpha; a \pmod{q}) := \sum_{n \equiv a \pmod{q}} \alpha(n) - \frac{1}{\varphi(q)} \sum_{(n, q)=1} \alpha(n). \tag{3}$$

---

<sup>[1]</sup>When  $a, b$  are real numbers, we will also need to use  $(a, b)$  and  $[a, b]$  to denote the open and closed intervals respectively with endpoints  $a, b$ . Unfortunately, this notation conflicts with the notation given above, but it should be clear from the context which notation is in use.

For any fixed  $0 < \vartheta < 1$ , let  $\text{EH}[\vartheta]$  denote the following claim:

**Claim 2.2** (Elliott-Halberstam conjecture,  $\text{EH}[\vartheta]$ ) *If  $Q \ll x^\vartheta$  and  $A \geq 1$  is fixed, then*

$$\sum_{q \leq Q} \sup_{a \in (\mathbb{Z}/q\mathbb{Z})^\times} |\Delta(\Lambda \mathbf{1}_{[x, 2x]}; a(q))| \ll x \log^{-A} x. \quad (4)$$

In [13] it was conjectured that  $\text{EH}[\vartheta]$  held for all  $0 < \vartheta < 1$ . (The conjecture fails at the endpoint case  $\vartheta = 1$ ; see [19], [20] for a more precise statement.) The following classical result of Bombieri [5] and Vinogradov [63] remains the best partial result of the form  $\text{EH}[\vartheta]$ :

**Theorem 2.3** (Bombieri-Vinogradov theorem) [5, 63]  *$\text{EH}[\vartheta]$  holds for every fixed  $0 < \vartheta < 1/2$ .*

In [25] it was shown that any estimate of the form  $\text{EH}[\vartheta]$  with some fixed  $\vartheta > 1/2$  would imply the finiteness of  $H_1$ . While such an estimate remains unproven, it was observed by Motohashi-Pintz [43] and by Zhang [65] that a certain weakened version of  $\text{EH}[\vartheta]$  would still suffice for this purpose. More precisely (and following the notation of our previous paper [52]), let  $\varpi, \delta > 0$  be fixed, and let  $\text{MPZ}[\varpi, \delta]$  be the following claim:

**Claim 2.4** (Motohashi-Pintz-Zhang estimate,  $\text{MPZ}[\varpi, \delta]$ ) *Let  $I \subset [1, x^\delta]$  and  $Q \ll x^{1/2+2\varpi}$ . Let  $P_I$  denote the product of all the primes in  $I$ , and let  $S_I$  denote the square-free natural numbers whose prime factors lie in  $I$ . If the residue class  $a \pmod{P_I}$  is primitive (and is allowed to depend on  $x$ ), and  $A \geq 1$  is fixed, then*

$$\sum_{\substack{q \leq Q \\ q \in S_I}} |\Delta(\Lambda \mathbf{1}_{[x, 2x]}; a(q))| \ll x \log^{-A} x, \quad (5)$$

where the implied constant depends only on the fixed quantities  $(A, \varpi, \delta)$ , but not on  $a$ .

It is clear that  $\text{EH}[\frac{1}{2} + 2\varpi]$  implies  $\text{MPZ}[\varpi, \delta]$  whenever  $\varpi, \delta \geq 0$ . The first non-trivial estimate of the form  $\text{MPZ}[\varpi, \delta]$  was established by Zhang [65], who (essentially) obtained  $\text{MPZ}[\varpi, \delta]$  whenever  $0 \leq \varpi, \delta < \frac{1}{1168}$ . In [52, Theorem 2.17], we improved this result to the following.

**Theorem 2.5**  *$\text{MPZ}[\varpi, \delta]$  holds for every fixed  $\varpi, \delta \geq 0$  with  $600\varpi + 180\delta < 7$ .*

In fact, a stronger result was established in [52], in which the moduli  $q$  were assumed to be *densely divisible* rather than smooth, but we will not exploit such improvements here. For our application, the most important thing is to get  $\varpi$  as large as possible; in particular, Theorem 2.5 allows one to get  $\varpi$  arbitrarily close to  $\frac{7}{600} \approx 0.01167$ .

In this paper, we will also study the following generalization of the Elliott-Halberstam conjecture for a fixed choice of  $0 < \vartheta < 1$ :

**Claim 2.6** (Generalized Elliott-Halberstam conjecture,  $\text{GEH}[\vartheta]$ ) *Let  $\varepsilon > 0$  and  $A \geq 1$  be fixed. Let  $N, M$  be quantities such that  $x^\varepsilon \ll N \ll x^{1-\varepsilon}$  and  $x^\varepsilon \ll M \ll x^{1-\varepsilon}$  with  $NM \asymp x$ , and let  $\alpha, \beta : \mathbb{N} \rightarrow \mathbb{R}$  be sequences supported on  $[N, 2N]$  and  $[M, 2M]$  respectively, such that one has the pointwise bounds*

$$|\alpha(n)| \ll \tau(n)^{O(1)} \log^{O(1)} x; \quad |\beta(m)| \ll \tau(m)^{O(1)} \log^{O(1)} x \quad (6)$$

for all natural numbers  $n, m$ . Suppose also that  $\beta$  obeys the Siegel-Walfisz type bound

$$|\Delta(\beta \mathbf{1}_{(\cdot, r)=1}; a(q))| \ll \tau(qr)^{O(1)} M \log^{-A} x \quad (7)$$

for any  $q, r \geq 1$ , any fixed  $A$ , and any primitive residue class  $a(q)$ . Then for any  $Q \ll x^\vartheta$ , we have

$$\sum_{q \leq Q} \sup_{a \in (\mathbb{Z}/q\mathbb{Z})^\times} |\Delta(\alpha \star \beta; a(q))| \ll x \log^{-A} x. \quad (8)$$

In [7, Conjecture 1] it was essentially conjectured<sup>[2]</sup> that  $\text{GEH}[\vartheta]$  was true for all  $0 < \vartheta < 1$ . This is stronger than the Elliott-Halberstam conjecture:

**Proposition 2.7** *For any fixed  $0 < \vartheta < 1$ ,  $\text{GEH}[\vartheta]$  implies  $\text{EH}[\vartheta]$ .*

*Proof (Sketch)* As this argument is standard, we give only a brief sketch. Let  $A > 0$  be fixed. For  $n \in [x, 2x]$ , we have Vaughan's identity<sup>[3]</sup> [62]

$$\Lambda(n) = \mu_{<} \star L(n) - \mu_{<} \star \Lambda_{<} \star 1(n) + \mu_{\geq} \star \Lambda_{\geq} \star 1(n),$$

where  $L(n) := \log(n)$ ,  $1(n) := 1$ , and

$$\Lambda_{\geq}(n) := \Lambda(n) \mathbf{1}_{n \geq x^{1/3}}, \quad \Lambda_{<}(n) := \Lambda(n) \mathbf{1}_{n < x^{1/3}} \quad (9)$$

$$\mu_{\geq}(n) := \mu(n) \mathbf{1}_{n \geq x^{1/3}}, \quad \mu_{<}(n) := \mu(n) \mathbf{1}_{n < x^{1/3}}. \quad (10)$$

By decomposing each of the functions  $\mu_{<}$ ,  $\mu_{\geq}$ ,  $1$ ,  $\Lambda_{<}$ ,  $\Lambda_{\geq}$  into  $O(\log^{A+1} x)$  functions supported on intervals of the form  $[N, (1 + \log^{-A} x)N]$ , and discarding those contributions which meet the boundary of  $[x, 2x]$  (cf. [16], [22], [7], [65]), and using  $\text{GEH}[\vartheta]$  (with  $A$  replaced by a much larger fixed constant  $A'$ ) to control all remaining contributions, we obtain the claim (using the Siegel-Walfisz theorem, see e.g. [58, Satz 4] or [34, Th. 5.29]).  $\square$

By modifying the proof of the Bombieri-Vinogradov theorem Motohashi [42] established the following generalization of that theorem (see also [24] for some related ideas):

**Theorem 2.8** (Generalized Bombieri-Vinogradov theorem) [42]  *$\text{GEH}[\vartheta]$  holds for every fixed  $0 < \vartheta < 1/2$ .*

One could similarly describe a generalization of the Motohashi-Pintz-Zhang estimate  $\text{MPZ}[\varpi, \delta]$ , but unfortunately the arguments in [65] or Theorem 2.5 do not extend to this setting unless one is in the “Type I/Type II” case in which  $N, M$  are constrained to be somewhat close to  $x^{1/2}$ , or if one has “Type III” structure to the convolution  $\alpha \star \beta$ , in the sense that it can refactored as a convolution involving several “smooth” sequences. In any event, our analysis would not be able to make much use of such incremental improvements to  $\text{GEH}[\vartheta]$ , as we only use this hypothesis effectively in the case when  $\vartheta$  is very close to 1. In particular, we will not directly use Theorem 2.8 in this paper.

<sup>[2]</sup>Actually, there are some differences between [7, Conjecture 1] and the claim here. Firstly, we need an estimate that is uniform for all  $a$ , whereas in [7] only the case of a fixed modulus  $a$  was asserted. On the other hand,  $\alpha, \beta$  were assumed to be controlled in  $\ell^2$  instead of via the pointwise bounds (6), and  $Q$  was allowed to be as large as  $x \log^{-C} x$  for some fixed  $C$  (although, in view of the negative results in [19], [20], this latter strengthening may be too ambitious).

<sup>[3]</sup>One could also use the Heath-Brown identity [31] here if desired.



### 3 Outline of the key ingredients

In this section we describe the key subtheorems used in the proof of Theorem 1.4, with the proofs of these subtheorems mostly being deferred to later sections.

We begin with a weak version of the Dickson-Hardy-Littlewood prime tuples conjecture [30], which (following Pintz [46]) we refer to as  $\text{DHL}[k, j]$ . Recall that for any  $k \in \mathbb{N}$ , an *admissible  $k$ -tuple* is a tuple  $\mathcal{H} = (h_1, \dots, h_k)$  of  $k$  increasing integers  $h_1 < \dots < h_k$  which avoids at least one residue class  $a_p \pmod{p} := \{a_p + np : n \in \mathbb{Z}\}$  for every  $p$ . For instance,  $(0, 2, 6)$  is an admissible 3-tuple, but  $(0, 2, 4)$  is not.

For any  $k \geq j \geq 2$ , we let  $\text{DHL}[k, j]$  denote the following claim:

**Claim 3.1** (Weak Dickson-Hardy-Littlewood conjecture,  $\text{DHL}[k, j]$ ) *For any admissible  $k$ -tuple  $\mathcal{H} = (h_1, \dots, h_k)$  there exist infinitely many translates  $n + \mathcal{H} = (n + h_1, \dots, n + h_k)$  of  $\mathcal{H}$  which contain at least  $j$  primes.*

The full Dickson-Hardy-Littlewood conjecture is then the assertion that  $\text{DHL}[k, k]$  holds for all  $k \geq 2$ . In our analysis we will focus on the case when  $j$  is much smaller than  $k$ ; in fact  $j$  will be of the order of  $\log k$ .

For any  $k$ , let  $H(k)$  denote the minimal diameter  $h_k - h_1$  of an admissible  $k$ -tuple; thus for instance  $H(3) = 6$ . It is clear that for any natural numbers  $m \geq 1$  and  $k \geq m + 1$ , the claim  $\text{DHL}[k, m + 1]$  implies that  $H_m \leq H(k)$  (and the claim  $\text{DHL}[k, k]$  would imply that  $H_{k-1} = H(k)$ ). We will therefore deduce Theorem 1.4 from a number of claims of the form  $\text{DHL}[k, j]$ . More precisely, we have:

**Theorem 3.2** *Unconditionally, we have the following claims:*

- (i)  $\text{DHL}[50, 2]$ .
- (ii)  $\text{DHL}[35\,410, 3]$ .
- (iii)  $\text{DHL}[1\,649\,821, 4]$ .
- (iv)  $\text{DHL}[75\,845\,707, 5]$ .
- (v)  $\text{DHL}[3\,473\,955\,908, 6]$ .
- (vi)  $\text{DHL}[k, m + 1]$  whenever  $m \geq 1$  and  $k \geq C \exp((4 - \frac{28}{157})m)$  for some sufficiently large absolute (and effective) constant  $C$ .

*Assume the Elliott-Halberstam conjecture  $\text{EH}[\vartheta]$  for all  $0 < \vartheta < 1$ . Then we have the following improvements:*

- (vii)  $\text{DHL}[54, 3]$ .
- (viii)  $\text{DHL}[5511, 4]$ .
- (ix)  $\text{DHL}[41\,588, 5]$ .
- (x)  $\text{DHL}[309\,661, 6]$ .
- (xi)  $\text{DHL}[k, m + 1]$  whenever  $m \geq 1$  and  $k \geq C \exp(2m)$  for some sufficiently large absolute (and effective) constant  $C$ .

*Assume the generalized Elliott-Halberstam conjecture  $\text{GEH}[\vartheta]$  for all  $0 < \vartheta < 1$ . Then*

- (xii)  $\text{DHL}[3, 2]$ .
- (xiii)  $\text{DHL}[51, 3]$ .

Theorem 1.4 then follows from Theorem 3.2 and the following bounds on  $H(k)$  (ordered by increasing value of  $k$ ):

**Theorem 3.3** (Bounds on  $H(k)$ )

- (xii)  $H(3) = 6$ .
- (i)  $H(50) = 246$ .



**Table 1** Results used to prove various components of Theorem 3.2. Note that Theorems 3.8, 3.10, 3.12, 3.14 are in turn proven using Theorems 3.5, 3.6, and Lemma 3.4.

| Theorem 3.2 | Results used             |
|-------------|--------------------------|
| (i)         | Theorems 2.3, 3.12, 3.13 |
| (ii)-(vi)   | Theorems 2.5, 3.10, 3.11 |
| (vii)-(xi)  | Theorems 3.8, 3.9        |
| (xii)       | Theorems 3.14, 3.15      |
| (xiii)      | Theorems 3.12, 3.13      |

$$(xiii) \ H(51) = 252.$$

$$(vii) \ H(54) = 270.$$

$$(viii) \ H(5511) \leq 52\,116.$$

$$(ii) \ H(35\,410) \leq 398\,130.$$

$$(ix) \ H(41\,588) \leq 474\,266.$$

$$(x) \ H(309\,661) \leq 4\,137\,854.$$

$$(iii) \ H(1\,649\,821) \leq 24\,797\,814.$$

$$(iv) \ H(75\,845\,707) \leq 1\,431\,556\,072.$$

$$(v) \ H(3\,473\,955\,908) \leq 80\,550\,202\,480.$$

(vi), (xi) *In the asymptotic limit  $k \rightarrow \infty$ , one has  $H(k) \leq k \log k + k \log \log k - k + o(k)$ , with the bounds on the decay rate  $o(k)$  being effective.*

We prove Theorem 3.3 in Section 10. In the opposite direction, an application of the Brun-Titchmarsh theorem gives  $H(k) \geq (\frac{1}{2} + o(1))k \log k$  as  $k \rightarrow \infty$ ; see [53, §3.9] for this bound, as well as with some slight refinements.

The proof of Theorem 3.2 follows the Goldston-Pintz-Yıldırım strategy that was also used in all previous progress on this problem (e.g. [25], [43], [65], [52], [38]), namely that of constructing a sieve function adapted to an admissible  $k$ -tuple with good properties. More precisely, we set

$$w := \log \log \log x$$

and

$$W := \prod_{p \leq w} p,$$

and observe the crude bound

$$W \ll \log \log^{O(1)} x. \tag{11}$$

We have the following simple “pigeonhole principle” criterion for DHL[ $k, m+1$ ] (cf. [52, Lemma 4.1], though the normalization here is slightly different):

**Lemma 3.4** (Criterion for DHL) *Let  $k \geq 2$  and  $m \geq 1$  be fixed integers, and define the normalization constant*

$$B := \frac{\varphi(W)}{W} \log x. \tag{12}$$

*Suppose that for each fixed admissible  $k$ -tuple  $(h_1, \dots, h_k)$  and each residue class  $b \pmod{W}$  such that  $b + h_i$  is coprime to  $W$  for all  $i = 1, \dots, k$ , one can find a non-negative weight function  $\nu: \mathbb{N} \rightarrow \mathbb{R}^+$  and fixed*

quantities  $\alpha > 0$  and  $\beta_1, \dots, \beta_k \geq 0$ , such that one has the asymptotic upper bound

$$\sum_{\substack{x \leq n \leq 2x \\ n \equiv b \pmod{W}}} \nu(n) \leq (\alpha + o(1)) B^{-k} \frac{x}{W}, \quad (13)$$

the asymptotic lower bound

$$\sum_{\substack{x \leq n \leq 2x \\ n \equiv b \pmod{W}}} \nu(n) \theta(n + h_i) \geq (\beta_i - o(1)) B^{1-k} \frac{x}{\varphi(W)} \quad (14)$$

for all  $i = 1, \dots, k$ , and the key inequality

$$\frac{\beta_1 + \dots + \beta_k}{\alpha} > m. \quad (15)$$

Then  $\text{DHL}[k, m + 1]$  holds.

*Proof* Let  $(h_1, \dots, h_k)$  be a fixed admissible  $k$ -tuple. Since it is admissible, there is at least one residue class  $b \pmod{W}$  such that  $(b + h_i, W) = 1$  for all  $h_i \in \mathcal{H}$ . For an arithmetic function  $\nu$  as in the lemma, we consider the quantity

$$N := \sum_{\substack{x \leq n \leq 2x \\ n \equiv b \pmod{W}}} \nu(n) \left( \sum_{i=1}^k \theta(n + h_i) - m \log 3x \right).$$

Combining (13) and (14), we obtain the lower bound

$$N \geq (\beta_1 + \dots + \beta_k - o(1)) B^{1-k} \frac{x}{\varphi(W)} - (m\alpha + o(1)) B^{-k} \frac{x}{W} \log 3x.$$

From (12) and the crucial condition (15), it follows that  $N > 0$  if  $x$  is sufficiently large.

On the other hand, the sum

$$\sum_{i=1}^k \theta(n + h_i) - m \log 3x$$

can be positive only if  $n + h_i$  is prime for *at least*  $m + 1$  indices  $i = 1, \dots, k$ . We conclude that, for all sufficiently large  $x$ , there exists some integer  $n \in [x, 2x]$  such that  $n + h_i$  is prime for at least  $m + 1$  values of  $i = 1, \dots, k$ .

Since  $(h_1, \dots, h_k)$  is an arbitrary admissible  $k$ -tuple,  $\text{DHL}[k, m + 1]$  follows.  $\square$

The objective is then to construct non-negative weights  $\nu$  whose associated ratio  $\frac{\beta_1 + \dots + \beta_k}{\alpha}$  has provable lower bounds that are as large as possible. Our sieve majorants will be a variant of the multidimensional Selberg sieves used in [38]. As with all Selberg sieves, the  $\nu$  are constructed as the square of certain (signed) divisor sums. The divisor sums we will use will be finite linear combinations of products of “one-dimensional” divisor sums. More precisely, for any fixed smooth compactly supported function  $F : [0, +\infty) \rightarrow \mathbb{R}$ , define the divisor sum  $\lambda_F : \mathbb{Z} \rightarrow \mathbb{R}$  by the formula

$$\lambda_F(n) := \sum_{d|n} \mu(d) F(\log_x d) \quad (16)$$

where  $\log_x$  denotes the base  $x$  logarithm

$$\log_x n := \frac{\log n}{\log x}. \quad (17)$$

One should think of  $\lambda_F$  as a smoothed out version of the indicator function to numbers  $n$  which are “almost prime” in the sense that they have no prime factors less than  $x^\varepsilon$  for some small fixed  $\varepsilon > 0$ ; see Proposition 4.2 for a more rigorous version of this heuristic.

The functions  $\nu$  we will use will take the form

$$\nu(n) = \left( \sum_{j=1}^J c_j \lambda_{F_{j,1}}(n+h_1) \dots \lambda_{F_{j,k}}(n+h_k) \right)^2 \quad (18)$$

for some fixed natural number  $J$ , fixed coefficients  $c_1, \dots, c_J \in \mathbb{R}$  and fixed smooth compactly supported functions  $F_{j,i} : [0, +\infty) \rightarrow \mathbb{R}$  with  $j = 1, \dots, J$  and  $i = 1, \dots, k$ . (One can of course absorb the constant  $c_j$  into one of the  $F_{j,i}$  if one wishes.) Informally,  $\nu$  is a smooth restriction to those  $n$  for which  $n+h_1, \dots, n+h_k$  are all almost prime.

Clearly,  $\nu$  is a (positive-definite) fixed linear combination of functions of the form

$$n \mapsto \prod_{i=1}^k \lambda_{F_i}(n+h_i) \lambda_{G_i}(n+h_i)$$

for various fixed smooth functions  $F_1, \dots, F_k, G_1, \dots, G_k : [0, +\infty) \rightarrow \mathbb{R}$ . The sum appearing in (13) can thus be decomposed into fixed linear combinations of sums of the form

$$\sum_{\substack{x \leq n \leq 2x \\ n=b \pmod{W}}} \prod_{i=1}^k \lambda_{F_i}(n+h_i) \lambda_{G_i}(n+h_i). \quad (19)$$

Also, if  $F$  is supported on  $[0, 1]$ , then from (16) we clearly have

$$\lambda_F(n) = F(0) \quad (20)$$

when  $n \geq x$  is prime, and so the sum appearing in (14) can be similarly decomposed in this case into fixed linear combinations of sums of the form

$$\sum_{\substack{x \leq n \leq 2x \\ n=b \pmod{W}}} \theta(n+h_i) \prod_{1 \leq i' \leq k; i' \neq i} \lambda_{F_{i'}}(n+h_{i'}) \lambda_{G_{i'}}(n+h_{i'}). \quad (21)$$

To estimate the sums (21), we use the following asymptotic, proven in Section 4. For each compactly supported  $F : [0, +\infty) \rightarrow \mathbb{R}$ , let

$$S(F) := \sup\{x \geq 0 : F(x) \neq 0\} \quad (22)$$

denote the upper range of the support of  $F$  (with the convention that  $S(0) = 0$ ).

**Theorem 3.5** (Asymptotic for prime sums) *Let  $k \geq 2$  be fixed, let  $(h_1, \dots, h_k)$  be a fixed admissible  $k$ -tuple, and let  $b \pmod{W}$  be such that  $b+h_i$  is coprime to  $W$  for each  $i = 1, \dots, k$ . Let  $1 \leq i_0 \leq k$  be fixed, and for each  $1 \leq i \leq k$  distinct from  $i_0$ , let  $F_i, G_i : [0, +\infty) \rightarrow \mathbb{R}$  be fixed smooth compactly supported functions. Assume one of the following hypotheses:*

(i) (Elliott-Halberstam) *There exists a fixed  $0 < \vartheta < 1$  such that  $\text{EH}[\vartheta]$  holds, and such that*

$$\sum_{1 \leq i \leq k; i \neq i_0} (S(F_i) + S(G_i)) < \vartheta. \quad (23)$$

(ii) (Motohashi-Pintz-Zhang) There exists fixed  $0 \leq \varpi < 1/4$  and  $\delta > 0$  such that  $\text{MPZ}[\varpi, \delta]$  holds, and such that

$$\sum_{1 \leq i \leq k; i \neq i_0} (S(F_i) + S(G_i)) < \frac{1}{2} + 2\varpi \quad (24)$$

and

$$\max_{1 \leq i \leq k; i \neq i_0} \{S(F_i), S(G_i)\} < \delta. \quad (25)$$

Then we have

$$\sum_{\substack{x \leq n \leq 2x \\ n=b \ (W)}} \theta(n + h_{i_0}) \prod_{1 \leq i \leq k; i \neq i_0} \lambda_{F_i}(n + h_i) \lambda_{G_i}(n + h_i) = (c + o(1)) B^{1-k} \frac{x}{\varphi(W)} \quad (26)$$

where  $B$  is given by (12) and

$$c := \prod_{1 \leq i \leq k; i \neq i_0} \left( \int_0^1 F'_i(t_i) G'_i(t_i) dt_i \right).$$

Here of course  $F'$  denotes the derivative of  $F$ .

To estimate the sums (19), we use the following asymptotic, also proven in Section 4.

**Theorem 3.6** (Asymptotic for non-prime sums) *Let  $k \geq 1$  be fixed, let  $(h_1, \dots, h_k)$  be a fixed admissible  $k$ -tuple, and let  $b \ (W)$  be such that  $b + h_i$  is coprime to  $W$  for each  $i = 1, \dots, k$ . For each fixed  $1 \leq i \leq k$ , let  $F_i, G_i : [0, +\infty) \rightarrow \mathbb{R}$  be fixed smooth compactly supported functions. Assume one of the following hypotheses:*

(i) (Trivial case) One has

$$\sum_{i=1}^k (S(F_i) + S(G_i)) < 1. \quad (27)$$

(ii) (Generalized Elliott-Halberstam) There exists a fixed  $0 < \vartheta < 1$  and  $i_0 \in \{1, \dots, k\}$  such that  $\text{GEH}[\vartheta]$  holds, and

$$\sum_{1 \leq i \leq k; i \neq i_0} (S(F_i) + S(G_i)) < \vartheta. \quad (28)$$

Then we have

$$\sum_{\substack{x \leq n \leq 2x \\ n=b \ (W)}} \prod_{i=1}^k \lambda_{F_i}(n + h_i) \lambda_{G_i}(n + h_i) = (c + o(1)) B^{-k} \frac{x}{W}, \quad (29)$$

where  $B$  is given by (12) and

$$c := \prod_{i=1}^k \left( \int_0^1 F'_i(t_i) G'_i(t_i) dt_i \right). \quad (30)$$

A key point in (ii) is that no upper bound on  $S(F_{i_0})$  or  $S(G_{i_0})$  is required (although, as we will see in Section 4.5, the result is a little easier to prove when one has  $S(F_{i_0}) + S(G_{i_0}) < 1$ ). This flexibility in the  $F_{i_0}, G_{i_0}$  functions will be particularly crucial to obtain part (xii) of Theorem 3.2 and Theorem 1.4.

**Remark 3.7** Theorems 3.5, 3.6 can be viewed as probabilistic assertions of the following form: if  $n$  is chosen uniformly at random from the set  $\{x \leq n \leq 2x : n = b(W)\}$ , then the random variables  $\theta(n + h_i)$  and  $\lambda_{F_j}(n + h_j)\lambda_{G_j}(n + h_j)$  for  $i, j = 1, \dots, k$  have mean  $(1 + o(1))\frac{W}{\varphi(W)}$  and  $(\int_0^1 F'_j(t)G'_j(t) dt + o(1))B^{-1}$  respectively, and furthermore these random variables enjoy a limited amount of independence, except for the fact (as can be seen from (20)) that  $\theta(n + h_i)$  and  $\lambda_{F_i}(n + h_i)\lambda_{G_i}(n + h_i)$  are highly correlated. Note though that we do not have asymptotics for any sum which involves two or more factors of  $\theta$ , as such estimates are of a difficulty at least as great as that of the twin prime conjecture (which is equivalent to the divergence of the sum  $\sum_n \theta(n)\theta(n + 2)$ ).

Theorems 3.5, 3.6 may be combined with Lemma 3.4 to reduce the task of establishing estimates of the form  $\text{DHL}[k, m + 1]$  to that of obtaining sufficiently good solutions to certain variational problems. For instance, in Section 5.1 we reprove the following result of Maynard [38, Proposition 4.2]:

**Theorem 3.8** (Sieving on the standard simplex) *Let  $k \geq 2$  and  $m \geq 1$  be fixed integers. For any fixed compactly supported square-integrable function  $F : [0, +\infty)^k \rightarrow \mathbb{R}$ , define the functionals*

$$I(F) := \int_{[0, +\infty)^k} F(t_1, \dots, t_k)^2 dt_1 \dots dt_k \quad (31)$$

and

$$J_i(F) := \int_{[0, +\infty)^{k-1}} \left( \int_0^\infty F(t_1, \dots, t_k) dt_i \right)^2 dt_1 \dots dt_{i-1} dt_{i+1} \dots dt_k \quad (32)$$

for  $i = 1, \dots, k$ , and let  $M_k$  be the supremum

$$M_k := \sup \frac{\sum_{i=1}^k J_i(F)}{I(F)} \quad (33)$$

over all square-integrable functions  $F$  that are supported on the simplex

$$\mathcal{R}_k := \{(t_1, \dots, t_k) \in [0, +\infty)^k : t_1 + \dots + t_k \leq 1\}$$

and are not identically zero (up to almost everywhere equivalence, of course). Suppose that there is a fixed  $0 < \vartheta < 1$  such that  $\text{EH}[\vartheta]$  holds, and such that

$$M_k > \frac{2m}{\vartheta}.$$

Then  $\text{DHL}[k, m + 1]$  holds.

Parts (vii)-(xi) of Theorem 3.2 (and hence Theorem 1.4) are then immediate from the following results, proven in Sections 6, 7, and ordered by increasing value of  $k$ :

**Theorem 3.9** (Lower bounds on  $M_k$ )

- (vii)  $M_{54} > 4.00238$ .
- (viii)  $M_{5511} > 6$ .
- (ix)  $M_{41588} > 8$ .
- (x)  $M_{309661} > 10$ .
- (xi) One has  $M_k \geq \log k - C$  for all  $k \geq C$ , where  $C$  is an absolute (and effective) constant.

For sake of comparison, in [38, Proposition 4.3] it was shown that  $M_5 > 2$ ,  $M_{105} > 4$ , and  $M_k \geq \log k - 2 \log \log k - 2$  for all sufficiently large  $k$ . As remarked in that paper, the sieves used on the bounded gap problem prior to the work in [38] would essentially correspond, in this notation, to the choice of functions  $F$  of the special form  $F(t_1, \dots, t_k) := f(t_1 + \dots + t_k)$ , which severely limits the size of the ratio in (33) (in particular, the analogue of  $M_k$  in this special case cannot exceed 4, as shown in [59]).

In the converse direction, in Corollary 6.4 we will also show the upper bound  $M_k \leq \frac{k}{k-1} \log k$  for all  $k \geq 2$ , which shows in particular that the bounds in (vii) and (xi) of the above theorem cannot be significantly improved. We remark that Theorem 3.9(vii) and the Bombieri-Vinogradov theorem also gives a weaker version DHL[54, 2] of Theorem 3.2(i).

We also have a variant of Theorem 3.8 which can accept inputs of the form  $\text{MPZ}[\varpi, \delta]$ :

**Theorem 3.10** (Sieving on a truncated simplex) *Let  $k \geq 2$  and  $m \geq 1$  be fixed integers. Let  $0 < \varpi < 1/4$  and  $0 < \delta < 1/2$  be such that  $\text{MPZ}[\varpi, \delta]$  holds. For any  $\alpha > 0$ , let  $M_k^{[\alpha]}$  be defined as in (33), but where the supremum now ranges over all square-integrable functions  $F$  supported in the truncated simplex*

$$\{(t_1, \dots, t_k) \in [0, \alpha]^k : t_1 + \dots + t_k \leq 1\} \quad (34)$$

*and are not identically zero. If*

$$M_k^{\lfloor \frac{\delta}{1/4+\varpi} \rfloor} > \frac{m}{1/4+\varpi},$$

*then DHL[ $k, m+1$ ] holds.*

In Section 6 we will establish the following variant of Theorem 3.9, which when combined with Theorem 2.5, allows one to use Theorem 3.10 to establish parts (ii)-(vi) of Theorem 3.2 (and hence Theorem 1.4):

**Theorem 3.11** (Lower bounds on  $M_k^{[\alpha]}$ )

- (ii) *There exist  $\delta, \varpi > 0$  with  $600\varpi + 180\delta < 7$  and  $M_{35\,410}^{\lfloor \frac{\delta}{1/4+\varpi} \rfloor} > \frac{2}{1/4+\varpi}$ .*
- (iii) *There exist  $\delta, \varpi > 0$  with  $600\varpi + 180\delta < 7$  and  $M_{1\,649\,821}^{\lfloor \frac{\delta}{1/4+\varpi} \rfloor} > \frac{3}{1/4+\varpi}$ .*
- (iv) *There exist  $\delta, \varpi > 0$  with  $600\varpi + 180\delta < 7$  and  $M_{75\,845\,707}^{\lfloor \frac{\delta}{1/4+\varpi} \rfloor} > \frac{4}{1/4+\varpi}$ .*
- (v) *There exist  $\delta, \varpi > 0$  with  $600\varpi + 180\delta < 7$  and  $M_{3\,473\,955\,908}^{\lfloor \frac{\delta}{1/4+\varpi} \rfloor} > \frac{5}{1/4+\varpi}$ .*
- (vi) *For all  $k \geq C$ , there exist  $\delta, \varpi > 0$  with  $600\varpi + 180\delta < 7$ ,  $\varpi \geq \frac{7}{600} - \frac{C}{\log k}$ , and  $M_k^{\lfloor \frac{\delta}{1/4+\varpi} \rfloor} \geq \log k - C$  for some absolute (and effective) constant  $C$ .*

The implication is clear for (ii)-(v). For (vi), observe that from Theorem 3.11(vi), Theorem 2.5, and Theorem 3.10, we see that DHL[ $k, m+1$ ] holds whenever  $k$  is sufficiently large and

$$m \leq (\log k - C) \left( \frac{1}{4} + \frac{7}{600} - \frac{C}{\log k} \right)$$

which is in particular implied by

$$m \leq \frac{\log k}{4 - \frac{28}{157}} - C'$$

for some absolute constant  $C'$ , giving Theorem 3.2(vi).

Now we give a more flexible variant of Theorem 3.8, in which the support of  $F$  is enlarged, at the cost of reducing the range of integration of the  $J_i$ .

**Theorem 3.12** (Sieving on an epsilon-enlarged simplex) *Let  $k \geq 2$  and  $m \geq 1$  be fixed integers, and let  $0 < \varepsilon < 1$  be fixed also. For any fixed compactly supported square-integrable function  $F : [0, +\infty)^k \rightarrow \mathbb{R}$ , define the functionals*

$$J_{i,1-\varepsilon}(F) := \int_{(1-\varepsilon) \cdot \mathcal{R}_{k-1}} \left( \int_0^\infty F(t_1, \dots, t_k) dt_i \right)^2 dt_1 \dots dt_{i-1} dt_{i+1} \dots dt_k$$

for  $i = 1, \dots, k$ , and let  $M_{k,\varepsilon}$  be the supremum

$$M_{k,\varepsilon} := \sup \frac{\sum_{i=1}^k J_{i,1-\varepsilon}(F)}{I(F)}$$

over all square-integrable functions  $F$  that are supported on the simplex

$$(1 + \varepsilon) \cdot \mathcal{R}_k = \{(t_1, \dots, t_k) \in [0, +\infty)^k : t_1 + \dots + t_k \leq 1 + \varepsilon\}$$

and are not identically zero. Suppose that there is a fixed  $0 < \vartheta < 1$ , such that one of the following two hypotheses holds:

- (i)  $\text{EH}[\vartheta]$  holds, and  $1 + \varepsilon < \frac{1}{\vartheta}$ .
- (ii)  $\text{GEH}[\vartheta]$  holds, and  $\varepsilon < \frac{1}{k-1}$ .

If

$$M_{k,\varepsilon} > \frac{2m}{\vartheta}$$

then  $\text{DHL}[k, m + 1]$  holds.

We prove this theorem in Section 5.3. We remark that due to the continuity of  $M_{k,\varepsilon}$  in  $\varepsilon$ , the strict inequalities in (i), (ii) of this theorem may be replaced by non-strict inequalities. Parts (i), (xiii) of Theorem 3.2, and a weaker version  $\text{DHL}[4, 2]$  of part (xii), then follow from Theorem 2.3 and the following computations, proven in Sections 7.2, 7.3:

**Theorem 3.13** (Lower bounds on  $M_{k,\varepsilon}$ )

- (i)  $M_{50,1/25} > 4.0043$ .
- (xii')  $M_{4,0.168} > 2.00558$ .
- (xiii)  $M_{51,1/50} > 4.00156$ .

We remark that computations in the proof of Theorem 3.13(xii') are simple enough that the bound may be checked by hand, without use of a computer. The computations used to establish the full strength of Theorem 3.2(xii) are however significantly more complicated.

In fact, we may enlarge the support of  $F$  further. We give a version corresponding to part (ii) of Theorem 3.12; there is also a version corresponding to part (i), but we will not give it here as we will not have any use for it.



**Theorem 3.14** (Going beyond the epsilon enlargement) *Let  $k \geq 2$  and  $m \geq 1$  be fixed integers, let  $0 < \vartheta < 1$  be a fixed quantity such that  $\text{GEH}[\vartheta]$  holds, and let  $0 < \varepsilon < \frac{1}{k-1}$  be fixed also. Suppose that there is a fixed non-zero square-integrable function  $F : [0, +\infty)^k \rightarrow \mathbb{R}$  supported in  $\frac{k}{k-1} \cdot \mathcal{R}_k$ , such that for  $i = 1, \dots, k$  one has the vanishing marginal condition*

$$\int_0^\infty F(t_1, \dots, t_k) dt_i = 0 \quad (35)$$

*whenever  $t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_k \geq 0$  are such that*

$$t_1 + \dots + t_{i-1} + t_{i+1} + \dots + t_k > 1 + \varepsilon.$$

*Suppose that we also have the inequality*

$$\frac{\sum_{i=1}^k J_{i,1-\varepsilon}(F)}{I(F)} > \frac{2m}{\vartheta}.$$

*Then  $\text{DHL}[k, m+1]$  holds.*

This theorem is proven in Section 5.4. Theorem 3.2(xii) is then an immediate consequence of Theorem 3.14 and the following numerical fact, established in Section 7.4.

**Theorem 3.15** (A piecewise polynomial cutoff) *Set  $\varepsilon := \frac{1}{4}$ . Then there exists a piecewise polynomial function  $F : [0, +\infty)^3 \rightarrow \mathbb{R}$  supported on the simplex*

$$\frac{3}{2} \cdot \mathcal{R}_3 = \left\{ (t_1, t_2, t_3) \in [0, +\infty)^3 : t_1 + t_2 + t_3 \leq \frac{3}{2} \right\}$$

*and symmetric in the  $t_1, t_2, t_3$  variables, such that  $F$  is not identically zero and obeys the vanishing marginal condition*

$$\int_0^\infty F(t_1, t_2, t_3) dt_3 = 0$$

*whenever  $t_1, t_2 \geq 0$  with  $t_1 + t_2 > 1 + \varepsilon$ , and such that*

$$\frac{3 \int_{t_1+t_2 \leq 1-\varepsilon} \left( \int_0^\infty F(t_1, t_2, t_3) dt_3 \right)^2 dt_1 dt_2}{\int_{[0, \infty)^3} F(t_1, t_2, t_3)^2 dt_1 dt_2 dt_3} > 2.$$

There are several other ways to combine Theorems 3.5, 3.6 with equidistribution theorems on the primes to obtain results of the form  $\text{DHL}[k, m+1]$ , but all of our attempts to do so either did not improve the numerology, or else were numerically infeasible to implement.

## 4 Multidimensional Selberg sieves

In this section we prove Theorems 3.5 and 3.6. A key asymptotic used in both theorems is the following:

**Lemma 4.1** (Asymptotic) *Let  $k \geq 1$  be a fixed integer, and let  $N$  be a natural number coprime to  $W$  with  $\log N = O(\log^{O(1)} x)$ . Let  $F_1, \dots, F_k, G_1, \dots, G_k : [0, +\infty) \rightarrow \mathbb{R}$  be fixed smooth compactly supported functions. Then*

$$\sum_{\substack{d_1, \dots, d_k, d'_1, \dots, d'_k \\ [d_1, d'_1], \dots, [d_k, d'_k], W, N \text{ coprime}}} \prod_{j=1}^k \frac{\mu(d_j) \mu(d'_j) F_j(\log_x d_j) G_j(\log_x d'_j)}{[d_j, d'_j]} = (c + o(1)) B^{-k} \frac{N^k}{\varphi(N)^k} \quad (36)$$

where  $B$  was defined in (12), and

$$c := \prod_{j=1}^k \int_0^\infty F'_j(t_j) G'_j(t_j) dt_j.$$

The same claim holds if the denominators  $[d_j, d'_j]$  are replaced by  $\varphi([d_j, d'_j])$ .

Such asymptotics are standard in the literature; see e.g. [27] for some similar computations. In older literature, it is common to establish these asymptotics via contour integration (e.g. via Perron's formula), but we will use the Fourier-analytic approach here. Of course, both approaches ultimately use the same input, namely the simple pole of the Riemann zeta function at  $s = 1$ .

*Proof* We begin with the first claim. For  $j = 1, \dots, k$ , the functions  $t \mapsto e^t F_j(t)$ ,  $t \mapsto e^t G_j(t)$  may be extended to smooth compactly supported functions on all of  $\mathbb{R}$ , and so we have Fourier expansions

$$e^t F_j(t) = \int_{\mathbb{R}} e^{-it\xi} f_j(\xi) d\xi \quad (37)$$

and

$$e^t G_j(t) = \int_{\mathbb{R}} e^{-it\xi} g_j(\xi) d\xi$$

for some fixed functions  $f_j, g_j : \mathbb{R} \rightarrow \mathbb{C}$  that are smooth and rapidly decreasing in the sense that  $f_j(\xi), g_j(\xi) = O((1 + |\xi|)^{-A})$  for any fixed  $A > 0$  and all  $\xi \in \mathbb{R}$  (here the implied constant is independent of  $\xi$  and depends only on  $A$ ).

We may thus write

$$F_j(\log x d_j) = \int_{\mathbb{R}} \frac{f_j(\xi_j)}{d_j^{\frac{1+i\xi_j}{\log x}}} d\xi_j$$

and

$$G_j(\log x d'_j) = \int_{\mathbb{R}} \frac{g_j(\xi'_j)}{(d'_j)^{\frac{1+i\xi'_j}{\log x}}} d\xi'_j$$

for all  $d_j, d'_j \geq 1$ . We note that

$$\begin{aligned} \sum_{d_j, d'_j} \frac{|\mu(d_j)\mu(d'_j)|}{[d_j, d'_j] d_j^{1/\log x} (d'_j)^{1/\log x}} &= \prod_p \left(1 + \frac{2}{p^{1+1/\log x}} + \frac{1}{p^{1+2/\log x}}\right) \\ &\leq \zeta \left(1 + \frac{1}{\log x}\right)^3 \\ &\ll \log^3 x. \end{aligned}$$

Therefore, if we substitute the Fourier expansions into the left-hand side of (36), the resulting expression is absolutely convergent. Thus we can apply Fubini's theorem, and the left-hand side of (36) can thus be rewritten as

$$\int_{\mathbb{R}} \dots \int_{\mathbb{R}} K(\xi_1, \dots, \xi_k, \xi'_1, \dots, \xi'_k) \prod_{j=1}^k f_j(\xi_j) g_j(\xi'_j) d\xi_j d\xi'_j, \quad (38)$$

where

$$K(\xi_1, \dots, \xi_k, \xi'_1, \dots, \xi'_k) := \sum_{\substack{d_1, \dots, d_k, d'_1, \dots, d'_k \\ [d_1, d'_1], \dots, [d_k, d'_k], W, N \text{ coprime}}} \prod_{j=1}^k \frac{\mu(d_j)\mu(d'_j)}{[d_j, d'_j] d_j^{\frac{1+i\xi_j}{\log x}} (d'_j)^{\frac{1+i\xi'_j}{\log x}}}.$$

This latter expression factorizes as an Euler product

$$K = \prod_{p \nmid WN} K_p,$$

where the local factors  $K_p$  are given by

$$K_p(\xi_1, \dots, \xi_k, \xi'_1, \dots, \xi'_k) := 1 + \frac{1}{p} \sum_{\substack{d_1, \dots, d_k, d'_1, \dots, d'_k \\ [d_1, \dots, d_k, d'_1, \dots, d'_k] = p \\ [d_1, d'_1], \dots, [d_k, d'_k] \text{ coprime}}} \prod_{j=1}^k \frac{\mu(d_j)\mu(d'_j)}{d_j^{\frac{1+i\xi_j}{\log x}} (d'_j)^{\frac{1+i\xi'_j}{\log x}}}. \quad (39)$$

We can estimate each Euler factor as

$$K_p(\xi_1, \dots, \xi_k, \xi'_1, \dots, \xi'_k) = \left(1 + O\left(\frac{1}{p^2}\right)\right) \prod_{j=1}^k \frac{\left(1 - p^{-1 - \frac{1+i\xi_j}{\log x}}\right) \left(1 - p^{-1 - \frac{1+i\xi'_j}{\log x}}\right)}{1 - p^{-1 - \frac{2+i\xi_j+i\xi'_j}{\log x}}}. \quad (40)$$

Since

$$\prod_{p: p > w} \left(1 + O\left(\frac{1}{p^2}\right)\right) = 1 + o(1),$$

we have

$$K(\xi_1, \dots, \xi_k, \xi'_1, \dots, \xi'_k) = (1 + o(1)) \prod_{j=1}^k \frac{\zeta_{WN}(1 + \frac{2+i\xi_j+i\xi'_j}{\log x})}{\zeta_{WN}(1 + \frac{1+i\xi_j}{\log x}) \zeta_{WN}(1 + \frac{1+i\xi'_j}{\log x})}$$

where the modified zeta function  $\zeta_{WN}$  is defined by the formula

$$\zeta_{WN}(s) := \prod_{p \nmid WN} \left(1 - \frac{1}{p^s}\right)^{-1}$$

for  $\operatorname{Re}(s) > 1$ .

For  $\operatorname{Re}(s) \geq 1 + \frac{1}{\log x}$  we have the crude bounds

$$|\zeta_{WN}(s)|, |\zeta_{WN}(s)|^{-1} \leq \zeta\left(1 + \frac{1}{\log x}\right) \ll \log x$$

where the first inequality comes from comparing the factors in the Euler product. Thus

$$K(\xi_1, \dots, \xi_k, \xi'_1, \dots, \xi'_k) = O(\log^{3k} x).$$

Combining this with the rapid decrease of  $f_j, g_j$ , we see that the contribution to (38) outside of the cube  $\{\max(|\xi_1|, \dots, |\xi_k|, |\xi'_1|, \dots, |\xi'_k|) \leq \sqrt{\log x}\}$  (say) is negligible. Thus it will suffice to show that

$$\int_{-\sqrt{\log x}}^{\sqrt{\log x}} \dots \int_{-\sqrt{\log x}}^{\sqrt{\log x}} K(\xi_1, \dots, \xi_k, \xi'_1, \dots, \xi'_k) \prod_{j=1}^k f_j(\xi_j) g_j(\xi'_j) d\xi_j d\xi'_j = (c + o(1)) B^{-k} \frac{N^k}{\varphi(N)^k}.$$

When  $|\xi_j| \leq \sqrt{\log x}$ , we see from the simple pole of the Riemann zeta function  $\zeta(s) = \prod_p (1 - \frac{1}{p^s})^{-1}$  at  $s = 1$  that

$$\zeta\left(1 + \frac{1 + i\xi_j}{\log x}\right) = (1 + o(1)) \frac{\log x}{1 + i\xi_j}.$$

For  $-\sqrt{\log x} \leq \xi_j \leq \sqrt{\log x}$ , we see that

$$1 - \frac{1}{p^{1 + \frac{1 + i\xi_j}{\log x}}} = 1 - \frac{1}{p} + O\left(\frac{\log p}{p\sqrt{\log x}}\right).$$

Since  $\log(WN) \ll \log^{O(1)} x$ , this gives

$$\prod_{p|WN} \left(1 - \frac{1}{p^{1 + \frac{1 + i\xi_j}{\log x}}}\right) = \frac{\varphi(WN)}{WN} \exp\left(O\left(\sum_{p|WN} \frac{\log p}{p\sqrt{\log x}}\right)\right) = (1 + o(1)) \frac{\varphi(WN)}{WN},$$

since the sum is maximized when  $WN$  is composed only of primes  $p \ll \log^{O(1)} x$ . Thus

$$\zeta_{WN}\left(1 + \frac{1 + i\xi_j}{\log x}\right) = \frac{(1 + o(1))B\varphi(N)}{(1 + i\xi_j)N}.$$

Similarly with  $1 + i\xi_j$  replaced by  $1 + i\xi'_j$  or  $2 + i\xi_j + i\xi'_j$ . We conclude that

$$K(\xi_1, \dots, \xi_k, \xi'_1, \dots, \xi'_k) = (1 + o(1))B^{-k} \frac{N^k}{\varphi(N)^k} \prod_{j=1}^k \frac{(1 + i\xi_j)(1 + i\xi'_j)}{2 + i\xi_j + i\xi'_j}. \quad (41)$$

Therefore it will suffice to show that

$$\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \prod_{j=1}^k \frac{(1 + i\xi_j)(1 + i\xi'_j)}{2 + i\xi_j + i\xi'_j} f_j(\xi_j) g_j(\xi'_j) d\xi_j d\xi'_j = c,$$

since the errors caused by the  $1 + o(1)$  multiplicative factor in (41) or the truncation  $|\xi_j|, |\xi'_j| \leq \sqrt{\log x}$  can be seen to be negligible using the rapid decay of  $f_j, g_j$ . By Fubini's theorem, it suffices to show that

$$\int_{\mathbb{R}} \int_{\mathbb{R}} \frac{(1 + i\xi)(1 + i\xi')}{2 + i\xi + i\xi'} f_j(\xi) g_j(\xi') d\xi d\xi' = \int_0^{+\infty} F'_j(t) G'_j(t) dt$$

for each  $j = 1, \dots, k$ . But from dividing (37) by  $e^t$  and differentiating under the integral sign, we have

$$F'_j(t) = - \int_{\mathbb{R}} (1 + i\xi) e^{-t(1 + i\xi)} f_j(\xi) d\xi,$$

and the claim then follows from Fubini's theorem.

Finally, suppose that we replace the denominators  $[d_j, d'_j]$  with  $\varphi([d_j, d'_j])$ . An inspection of the above argument shows that the only change that occurs is that the  $\frac{1}{p}$  term in (39) is replaced by  $\frac{1}{p-1}$ ; but this modification may be absorbed into the  $1 + O(\frac{1}{p^2})$  factor in (40), and the rest of the argument continues as before.  $\square$

#### 4.1 The trivial case

We can now prove the easiest case of the two theorems, namely case (i) of Theorem 3.6; a closely related estimate also appears in [38, Lemma 6.2]. We may assume that  $x$  is sufficiently large depending on all

fixed quantities. By (16), the left-hand side of (29) may be expanded as

$$\sum_{d_1, \dots, d_k, d'_1, \dots, d'_k} \left( \prod_{i=1}^k \mu(d_i) \mu(d'_i) F_i(\log_x d_i) G_i(\log_x d'_i) \right) S(d_1, \dots, d_k, d'_1, \dots, d'_k) \quad (42)$$

where

$$S(d_1, \dots, d_k, d'_1, \dots, d'_k) := \sum_{\substack{x \leq n \leq 2x \\ n \equiv b \pmod{W} \\ n + h_i \equiv 0 \pmod{[d_i, d'_i]} \quad \forall i}} 1.$$

By hypothesis,  $b + h_i$  is coprime to  $W$  for all  $i = 1, \dots, k$ , and  $|h_i - h_j| < w$  for all distinct  $i, j$ . Thus,  $S(d_1, \dots, d_k, d'_1, \dots, d'_k)$  vanishes unless the  $[d_i, d'_i]$  are coprime to each other and to  $W$ . In this case,  $S(d_1, \dots, d_k, d'_1, \dots, d'_k)$  is summing the constant function 1 over an arithmetic progression in  $[x, 2x]$  of spacing  $W[d_1, d'_1] \dots [d_k, d'_k]$ , and so

$$S(d_1, \dots, d_k, d'_1, \dots, d'_k) = \frac{x}{W[d_1, d'_1] \dots [d_k, d'_k]} + O(1).$$

By Lemma 4.1, the contribution of the main term  $\frac{x}{W[d_1, d'_1] \dots [d_k, d'_k]}$  to (29) is  $(c + o(1))B^{-k} \frac{x}{W}$ ; note that the restriction of the integrals in (30) to  $[0, 1]$  instead of  $[0, +\infty)$  is harmless since  $S(F_i), S(G_i) < 1$  for all  $i$ . Meanwhile, the contribution of the  $O(1)$  error is then bounded by

$$O\left( \sum_{d_1, \dots, d_k, d'_1, \dots, d'_k} \left( \prod_{i=1}^k |F_i(\log_x d_i)| |G_i(\log_x d'_i)| \right) \right).$$

By the hypothesis in Theorem 3.6(i), we see that for  $d_1, \dots, d_k, d'_1, \dots, d'_k$  contributing a non-zero term here, one has

$$[d_1, d'_1] \dots [d_k, d'_k] \ll x^{1-\varepsilon}$$

for some fixed  $\varepsilon > 0$ . From the divisor bound (1) we see that each choice of  $[d_1, d'_1] \dots [d_k, d'_k]$  arises from  $\ll 1$  choices of  $d_1, \dots, d_k, d'_1, \dots, d'_k$ . We conclude that the net contribution of the  $O(1)$  error to (29) is  $\ll x^{1-\varepsilon}$ , and the claim follows.

#### 4.2 The Elliott-Halberstam case

Now we show case (i) of Theorem 3.5. For sake of notation we take  $i_0 = k$ , as the other cases are similar. We use (16) to rewrite the left-hand side of (26) as

$$\sum_{d_1, \dots, d_{k-1}, d'_1, \dots, d'_{k-1}} \left( \prod_{i=1}^{k-1} \mu(d_i) \mu(d'_i) F_i(\log_x d_i) G_i(\log_x d'_i) \right) \tilde{S}(d_1, \dots, d_{k-1}, d'_1, \dots, d'_{k-1}) \quad (43)$$

where

$$\tilde{S}(d_1, \dots, d_{k-1}, d'_1, \dots, d'_{k-1}) := \sum_{\substack{x \leq n \leq 2x \\ n \equiv b \pmod{W} \\ n + h_i \equiv 0 \pmod{[d_i, d'_i]} \quad \forall i=1, \dots, k-1}} \theta(n + h_k).$$

As in the previous case,  $\tilde{S}(d_1, \dots, d_{k-1}, d'_1, \dots, d'_{k-1})$  vanishes unless the  $[d_i, d'_i]$  are coprime to each other and to  $W$ , and so the summand in (43) vanishes unless the modulus  $q_{W, d_1, \dots, d'_{k-1}}$  defined by

$$q_{W, d_1, \dots, d'_{k-1}} := W[d_1, d'_1] \dots [d_{k-1}, d'_{k-1}] \quad (44)$$

is squarefree. In that case, we may use the Chinese remainder theorem to concatenate the congruence conditions on  $n$  into a single primitive congruence condition

$$n + h_k = a_{W,d_1,\dots,d'_{k-1}}(q_{W,d_1,\dots,d'_{k-1}})$$

for some  $a_{W,d_1,\dots,d'_{k-1}}$  depending on  $W, d_1, \dots, d_{k-1}, d'_1, \dots, d'_{k-1}$ , and conclude using (3) that

$$\begin{aligned} \tilde{S}(d_1, \dots, d_{k-1}, d'_1, \dots, d'_{k-1}) &= \frac{1}{\varphi(q_{W,d_1,\dots,d'_{k-1}})} \sum_{x+h_k \leq n \leq 2x+h_k} \theta(n) \\ &\quad + \Delta(\mathbf{1}_{[x+h_k, 2x+h_k]} \theta; a_{W,d_1,\dots,d'_{k-1}}(q_{W,d_1,\dots,d'_{k-1}})). \end{aligned} \quad (45)$$

From the prime number theorem we have

$$\sum_{x+h_k \leq n \leq 2x+h_k} \theta(n) = (1 + o(1))x$$

and this expression is clearly independent of  $d_1, \dots, d'_{k-1}$ . Thus by Lemma 4.1, the contribution of the main term in (45) to (43) is  $(c + o(1))B^{1-k} \frac{x}{\varphi(W)}$ . By (11) and (12), it thus suffices to show that for any fixed  $A$  we have

$$\sum_{d_1, \dots, d_{k-1}, d'_1, \dots, d'_{k-1}} \left( \prod_{i=1}^{k-1} |F_i(\log_x d_i)| |G_i(\log_x d'_i)| \right) |\Delta(\mathbf{1}_{[x+h_k, 2x+h_k]} \theta; a(q))| \ll x \log^{-A} x, \quad (46)$$

where  $a = a_{W,d_1,\dots,d'_{k-1}}$  and  $q = q_{W,d_1,\dots,d'_{k-1}}$ . For future reference we note that we may restrict the summation here to those  $d_1, \dots, d'_{k-1}$  for which  $q_{W,d_1,\dots,d'_{k-1}}$  is square-free.

From the hypotheses of Theorem 3.5(i), we have

$$q_{W,d_1,\dots,d'_{k-1}} \ll x^\vartheta$$

whenever the summand in (43) is non-zero, and each choice  $q$  of  $q_{W,d_1,\dots,d'_{k-1}}$  is associated to  $O(\tau(q)^{O(1)})$  choices of  $d_1, \dots, d_{k-1}, d'_1, \dots, d'_{k-1}$ . Thus this contribution is

$$\ll \sum_{q \ll x^\vartheta} \tau(q)^{O(1)} \sup_{a \in (\mathbb{Z}/q\mathbb{Z})^\times} |\Delta(\mathbf{1}_{[x+h_k, 2x+h_k]} \theta; a(q))|.$$

Using the crude bound

$$|\Delta(\mathbf{1}_{[x+h_k, 2x+h_k]} \theta; a(q))| \ll \frac{x}{q} \log^{O(1)} x$$

and (2), we have

$$\sum_{q \ll x^\vartheta} \tau(q)^C \sup_{a \in (\mathbb{Z}/q\mathbb{Z})^\times} |\Delta(\mathbf{1}_{[x+h_k, 2x+h_k]} \theta; a(q))| \ll x \log^{O(1)} x$$

for any fixed  $C > 0$ . By the Cauchy-Schwarz inequality it suffices to show that

$$\sum_{q \ll x^\vartheta} \sup_{a \in (\mathbb{Z}/q\mathbb{Z})^\times} |\Delta(\mathbf{1}_{[x+h_k, 2x+h_k]} \theta; a(q))| \ll x \log^{-A} x$$

for any fixed  $A > 0$ . However, since  $\theta$  only differs from  $\Lambda$  on powers  $p^j$  of primes with  $j > 1$ , it is not difficult to show that

$$|\Delta(\mathbf{1}_{[x+h_k, 2x+h_k]} \theta; a(q)) - \Delta(\mathbf{1}_{[x+h_k, 2x+h_k]} \Lambda; a(q))| \ll \sqrt{\frac{x}{q}},$$

so the net error in replacing  $\theta$  here by  $\Lambda$  is  $\ll x^{1-(1-\vartheta)/2}$ , which is certainly acceptable. The claim now follows from the hypothesis  $\text{EH}[\vartheta]$ , thanks to Claim 2.2.

### 4.3 The Motohashi-Pintz-Zhang case

Now we show case (ii) of Theorem 3.5. We repeat the arguments from Section 4.2, with the only difference being in the derivation of (46). As observed previously, we may restrict  $q_{W,d_1,\dots,d'_{k-1}}$  to be squarefree. From the hypotheses in Theorem 3.5(ii), we also see that

$$q_{W,d_1,\dots,d'_{k-1}} \ll x^{1/2+2\varpi}$$

and that all the prime factors of  $q_{W,d_1,\dots,d'_{k-1}}$  are at most  $x^\delta$ . Thus, if we set  $I := [1, x^\delta]$ , we see (using the notation from Claim 2.4) that  $q_{W,d_1,\dots,d'_{k-1}}$  lies in  $\mathcal{S}_I$ , and is thus a factor of  $P_I$ . If we then let  $\mathcal{A} \subset \mathbb{Z}/P_I\mathbb{Z}$  denote all the primitive residue classes  $a \pmod{P_I}$  with the property that  $a \equiv b \pmod{W}$ , and such that for each prime  $w < p \leq x^\delta$ , one has  $a + h_i \equiv 0 \pmod{p}$  for some  $i = 1, \dots, k$ , then we see that  $a_{W,d_1,\dots,d'_{k-1}}$  lies in the projection of  $\mathcal{A}$  to  $\mathbb{Z}/q_{W,d_1,\dots,d'_{k-1}}\mathbb{Z}$ . Each  $q \in \mathcal{S}_I$  is equal to  $q_{W,d_1,\dots,d'_{k-1}}$  for  $O(\tau(q)^{O(1)})$  choices of  $d_1, \dots, d'_{k-1}$ . Thus the left-hand side of (46) is

$$\ll \sum_{q \in \mathcal{S}_I: q \ll x^{1/2+2\varpi}} \tau(q)^{O(1)} \sup_{a \in \mathcal{A}} |\Delta(\mathbf{1}_{[x+h_k, 2x+h_k]}\theta; a \pmod{q})|.$$

Note from the Chinese remainder theorem that for any given  $q$ , if one lets  $a$  range uniformly in  $\mathcal{A}$ , then  $a \pmod{q}$  is uniformly distributed among  $O(\tau(q)^{O(1)})$  different moduli. Thus we have

$$\sup_{a \in \mathcal{A}} |\Delta(\mathbf{1}_{[x+h_k, 2x+h_k]}\theta; a \pmod{q})| \ll \frac{\tau(q)^{O(1)}}{|\mathcal{A}|} \sum_{a \in \mathcal{A}} |\Delta(\mathbf{1}_{[x+h_k, 2x+h_k]}\theta; a \pmod{q})|,$$

and so it suffices to show that

$$\sum_{q \in \mathcal{S}_I: q \ll x^{1/2+2\varpi}} \frac{\tau(q)^{O(1)}}{|\mathcal{A}|} \sum_{a \in \mathcal{A}} |\Delta(\mathbf{1}_{[x+h_k, 2x+h_k]}\theta; a \pmod{q})| \ll x \log^{-A} x$$

for any fixed  $A > 0$ . We see it suffices to show that

$$\sum_{q \in \mathcal{S}_I: q \ll x^{1/2+2\varpi}} \tau(q)^{O(1)} |\Delta(\mathbf{1}_{[x+h_k, 2x+h_k]}\theta; a \pmod{q})| \ll x \log^{-A} x$$

for any given  $a \in \mathcal{A}$ . But this follows from the hypothesis  $\text{MPZ}[\varpi, \delta]$  by repeating the arguments of Section 4.2.

### 4.4 Crude estimates on divisor sums

To proceed further, we will need some additional information on the divisor sums  $\lambda_F$  (defined in (16)), namely that these sums are concentrated on “almost primes”; results of this type have also appeared in [44].

**Proposition 4.2** (Almost primality) *Let  $k \geq 1$  be fixed, let  $(h_1, \dots, h_k)$  be a fixed admissible  $k$ -tuple, and let  $b \pmod{W}$  be such that  $b + h_i$  is coprime to  $W$  for each  $i = 1, \dots, k$ . Let  $F_1, \dots, F_k : [0, +\infty) \rightarrow \mathbb{R}$  be fixed smooth compactly supported functions, and let  $m_1, \dots, m_k \geq 0$  and  $a_1, \dots, a_k \geq 1$  be fixed natural numbers. Then*

$$\sum_{x \leq n \leq 2x: n \equiv b \pmod{W}} \prod_{j=1}^k \left( |\lambda_{F_j}(n + h_j)|^{a_j} \tau(n + h_j)^{m_j} \right) \ll B^{-k} \frac{x}{W}. \quad (47)$$

Furthermore, if  $1 \leq j_0 \leq k$  is fixed and  $p_0$  is a prime with  $p_0 \leq x^{\frac{1}{10k}}$ , then we have the variant

$$\sum_{x \leq n \leq 2x: n=b \ (W)} \prod_{j=1}^k \left( |\lambda_{F_j}(n+h_j)|^{a_j} \tau(n+h_j)^{m_j} \right) \mathbf{1}_{p_0 | n+h_{j_0}} \ll \frac{\log_x p_0}{p_0} B^{-k} \frac{x}{W}. \quad (48)$$

As a consequence, we have

$$\sum_{x \leq n \leq 2x: n=b \ (W)} \prod_{j=1}^k \left( |\lambda_{F_j}(n+h_j)|^{a_j} \tau(n+h_j)^{m_j} \right) \mathbf{1}_{p(n+h_{j_0}) \leq x^\varepsilon} \ll \varepsilon B^{-k} \frac{x}{W}, \quad (49)$$

for any  $\varepsilon > 0$ , where  $p(n)$  denotes the least prime factor of  $n$ .

The exponent  $\frac{1}{10k}$  can certainly be improved here, but for our purposes any fixed positive exponent depending only on  $k$  will suffice.

*Proof* The strategy is to estimate the alternating divisor sums  $\lambda_{F_j}(n+h_j)$  by non-negative expressions involving prime factors of  $n+h_j$ , which can then be bounded combinatorially using standard tools.

We first prove (47). As in the proof of Proposition 4.1, we can use Fourier expansion to write

$$F_j(\log_x d) = \int_{\mathbb{R}} \frac{f_j(\xi)}{d^{\frac{1+i\xi}{\log x}}} d\xi$$

for some rapidly decreasing  $f_j : \mathbb{R} \rightarrow \mathbb{C}$  and all natural numbers  $d$ . Thus

$$\lambda_{F_j}(n) = \int_{\mathbb{R}} \left( \sum_{d|n} \frac{\mu(d)}{d^{\frac{1+i\xi}{\log x}}} \right) f_j(\xi) d\xi,$$

which factorizes using Euler products as

$$\lambda_{F_j}(n) = \int_{\mathbb{R}} \prod_{p|n} \left( 1 - \frac{1}{p^{\frac{1+i\xi}{\log x}}} \right) f_j(\xi) d\xi.$$

The function  $s \mapsto p^{\frac{-s}{\log x}}$  has a magnitude of  $O(1)$  and a derivative of  $O(\log_x p)$  when  $\operatorname{Re}(s) > 1$ , and thus

$$1 - \frac{1}{p^{\frac{1+i\xi}{\log x}}} = O\left(\min((1+|\xi|)\log_x p, 1)\right).$$

From the rapid decrease of  $f_j$  and the triangle inequality, we conclude that

$$|\lambda_{F_j}(n)| \ll \int_{\mathbb{R}} \left( \prod_{p|n} O\left(\min((1+|\xi|)\log_x p, 1)\right) \right) \frac{d\xi}{(1+|\xi|)^A}$$

for any fixed  $A > 0$ . Thus, noting that  $\prod_{p|n} O(1) \ll \tau(n)^{O(1)}$ , we have

$$|\lambda_{F_j}(n)|^{a_j} \ll \tau(n)^{O(1)} \int_{\mathbb{R}} \dots \int_{\mathbb{R}} \left( \prod_{p|n} \prod_{l=1}^{a_j} \min((1+|\xi_l|)\log_x p, 1) \right) \frac{d\xi_1 \dots d\xi_{a_j}}{(1+|\xi_1|)^A \dots (1+|\xi_{a_j}|)^A}$$

for any fixed  $a_j, A$ . However, we have

$$\prod_{i=1}^{a_j} \min((1+|\xi_i|)\log_x p, 1) \leq \min((1+|\xi_1| + \dots + |\xi_{a_j}|)\log_x p, 1),$$



and so

$$|\lambda_{F_j}(n)|^{a_j} \ll \tau(n)^{O(1)} \int_{\mathbb{R}} \dots \int_{\mathbb{R}} \frac{(\prod_{p|n} \min((1 + |\xi_1| + \dots + |\xi_{a_j}|) \log_x p, 1)) d\xi_1 \dots d\xi_{a_j}}{(1 + |\xi_1| + \dots + |\xi_{a_j}|)^A}.$$

Making the change of variables  $\sigma := 1 + |\xi_1| + \dots + |\xi_{a_j}|$ , we obtain

$$|\lambda_{F_j}(n)|^{a_j} \ll \tau(n)^{O(1)} \int_1^\infty \left( \prod_{p|n} \min(\sigma \log_x p, 1) \right) \frac{d\sigma}{\sigma^A}$$

for any fixed  $A > 0$ . In view of this bound and the Fubini-Tonelli theorem, it suffices to show that

$$\sum_{x \leq n \leq 2x: n=b \ (W)} \prod_{j=1}^k \left( \tau(n + h_j)^{O(1)} \prod_{p|n+h_j} \min(\sigma_j \log_x p, 1) \right) \ll B^{-k} \frac{x}{W} (\sigma_1 + \dots + \sigma_k)^{O(1)}$$

for all  $\sigma_1, \dots, \sigma_k \geq 1$ . By setting  $\sigma := \sigma_1 + \dots + \sigma_k$ , it suffices to show that

$$\sum_{x \leq n \leq 2x: n=b \ (W)} \prod_{j=1}^k \left( \tau(n + h_j)^{O(1)} \prod_{p|n+h_j} \min(\sigma \log_x p, 1) \right) \ll B^{-k} \frac{x}{W} \sigma^{O(1)} \quad (50)$$

for any  $\sigma \geq 1$ .

To proceed further, we factorize  $n + h_j$  as a product

$$n + h_j = p_1 \dots p_r$$

of primes  $p_1 \leq \dots \leq p_r$  in increasing order, and then write

$$n + h_j = d_j m_j$$

where  $d_j := p_1 \dots p_{i_j}$  and  $i_j$  is the largest index for which  $p_1 \dots p_{i_j} < x^{\frac{1}{10k}}$ , and  $m_j := p_{i_j+1} \dots p_r$ . By construction, we see that  $0 \leq i_j < r$ ,  $d_j \leq x^{\frac{1}{10k}}$ . Also, we have

$$p_{i_j+1} \geq (p_1 \dots p_{i_j+1})^{\frac{1}{i_j+1}} \geq x^{\frac{1}{10k(i_j+1)}}.$$

Since  $n \leq 2x$ , this implies that

$$r = O(i_j + 1)$$

and so

$$\tau(n + h_j) \leq 2^{O(1+\Omega(d_j))},$$

where we recall that  $\Omega(d_j) = i_j$  denotes the number of prime factors of  $d_j$ , counting multiplicity. We also see that

$$p(m_j) \geq x^{\frac{1}{10k(1+\Omega(d_j))}} \geq x^{\frac{1}{10k(1+\Omega(d_1 \dots d_k))}} =: R,$$

where  $p(n)$  denotes the least prime factor of  $n$ . Finally, we have that

$$\prod_{p|n+h_j} \min(\sigma \log_x p, 1) \leq \prod_{p|d_j} \min(\sigma \log_x p, 1),$$

and we see the  $d_1, \dots, d_k, W$  are coprime. We may thus estimate the left-hand side of (50) by

$$\ll \sum_{*} \left( \prod_{j=1}^k 2^{O(1+\Omega(d_j))} \prod_{p|d_j} \min(\sigma \log_x p, 1) \right) \sum_{**} 1$$

where the outer sum  $\sum_{*}$  is over  $d_1, \dots, d_k \leq x^{\frac{1}{10k}}$  with  $d_1, \dots, d_k, W$  coprime, and the inner sum  $\sum_{**}$  is over  $x \leq n \leq 2x$  with  $n = b(W)$  and  $n + h_j = 0 \pmod{d_j}$  for each  $j$ , with  $p(\frac{n+h_j}{d_j}) \geq R$  for each  $j$ .

We bound the inner sum  $\sum_{**} 1$  using a Selberg sieve upper bound. Let  $G$  be a smooth function supported on  $[0, 1]$  with  $G(0) = 1$ , and let  $d = d_1 \dots d_k$ . We see that

$$\sum_{**} 1 \leq \sum_{\substack{x \leq n \leq 2x \\ n+h_i=0 \pmod{d_i} \\ n \equiv b \pmod{W}}} \prod_{i=1}^k \left( \sum_{\substack{e|n+h_i \\ (e,dW)=1}} \mu(e) G(\log_R e) \right)^2,$$

since the product is  $G(0)^{2k} = 1$  if  $p(\frac{n+h_j}{d_j}) \geq R$ , and non-negative otherwise. The right hand side may be expanded as

$$\sum_{\substack{e_1, \dots, e_k, e'_1, \dots, e'_k \\ (e_i e'_i, dW) = 1 \forall i}} \left( \prod_{i=1}^k \mu(e_i) \mu(e'_i) G(\log_R e_i) G(\log_R e'_i) \right) \sum_{\substack{x \leq n \leq 2x \\ n+h_i=0 \pmod{d_i[e_i, e'_i]} \\ n \equiv b \pmod{W}}} 1.$$

As in Section 4.1, the inner sum vanishes unless the  $e_i e'_i$  are coprime to each other and  $dW$ , in which case it is

$$\frac{x}{dW[e_1, e'_1] \dots [e_k, e'_k]} + O(1).$$

The  $O(1)$  term contributes  $\ll R^k \ll x^{1/10}$ , which is negligible. By Lemma 4.1, if  $\Omega(d) \ll \log^{1/2} x$  then the main term contributes

$$\ll \left( \frac{d}{\varphi(d)} \right)^k \frac{x}{dW} (\log R)^{-k} \ll 2^{\Omega(d)} B^{-k} \frac{x}{dW}.$$

We see that this final bound applies trivially if  $\Omega(d) \gg \log^{1/2} x$ . The bound (50) thus reduces to

$$\sum_{*} \left( \prod_{j=1}^k \frac{2^{O(1+\Omega(d_j))}}{d_j} \prod_{p|d_j} \min(\sigma \log_x p, 1) \right) \ll \sigma^{O(1)}. \quad (51)$$

Ignoring the coprimality conditions on the  $d_j$  for an upper bound, we see this is bounded by

$$\prod_{w < p \leq x^{\frac{1}{10k}}} \left( 1 + \frac{O(\min(\sigma \log_x(p), 1))}{p} \sum_{j \geq 0} \frac{O(1)^j}{p^j} \right)^k \ll \exp \left( O \left( \sum_{p \leq x} \frac{(\min(\sigma \log_x(p), 1))}{p} \right) \right).$$

But from Mertens' theorem we have

$$\sum_{p \leq x} \frac{\min(\sigma \log_x p, 1)}{p} = O \left( \log \frac{1}{\sigma} \right),$$

and the claim (47) follows.

The proof of (48) is a minor modification of the argument above used to prove (47). Namely, the variable  $d_{j_0}$  is now replaced by  $[d_0, p_0] < x^{1/5k}$ , which upon factoring out  $p_0$  has the effect of multiplying the upper bound for (51) by  $O(\frac{\sigma \log_x p_0}{p_0})$  (at the negligible cost of deleting the prime  $p_0$  from the sum  $\sum_{p \leq x}$ ), giving the claim; we omit the details.

Finally, (49) follows immediately from (47) when  $\varepsilon > \frac{1}{10k}$ , and from (48) and Mertens' theorem when  $\varepsilon \leq \frac{1}{10k}$ .  $\square$

**Remark 4.3** As in [44], one can use Proposition 4.2, together with the observation that the quantity  $\lambda_F(n)$  is bounded whenever  $n = O(x)$  and  $p(n) \geq x^\varepsilon$ , to conclude that whenever the hypotheses of Lemma 3.4 are obeyed for some  $\nu$  of the form (18), then there exists a fixed  $\varepsilon > 0$  such that for all sufficiently large  $x$ , there are  $\gg \frac{x}{\log^k x}$  elements  $n$  of  $[x, 2x]$  such that  $n + h_1, \dots, n + h_k$  have no prime factor less than  $x^\varepsilon$ , and that at least  $m$  of the  $n + h_1, \dots, n + h_k$  are prime.

#### 4.5 The generalized Elliott-Halberstam case

Now we show case (ii) of Theorem 3.6. For sake of notation we shall take  $i_0 = k$ , as the other cases are similar; thus we have

$$\sum_{i=1}^{k-1} (S(F_i) + S(G_i)) < \vartheta. \quad (52)$$

The basic idea is to view the sum (29) as a variant of (26), with the role of the function  $\theta$  now being played by the product divisor sum  $\lambda_{F_k} \lambda_{G_k}$ , and to repeat the arguments in Section 4.2. To do this we rely on Proposition 4.2 to restrict  $n + h_i$  to the almost primes.

We turn to the details. Let  $\varepsilon > 0$  be an arbitrary fixed quantity. From (49) and Cauchy-Schwarz one has

$$\sum_{\substack{x \leq n \leq 2x \\ n=b(W)}} \left( \prod_{i=1}^k \lambda_{F_i}(n + h_i) \lambda_{G_i}(n + h_i) \right) \mathbf{1}_{p(n+h_k) \leq x^\varepsilon} = O\left(\varepsilon B^{-k} \frac{x}{W}\right)$$

with the implied constant uniform in  $\varepsilon$ , so by the triangle inequality and a limiting argument as  $\varepsilon \rightarrow 0$  it suffices to show that

$$\sum_{\substack{x \leq n \leq 2x \\ n=b(W)}} \left( \prod_{i=1}^k \lambda_{F_i}(n + h_i) \lambda_{G_i}(n + h_i) \right) \mathbf{1}_{p(n+h_k) > x^\varepsilon} = (c_\varepsilon + o(1)) B^{-k} \frac{x}{W} \quad (53)$$

where  $c_\varepsilon$  is a quantity depending on  $\varepsilon$  but not on  $x$ , such that

$$\lim_{\varepsilon \rightarrow 0} c_\varepsilon = \prod_{i=1}^k \int_0^1 F'_i(t) G'_i(t) dt.$$

We use (16) to expand out  $\lambda_{F_i}, \lambda_{G_i}$  for  $i = 1, \dots, k-1$ , but *not* for  $i = k$ , so that the left-hand side of (29) becomes

$$\sum_{d_1, \dots, d_{k-1}, d'_1, \dots, d'_{k-1}} \left( \prod_{i=1}^k \mu(d_i) \mu(d'_i) F_i(\log_x d_i) G_i(\log_x d'_i) \right) S'(d_1, \dots, d_{k-1}, d'_1, \dots, d'_{k-1}) \quad (54)$$

where

$$S'(d_1, \dots, d_{k-1}, d'_1, \dots, d'_{k-1}) := \sum_{\substack{x \leq n \leq 2x \\ n=b(W) \\ n+h_i=0 \pmod{[d_i, d'_i]} \quad \forall i=1, \dots, k-1}} \lambda_{F_k}(n + h_k) \lambda_{G_k}(n + h_k) \mathbf{1}_{p(n+h_k) > x^\varepsilon}.$$

As before, the summand in (54) vanishes unless the modulus<sup>[4]</sup>  $q_{W,d_1,\dots,d'_{k-1}}$  defined in (44) is squarefree, in which case we have the analogue

$$\begin{aligned} S'(d_1, \dots, d_{k-1}, d'_1, \dots, d'_{k-1}) &= \frac{1}{\varphi(q)} \sum_{\substack{x+h_k \leq n \leq 2x+h_k \\ (n,q)=1}} \lambda_{F_k}(n) \lambda_{G_k}(n) \mathbf{1}_{p(n) > x^\varepsilon} \\ &\quad + \Delta(\mathbf{1}_{[x+h_k, 2x+h_k]} \lambda_{F_k} \lambda_{G_k} \mathbf{1}_{p(\cdot) > x^\varepsilon}; a(q)) \end{aligned} \quad (55)$$

of (45). Here we have put  $q = q_{W,d_1,\dots,d'_{k-1}}$  and  $a = a_{W,d_1,\dots,d'_{k-1}}$  for convenience. We thus split

$$S' = S'_1 - S'_2 + S'_3,$$

where,

$$S'_1(d_1, \dots, d_{k-1}, d'_1, \dots, d'_{k-1}) = \frac{1}{\varphi(q)} \sum_{x+h_k \leq n \leq 2x+h_k} \lambda_{F_k}(n) \lambda_{G_k}(n) \mathbf{1}_{p(n) > x^\varepsilon}, \quad (56)$$

$$S'_2(d_1, \dots, d_{k-1}, d'_1, \dots, d'_{k-1}) = \frac{1}{\varphi(q)} \sum_{x+h_k \leq n \leq 2x+h_k; (n,q) > 1} \lambda_{F_k}(n) \lambda_{G_k}(n) \mathbf{1}_{p(n) > x^\varepsilon}, \quad (57)$$

$$S'_3(d_1, \dots, d_{k-1}, d'_1, \dots, d'_{k-1}) = \Delta(\mathbf{1}_{[x+h_k, 2x+h_k]} \lambda_{F_k} \lambda_{G_k} \mathbf{1}_{p(\cdot) > x^\varepsilon}; a(q)), \quad (58)$$

when  $q = q_{W,d_1,\dots,d'_{k-1}}$  is squarefree, with  $S'_1 = S'_2 = S'_3 = 0$  otherwise.

For  $j \in \{1, 2, 3\}$ , let

$$\Sigma_j = \sum_{d_1, \dots, d_{k-1}, d'_1, \dots, d'_{k-1}} \left( \prod_{i=1}^k \mu(d_i) \mu(d'_i) F_i(\log_x d_i) G_i(\log_x d'_i) \right) S'_j(d_1, \dots, d_{k-1}, d'_1, \dots, d'_{k-1}). \quad (59)$$

To show (53), it thus suffices to show the main term estimate

$$\Sigma_1 = (c_\varepsilon + o(1)) B^{-k} \frac{x}{W}, \quad (60)$$

the first error term estimate

$$\Sigma_2 \ll x^{1-\varepsilon}, \quad (61)$$

and the second error term estimate

$$\Sigma_3 \ll x \log^{-A} x \quad (62)$$

for any fixed  $A > 0$ .

We begin with (61). Observe that if  $p(n) > x^\varepsilon$ , then the only way that  $(n, q_{W,d_1,\dots,d'_{k-1}})$  can exceed 1 is if there is a prime  $x^\varepsilon < p \ll x$  which divides both  $n$  and one of  $d_1, \dots, d'_{k-1}$ ; in particular, this case can only occur when  $k > 1$ . For sake of notation we will just consider the contribution when there is a prime that divides  $n$  and  $d_1$ , as the other  $2k-3$  cases are similar. By (57), this contribution to  $\Sigma_2$  can

---

<sup>[4]</sup>In the  $k=1$  case, we of course just have  $q_{W,d_1,\dots,d'_{k-1}} = W$ .

then be crudely bounded (using (1)) by

$$\begin{aligned}
\Sigma_2 &\ll \sum_{x^\varepsilon < p \ll x} \sum_{d_1, \dots, d_{k-1}, d'_1, \dots, d'_{k-1} \leq x; p|d_1} \frac{1}{[d_1, d'_1] \dots [d_{k-1}, d'_{k-1}]} \sum_{n \ll x; p|n} 1 \\
&\ll \sum_{x^\varepsilon < p \ll x} \frac{x}{p} \left( \sum_{e_1 \leq x^2; p|e_1} \frac{\tau(e_1)}{e_1} \right) \prod_{i=2}^{k-1} \left( \sum_{e_i \leq x^2} \frac{\tau(e_i)}{e_i} \right) \\
&\ll \sum_{x^\varepsilon < p \ll x} \frac{x}{p^2} \\
&\ll x^{1-\varepsilon}
\end{aligned}$$

as required, where we have made the change of variables  $e_i := [d_i, d'_i]$ , using the divisor bound to control the multiplicity.

Now we show (62). From the hypothesis (28) we have  $q_{W, d_1, \dots, d'_{k-1}} \ll x^\vartheta$  whenever the summand in (62) is non-zero. From the divisor bound, for each  $q \ll x^\vartheta$  there are  $O(\tau(q)^{O(1)})$  choices of  $d_1, \dots, d'_{k-1}$  with  $q_{W, d_1, \dots, d'_{k-1}} = q$ . We see the product in (59) is  $O(1)$ . Thus by (58), we may bound  $\Sigma_3$  by

$$\Sigma_3 \ll \sum_{q \ll x^\vartheta} \tau(q)^{O(1)} \sup_{a \in (\mathbb{Z}/q\mathbb{Z})^\times} |\Delta(\mathbf{1}_{[x+h_k, 2x+h_k]} \lambda_{F_k} \lambda_{G_k} \mathbf{1}_{p(\cdot) > x^\varepsilon}; a(q))|.$$

From (2) we easily obtain the bound

$$\Sigma_3 \ll \sum_{q \ll x^\vartheta} \tau(q)^{O(1)} \sup_{a \in (\mathbb{Z}/q\mathbb{Z})^\times} |\Delta(\mathbf{1}_{[x+h_k, 2x+h_k]} \lambda_{F_k} \lambda_{G_k} \mathbf{1}_{p(\cdot) > x^\varepsilon}; a(q))| \ll x \log^{O(1)} x,$$

so by Cauchy-Schwarz it suffices to show that

$$\sum_{q \ll x^\vartheta} \sup_{a \in (\mathbb{Z}/q\mathbb{Z})^\times} |\Delta(\mathbf{1}_{[x+h_k, 2x+h_k]} \lambda_{F_k} \lambda_{G_k} \mathbf{1}_{p(\cdot) > x^\varepsilon}; a(q))| \ll x \log^{-A} x \quad (63)$$

for any fixed  $A > 0$ .

If we had the additional hypothesis  $S(F_k) + S(G_k) < 1$ , then this would follow easily from the hypothesis  $\text{GEH}[\vartheta]$  thanks to Claim 2.6, since one can write  $\lambda_{F_k} \lambda_{G_k} \mathbf{1}_{p(\cdot) > x^\varepsilon} = \alpha \star \beta$  with

$$\alpha(n) := \mathbf{1}_{p(n) > x^\varepsilon} \sum_{d, d': [d, d'] = n} \mu(d) F_k(\log_x d) \mu(d') G_k(\log_x d')$$

and

$$\beta(n) := \mathbf{1}_{p(n) > x^\varepsilon}.$$

But even in the absence of the hypothesis  $S(F_k) + S(G_k) < 1$ , we can still invoke  $\text{GEH}[\vartheta]$  after appealing to the fundamental theorem of arithmetic. Indeed, if  $n \in [x + h_k, 2x + h_k]$  with  $p(\cdot) > \varepsilon$ , then we have

$$n = p_1 \dots p_r$$

for some primes  $x^\varepsilon < p_1 \leq \dots \leq p_r \leq 2x + h_k$ , which forces  $r \leq \frac{1}{\varepsilon} + 1$ . If we then partition  $[x^\varepsilon, 2x + h_k]$  by  $O(\log^{A+1} x)$  intervals  $I_1, \dots, I_m$ , with each  $I_j$  contained in an interval of the form  $[N, (1 + \log^{-A} x)N]$ , then we have  $p_i \in I_{j_i}$  for some  $1 \leq j_1 \leq \dots \leq j_r \leq m$ , with the product interval  $I_{j_1} \dots I_{j_r}$  intersecting  $[x + h_k, 2x + h_k]$ . For fixed  $r$ , there are  $O(\log^{Ar+r} x)$  such tuples  $(j_1, \dots, j_r)$ , and a simple application of the prime number theorem with classical error term (and crude estimates on the discrepancy  $\Delta$ )

shows that each tuple contributes  $O(x \log^{-Ar+O(1)} x)$  to (63) (here, and for the rest of this section, implied constants will be independent of  $A$  unless stated otherwise). In particular, the  $O(\log^{A(r-1)} x)$  tuples  $(j_1, \dots, j_r)$  with one repeated  $j_i$ , or for which the interval  $I_{j_1} \cdots I_{j_r}$  meets the boundary of  $[x + h_k, 2x + h_k]$ , contribute a total of  $O(\log^{-A+O(1)} x)$ . This is an acceptable error to (63), and so these tuples may be removed. Thus it suffices to show that

$$\sum_{q \ll x^\vartheta} \sup_{a \in (\mathbb{Z}/q\mathbb{Z})^\times} |\Delta(\lambda_{F_k} \lambda_{G_k} \mathbf{1}_{A_{j_1, \dots, j_r}}; a(q))| \ll x \log^{-A(r+1)+O(1)} x$$

for any  $1 \leq r \leq \frac{1}{\varepsilon} + 1$  and  $1 \leq j_1 < \dots < j_r \leq m$  with  $I_{j_1} \cdots I_{j_r}$  contained in  $[x + h_k, x + 2h_k]$ , where  $A_{j_1, \dots, j_r}$  is the set of all products  $p_1 \dots p_r$  with  $p_i \in I_{j_i}$  for  $i = 1, \dots, r$ , and where we allow implied constants in the  $\ll$  notation to depend on  $\varepsilon$ . But for  $n$  in  $A_{j_1, \dots, j_r}$ , the  $2^r$  factors of  $n$  are just the products of subsets of  $\{p_1, \dots, p_r\}$ , and from the smoothness of  $F_k, G_k$  we see that  $\lambda_{F_k}(n)$  is equal to some bounded constant (depending on  $j_1, \dots, j_r$ , but independent of  $p_1, \dots, p_r$ ), plus an error of  $O(\log^{-A} x)$ . As before, the contribution of this error is  $O(\log^{-A(r+1)+O(1)} x)$ , so it suffices to show that

$$\sum_{q \ll x^\vartheta} \sup_{a \in (\mathbb{Z}/q\mathbb{Z})^\times} |\Delta(\mathbf{1}_{A_{j_1, \dots, j_r}}; a(q))| \ll x \log^{-A(r+1)+O(1)} x.$$

But one can write  $\mathbf{1}_{A_{j_1, \dots, j_r}}$  as a convolution  $\mathbf{1}_{A_{j_1}} \star \dots \star \mathbf{1}_{A_{j_r}}$ , where  $A_{j_i}$  denotes the primes in  $I_{j_i}$ ; assigning  $A_{j_r}$  (for instance) to be  $\beta$  and the remaining portion of the convolution to be  $\alpha$ , the claim now follows from the hypothesis  $\text{GEH}[\vartheta]$ , thanks to the Siegel-Walfisz theorem (see e.g. [58, Satz 4] or [34, Th. 5.29]).

Finally, we show (60). By Lemma 4.1 we have

$$\sum_{\substack{d_1, \dots, d_{k-1}, d'_1, \dots, d'_{k-1} \\ d_1 d'_1, \dots, d_{k-1} d'_{k-1}, W \text{ coprime}}} \frac{\prod_{i=1}^{k-1} \mu(d_i) \mu(d'_i) F_i(\log_x d_i) G_i(\log_x d'_i)}{\varphi(qW, d_1, \dots, d'_{k-1})} = \frac{1}{\varphi(W)} (c' + o(1)) B^{-k+1},$$

where

$$c' := \prod_{i=1}^{k-1} \int_0^1 F'_i(t) G'_i(t) dt$$

(note that  $F_i, G_i$  are supported on  $[0, 1]$  by hypothesis), so by (56) it suffices to show that

$$\sum_{x+h_k \leq n \leq 2x+h_k} \lambda_{F_k}(n) \lambda_{G_k}(n) \mathbf{1}_{p(n) > x^\varepsilon} = (c''_\varepsilon + o(1)) \frac{x}{\log x}, \quad (64)$$

where  $c''_\varepsilon$  is a quantity depending on  $\varepsilon$  but not on  $x$  such that

$$\lim_{\varepsilon \rightarrow 0} c''_\varepsilon = \int_0^1 F'_k(t) G'_k(t) dt.$$

In the case  $S(F_k) + S(G_k) < 1$ , this would follow easily from (the  $k = 1$  case of) Theorem 3.6(i) and Proposition 4.2. In the general case, we may appeal once more to the fundamental theorem of arithmetic. As before, we may factor  $n = p_1 \dots p_r$  for some  $x^\varepsilon \leq p_1 \leq \dots \leq p_r \leq 2x + h_k$  and  $r \leq \frac{1}{\varepsilon} + 1$ . The contribution of those  $n$  with a repeated prime factor  $p_i = p_{i+1}$  can easily be shown to be  $\ll x^{1-\varepsilon}$  in the same manner we dealt with  $\Sigma_2$ , so we may restrict attention to the square-free  $n$ , for which the  $p_i$  are strictly increasing. In that case, one can write

$$\lambda_{F_k}(n) = (-1)^r \partial_{(\log_x p_1)} \dots \partial_{(\log_x p_r)} F_k(0)$$

and

$$\lambda_{G_k}(n) = (-1)^r \partial_{(\log_x p_1)} \dots \partial_{(\log_x p_r)} G_k(0)$$

where  $\partial_{(h)} F(x) := F(x+h) - F(x)$ . On the other hand, a standard application of Mertens' theorem and the prime number theorem (and an induction on  $r$ ) shows that for any fixed  $r \geq 1$  and any fixed continuous function  $f : \mathbb{R}^r \rightarrow \mathbb{R}$ , we have

$$\sum_{x^\varepsilon \leq p_1 < \dots < p_r : x+h_k \leq p_1 \dots p_r \leq 2x+h_k} f(\log_x p_1, \dots, \log_x p_r) = (c_f + o(1)) \frac{x}{\log x}$$

where  $c_f$  is the quantity

$$c_f := \int_{\varepsilon \leq t_1 < \dots < t_r : t_1 + \dots + t_r = 1} f(t_1, \dots, t_r) \frac{dt_1 \dots dt_{r-1}}{t_1 \dots t_r}$$

where we lift Lebesgue measure  $dt_1 \dots dt_{r-1}$  up to the hyperplane  $t_1 + \dots + t_r = 1$ , thus

$$\int_{t_1 + \dots + t_r = 1} F(t_1, \dots, t_r) dt_1 \dots dt_{r-1} := \int_{\mathbb{R}^{r-1}} F(t_1, \dots, t_{r-1}, 1 - t_1 - \dots - t_{r-1}) dt_1 \dots dt_{r-1}.$$

Putting all this together, we see that we obtain an asymptotic (64) with

$$c_\varepsilon'' := \sum_{1 \leq r \leq \frac{1}{\varepsilon} + 1} \int_{\varepsilon \leq t_1 < \dots < t_r : t_1 + \dots + t_r = 1} \partial_{(t_1)} \dots \partial_{(t_r)} F_k(0) \partial_{(t_1)} \dots \partial_{(t_r)} G_k(0) \frac{dt_1 \dots dt_{r-1}}{t_1 \dots t_r}.$$

Comparing (64) with the first part of Proposition 4.2 we see that  $c_\varepsilon'' = O(1)$  uniformly in  $\varepsilon$ ; subtracting two instances of (64) and comparing with the last part of Proposition 4.2 we see that  $|c_{\varepsilon_1}'' - c_{\varepsilon_2}''| \ll \varepsilon_1 + \varepsilon_2$  for any  $\varepsilon_1, \varepsilon_2 > 0$ . We conclude that  $c_\varepsilon''$  converges to a limit as  $\varepsilon \rightarrow 0$  for any  $F, G$ . This implies the absolute convergence

$$\sum_{r>0} \int_{0 < t_1 < \dots < t_r : t_1 + \dots + t_r = 1} |\partial_{(t_1)} \dots \partial_{(t_r)} F_k(0)| |\partial_{(t_1)} \dots \partial_{(t_r)} G_k(0)| \frac{dt_1 \dots dt_{r-1}}{t_1 \dots t_r} < \infty; \quad (65)$$

indeed, by the Cauchy-Schwarz inequality it suffices to establish this for  $F = G$ , at which point we may remove the absolute value signs and use the boundedness of  $c_\varepsilon''$ . By the dominated convergence theorem, it therefore suffices to establish the identity

$$\sum_{r>0} \int_{0 < t_1 < \dots < t_r : t_1 + \dots + t_r = 1} \partial_{(t_1)} \dots \partial_{(t_r)} F_k(0) \partial_{(t_1)} \dots \partial_{(t_r)} G_k(0) \frac{dt_1 \dots dt_{r-1}}{t_1 \dots t_r} = \int_0^1 F'_k(t) G'_k(t) dt. \quad (66)$$

It will suffice to show the identity

$$\sum_{r>0} \int_{0 < t_1 < \dots < t_r : t_1 + \dots + t_r = 1} |\partial_{(t_1)} \dots \partial_{(t_r)} F(0)|^2 \frac{dt_1 \dots dt_{r-1}}{t_1 \dots t_r} = \int_0^1 |F'(t)|^2 dt \quad (67)$$

for any smooth  $F : [0, +\infty) \rightarrow \mathbb{R}$ , since (66) follows by replacing  $F$  with  $F_k + G_k$  and  $F_k - G_k$  and then subtracting.

At this point we use the following identity:

**Lemma 4.4** *For any positive reals  $t_1, \dots, t_r$  with  $r \geq 1$ , we have*

$$\frac{1}{t_1 \dots t_r} = \sum_{\sigma \in S_r} \frac{1}{\prod_{i=1}^r (\sum_{j=i}^r t_{\sigma(j)})}. \quad (68)$$

Thus, for instance, when  $r = 2$  we have

$$\frac{1}{t_1 t_2} = \frac{1}{(t_1 + t_2)t_1} + \frac{1}{(t_1 + t_2)t_2}.$$

*Proof* If the right-hand side of (68) is denoted  $f_r(t_1, \dots, t_r)$ , then one easily verifies the identity

$$f_r(t_1, \dots, t_r) = \frac{1}{t_1 + \dots + t_r} \sum_{i=1}^r f_{r-1}(t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_r)$$

for any  $r > 1$ ; but the left-hand side of (68) also obeys this identity, and the claim then follows from induction.  $\square$

From this lemma and symmetrisation, we may rewrite the left-hand side of (67) as

$$\sum_{r>0} \int_{\substack{t_1, \dots, t_r \geq 0 \\ t_1 + \dots + t_r = 1}} |\partial_{(t_1)} \dots \partial_{(t_r)} F(0)|^2 \frac{dt_1 \dots dt_{r-1}}{\prod_{i=1}^r (\sum_{j=i}^r t_j)}.$$

Let

$$I_a(F) := \int_0^a F'(t)^2 dt,$$

and

$$J_a(F) := (\partial_{(a)} F(0))^2.$$

One can then rewrite (67) as the identity

$$I_1(F) = \sum_{r=1}^{\infty} K_{1,r}(F), \tag{69}$$

where

$$K_{a,r}(F) := \int_{\substack{t_1, \dots, t_r \geq 0 \\ t_1 + \dots + t_r = a}} J_{t_r}(\partial_{(t_1)} \dots \partial_{(t_{r-1})} F) \frac{dt_1 \dots dt_{r-1}}{a(a-t_1) \dots (a-t_1 - \dots - t_{r-1})}.$$

To prove this, we first observe the identity

$$I_a(F) = \frac{1}{a} J_a(F) + \int_{0 \leq t \leq a} I_{a-t}(\partial_{(t)} F) \frac{dt}{a}$$

for any  $a > 0$ ; indeed, we have

$$\begin{aligned} \int_{0 \leq t \leq a} I_{a-t}(\partial_{(t)} F) \frac{dt}{a} &= \int_{0 \leq t \leq a; 0 \leq u \leq a-t} |F'(t+u) - F'(t)|^2 \frac{dudt}{a} \\ &= \int_{0 \leq t \leq s \leq a} |F'(s) - F'(t)|^2 \frac{dsdt}{a} \\ &= \frac{1}{2} \int_0^a \int_0^a |F'(s) - F'(t)|^2 \frac{dsdt}{a} \\ &= \int_0^a |F'(s)|^2 ds - \frac{1}{a} \left( \int_0^a F'(s) ds \right) \left( \int_0^a F'(t) dt \right) \\ &= I_a(F) - \frac{1}{a} J_a(F), \end{aligned}$$



and the claim follows. Iterating this identity  $k$  times, we see that

$$I_a(F) = \sum_{r=1}^k K_{a,r}(F) + L_{a,k}(F) \quad (70)$$

for any  $k \geq 1$ , where

$$L_{a,k}(F) := \int_{\substack{t_1, \dots, t_k \geq 0 \\ t_1 + \dots + t_k \leq a}} I_{1-t_1-\dots-t_k}(\partial_{(t_1)} \dots \partial_{(t_k)} F) \frac{dt_1 \dots dt_k}{a(a-t_1) \dots (a-t_1-\dots-t_{k-1})}.$$

In particular, dropping the  $L_{a,k}(F)$  term and sending  $k \rightarrow \infty$  yields the lower bound

$$\sum_{r=1}^{\infty} K_{a,r}(F) \leq I_a(F). \quad (71)$$

On the other hand, we can expand  $L_{a,k}(F)$  as

$$\int_{\substack{t_1, \dots, t_k, t \geq 0 \\ t_1 + \dots + t_k + t \leq a}} |\partial_{(t_1)} \dots \partial_{(t_k)} F'(t)|^2 \frac{dt_1 \dots dt_k dt}{a(a-t_1) \dots (a-t_1-\dots-t_{k-1})}.$$

Writing  $s := t_1 + \dots + t_k$ , we obtain the upper bound

$$L_{a,k}(F) \leq \int_{s, t \geq 0: s+t \leq a} K_{s,k}(F'_t) dt,$$

where  $F'_t(x) := F(x+t)$ . Summing this and using (71) and the monotone convergence theorem, we conclude that

$$\sum_{k=1}^{\infty} L_{a,k}(F) \leq \int_{s, t \geq 0: s+t \leq a} I_s(F_t) dt < \infty,$$

and in particular  $L_{a,k}(F) \rightarrow 0$  as  $k \rightarrow \infty$ . Sending  $k \rightarrow \infty$  in (70), we obtain (69) as desired.

## 5 Reduction to a variational problem

Now that we have proven Theorems 3.5 and 3.6, we can establish Theorems 3.8, 3.10, 3.12, 3.14. The main technical difficulty is to take the multidimensional measurable functions  $F$  appearing in these functions and approximate them by tensor products of smooth functions, for which Theorems 3.5 and 3.6 may be applied.

### 5.1 Proof of Theorem 3.8

We now prove Theorem 3.8. Let  $k, m, \vartheta$  obey the hypotheses of that theorem, thus we may find a fixed square-integrable function  $F : [0, +\infty)^k \rightarrow \mathbb{R}$  supported on the simplex

$$\mathcal{R}_k := \{(t_1, \dots, t_k) \in [0, +\infty)^k : t_1 + \dots + t_k \leq 1\}$$

and not identically zero and with

$$\frac{\sum_{i=1}^k J_i(F)}{I(F)} > \frac{2m}{\vartheta}. \quad (72)$$

We now perform a number of technical steps to further improve the structure of  $F$ . Our arguments here will be somewhat convoluted, and are not the most efficient way to prove Theorem 3.8 (which in

any event was already established in [38]), but they will motivate the similar arguments given below to prove the more difficult results in Theorems 3.10, 3.12, 3.14. In particular, we will use regularisation techniques which are compatible with the vanishing marginal condition (35) that is a key hypothesis in Theorem 3.14.

We first need to rescale and retreat a little bit from the slanted boundary of the simplex  $\mathcal{R}_k$ . Let  $\delta_1 > 0$  be a sufficiently small fixed quantity, and write  $F_1 : [0, +\infty)^k \rightarrow \mathbb{R}$  to be the rescaled function

$$F_1(t_1, \dots, t_k) := F\left(\frac{t_1}{\vartheta/2 - \delta_1}, \dots, \frac{t_k}{\vartheta/2 - \delta_1}\right).$$

Thus  $F_1$  is a fixed square-integrable measurable function supported on the rescaled simplex

$$(\vartheta/2 - \delta_1) \cdot \mathcal{R}_k = \{(t_1, \dots, t_k) \in [0, +\infty)^k : t_1 + \dots + t_k \leq \vartheta/2 - \delta_1\}.$$

From (72), we see that if  $\delta_1$  is small enough, then  $F_1$  is not identically zero and

$$\frac{\sum_{i=1}^k J_i(F_1)}{I(F_1)} > m. \quad (73)$$

Let  $\delta_1$  and  $F_1$  be as above. Next, let  $\delta_2 > 0$  be a sufficiently small fixed quantity (smaller than  $\delta_1$ ), and write  $F_2 : [0, +\infty)^k \rightarrow \mathbb{R}$  to be the shifted function, defined by setting

$$F_2(t_1, \dots, t_k) := F_1(t_1 - \delta_2, \dots, t_k - \delta_2)$$

when  $t_1, \dots, t_k \geq \delta_2$ , and  $F_2(t_1, \dots, t_k) = 0$  otherwise. As  $F_1$  was square-integrable, compactly supported, and not identically zero, and because spatial translation is continuous in the strong operator topology on  $L^2$ , it is easy to see that we will have  $F_2$  not identically zero and that

$$\frac{\sum_{i=1}^k J_i(F_2)}{I(F_2)} > m \quad (74)$$

for  $\delta_2$  small enough (after restricting  $F_2$  back to  $[0, +\infty)^k$ , of course). For  $\delta_2$  small enough, this function will be supported on the region

$$\{(t_1, \dots, t_k) \in \mathbb{R}^k : t_1 \dots + t_k \leq \vartheta/2 - \delta_2; t_1, \dots, t_k \geq \delta_2\},$$

thus the support of  $F_2$  stays away from all the boundary faces of  $\mathcal{R}_k$ .

By convolving  $F_2$  with a smooth approximation to the identity that is supported sufficiently close to the origin, one may then find a *smooth* function  $F_3 : [0, +\infty)^k \rightarrow \mathbb{R}$ , supported on

$$\{(t_1, \dots, t_k) \in \mathbb{R}^k : t_1 \dots + t_k \leq \vartheta/2 - \delta_2/2; t_1, \dots, t_k \geq \delta_2/2\},$$

which is not identically zero, and such that

$$\frac{\sum_{i=1}^k J_i(F_3)}{I(F_3)} > m. \quad (75)$$

We extend  $F_3$  by zero to all of  $\mathbb{R}^k$ , and then define the function  $f_3 : \mathbb{R}^k \rightarrow \mathbb{R}$  by

$$f_3(t_1, \dots, t_k) := \int_{s_1 \geq t_1, \dots, s_k \geq t_k} F_3(s_1, \dots, s_k) ds_1 \dots ds_k,$$

thus  $f_3$  is smooth, not identically zero and supported on the region

$$\{(t_1, \dots, t_k) \in \mathbb{R}^k : \sum_{i=1}^k \max(t_i, \delta_2/2) \leq \vartheta/2 - \delta_2/2\}. \quad (76)$$

From the fundamental theorem of calculus we have

$$F_3(t_1, \dots, t_k) := (-1)^k \frac{\partial^k}{\partial t_1 \dots \partial t_k} f_3(t_1, \dots, t_k), \quad (77)$$

and so  $I(F_3) = \tilde{I}(f_3)$  and  $J_i(F_3) = \tilde{J}_i(f_3)$  for  $i = 1, \dots, k$ , where

$$\tilde{I}(f_3) := \int_{[0, +\infty)^k} \left| \frac{\partial^k}{\partial t_1 \dots \partial t_k} f_3(t_1, \dots, t_k) \right|^2 dt_1 \dots dt_k \quad (78)$$

and

$$\tilde{J}_i(f_3) := \int_{[0, +\infty)^{k-1}} \left| \frac{\partial^{k-1}}{\partial t_1 \dots \partial t_{i-1} \partial t_{i+1} \dots \partial t_k} f_3(t_1, \dots, t_{i-1}, 0, t_{i+1}, \dots, t_k) \right|^2 dt_1 \dots dt_{i-1} dt_{i+1} \dots dt_k. \quad (79)$$

In particular,

$$\frac{\sum_{i=1}^k \tilde{J}_i(f_3)}{\tilde{I}(f_3)} > m. \quad (80)$$

Now we approximate  $f_3$  by linear combinations of tensor products. By the Stone-Weierstrass theorem, we may express  $f_3$  (on  $[0, +\infty)^k$ ) as the uniform limit of functions of the form

$$(t_1, \dots, t_k) \mapsto \sum_{j=1}^J c_j f_{1,j}(t_1) \dots f_{k,j}(t_k) \quad (81)$$

where  $c_1, \dots, c_J$  are real scalars, and  $f_{i,j} : \mathbb{R} \rightarrow \mathbb{R}$  are smooth compactly supported functions. Since  $f_3$  is supported in (76), we can ensure that all the components  $f_{1,j}(t_1) \dots f_{k,j}(t_k)$  are supported in the slightly larger region

$$\{(t_1, \dots, t_k) \in \mathbb{R}^k : \sum_{i=1}^k \max(t_i, \delta_2/4) \leq \vartheta/2 - \delta_2/4\}.$$

Observe that if one convolves a function of the form (81) with a smooth approximation to the identity which is of tensor product form  $(t_1, \dots, t_k) \mapsto \varphi_1(t_1) \dots \varphi_1(t_k)$ , one obtains another function of this form. Such a convolution converts a uniformly convergent sequence of functions to a *uniformly smoothly* convergent sequence of functions (that is to say, all derivatives of the functions converge uniformly). From this, we conclude that  $f_3$  can be expressed (on  $[0, +\infty)^k$ ) as the *smooth* limit of functions of the form (81), with each component  $f_{1,j}(t_1) \dots f_{k,j}(t_k)$  supported in the region

$$\{(t_1, \dots, t_k) \in \mathbb{R}^k : \sum_{i=1}^k \max(t_i, \delta_2/8) \leq \vartheta/2 - \delta_2/8\}.$$

Thus, we may find such a linear combination

$$f_4(t_1, \dots, t_k) = \sum_{j=1}^J c_j f_{1,j}(t_1) \dots f_{k,j}(t_k) \quad (82)$$

with  $J$ ,  $c_j$ ,  $f_{i,j}$  fixed and  $f_4$  not identically zero, with

$$\frac{\sum_{i=1}^k \tilde{J}_i(f_4)}{\tilde{I}(f_4)} > m. \quad (83)$$

Furthermore, by construction we have

$$S(f_{1,j}) + \dots + S(f_{k,j}) < \frac{\vartheta}{2} \leq \frac{1}{2} \quad (84)$$

for all  $j = 1, \dots, J$ , where  $S()$  was defined in (22).

Now we construct the sieve weight  $\nu : \mathbb{N} \rightarrow \mathbb{R}$  by the formula

$$\nu(n) := \left( \sum_{j=1}^J c_j \lambda_{f_{1,j}}(n + h_1) \dots \lambda_{f_{k,j}}(n + h_k) \right)^2, \quad (85)$$

where the divisor sums  $\lambda_f$  were defined in (16).

Clearly  $\nu$  is non-negative. Expanding out the square and using Theorem 3.6(i) and (84), we see that

$$\sum_{\substack{x \leq n \leq 2x \\ n=b \pmod{W}}} \nu(n) = (\alpha + o(1)) B^{-k} \frac{x}{\log x}$$

where

$$\alpha := \sum_{j=1}^J \sum_{j'=1}^J c_j c_{j'} \prod_{i=1}^k \int_0^\infty f'_{i,j}(t_i) f'_{i,j'}(t_i) dt_i$$

which factorizes using (82), (78) as

$$\begin{aligned} \alpha &= \int_{[0,+\infty)^k} \left| \frac{\partial^{k-1}}{\partial t_1 \dots \partial t_k} f_4(t_1, \dots, t_k) \right|^2 dt_1 \dots dt_k \\ &= \tilde{I}(f_4). \end{aligned}$$

Now consider the sum

$$\sum_{\substack{x \leq n \leq 2x \\ n=b \pmod{W}}} \nu(n) \theta(n + h_k).$$

By (20), one has

$$\lambda_{f_{k,j}}(n + h_k) = f_{k,j}(0)$$

whenever  $n$  gives a non-zero contribution to the above sum. Expanding out the square in (85) again and using Theorem 3.5(i) and (84) (and the hypothesis  $\text{EH}[\vartheta]$ ), we thus see that

$$\sum_{\substack{x \leq n \leq 2x \\ n=b \pmod{W}}} \nu(n) \theta(n + h_k) = (\beta_k + o(1)) B^{1-k} \frac{x}{\varphi(W)}$$

where

$$\beta_k := \sum_{j=1}^J \sum_{j'=1}^J c_j c_{j'} f_{i,j}(0) f_{i,j'}(0) \prod_{i=1}^{k-1} \int_0^\infty f'_{i,j}(t_i) f'_{i,j'}(t_i) dt_i$$

which factorizes using (82), (79) as

$$\begin{aligned} \beta_k &= \int_{[0,+\infty)^k} \left| \frac{\partial^k}{\partial t_1 \dots \partial t_{k-1}} f_4(t_1, \dots, t_{k-1}, 0) \right|^2 dt_1 \dots dt_{k-1} \\ &= \tilde{J}_k(f_4). \end{aligned}$$

More generally, we see that

$$\sum_{\substack{x \leq n \leq 2x \\ n \equiv b \pmod{W}}} \nu(n) \theta(n + h_i) = (\beta_i + o(1)) B^{1-k} \frac{x}{\varphi(W)}$$

for  $i = 1, \dots, k$ , with  $\beta_i := \tilde{J}_i(f_4)$ . Applying Lemma 3.4 and (75), we obtain  $\text{DHL}[k, m+1]$  as required.

## 5.2 Proof of Theorem 3.10

Now we prove Theorem 3.10, which uses a very similar argument to that of the previous section. Let  $k, m, \varpi, \delta, F$  be as in Theorem 3.10. By performing the same rescaling as in the previous section (but with  $1/2 + 2\varpi$  playing the role of  $\vartheta$ ), we see that we can find a fixed square-integrable measurable function  $F_1$  supported on the rescaled truncated simplex

$$\{(t_1, \dots, t_k) \in [0, +\infty)^k : t_1 + \dots + t_k \leq \frac{1}{4} + \varpi - \delta_1; t_1, \dots, t_k < \delta - \delta_1\}$$

for some sufficiently small fixed  $\delta_1 > 0$ , such that (73) holds. By repeating the arguments of the previous section we may eventually arrive at a smooth function  $f_4 : \mathbb{R}^k \rightarrow \mathbb{R}$  of the form (82), which is not identically zero and obeys (83), and such that each component  $f_{1,j}(t_1) \dots f_{k,j}(t_k)$  is supported in the region

$$\{(t_1, \dots, t_k) \in \mathbb{R}^k : \sum_{i=1}^k \max(t_i, \delta_2/8) \leq \frac{1}{4} + \varpi - \delta_2/8; t_1, \dots, t_k < \delta - \delta_2/8\}$$

for some sufficiently small  $\delta_2 > 0$ . In particular, one has

$$S(f_{1,j}) + \dots + S(f_{k,j}) < \frac{1}{4} + \varpi \leq \frac{1}{2}$$

and

$$S(f_{1,j}), \dots, S(f_{k,j}) < \delta$$

for all  $j = 1, \dots, J$ . If we then define  $\nu$  by (85) as before, and repeat all of the above arguments (but use Theorem 3.5(ii) and  $\text{MPZ}[\varpi, \delta]$  in place of Theorem 3.5(i) and  $\text{EH}[\vartheta]$ ), we obtain the claim; we leave the details to the interested reader.

## 5.3 Proof of Theorem 3.12

Now we prove Theorem 3.12. Let  $k, m, \varepsilon, \vartheta$  be as in that theorem. Then one may find a square-integrable function  $F : [0, +\infty)^k \rightarrow \mathbb{R}$  supported on  $(1 + \varepsilon) \cdot \mathcal{R}_k$  which is not identically zero, and with

$$\frac{\sum_{i=1}^k J_{i,1-\varepsilon}(F)}{I(F)} > \frac{2m}{\vartheta}.$$

By truncating and rescaling as in Section 5.1, we may find a fixed bounded measurable function  $F_1 : [0, +\infty)^k \rightarrow \mathbb{R}$  on the simplex  $(1 + \varepsilon)(\frac{\vartheta}{2} - \delta_1) \cdot \mathcal{R}_k$  such that

$$\frac{\sum_{i=1}^k J_{i, (1-\varepsilon)\frac{\vartheta}{2}}(F_1)}{I(F_1)} > m.$$

By repeating the arguments in Section 5.1, we may eventually arrive at a smooth function  $f_4 : \mathbb{R}^k \rightarrow \mathbb{R}$  of the form (82), which is not identically zero and obeys

$$\frac{\sum_{i=1}^k \tilde{J}_{i, (1-\varepsilon)\frac{\vartheta}{2}}(f_4)}{\tilde{I}(f_4)} > m \quad (86)$$

with

$$\tilde{J}_{i, (1-\varepsilon)\frac{\vartheta}{2}}(f_4) := \int_{(1-\varepsilon)\frac{\vartheta}{2} \cdot \mathcal{R}_{k-1}} \left| \frac{\partial^{k-1}}{\partial t_1 \dots \partial t_{i-1} \partial t_{i+1} \dots \partial t_k} f_4(t_1, \dots, t_{i-1}, 0, t_{i+1}, \dots, t_k) \right|^2 dt_1 \dots dt_{i-1} dt_{i+1} \dots dt_k,$$

and such that each component  $f_{1,j}(t_1) \dots f_{k,j}(t_k)$  is supported in the region

$$\left\{ (t_1, \dots, t_k) \in \mathbb{R}^k : \sum_{i=1}^k \max(t_i, \delta_2/8) \leq (1 + \varepsilon)\frac{\vartheta}{2} - \frac{\delta_2}{8} \right\}$$

for some sufficiently small  $\delta_2 > 0$ . In particular, we have

$$S(f_{1,j}) + \dots + S(f_{k,j}) \leq (1 + \varepsilon)\frac{\vartheta}{2} - \frac{\delta_2}{8} \quad (87)$$

for all  $1 \leq j \leq J$ .

Let  $\delta_3 > 0$  be a sufficiently small fixed quantity (smaller than  $\delta_1$  or  $\delta_2$ ). By a smooth partitioning, we may assume that all of the  $f_{i,j}$  are supported in intervals of length at most  $\delta_3$ , while keeping the sum

$$\sum_{j=1}^J |c_j| |f_{1,j}(t_1)| \dots |f_{k,j}(t_k)| \quad (88)$$

bounded uniformly in  $t_1, \dots, t_k$  and in  $\delta_3$ .

Now let  $\nu$  be as in (85), and consider the expression

$$\sum_{\substack{x \leq n \leq 2x \\ n=b \ (W)}} \nu(n).$$

This expression expands as a linear combination of the expressions

$$\sum_{\substack{x \leq n \leq 2x \\ n=b \ (W)}} \prod_{i=1}^k \lambda_{f_{i,j}}(n + h_i) \lambda_{f_{i,j'}}(n + h_i)$$

for various  $1 \leq j, j' \leq J$ . We claim that this sum is equal to

$$\left( \prod_{i=1}^k \int_0^1 f'_{i,j}(t_i) f'_{i,j'}(t_i) dt_i + o(1) \right) B^{-k} \frac{x}{W}.$$

To see this, we divide into two cases. First suppose that hypothesis (i) from Theorem 3.12 holds. Then from (87) we have

$$\sum_{i=1}^k (S(f_{i,j}) + S(f_{i,j'})) < (1 + \varepsilon)\vartheta < 1$$

and the claim follows from Theorem 3.6(i). Now suppose instead that hypothesis (ii) from Theorem 3.12 holds, then from (87) one has

$$\sum_{i=1}^k (S(f_{i,j}) + S(f_{i,j'})) < (1 + \varepsilon)\vartheta < \frac{k}{k-1}\vartheta,$$

and so from the pigeonhole principle we have

$$\sum_{1 \leq i \leq k: i \neq i_0} (S(f_{i,j}) + S(f_{i,j'})) < \vartheta$$

for some  $1 \leq i_0 \leq k$ . The claim now follows from Theorem 3.6(ii).

Putting this together as in Section 5.1, we conclude that

$$\sum_{\substack{x \leq n \leq 2x \\ n=b \ (W)}} \nu(n) = (\alpha + o(1))B^{-k} \frac{x}{W}$$

where

$$\alpha := \tilde{I}(f_4).$$

Now we consider the sum

$$\sum_{\substack{x \leq n \leq 2x \\ n=b \ (W)}} \nu(n)\theta(n + h_k). \tag{89}$$

From Proposition 2.7 we see that we have  $\text{EH}[\vartheta]$  as a consequence of the hypotheses of Theorem 3.12. However, this combined with Theorem 3.5 is not strong enough to obtain an asymptotic for the sum (89), as there is an epsilon loss in (87). But observe that Lemma 3.4 only requires a *lower* bound on the sum (89), rather than an asymptotic.

To obtain this lower bound, we partition  $\{1, \dots, J\}$  into  $\mathcal{J}_1 \cup \mathcal{J}_2$ , where  $\mathcal{J}_1$  consists of those indices  $j \in \{1, \dots, J\}$  with

$$S(f_{1,j}) + \dots + S(f_{k-1,j}) < (1 - \varepsilon)\frac{\vartheta}{2} \tag{90}$$

and  $\mathcal{J}_2$  is the complement. From the elementary inequality

$$(x_1 + x_2)^2 = x_1^2 + 2x_1x_2 + x_2^2 \geq (x_1 + 2x_2)x_1$$

we obtain the pointwise lower bound

$$\nu(n) \geq \left( \left( \sum_{j \in \mathcal{J}_1} + 2 \sum_{j \in \mathcal{J}_2} \right) c_j \lambda_{f_{1,j}}(n + h_1) \dots \lambda_{f_{k,j}}(n + h_k) \right) \left( \sum_{j' \in \mathcal{J}_1} c_{j'} \lambda_{f_{1,j'}}(n + h_1) \dots \lambda_{f_{k,j'}}(n + h_k) \right).$$

The point of performing this lower bound is that if  $j \in \mathcal{J}_1 \cup \mathcal{J}_2$  and  $j' \in \mathcal{J}_1$ , then from (87), (90) one has

$$\sum_{i=1}^{k-1} (S(f_{i,j}) + S(f_{i,j'})) < \vartheta$$

which makes Theorem 3.5(i) available for use. Indeed, for any  $j \in \{1, \dots, J\}$  and  $i = 1, \dots, k$ , we have from (87) that

$$S(f_{i,j}) \leq (1 + \varepsilon) \frac{\vartheta}{2} < \vartheta < 1$$

and so by (20) we have

$$\begin{aligned} \nu(n)\theta(n + h_k) &\geq \left( \left( \sum_{j \in \mathcal{J}_1} + 2 \sum_{j \in \mathcal{J}_2} \right) c_j \lambda_{f_{1,j}}(n + h_1) \dots \lambda_{f_{k-1,j}}(n + h_{k-1}) f_{k,j}(0) \right) \\ &\quad \times \left( \sum_{j' \in \mathcal{J}_1} c_{j'} \lambda_{f_{1,j'}}(n + h_1) \dots \lambda_{f_{k-1,j'}}(n + h_{k-1}) f_{k,j'}(0) \right) \theta(n + h_k) \end{aligned} \quad (91)$$

for  $x \leq n \leq 2x$ . If we then apply Theorem 3.5(i) and the hypothesis  $\text{EH}[\vartheta]$ , we obtain the lower bound

$$\sum_{\substack{x \leq n \leq 2x \\ n=b(W)}} \nu(n)\theta(n + h_k) \geq (\beta_k - o(1)) B^{1-k} \frac{x}{\varphi(W)}$$

with

$$\beta_k := \left( \sum_{j \in \mathcal{J}_1} + 2 \sum_{j \in \mathcal{J}_2} \right) \sum_{j' \in \mathcal{J}_1} c_j c_{j'} f_{k,j}(0) f_{k,j'}(0) \prod_{i=1}^{k-1} \int_0^\infty f'_{i,j}(t_i) f'_{i,j'}(t_i) dt_i$$

which we can rearrange as

$$\begin{aligned} \beta_k &= \int_{[0,+\infty)^{k-1}} \left( \frac{\partial^{k-1}}{\partial t_1 \dots \partial t_{k-1}} f_{4,1}(t_1, \dots, t_{k-1}, 0) + 2 \frac{\partial^{k-1}}{\partial t_1 \dots \partial t_{k-1}} f_{4,2}(t_1, \dots, t_{k-1}, 0) \right) \\ &\quad \frac{\partial^{k-1}}{\partial t_1 \dots \partial t_{k-1}} f_{4,1}(t_1, \dots, t_{k-1}, 0) dt_1 \dots dt_{k-1} \end{aligned}$$

where

$$f_{4,l}(t_1, \dots, t_k) := \sum_{j \in \mathcal{J}_l} c_j f_{1,j}(t_1) \dots f_{k,j}(t_k)$$

for  $l = 1, 2$ . Note that  $f_{4,1}, f_{4,2}$  are both bounded pointwise by (88), and their supports only overlap on a set of measure  $O(\delta_3)$ . We conclude that

$$\beta_k = \tilde{J}_k(f_{4,1}) + O(\delta_3)$$

with the implied constant independent of  $\delta_3$ , and thus

$$\beta_k = \tilde{J}_{k, (1-\varepsilon)\frac{\vartheta}{2}}(f_4) + O(\delta_3).$$

A similar argument gives

$$\sum_{\substack{x \leq n \leq 2x \\ n=b(W)}} \nu(n)\theta(n + h_i) \geq (\beta_i - o(1)) B^{1-k} \frac{x}{\varphi(W)}$$



for  $i = 1, \dots, k$  with

$$\beta_i = \tilde{J}_{i,(1-\varepsilon)\frac{\vartheta}{2}}(f_4) + O(\delta_3).$$

If we choose  $\delta_3$  small enough, then the claim  $\text{DHL}[k, m+1]$  now follows from Lemma 3.4 and (86).

#### 5.4 Proof of Theorem 3.14

Finally, we prove Theorem 3.14. Let  $k, m, \varepsilon, F$  be as in that theorem. By rescaling as in previous sections, we may find a square-integrable function  $F_1 : [0, +\infty)^k \rightarrow \mathbb{R}$  supported on  $(\frac{k}{k-1}\frac{\vartheta}{2} - \delta_1) \cdot \mathcal{R}_k$  for some sufficiently small fixed  $\delta_1 > 0$ , which is not identically zero, which obeys the bound

$$\frac{\sum_{i=1}^k J_{i,(1-\varepsilon)\frac{\vartheta}{2}}(F_1)}{I(F_1)} > m$$

and also obeys the vanishing marginal condition (35) whenever  $t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_k \geq 0$  are such that

$$t_1 + \dots + t_{i-1} + t_{i+1} + \dots + t_k > (1 + \varepsilon)\frac{\vartheta}{2} - \delta_1.$$

As before, we pass from  $F_1$  to  $F_2$  by a spatial translation, and from  $F_2$  to  $F_3$  by a regularisation; crucially, we note that both of these operations interact well with the vanishing marginal condition (35), with the end product being that we obtain a smooth function  $F_3 : [0, +\infty)^k \rightarrow \mathbb{R}$ , supported on the region

$$\{(t_1, \dots, t_k) \in \mathbb{R}^k : t_1 \dots + t_k \leq \frac{k}{k-1}\frac{\vartheta}{2} - \frac{\delta_2}{2}; t_1, \dots, t_k \geq \frac{\delta_2}{2}\}$$

for some sufficiently small  $\delta_2 > 0$ , which is not identically zero, obeying the bound

$$\frac{\sum_{i=1}^k J_{i,(1-\varepsilon)\frac{\vartheta}{2}}(F_3)}{I(F_3)} > m$$

and also obeying the vanishing marginal condition (35) whenever  $t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_k \geq 0$  are such that

$$t_1 + \dots + t_{i-1} + t_{i+1} + \dots + t_k > (1 + \varepsilon)\frac{\vartheta}{2} - \frac{\delta_2}{2}.$$

As before, we now define the function  $f_3 : \mathbb{R}^k \rightarrow \mathbb{R}$  by

$$f_3(t_1, \dots, t_k) := \int_{s_1 \geq t_1, \dots, s_k \geq t_k} F_3(s_1, \dots, s_k) ds_1 \dots ds_k,$$

thus  $f_3$  is smooth, not identically zero and supported on the region

$$\left\{ (t_1, \dots, t_k) \in \mathbb{R}^k : \sum_{i=1}^k \max(t_i, \delta_2/2) \leq \frac{k}{k-1}\frac{\vartheta}{2} - \frac{\delta_2}{2} \right\}.$$

Furthermore, from the vanishing marginal condition we see that we also have

$$f_3(t_1, \dots, t_k) = 0$$

whenever we have some  $1 \leq i \leq k$  for which  $t_i \leq \delta_2/2$  and

$$t_1 + \dots + t_{i-1} + t_{i+1} + \dots + t_k \geq (1 + \varepsilon)\frac{\vartheta}{2} - \frac{\delta_2}{2}.$$

From the fundamental theorem of calculus as before, we have

$$\frac{\sum_{i=1}^k \tilde{J}_{i,(1-\varepsilon)\frac{\vartheta}{2}}(f_3)}{\tilde{I}(f_3)} > m.$$

Using the Stone-Weierstrass theorem as before, we can then find a function  $f_4$  of the form

$$(t_1, \dots, t_k) \mapsto \sum_{j=1}^J c_j f_{1,j}(t_1) \dots f_{k,j}(t_k) \quad (92)$$

where  $c_1, \dots, c_J$  are real scalars, and  $f_{i,j} : \mathbb{R} \rightarrow \mathbb{R}$  are smooth functions supported on intervals of length at most  $\delta_3 > 0$  for some sufficiently small  $\delta_3 > 0$ , with each component  $f_{1,j}(t_1) \dots f_{k,j}(t_k)$  supported in the region

$$\left\{ (t_1, \dots, t_k) \in \mathbb{R}^k : \sum_{i=1}^k \max(t_i, \delta_2/8) \leq \frac{k}{k-1} \frac{\vartheta}{2} - \delta_2/8 \right\}$$

and avoiding the regions

$$\left\{ (t_1, \dots, t_k) \in \mathbb{R}^k : t_i \leq \delta_2/8; \quad t_1 + \dots + t_{i-1} + t_{i+1} + \dots + t_k \geq (1+\varepsilon)\frac{\vartheta}{2} - \delta_2/8 \right\}$$

for each  $i = 1, \dots, k$ , and such that

$$\frac{\sum_{i=1}^k \tilde{J}_{i,(1-\varepsilon)\frac{\vartheta}{2}}(f_4)}{\tilde{I}(f_4)} > m.$$

In particular, for any  $j = 1, \dots, J$  we have

$$S(f_{1,j}) + \dots + S(f_{k,j}) < \frac{k}{k-1} \frac{\vartheta}{2} < \frac{1}{2} \frac{k}{k-1} \leq 1 \quad (93)$$

and for any  $i = 1, \dots, k$  with  $f_{k,i}$  not vanishing at zero, we have

$$S(f_{1,j}) + \dots + S(f_{k,i-1}) + S(f_{k,i+1}) + \dots + S(f_{k,j}) < (1+\varepsilon)\frac{\vartheta}{2}. \quad (94)$$

Let  $\nu$  be defined by (85). From (93), the hypothesis  $\text{GEH}[\vartheta]$ , and the argument from the previous section used to prove Theorem 3.12(ii), we have

$$\sum_{\substack{x \leq n \leq 2x \\ n=b(W)}} \nu(n) = (\alpha + o(1)) B^{-k} \frac{x}{W}$$

where

$$\alpha := \tilde{I}(f_4).$$

Similarly, from (94) (and the upper bound  $S(f_{i,j}) < 1$  from (93)), the hypothesis  $\text{EH}[\vartheta]$  (which is available by Proposition 2.7), and the argument from the previous section we have

$$\sum_{\substack{x \leq n \leq 2x \\ n=b(W)}} \nu(n) \theta(n + h_i) \geq (\beta_i - o(1)) B^{1-k} \frac{x}{\varphi(W)}$$

for  $i = 1, \dots, k$  with

$$\beta_i = \tilde{J}_{i, (1-\varepsilon)^{\frac{p}{2}}}(f_4) + O(\delta_3).$$

Setting  $\delta_3$  small enough, the claim  $\text{DHL}[k, m+1]$  now follows from Lemma 3.4.

## 6 Asymptotic analysis

We now establish upper and lower bounds on the quantity  $M_k$  defined in (33), as well as for the related quantities appearing in Theorem 3.10.

To obtain an upper bound on  $M_k$ , we use the following consequence of the Cauchy-Schwarz inequality.

**Lemma 6.1** (Cauchy-Schwarz) *Let  $k \geq 2$ , and suppose that there exist positive measurable functions  $G_i : \mathcal{R}_k \rightarrow (0, +\infty)$  for  $i = 1, \dots, k$  such that*

$$\int_0^\infty G_i(t_1, \dots, t_k) dt_i \leq 1 \quad (95)$$

for all  $t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_k \geq 0$ , where we extend  $G_i$  by zero to all of  $[0, +\infty)^k$ . Then we have

$$M_k \leq \text{ess sup}_{(t_1, \dots, t_k) \in \mathcal{R}_k} \sum_{i=1}^k \frac{1}{G_i(t_1, \dots, t_k)}. \quad (96)$$

Here  $\text{ess sup}$  refers to essential supremum (thus, we may ignore a subset of  $\mathcal{R}_k$  of measure zero in the supremum).

*Proof* Let  $F : [0, +\infty)^k \rightarrow \mathbb{R}$  be a square-integrable function supported on  $\mathcal{R}_k$ . From the Cauchy-Schwarz inequality and (95), we have

$$\left( \int_0^\infty F(t_1, \dots, t_k) dt_i \right)^2 \leq \int_0^\infty \frac{F(t_1, \dots, t_k)^2}{G_i(t_1, \dots, t_k)} dt_i$$

for any  $t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_k \geq 0$ , with  $F^2/G$  extended by zero outside of  $\mathcal{R}_k$ . Inserting this into (32) and integrating, we conclude that

$$J_i(F) \leq \int_{\mathcal{R}_k} \frac{F(t_1, \dots, t_k)^2}{G_i(t_1, \dots, t_k)} dt_1 \dots dt_k.$$

Summing in  $i$  and using (31), (33), (96) we obtain the claim.  $\square$

As a corollary, we can compute  $M_k$  exactly if we can locate a positive eigenfunction:

**Corollary 6.2** *Let  $k \geq 2$ , and suppose that there exists a positive function  $F : \mathcal{R}_k \rightarrow (0, +\infty)$  obeying the eigenfunction equation*

$$\lambda F(t_1, \dots, t_k) = \sum_{i=1}^k \int_0^\infty F(t_1, \dots, t_{i-1}, t'_i, t_{i+1}, \dots, t_k) dt'_i \quad (97)$$

for some  $\lambda > 0$  and all  $(t_1, \dots, t_k) \in \mathcal{R}_k$ , where we extend  $F$  by zero to all of  $[0, +\infty)^k$ . Then  $\lambda = M_k$ .

*Proof* On the one hand, if we integrate (97) against  $F$  and use (31), (32) we see that

$$\lambda I(F) = \sum_{i=1}^k J_i(F)$$

and thus by (33) we see that  $M_k \geq \lambda$ . On the other hand, if we apply Lemma 6.1 with

$$G_i(t_1, \dots, t_k) := \frac{F(t_1, \dots, t_k)}{\int_0^\infty F(t_1, \dots, t_{i-1}, t'_i, t_{i+1}, \dots, t_k) dt'_i}$$

we see that  $M_k \leq \lambda$ , and the claim follows.  $\square$

This allows for an exact calculation of  $M_2$ :

**Corollary 6.3** (Computation of  $M_2$ ) *We have*

$$M_2 = \frac{1}{1 - W(1/e)} = 1.38593 \dots$$

where the Lambert  $W$ -function  $W(x)$  is defined for positive  $x$  as the unique positive solution to  $x = W(x)e^{W(x)}$ .

*Proof* If we set  $\lambda := \frac{1}{1 - W(1/e)} = 1.38593 \dots$ , then a brief calculation shows that

$$2\lambda - 1 = \lambda \log \lambda - \lambda \log(\lambda - 1). \quad (98)$$

Now if we define the function  $f : [0, 1] \rightarrow [0, +\infty)$  by the formula

$$f(x) := \frac{1}{\lambda - 1 + x} + \frac{1}{2\lambda - 1} \log \frac{\lambda - x}{\lambda - 1 + x}$$

then a further brief calculation shows that

$$\int_0^{1-x} f(y) dy = \frac{\lambda - 1 + x}{2\lambda - 1} \log \frac{\lambda - x}{\lambda - 1 + x} + \frac{\lambda \log \lambda - \lambda \log(\lambda - 1)}{2\lambda - 1}$$

for any  $0 \leq x \leq 1$ , and hence by (98) that

$$\int_0^{1-x} f(y) dy = (\lambda - 1 + x)f(x).$$

If we then define the function  $F : \mathcal{R}_2 \rightarrow (0, +\infty)$  by  $F(x, y) := f(x) + f(y)$ , we conclude that

$$\int_0^{1-x} F(x', y) dx' + \int_0^{1-y} F(x, y') dy' = \lambda F(x, y)$$

for all  $(x, y) \in \mathcal{R}_2$ , and the claim now follows from Corollary 6.2.  $\square$

We conjecture that a positive eigenfunction for  $M_k$  exists for all  $k \geq 2$ , not just for  $k = 2$ ; however, we were unable to produce any such eigenfunctions for  $k > 2$ . Nevertheless, Lemma 6.1 still gives us a general upper bound:

**Corollary 6.4** *We have  $M_k \leq \frac{k}{k-1} \log k$  for any  $k \geq 2$ .*

Thus for instance one has  $M_2 \leq 2 \log 2 = 1.38629\dots$ , which compares well with Corollary 6.3. On the other hand, Corollary 6.4 also gives

$$M_4 \leq \frac{4}{3} \log 4 = 1.8454\dots,$$

so that one cannot hope to establish DHL[4, 2] (or DHL[3, 2]) solely through Theorem 3.8 even when assuming GEH, and must rely instead on more sophisticated criteria for DHL[ $k, m$ ] such as Theorem 3.12 or Theorem 3.14.

*Proof* If we set  $G_i : \mathcal{R}_k \rightarrow (0, +\infty)$  for  $i = 1, \dots, k$  to be the functions

$$G_i(t_1, \dots, t_k) := \frac{k-1}{\log k} \frac{1}{1-t_1-\dots-t_k+kt_i}$$

then direct calculation shows that

$$\int_0^\infty G_i(t_1, \dots, t_k) dt_i \leq 1$$

for all  $t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_k \geq 0$ , where we extend  $G_i$  by zero to all of  $[0, +\infty)^k$ . On the other hand, we have

$$\sum_{i=1}^k \frac{1}{G_i(t_1, \dots, t_k)} = \frac{k}{k-1} \log k$$

for all  $(t_1, \dots, t_k) \in \mathcal{R}_k$ . The claim now follows from Lemma 6.1.  $\square$

The upper bound arguments for  $M_k$  can be extended to other quantities such as  $M_{k,\varepsilon}$ , although the bounds do not appear to be as sharp in that case. For instance, we have the following variant of Lemma 6.4, which shows that the improvement in constants when moving from  $M_k$  to  $M_{k,\varepsilon}$  is asymptotically modest:

**Proposition 6.5** *For any  $k \geq 2$  and  $0 \leq \varepsilon < 1$  we have*

$$M_{k,\varepsilon} \leq \frac{k}{k-1} \log(2k-1).$$

*Proof* Let  $F : [0, +\infty)^k \rightarrow \mathbb{R}$  be a square-integrable function supported on  $(1+\varepsilon) \cdot \mathcal{R}_k$ . If  $i = 1, \dots, k$  and  $(t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_k) \in (1-\varepsilon) \cdot \mathcal{R}_k$ , then if we write  $s := 1 - t_1 - \dots - t_{i-1} - t_{i+1} - \dots - t_k$ , we have  $s \geq \varepsilon$  and hence

$$\begin{aligned} \int_0^{1-t_1-\dots-t_{i-1}-t_{i+1}-\dots-t_k+\varepsilon} \frac{1}{1-t_1-\dots-t_k+kt_i} dt_i &= \int_0^{s+\varepsilon} \frac{1}{s+(k-1)t_i} dt_i \\ &= \frac{1}{k-1} \log \frac{ks+(k-1)\varepsilon}{s} \\ &\leq \frac{1}{k-1} \log(2k-1). \end{aligned}$$

By Cauchy-Schwarz, we conclude that

$$\left( \int_0^\infty F(t_1, \dots, t_k) dt_i \right)^2 \leq \frac{1}{k-1} \log(2k-1) \int_0^\infty (1-t_1-\dots-t_k+kt_i) F(t_1, \dots, t_k)^2 dt_i.$$

Integrating in  $t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_k$  and summing in  $i$ , we obtain the claim.  $\square$

**Remark 6.6** The same argument, using the weight  $1 + a(-t_1 - \dots - t_k + kt_i)$ , gives the more general inequality

$$M_{k,\varepsilon} \leq \frac{k}{a(k-1)} \log \left( k + \frac{(a(1+\varepsilon)-1)(k-1)}{1-a(1-\varepsilon)} \right)$$

whenever  $\frac{1}{1+\varepsilon} < a < \frac{1}{1-\varepsilon}$ ; the case  $a = 1$  is Proposition 6.5, and the limiting case  $a = \frac{1}{1+\varepsilon}$  recovers Lemma 6.4 when one sends  $\varepsilon$  to zero.

One can also adapt the computations in Corollary 6.3 to obtain exact expressions for  $M_{2,\varepsilon}$ , although the calculations are rather lengthy and will only be summarized here. For fixed  $0 < \varepsilon < 1$ , the eigenfunctions  $F$  one seeks should take the form

$$F(x, y) := f(x) + f(y)$$

for  $x, y \geq 0$  and  $x + y \leq 1 + \varepsilon$ , where

$$f(x) := \mathbf{1}_{x \leq 1-\varepsilon} \int_0^{1+\varepsilon-x} F(x, t) dt.$$

In the regime  $0 < \varepsilon < 1/3$ , one can calculate that  $f$  will (up to scalar multiples) take the form

$$f(x) = \mathbf{1}_{x \leq 2\varepsilon} \frac{C_1}{\lambda - 1 - \varepsilon + x} + \mathbf{1}_{2\varepsilon \leq x \leq 1-\varepsilon} \left( \frac{\log(\lambda - x) - \log(\lambda - 1 - \varepsilon + x)}{2\lambda - 1 - \varepsilon} + \frac{1}{\lambda - 1 - \varepsilon + x} \right)$$

where

$$C_1 := \frac{\log(\lambda - 2\varepsilon) - \log(\lambda - 1 + \varepsilon)}{1 - \log(\lambda - 1 + \varepsilon) + \log(\lambda - 1 - \varepsilon)}$$

and  $\lambda$  is the largest root of the equation

$$1 = C_1(\log(\lambda - 1 + \varepsilon) - \log(\lambda - 1 - \varepsilon)) - \log(\lambda - 1 + \varepsilon) + \frac{(\lambda - 1 + \varepsilon)\log(\lambda - 1 + \varepsilon) - (\lambda - 2\varepsilon)\log(\lambda - 2\varepsilon)}{2\lambda - 1 - \varepsilon}.$$

In the regime  $1/3 \leq \varepsilon < 1$ , the situation is significantly simpler, and one has the exact expressions

$$f(x) = \frac{\mathbf{1}_{x \leq 1-\varepsilon}}{\lambda - 1 - \varepsilon + x}$$

and

$$\lambda = \frac{e(1+\varepsilon) - 2\varepsilon}{e - 1}.$$

In both cases, a variant of Corollary 6.2 can be used to show that  $M_{2,\varepsilon}$  will be equal to  $\lambda$ ; thus for instance

$$M_{2,\varepsilon} = \frac{e(1+\varepsilon) - 2\varepsilon}{e - 1}$$

for  $1/3 \leq \varepsilon < 1$ . In particular,  $M_{2,\varepsilon}$  increases to 2 in the limit  $\varepsilon \rightarrow 1$ ; the lower bound  $\liminf_{\varepsilon \rightarrow 1} M_{2,\varepsilon} \geq 2$  can also be established by testing with the function  $F(x, y) := \mathbf{1}_{x \leq \delta, y \leq 1+\varepsilon-\delta} + \mathbf{1}_{y \leq \delta, x \leq 1+\varepsilon-\delta}$  for some sufficiently small  $\delta > 0$ .

Now we turn to lower bounds on  $M_k$ , which are of more relevance for the purpose of establishing results such as Theorem 3.9. If one restricts attention to those functions  $F : \mathcal{R}_k \rightarrow \mathbb{R}$  of the special form  $F(t_1, \dots, t_k) = f(t_1 + \dots + t_k)$  for some function  $f : [0, 1] \rightarrow \mathbb{R}$  then the resulting variational problem has been optimized in previous works [14], [53] (and originally in unpublished work of Conrey), giving rise to the lower bound

$$M_k \geq \frac{4k(k-1)}{j_{k-2}^2}$$

where  $j_{k-2}$  is the first positive zero of the Bessel function  $J_{k-2}$ . This lower bound is reasonably strong for small  $k$ ; for instance, when  $k = 2$  it shows that

$$M_2 \geq 1.383 \dots$$

which compares well with Corollary 6.3, and also shows that  $M_6 > 2$ , recovering the result of Goldston, Pintz, and Yıldırım that DHL[6, 2] (and hence  $H_1 \leq 16$ ) was true on the Elliott-Halberstam conjecture. However, one can show that  $\frac{4k(k-1)}{j_{k-2}^2} < 4$  for all  $k$  (see [59]), so this lower bound cannot be used to force  $M_k$  to be larger than 4.

In [38] the lower bound

$$M_k \geq \log k - 2 \log \log k - 2 \tag{99}$$

was established for all sufficiently large  $k$ . In fact, the arguments in [38] can be used to show this bound for all  $k \geq 200$  (for  $k < 200$ , the right-hand side of (99) is either negative or undefined). Indeed, if we use the bound [38, (7.19)] with  $A$  chosen so that  $A^2 e^A = k$ , then  $3 < A < \log k$  when  $k \geq 200$ , hence  $e^A = k/A^2 > k/\log^2 k$  and so  $A \geq \log k - 2 \log \log k$ . By using the bounds  $\frac{A}{e^A - 1} < \frac{1}{6}$  (since  $A > 3$ ) and  $e^A/k = 1/A^2 < 1/9$ , we see that the right-hand side of [38, (8.17)] exceeds  $A - \frac{1}{(1-1/6-1/9)^2} \geq A - 2$ , which gives (99).

We will remove the  $\log \log k$  term in (99) via the following explicit estimate.

**Theorem 6.7** *Let  $k \geq 2$ , and let  $c, T, \tau > 0$  be parameters. Define the function  $g : [0, T] \rightarrow \mathbb{R}$  by*

$$g(t) := \frac{1}{c + (k-1)t} \tag{100}$$

*and the quantities*

$$m_2 := \int_0^T g(t)^2 dt \tag{101}$$

$$\mu := \frac{1}{m_2} \int_0^T t g(t)^2 dt \tag{102}$$

$$\sigma^2 := \frac{1}{m_2} \int_0^T t^2 g(t)^2 dt - \mu^2. \tag{103}$$

*Assume the inequalities*

$$k\mu \leq 1 - \tau \tag{104}$$

$$k\mu < 1 - T \tag{105}$$

$$k\sigma^2 < (1 + \tau - k\mu)^2. \tag{106}$$

Then one has

$$\frac{k}{k-1} \log k - M_k^{[T]} \leq \frac{k}{k-1} \frac{Z + Z_3 + WX + VU}{(1 + \tau/2)(1 - \frac{k\sigma^2}{(1+\tau-k\mu)^2})} \quad (107)$$

where  $Z, Z_3, W, X, V, U$  are the explicitly computable quantities

$$Z := \frac{1}{\tau} \int_1^{1+\tau} \left( r \left( \log \frac{r-k\mu}{T} + \frac{k\sigma^2}{4(r-k\mu)^2 \log \frac{r-k\mu}{T}} \right) + \frac{r^2}{4kT} \right) dr \quad (108)$$

$$Z_3 := \frac{1}{m_2} \int_0^T kt \log(1 + \frac{t}{T}) g(t)^2 dt \quad (109)$$

$$W := \frac{1}{m_2} \int_0^T \log(1 + \frac{\tau}{kt}) g(t)^2 dt \quad (110)$$

$$X := \frac{\log k}{\tau} c^2 \quad (111)$$

$$V := \frac{c}{m_2} \int_0^T \frac{1}{2c + (k-1)t} g(t)^2 dt \quad (112)$$

$$U := \frac{\log k}{c} \int_0^1 ((1+u\tau - (k-1)\mu - c)^2 + (k-1)\sigma^2) du. \quad (113)$$

Of course, since  $M_k^{[T]} \leq M_k$ , the bound (107) also holds with  $M_k^{[T]}$  replaced by  $M_k$ .

*Proof* From (33) we have

$$\sum_{i=1}^k J_i(F) \leq M_k^{[T]} I(F)$$

whenever  $F : [0, +\infty)^k \rightarrow \mathbb{R}$  is square-integrable and supported on  $[0, T]^k \cap \mathcal{R}_k$ . By rescaling, we conclude that

$$\sum_{i=1}^k J_i(F) \leq r M_k^{[T]} I(F)$$

whenever  $r > 0$  and  $F : [0, +\infty)^k \rightarrow \mathbb{R}$  is square-integrable and supported on  $[0, rT]^k \cap r \cdot \mathcal{R}_k$ . We apply this inequality with the function

$$F(t_1, \dots, t_k) := \mathbf{1}_{t_1 + \dots + t_k \leq r} g(t_1) \dots g(t_k)$$

where  $r > 1$  is a parameter which we will eventually average over, and  $g$  is extended by zero to  $[0, +\infty)$ . We thus have

$$I(F) = m_2^k \int_0^\infty \dots \int_0^\infty \mathbf{1}_{t_1 + \dots + t_k \leq r} \prod_{i=1}^k \frac{g(t_i)^2}{m_2} dt_i.$$

We can interpret this probabilistically as

$$I(F) = m_2^k \mathbb{P}(X_1 + \dots + X_k \leq r)$$

where  $X_1, \dots, X_k$  are independent random variables taking values in  $[0, T]$  with probability distribution  $\frac{1}{m_2} g(t)^2 dt$ . In a similar fashion, we have

$$J_k(F) = m_2^{k-1} \int_0^\infty \dots \int_0^\infty \left( \int_{[0, r-t_1-\dots-t_{k-1}]} g(t) dt \right)^2 \prod_{i=1}^{k-1} \frac{g(t_i)^2}{m_2} dt_i,$$



where we adopt the convention that  $\int_{[a,b]}$  vanishes when  $b < a$ . In probabilistic language, we thus have

$$J_k(F) = m_2^{k-1} \mathbb{E} \left( \int_{[0, r-X_1-\dots-X_{k-1}]} g(t) dt \right)^2$$

where we adopt the convention that the expectation operator  $\mathbb{E}$  applies to the entire expression to the right of that operator unless explicitly restricted by parentheses. Also by symmetry we see that  $J_i(F) = J_k(F)$  for all  $i = 1, \dots, k$ . Putting all this together, we conclude that

$$\mathbb{E} \left( \int_0^{r-X_1-\dots-X_{k-1}} g(t) dt \right)^2 \leq \frac{m_2 M_k^{[T]} r}{k} \mathbb{P}(X_1 + \dots + X_k \geq r)$$

for all  $r > 1$ . Writing  $S_i := X_1 + \dots + X_i$ , we abbreviate this as

$$\mathbb{E} \left( \int_{[0, r-S_{k-1}]} g(t) dt \right)^2 \leq \frac{m_2 M_k^{[T]} r}{k} \mathbb{P}(S_k \geq r). \quad (114)$$

Now we run a variant of the Cauchy-Schwarz argument used to prove Corollary 6.4. If, for fixed  $r > 0$ , we introduce the random function  $h : (0, +\infty) \rightarrow \mathbb{R}$  by the formula

$$h(t) := \frac{1}{r - S_{k-1} + (k-1)t} \mathbf{1}_{S_{k-1} < r} \quad (115)$$

and observe that whenever  $S_{k-1} < r$ , we have

$$\int_{[0, r-S_{k-1}]} h(t) dt = \frac{\log k}{k-1} \quad (116)$$

and thus by the Legendre identity we have

$$\left( \int_{[0, r-S_{k-1}]} g(t) dt \right)^2 = \frac{\log k}{k-1} \int_{[0, r-S_{k-1}]} \frac{g(t)^2}{h(t)} dt - \frac{1}{2} \int_{[0, r-S_{k-1}]} \int_{[0, r-S_{k-1}]} \frac{(g(s)h(t) - g(t)h(s))^2}{h(s)h(t)} ds dt$$

for  $S_{k-1} < r$ ; but the claim also holds when  $r \leq S_{k-1}$  since all integrals vanish in that case. On the other hand, we have

$$\begin{aligned} \mathbb{E} \int_{[0, r-S_{k-1}]} \frac{g(t)^2}{h(t)} dt &= m_2 \mathbb{E}(r - S_{k-1} + (k-1)X_k) \mathbf{1}_{X_k \leq r-S_{k-1}} \\ &= m_2 \mathbb{E}(r - S_k + kX_k) \mathbf{1}_{S_k \leq r} \\ &= m_2 \mathbb{E} r \mathbf{1}_{S_k \leq r} \\ &= m_2 r \mathbb{P}(S_k \leq r) \end{aligned}$$

where we have used symmetry to get the third equality. We conclude that

$$\mathbb{E} \left( \int_{[0, r-S_{k-1}]} g(t) dt \right)^2 = \frac{\log k}{k-1} m_2 r \mathbb{P}(S_k \leq r) - \frac{1}{2} \mathbb{E} \int_{[0, r-S_{k-1}]} \int_{[0, r-S_{k-1}]} \frac{(g(s)h(t) - g(t)h(s))^2}{h(s)h(t)} ds dt.$$

Combining this with (114), we conclude that

$$\Delta r \mathbb{P}(S_k \leq r) \leq \frac{k}{2m_2} \mathbb{E} \int_{[0, r-S_{k-1}]} \int_{[0, r-S_{k-1}]} \frac{(g(s)h(t) - g(t)h(s))^2}{h(s)h(t)} ds dt$$

where

$$\Delta := \frac{k}{k-1} \log k - M_k^{[T]}.$$

Splitting into regions where  $s, t$  are less than  $T$  or greater than  $T$ , and noting that  $g(s)$  vanishes for  $s > T$ , we conclude that

$$\Delta r \mathbb{P}(S_k \leq r) \leq Y_1(r) + Y_2(r)$$

where

$$Y_1(r) := \frac{k}{m_2} \mathbb{E} \int_{[0, T]} \int_{[T, r-S_{k-1}]} \frac{g(t)^2}{h(t)} h(s) \, ds dt$$

and

$$Y_2(r) := \frac{k}{2m_2} \mathbb{E} \int_{[0, \min(T, r-S_{k-1})]} \int_{[0, \min(T, r-S_{k-1})]} \frac{(g(s)h(t) - g(t)h(s))^2}{h(s)h(t)} \, ds dt.$$

We average this from  $r = 1$  to  $r = 1 + \tau$ , to conclude that

$$\Delta \left( \frac{1}{\tau} \int_1^{1+\tau} r \mathbb{P}(S_k \leq r) \, dr \right) \leq \frac{1}{\tau} \int_1^{1+\tau} Y_1(r) \, dr + \frac{1}{\tau} \int_1^{1+\tau} Y_2(r) \, dr.$$

Thus to prove (107), it suffices (by (106)) to establish the bounds

$$\frac{1}{\tau} \int_1^{1+\tau} r \mathbb{P}(S_k \leq r) \, dr \geq (1 + \tau/2) \left( 1 - \frac{k\sigma^2}{(1 + \tau - k\mu)^2} \right), \quad (117)$$

$$\frac{k}{k-1} Y_1(r) \leq Z + Z_3 \quad (118)$$

for all  $1 < r \leq 1 + \tau$ , and

$$\frac{1}{\tau} \int_1^{1+\tau} Y_2(r) \, dr \leq \frac{k}{k-1} (WX + VU). \quad (119)$$

We begin with (117). Since

$$\frac{1}{\tau} \int_1^{1+\tau} r \, dr = 1 + \frac{\tau}{2}$$

it suffices to show that

$$\frac{1}{\tau} \int_1^{1+\tau} r \mathbb{P}(S_k > r) \leq \left( 1 + \frac{\tau}{2} \right) \frac{k\sigma^2}{(1 + \tau - k\mu)^2}.$$

But, from (102), (103), we see that each  $X_i$  has mean  $\mu$  and variance  $\sigma^2$ , so  $S_k$  has mean  $k\mu$  and variance  $k\sigma^2$ . It thus suffices to show the pointwise bound

$$\frac{1}{\tau} \int_1^{1+\tau} r 1_{x>r} \leq \left( 1 + \frac{\tau}{2} \right) \frac{(x - k\mu)^2}{(1 + \tau - k\mu)^2}$$

for any  $x$ . It suffices to verify this in the range  $1 \leq x \leq 1 + \tau$ . But in this range, the left-hand side is convex, equals 0 at 1 and  $1 + \tau/2$  at  $1 + \tau$ , while the right-hand side is convex, and equals  $1 + \tau/2$  at  $1 + \tau$  with slope at least  $(1 + \tau/2)/\tau$  there thanks to (104). The claim follows.

Now we show (118). The quantity  $Y_1(r)$  is vanishing unless  $r - S_{k-1} \geq T$ . Using the crude bound  $h(s) \leq \frac{1}{(k-1)s}$  from (115), we see that

$$\int_{[T, r-S_{k-1}]} h(s) ds \leq \frac{1}{k-1} \log_+ \frac{r - S_{k-1}}{T}$$

where  $\log_+(x) := \max(\log x, 0)$ . We conclude that

$$Y_1(r) \leq \frac{k}{k-1} \frac{1}{m_2} \mathbb{E} \int_{[0, T]} \frac{g(t)^2}{h(t)} dt \log_+ \frac{r - S_{k-1}}{T}.$$

We can rewrite this as

$$Y_1(r) \leq \frac{k}{k-1} \mathbb{E} \frac{\mathbf{1}_{S_k \leq r}}{h(X_k)} \log_+ \frac{r - S_{k-1}}{T}.$$

By (115), we have

$$\frac{\mathbf{1}_{S_k \leq r}}{h(X_k)} = (r - S_k + kX_k) \mathbf{1}_{S_k \leq r}.$$

Also, from the elementary bound  $\log_+(x + y) \leq \log_+ x + \log(1 + y)$  for any  $x, y \geq 0$ , we see that

$$\log_+ \frac{r - S_{k-1}}{T} \leq \log_+ \frac{r - S_k}{T} + \log \left( 1 + \frac{X_k}{T} \right).$$

We conclude that

$$\begin{aligned} Y_1(r) &\leq \frac{k}{k-1} \mathbb{E} (r - S_k + kX_k) \left( \log_+ \frac{r - S_k}{T} + \log \left( 1 + \frac{X_k}{T} \right) \right) \mathbf{1}_{S_k \leq r} \\ &\leq \frac{k}{k-1} \left( \mathbb{E} (r - S_k + kX_k) \log_+ \frac{r - S_k}{T} + \max(r - S_k, 0) \frac{X_k}{T} + kX_k \log \left( 1 + \frac{X_k}{T} \right) \right) \end{aligned}$$

using the elementary bound  $\log(1 + y) \leq y$ . Symmetrizing in the  $X_1, \dots, X_k$ , we conclude that

$$Y_1(r) \leq \frac{k}{k-1} (Z_1(r) + Z_2(r) + Z_3) \tag{120}$$

where

$$\begin{aligned} Z_1(r) &:= \mathbb{E} r \log_+ \frac{r - S_k}{T} \\ Z_2(r) &:= \mathbb{E} (r - S_k) \mathbf{1}_{S_k \leq r} \frac{S_k}{kT} \end{aligned}$$

and  $Z_3$  was defined in (109).

For the minor error term  $Z_2$ , we use the crude bound  $(r - S_k) \mathbf{1}_{S_k \leq r} S_k \leq \frac{r^2}{4}$ , so

$$Z_2(r) \leq \frac{r^2}{4kT}. \tag{121}$$

For  $Z_1$ , we upper bound  $\log_+ x$  by a quadratic expression in  $x$ . More precisely, we observe the inequality

$$\log_+ x \leq \frac{(x - 2a \log a - a)^2}{4a^2 \log a}$$

for any  $a > 1$  and  $x \in \mathbb{R}$ , since the left-hand side is concave in  $x$  for  $x \geq 1$ , while the right-hand side is convex in  $x$ , non-negative, and tangent to the left-hand side at  $x = a$ . We conclude that

$$\log_+ \frac{r - S_k}{T} \leq \frac{(r - S_k - 2aT \log a - aT)^2}{4a^2 T^2 \log a}.$$

On the other hand, from (102), (103), we see that each  $X_i$  has mean  $\mu$  and variance  $\sigma^2$ , so  $S_k$  has mean  $k\mu$  and variance  $k\sigma^2$ . We conclude that

$$Z_1(r) \leq r \frac{(r - k\mu - 2aT \log a - aT)^2 + k\sigma^2}{4a^2 T^2 \log a}$$

for any  $a > 1$ .

From (105) and the assumption  $r > 1$ , we may choose  $a := \frac{r-k\mu}{T}$  here, leading to the simplified formula

$$Z_1(r) \leq r \left( \log \frac{r - k\mu}{T} + \frac{k\sigma^2}{4(r - k\mu)^2 \log \frac{r-k\mu}{T}} \right). \quad (122)$$

From (120), (121), (122), (108) we conclude (118).

Finally, we prove (119). Here, we finally use the specific form (100) of the function  $g$ . Indeed, from (100), (115) we observe the identity

$$g(t) - h(t) = (r - S_{k-1} - c)g(t)h(t)$$

for  $t \in [0, \min(r - S_{k-1}, T)]$ . Thus

$$\begin{aligned} Y_2(r) &= \frac{k}{2m_2} \mathbb{E} \int_{[0, \min(r - S_{k-1}, T)]} \int_{[0, \min(r - S_{k-1}, T)]} \frac{((g - h)(s)h(t) - (g - h)(t)h(s))^2}{h(s)h(t)} ds dt \\ &= \frac{k}{2m_2} \mathbb{E}(r - S_{k-1} - c)^2 \int_{[0, \min(r - S_{k-1}, T)]} \int_{[0, \min(r - S_{k-1}, T)]} (g(s) - g(t))^2 h(s)h(t) ds dt. \end{aligned}$$

Using the crude bound  $(g(s) - g(t))^2 \leq g(s)^2 + g(t)^2$  and using symmetry, we conclude

$$Y_2(r) \leq \frac{k}{m_2} \mathbb{E}(r - S_{k-1} - c)^2 \int_{[0, \min(r - S_{k-1}, T)]} \int_{[0, \min(r - S_{k-1}, T)]} g(s)^2 h(s)h(t) ds dt.$$

From (116), (115) we conclude that

$$Y_2(r) \leq \frac{k}{k-1} Z_4(r)$$

where

$$Z_4(r) := \frac{\log k}{m_2} \mathbb{E} \left( (r - S_{k-1} - c)^2 \int_{[0, \min(r - S_{k-1}, T)]} \frac{g(s)^2}{r - S_{k-1} + (k-1)s} ds \right).$$

To prove (119), it thus suffices (after making the change of variables  $r = 1 + u\tau$ ) to show that

$$\int_0^1 Z_4(1 + u\tau) du \leq WX + VU. \quad (123)$$

We will exploit the averaging in  $u$  to deal with the singular nature of the factor  $\frac{1}{r-S_{k-1}+(k-1)s}$ . By Fubini's theorem, the left-hand side of (123) may be written as

$$\frac{\log k}{m_2} \mathbb{E} \int_0^1 Q(u) \, du$$

where  $Q(u)$  is the random variable

$$Q(u) := (1 + u\tau - S_{k-1} - c)^2 \int_{[0, \min(1+u\tau-S_{k-1}, T)]} \frac{g(s)^2}{1 + u\tau - S_{k-1} + (k-1)s} \, ds.$$

Note that  $Q(u)$  vanishes unless  $1 + u\tau - S_{k-1} > 0$ . Consider first the contribution of those  $Q(u)$  for which

$$0 < 1 + u\tau - S_{k-1} \leq 2c.$$

In this regime we may bound

$$(1 + u\tau - S_{k-1} - c)^2 \leq c^2,$$

so this contribution to (123) may be bounded by

$$\frac{\log k}{m_2} c^2 \mathbb{E} \int_{[0, T]} g(s)^2 \left( \int_0^1 \frac{\mathbf{1}_{1+u\tau-S_{k-1} \geq s}}{1 + u\tau - S_{k-1} + (k-1)s} \, du \right) \, ds.$$

Observe on making the change of variables  $v := 1 + u\tau - S_{k-1} + (k-1)s$  that

$$\begin{aligned} \int_0^1 \frac{\mathbf{1}_{1+u\tau-S_{k-1} \geq s}}{1 + u\tau - S_{k-1} + (k-1)s} \, du &= \frac{1}{\tau} \int_{[\max(ks, 1-S_{k-1}+(k-1)s), 1-S_{k-1}+\tau+(k-1)s]} \frac{dv}{v} \\ &\leq \frac{1}{\tau} \log \frac{ks + \tau}{ks} \end{aligned}$$

and so this contribution to (123) is bounded by  $WX$ , where  $W, X$  are defined in (110), (111).

Now we consider the contribution to (123) when<sup>[5]</sup>

$$1 + u\tau - S_{k-1} > 2c.$$

In this regime we bound

$$\frac{1}{1 + u\tau - S_{k-1} + (k-1)s} \leq \frac{1}{2c + (k-1)t},$$

and so this portion of  $\int_0^1 Z_4[1 + u\tau] \, du$  may be bounded by

$$\int_0^1 \frac{\log k}{c} \mathbb{E} (1 + u\tau - S_{k-1} - c)^2 V \, du = VU$$

where  $V, U$  are defined in (112), (113). The proof of the theorem is now complete.  $\square$

---

<sup>[5]</sup>One could obtain a small improvement to the bounds here by replacing the threshold  $2c$  with a parameter to be optimized over.

We can now perform an asymptotic analysis in the limit  $k \rightarrow \infty$  to establish Theorem 3.9(xi) and Theorem 3.11(vi). For  $k$  sufficiently large, we select the parameters

$$\begin{aligned} c &:= \frac{1}{\log k} + \frac{\alpha}{\log^2 k} \\ T &:= \frac{\beta}{\log k} \\ \tau &:= \frac{\gamma}{\log k} \end{aligned}$$

for some real parameters  $\alpha \in \mathbb{R}$  and  $\beta, \gamma > 0$  independent of  $k$  to be optimized in later. From (100), (101) we have

$$\begin{aligned} m_2 &= \frac{1}{k-1} \left( \frac{1}{c} - \frac{1}{c + (k-1)T} \right) \\ &= \frac{\log k}{k} \left( 1 - \frac{\alpha}{\log k} + o\left(\frac{1}{\log k}\right) \right) \end{aligned}$$

where we use  $o(f(k))$  to denote a function  $g(k)$  of  $k$  with  $g(k)/f(k) \rightarrow 0$  as  $k \rightarrow \infty$ . On the other hand, we have from (100), (102) that

$$\begin{aligned} m_2(c + (k-1)\mu) &= \int_0^T (c + (k-1)t)g(t)^2 dt \\ &= \frac{1}{k-1} \log \frac{c + (k-1)T}{c} \\ &= \frac{\log k}{k} \left( 1 + \frac{\log \beta}{\log k} + o\left(\frac{1}{\log k}\right) \right) \end{aligned}$$

and thus

$$\begin{aligned} k\mu &= \frac{k}{k-1} \left( 1 + \frac{\log \beta + \alpha}{\log k} + o\left(\frac{1}{\log k}\right) \right) - \frac{kc}{k-1} \\ &= 1 + \frac{\log \beta + \alpha}{\log k} + o\left(\frac{1}{\log k}\right) - \left( \frac{1}{\log k} + o\left(\frac{1}{\log k}\right) \right) \\ &= 1 + \frac{\log \beta + \alpha - 1}{\log k} + o\left(\frac{1}{\log k}\right). \end{aligned}$$

Similarly, from (100), (102), (103) we have

$$\begin{aligned} m_2(c^2 + 2c(k-1)\mu + (k-1)^2(\mu^2 + \sigma^2)) &= \int_0^T (c + (k-1)t)^2 g(t)^2 dt \\ &= T \end{aligned}$$

and thus

$$\begin{aligned} k\sigma^2 &= \frac{k}{(k-1)^2} \left( \frac{T}{m_2} - c^2 - 2c(k-1)\mu \right) - k\mu^2 \\ &= \frac{\beta}{\log^2 k} + o\left(\frac{1}{\log^2 k}\right). \end{aligned}$$

We conclude that the hypotheses (104), (105), (106) will be obeyed for sufficiently large  $k$  if we have

$$\begin{aligned}\log \beta + \alpha + \gamma &< 1 \\ \log \beta + \alpha + \beta &< 1 \\ \beta &< (1 + \gamma - \alpha - \log \beta)^2.\end{aligned}$$

These conditions can be simultaneously obeyed, for instance by setting  $\beta = \gamma = 1$  and  $\alpha = -1$ .

Now we crudely estimate the quantities  $Z, Z_3, W, X, V, U$  in (108)-(113). For  $1 \leq r \leq 1 + \tau$ , we have  $r - k\mu \asymp 1/\log k$ , and so

$$\frac{r - k\mu}{T} \asymp 1; \quad \frac{k\sigma^2}{(r - k\mu)^2} \asymp 1; \quad \frac{r^2}{4kT} = o(1)$$

and so by (108)  $Z = O(1)$ . Using the crude bound  $\log(1 + \frac{t}{T}) = O(1)$  for  $0 \leq t \leq T$ , we see from (109), (102) that  $Z_3 = O(k\mu) = O(1)$ . It is clear that  $X = O(1)$ , and using the crude bound  $\frac{1}{2c + (k-1)t} \leq \frac{1}{c}$  we see from (112), (101) that  $V = O(1)$ . For  $0 \leq u \leq 1$  we have  $1 + u\tau - (k-1)\mu - c = O(1/\log k)$ , so from (113) we have  $U = O(1)$ . Finally, from (110) and the change of variables  $t = \frac{s}{k \log k}$  we have

$$\begin{aligned}W &= \frac{\log k}{km_2} \int_0^{kT \log k} \log \left(1 + \frac{\gamma}{s}\right) \frac{ds}{\left(1 + \frac{\alpha}{\log k} + \frac{k-1}{k}s\right)^2} \\ &= O \left( \int_0^\infty \log \left(1 + \frac{\gamma}{s}\right) \frac{ds}{(1 + o(1))(1 + s)^2} \right) \\ &= O(1).\end{aligned}$$

Finally we have

$$1 - \frac{k\sigma^2}{(1 + \tau - k\mu)^2} \asymp 1.$$

Putting all this together, we see from (107) that

$$M_k \geq M_k^{[T]} \geq \frac{k}{k-1} \log k - O(1)$$

giving Theorem 3.9(xi). Furthermore, if we set

$$\varpi := \frac{7}{600} - \frac{C}{\log k}$$

and

$$\delta := \left( \frac{1}{4} + \frac{7}{600} \right) \frac{\beta}{\log k}$$

then we will have  $600\varpi + 180\delta < 7$  for  $C$  large enough, and Theorem 3.11(vi) also follows (as one can verify from inspection that all implied constants here are effective).

Finally, Theorem 3.9(viii), (ix), (x) follow by setting

$$\begin{aligned}c &:= \frac{\theta}{\log k} \\ T &:= \frac{\beta}{\log k} \\ \tau &= 1 - k\mu\end{aligned}$$

**Table 2** Parameter choices for Theorems 3.9, 3.11.

| $k$        | $\theta$  | $\beta$   | $M$         |
|------------|-----------|-----------|-------------|
| 5511       | 0.965     | 0.973     | 6.000048609 |
| 35410      | 0.99479   | 0.85213   | 7.829849259 |
| 41588      | 0.97878   | 0.94319   | 8.000001401 |
| 309661     | 0.98627   | 0.92091   | 10.00000032 |
| 1649821    | 1.00422   | 0.80148   | 11.65752556 |
| 75845707   | 1.00712   | 0.77003   | 15.48125090 |
| 3473955908 | 1.0079318 | 0.7490925 | 19.30374872 |

with  $\theta, \beta$  given by Table 2, with (107) then giving the bound  $M_k^{[T]} > M$  with  $M$  as given by the table, after verifying of course that the conditions (104), (105), (106) are obeyed. Similarly, Theorem 3.11 (ii), (iii), (iv), (v) follows with  $\theta, \beta$  given by the same table, with  $\varpi$  chosen so that

$$M = \frac{m}{\frac{1}{4} + \varpi}$$

with  $m = 2, 3, 4, 5$  for (ii), (iii), (iv), (v) respectively, and  $\delta$  chosen by the formula

$$\delta := T\left(\frac{1}{4} + \varpi\right).$$

## 7 The case of small and medium dimension

In this section we establish lower bounds for  $M_k$  (and related quantities, such as  $M_{k,\varepsilon}$ ) both for small values of  $k$  (in particular,  $k = 3$  and  $k = 4$ ) and medium values of  $k$  (in particular,  $k = 50$  and  $k = 54$ ). Specifically, we will establish Theorem 3.9(vii), Theorem 3.13, and Theorem 3.15.

### 7.1 Bounding $M_k$ for medium $k$

We begin with the problem of lower bounding  $M_k$ . We first formalize an observation<sup>[6]</sup> of Maynard [38] that one may restrict without loss of generality to symmetric functions:

**Lemma 7.1** *For any  $k \geq 2$ , one has*

$$M_k := \sup \frac{k J_1(F)}{I(F)}$$

where  $F$  ranges over symmetric square-integrable functions on  $\mathcal{R}_k$  that are not identically zero.

*Proof* Firstly, observe that if one replaces a square-integrable function  $F : [0, +\infty)^k \rightarrow \mathbb{R}$  with its absolute value  $|F|$ , then  $I(|F|) = I(F)$  and  $J_i(|F|) \geq J_i(F)$ . Thus one may restrict the supremum in (33) to non-negative functions without loss of generality. We may thus find a sequence  $F_n$  of square-integrable non-negative functions on  $\mathcal{R}_k$ , normalized so that  $I(F_n) = 1$ , and such that  $\sum_{i=1}^k J_i(F_n) \rightarrow M_k$  as  $n \rightarrow \infty$ .

Now let

$$\overline{F}_n(t_1, \dots, t_k) := \frac{1}{k!} \sum_{\sigma \in S_k} F_n(t_{\sigma(1)}, \dots, t_{\sigma(k)})$$

be the symmetrization of  $F_n$ . Since the  $F_n$  are non-negative with  $I(F_n) = 1$ , we see that

$$I(\overline{F}_n) \geq I\left(\frac{1}{k!} F_n\right) = \frac{1}{(k!)^2}$$

<sup>[6]</sup>The arguments in [38] are rigorous under the assumption of a positive eigenfunction as in Corollary 6.2, but the existence of such an eigenfunction remains open for  $k \geq 3$ .



and so  $I(\overline{F_n})$  is bounded away from zero. Also, from (33), we know that the quadratic form

$$Q(F) := M_k I(F) - \sum_{i=1}^k J_i(F)$$

is positive semi-definite and is also invariant with respect to symmetries, and so from the triangle inequality for inner product spaces we conclude that

$$Q(\overline{F_n}) \leq Q(F_n).$$

By construction,  $Q(F_n)$  goes to zero as  $n \rightarrow \infty$ , and thus  $Q(\overline{F_n})$  also goes to zero. We conclude that

$$\frac{k J_1(\overline{F_n})}{I(\overline{F_n})} = \frac{\sum_{i=1}^k J_i(\overline{F_n})}{I(\overline{F_n})} \rightarrow M_k$$

as  $n \rightarrow \infty$ , and so

$$M_k \geq \sup \frac{k J_1(F)}{I(F)}.$$

The reverse inequality is immediate from (33), and the claim follows.  $\square$

To establish a lower bound of the form  $M_k > C$  for some  $C > 0$ , one thus seeks to locate a symmetric function  $F : [0, +\infty)^k \rightarrow \mathbb{R}$  supported on  $\mathcal{R}_k$  such that

$$k J_1(F) > C I(F). \quad (124)$$

To do this numerically, we follow [38] (see also [25] for some related ideas) and can restrict attention to functions  $F$  that are linear combinations

$$F = \sum_{i=1}^n a_i b_i$$

of some explicit finite set of symmetric square-integrable functions  $b_1, \dots, b_n : [0, +\infty)^k \rightarrow \mathbb{R}$  supported on  $\mathcal{R}_k$ , and some real scalars  $a_1, \dots, a_n$  that we may optimize in. The condition (124) then may be rewritten as

$$\mathbf{a}^T \mathbf{M}_2 \mathbf{a} - C \mathbf{a}^T \mathbf{M}_1 \mathbf{a} > 0 \quad (125)$$

where  $\mathbf{a}$  is the vector

$$\mathbf{a} := \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

and  $\mathbf{M}_1, \mathbf{M}_2$  are the real symmetric and positive semi-definite  $n \times n$  matrices

$$\mathbf{M}_1 = \left( \int_{\mathbb{R}^k} b_i(t_1, \dots, t_k) b_j(t_1, \dots, t_k) dt_1 \dots dt_k \right)_{1 \leq i, j \leq n} \quad (126)$$

$$\mathbf{M}_2 = \left( k \int_{\mathbb{R}^{k+1}} b_i(t_1, \dots, t_k) b_j(t_1, \dots, t_{k-1}, t'_k) dt_1 \dots dt_k dt'_k \right)_{1 \leq i, j \leq n}. \quad (127)$$

If the  $b_1, \dots, b_n$  are linearly independent in  $L^2(\mathcal{R}_k)$ , then  $\mathbf{M}_1$  is strictly positive definite, and (as observed in [38, Lemma 8.3]), one can find  $\mathbf{a}$  obeying (125) if and only if the largest eigenvalue of  $\mathbf{M}_2 \mathbf{M}_1^{-1}$  exceeds  $C$ . This is a criterion that can be numerically verified for medium-sized values of  $n$ , if the  $b_1, \dots, b_n$  are chosen so that the matrix coefficients of  $\mathbf{M}_1, \mathbf{M}_2$  are explicitly computable.

In order to facilitate computations, it is natural to work with bases  $b_1, \dots, b_n$  of symmetric polynomials. We have the following basic integration identity:

**Lemma 7.2** (Beta function identity) *For any non-negative  $a, a_1, \dots, a_k$ , we have*

$$\int_{\mathcal{R}_k} (1 - t_1 - \dots - t_k)^a t_1^{a_1} \dots t_k^{a_k} dt_1 \dots dt_k = \frac{\Gamma(a+1)\Gamma(a_1+1)\dots\Gamma(a_k+1)}{\Gamma(a_1+\dots+a_k+k+a+1)}$$

where  $\Gamma(s) := \int_0^\infty t^{s-1} e^{-t} dt$  is the Gamma function. In particular, if  $a_1, \dots, a_k$  are natural numbers, then

$$\int_{\mathcal{R}_k} (1 - t_1 - \dots - t_k)^a t_1^{a_1} \dots t_k^{a_k} dt_1 \dots dt_k = \frac{a! a_1! \dots a_k!}{(a_1 + \dots + a_k + k + a)!}.$$

*Proof* Since

$$\int_{\mathcal{R}_k} (1 - t_1 - \dots - t_k)^a t_1^{a_1} \dots t_k^{a_k} dt_1 \dots dt_k = a \int_{\mathcal{R}_{k+1}} t_1^{a_1} \dots t_k^{a_k} t_{k+1}^{a-1} dt_1 \dots dt_{k+1}$$

we see that to establish the lemma it suffices to do so in the case  $a = 0$ .

If we write

$$X := \int_{t_1+\dots+t_k=1} t_1^{a_1} \dots t_k^{a_k} dt_1 \dots dt_{k-1}$$

then by homogeneity we have

$$r^{a_1+\dots+a_k+k-1} X = \int_{t_1+\dots+t_k=r} t_1^{a_1} \dots t_k^{a_k} dt_1 \dots dt_{k-1}$$

for any  $r > 0$ , and hence on integrating  $r$  from 0 to 1 we conclude that

$$\frac{X}{a_1 + \dots + a_k + k} = \int_{\mathcal{R}_k} t_1^{a_1} \dots t_k^{a_k} dt_1 \dots dt_k.$$

On the other hand, if we multiply by  $e^{-r}$  and integrate  $r$  from 0 to  $\infty$ , we obtain instead

$$\int_0^\infty r^{a_1+\dots+a_k+k-1} X e^{-r} dr = \int_{[0,+\infty)^k} t_1^{a_1} \dots t_k^{a_k} e^{-t_1-\dots-t_k} dt_1 \dots dt_k.$$

Using the definition of the Gamma function, this becomes

$$\Gamma(a_1 + \dots + a_k + k) X = \Gamma(a_1 + 1) \dots \Gamma(a_k + 1)$$

and the claim follows.  $\square$

Define a *signature* to be a non-increasing sequence  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)$  of natural numbers; for brevity we omit zeroes, thus for instance if  $k = 6$ , then  $(2, 2, 1, 1, 0, 0)$  will be abbreviated as  $(2, 2, 1, 1)$ . The number of non-zero elements of  $\alpha$  will be called the *length* of the signature  $\alpha$ , and as usual the

degree of  $\alpha$  will be  $\alpha_1 + \dots + \alpha_k$ . For each signature  $\alpha$ , we then define the symmetric polynomials  $P_\alpha = P_\alpha^{(k)}$  by the formula

$$P_\alpha(t_1, \dots, t_k) = \sum_{a: s(a) = \alpha} t_1^{a_1} \dots t_k^{a_k}$$

where the summation is over all tuples  $a = (a_1, \dots, a_k)$  whose non-increasing rearrangement  $s(a)$  is equal to  $\alpha$ . Thus for instance

$$\begin{aligned} P_{(1)}(t_1, \dots, t_k) &= t_1 + \dots + t_k \\ P_{(2)}(t_1, \dots, t_k) &= t_1^2 + \dots + t_k^2 \\ P_{(1,1)}(t_1, \dots, t_k) &= \sum_{1 \leq i < j \leq k} t_i t_j \\ P_{(2,1)}(t_1, \dots, t_k) &= \sum_{1 \leq i < j \leq k} t_i^2 t_j + t_i t_j^2 \end{aligned}$$

and so forth. Clearly, the  $P_\alpha$  form a linear basis for the symmetric polynomials of  $t_1, \dots, t_k$ . Observe that if  $\alpha = (\alpha', 1)$  is a signature containing 1, then one can express  $P_\alpha$  as  $P_{(1)}P_{\alpha'}$  minus a linear combination of polynomials  $P_\beta$  with the length of  $\beta$  less than that of  $\alpha$ . This implies that the functions  $P_{(1)}^a P_\alpha$ , with  $a \geq 0$  and  $\alpha$  avoiding 1, are also a basis for the symmetric polynomials. Equivalently, the functions  $(1 - P_{(1)})^a P_\alpha$  with  $a \geq 0$  and  $\alpha$  avoiding 1 form a basis.

After extensive experimentation, we have discovered that a good basis  $b_1, \dots, b_n$  to use for the above problem comes by setting the  $b_i$  to be all the symmetric polynomials of the form  $(1 - P_{(1)})^a P_\alpha$ , where  $a \geq 0$  and  $\alpha$  consists entirely of even numbers, whose total degree  $a + \alpha_1 + \dots + \alpha_k$  is less than or equal to some chosen threshold  $d$ . For such functions, the coefficients of  $\mathbf{M}_1, \mathbf{M}_2$  can be computed exactly using Lemma 7.2.

More explicitly, first we quickly compute a look-up table for the structure constants  $c_{\alpha, \beta, \gamma} \in \mathbb{Z}$  derived from simple products of the form

$$P_\alpha P_\beta = \sum_{\gamma} c_{\alpha, \beta, \gamma} P_\gamma$$

where  $\deg(\alpha) + \deg(\beta) \leq d$ . Using this look-up table we rewrite the integrands of the entries of the matrices in (126) and (127) as integer linear combinations of nearly “pure” monomials of the form  $(1 - P_{(1)})^a t_1^{a_1} \dots t_k^{a_k}$ . We then calculate the entries of  $\mathbf{M}_1$  and  $\mathbf{M}_2$ , as exact rational numbers, using Lemma 7.2.

We next run a generalized eigenvector routine on (real approximations to)  $\mathbf{M}_1$  and  $\mathbf{M}_2$  to find a vector  $\mathbf{a}'$  which nearly maximizes the quantity  $C$  in (125). Taking a rational approximation  $\mathbf{a}$  to  $\mathbf{a}'$ , we then do the quick (and exact) arithmetic to verify that (125) holds for some constant  $C > 4$ . This generalized eigenvector routine is time-intensive when the sizes of  $\mathbf{M}_1$  and  $\mathbf{M}_2$  are large (say, bigger than  $1500 \times 1500$ ), and in practice is the most computationally intensive step of our calculation. When one does not care about an exact arithmetic proof that  $C > 4$ , instead one can run a test for positive-definiteness for the matrix  $C\mathbf{M}_1 - \mathbf{M}_2$ , which is usually much faster and less RAM intensive.

Using this method, we were able to demonstrate  $M_{54} > 4.00238$ , thus establishing Theorem 3.9(vii). We took  $d = 23$  and imposed the restriction on signatures  $\alpha$  that they be composed only of even numbers. It is likely that  $d = 22$  would suffice in the absence of this restriction on signatures, but we found that the gain in  $M_{54}$  from lifting this restriction is typically only in the region of 0.005, whereas the execution time is increased by a large factor. We do not have a good understanding of why this particular restriction on signatures is so inexpensive in terms of the trade-off between the accuracy of  $M$ -values and computational complexity. The total run-time for this computation was under one hour.

We now describe a second choice for the basis elements  $b_1, \dots, b_n$ , which uses the Krylov subspace method; it gives faster and more efficient numerical results than the previous basis, but does not seem to extend as well to more complicated variational problems such as  $M_{k,\varepsilon}$ . We introduce the linear operator  $\mathcal{L} : L^2(\mathcal{R}_k) \rightarrow L^2(\mathcal{R}_k)$  defined by

$$\mathcal{L}f(t_1, \dots, t_k) := \sum_{i=1}^k \int_0^{1-t_1-\dots-t_{i-1}-t_{i+1}-\dots-t_k} f(t_1, \dots, t_{i-1}, t'_i, t_{i+1}, \dots, t_k) dt'_i.$$

This is a self-adjoint and positive semi-definite operator on  $L^2(\mathcal{R}_k)$ . For symmetric  $b_1, \dots, b_n \in L^2(\mathcal{R}_k)$ , one can then write

$$\begin{aligned} \mathbf{M}_1 &= (\langle b_i, b_j \rangle)_{1 \leq i, j \leq n} \\ \mathbf{M}_2 &= (\langle \mathcal{L}b_i, b_j \rangle)_{1 \leq i, j \leq n}. \end{aligned}$$

If we then choose

$$b_i := \mathcal{L}^{i-1}1$$

where 1 is the unit constant function on  $\mathcal{R}_k$ , then the matrices  $\mathbf{M}_1, \mathbf{M}_2$  take the Hankel form

$$\begin{aligned} \mathbf{M}_1 &= (\langle \mathcal{L}^{i+j-2}1, 1 \rangle)_{1 \leq i, j \leq n} \\ \mathbf{M}_2 &= (\langle \mathcal{L}^{i+j-1}1, 1 \rangle)_{1 \leq i, j \leq n}, \end{aligned}$$

and so can be computed entirely in terms of the  $2n$  numbers  $\langle \mathcal{L}^i 1, 1 \rangle$  for  $i = 0, \dots, 2n-1$ .

The operator  $\mathcal{L}$  maps symmetric polynomials to symmetric polynomials; for instance, one has

$$\begin{aligned} \mathcal{L}1 &= k - (k-1)P_{(1)} \\ \mathcal{L}P_{(1)} &= \frac{k}{2} - \frac{k-1}{2}P_{(2)} - (k-2)P_{(1,1)} \end{aligned}$$

and so forth. From this and Lemma 7.2, the quantities  $\langle \mathcal{L}^i 1, 1 \rangle$  are explicitly computable rational numbers; for instance, one can calculate

$$\begin{aligned} \langle 1, 1 \rangle &= \frac{1}{k!} \\ \langle \mathcal{L}1, 1 \rangle &= \frac{2k}{(k+1)!} \\ \langle \mathcal{L}^2 1, 1 \rangle &= \frac{k(5k+1)}{(k+2)!} \\ \langle \mathcal{L}^3 1, 1 \rangle &= \frac{2k^2(7k+5)}{(k+3)!} \end{aligned}$$

and so forth.

With Maple, we were able to compute  $\langle \mathcal{L}^i 1, 1 \rangle$  for  $i \leq 50$  and  $k \leq 100$ , leading to lower bounds on  $M_k$  for these values of  $k$ , a selection of which are given in Table 3.

## 7.2 Bounding $M_{k,\varepsilon}$ for medium $k$

When bounding  $M_{k,\varepsilon}$ , we have not been able to implement the Krylov method, because the analogue of  $\mathcal{L}^i 1$  in this context is piecewise polynomial instead of polynomial, and we were only able to compute it explicitly for very small values of  $i$ , such as  $i = 1, 2, 3$ , which are insufficient for good numerics.

**Table 3** Selected lower bounds on  $M_k$  obtained from the Krylov subspace method, with the  $\frac{k}{k-1} \log k$  upper bound displayed for comparison.

| $k$ | Lower bound on $M_k$ | $\frac{k}{k-1} \log k$ |
|-----|----------------------|------------------------|
| 2   | 1.38593              | 1.38630                |
| 3   | 1.64644              | 1.64792                |
| 4   | 1.84540              | 1.84840                |
| 5   | 2.00714              | 2.01180                |
| 10  | 2.54547              | 2.55843                |
| 20  | 3.12756              | 3.15341                |
| 30  | 3.48313              | 3.51849                |
| 40  | 3.73919              | 3.78347                |
| 50  | 3.93586              | 3.99187                |
| 53  | 3.98621              | 4.04665                |
| 54  | 4.00223              | 4.06425                |
| 60  | 4.09101              | 4.16375                |
| 100 | 4.46424              | 4.65169                |

Thus, we rely on the previously discussed approach, in which symmetric polynomials are used for the basis functions. Instead of computing integrals over the region  $\mathcal{R}_k$  we pass to the regions  $(1 \pm \varepsilon)\mathcal{R}_k$ . In order to apply Lemma 7.2 over these regions, this necessitates working with a slightly different basis of polynomials. We chose to work with those polynomials of the form  $(1 + \varepsilon - P_{(1)})^a P_\alpha$ , where  $\alpha$  is a signature with no 1's. Over the region  $(1 + \varepsilon)\mathcal{R}_k$ , a single change of variables converts the needed integrals into those of the form in Lemma 7.2, and we can then compute the entries of  $\mathbf{M}_1$ .

On the other hand, over the region  $(1 - \varepsilon)\mathcal{R}_k$  we instead want to work with polynomials of the form  $(1 - \varepsilon - P_{(1)})^a P_\alpha$ . Since  $(1 + \varepsilon - P_{(1)})^a = (2\varepsilon + (1 - \varepsilon - P_{(1)}))^a$ , an expansion using the binomial theorem allows us to convert from our given basis to polynomials of the needed form.

With these modifications, and calculating as in the previous section, we find that  $M_{50,1/25} > 4.00124$  if  $d = 25$  and  $M_{50,1/25} > 4.0043$  if  $d = 27$ , thus establishing Theorem 3.13(i). As before, we found it optimal to restrict signatures to contain only even entries, which greatly reduced execution time while only reducing  $M$  by a few thousandths.

One surprising additional computational difficulty introduced by allowing  $\varepsilon > 0$  is that the “complexity” of  $\varepsilon$  as a rational number affects the run-time of the calculations. We found that choosing  $\varepsilon = 1/m$  (where  $m \in \mathbb{Z}$  has only small prime factors) reduces this effect.

A similar argument gives  $M_{51,1/50} > 4.00156$ , thus establishing Theorem 3.13(xiii). In this case our polynomials were of maximum degree  $d = 22$ .

Code and data for these calculations may be found at

[www.dropbox.com/sh/0xb4xrsx4qmua7u/W0huo2Gx7f/Polymath8b](http://www.dropbox.com/sh/0xb4xrsx4qmua7u/W0huo2Gx7f/Polymath8b).

### 7.3 Bounding $M_{4,\varepsilon}$

We now prove Theorem 3.13(xii'), which can be established by a direct numerical calculation. We introduce the explicit function  $F : [0, +\infty)^4 \rightarrow \mathbb{R}$  defined by

$$F(t_1, t_2, t_3, t_4) := (1 - \alpha(t_1 + t_2 + t_3 + t_4)) \mathbf{1}_{t_1+t_2+t_3+t_4 \leq 1+\varepsilon}$$

with  $\varepsilon := 0.168$  and  $\alpha := 0.784$ . As  $F$  is symmetric in  $t_1, t_2, t_3, t_4$ , we have  $J_{i,1-\varepsilon}(F) = J_{1,1-\varepsilon}(F)$ , so to show Theorem 3.13(xii') it will suffice to show that

$$\frac{4J_{1,1-\varepsilon}(F)}{I(F)} > 2.00558. \quad (128)$$

By making the change of variables  $s = t_1 + t_2 + t_3 + t_4$  we see that

$$\begin{aligned}
 I(F) &= \int_{t_1+t_2+t_3+t_4 \leq 1+\varepsilon} (1 - \alpha(t_1 + t_2 + t_3 + t_4))^2 dt_1 dt_2 dt_3 dt_4 \\
 &= \int_0^{1+\varepsilon} (1 - \alpha s)^2 \frac{s^3}{3!} ds \\
 &= \alpha^2 \frac{(1 + \varepsilon)^6}{36} - \alpha \frac{(1 + \varepsilon)^5}{15} + \frac{(1 + \varepsilon)^4}{24} \\
 &= 0.00728001347 \dots
 \end{aligned}$$

and similarly by making the change of variables  $u = t_1 + t_2 + t_3$

$$\begin{aligned}
 J_{1,1-\varepsilon}(F) &= \int_{t_1+t_2+t_3 \leq 1-\varepsilon} \left( \int_0^{1+\varepsilon-t_1-t_2-t_3} (1 - \alpha(t_1 + t_2 + t_3 + t_4)) dt_4 \right)^2 dt_1 dt_2 dt_3 \\
 &= \int_0^{1-\varepsilon} \left( \int_0^{1+\varepsilon-u} (1 - \alpha(u + t_4)) dt_4 \right)^2 \frac{u^2}{2!} du \\
 &= \int_0^{1-\varepsilon} (1 + \varepsilon - u)^2 \left( 1 - \alpha \frac{1 + \varepsilon + u}{2} \right)^2 \frac{u^2}{2} du \\
 &= 0.003650160667 \dots
 \end{aligned}$$

and so (128) follows.

**Remark 7.3** If one uses the truncated function

$$\tilde{F}(t_1, t_2, t_3, t_4) := F(t_1, t_2, t_3, t_4) \mathbf{1}_{t_1, t_2, t_3, t_4 \leq 1}$$

in place of  $F$ , and sets  $\varepsilon$  to 0.18 instead of 0.168, one can compute that

$$\frac{4J_{1,1-\varepsilon}(\tilde{F})}{I(\tilde{F})} > 2.00235.$$

Thus it is possible to establish Theorem 3.13(xii') using a cutoff function  $F'$  that is also supported in the unit cube  $[0, 1]^4$ . This allows for a slight simplification to the proof of DHL[4, 2] assuming GEH, as one can add the additional hypothesis  $S(F_{i_0}) + S(G_{i_0}) < 1$  to Theorem 3.6(ii) in that case.

**Remark 7.4** By optimising in  $\varepsilon$  and taking  $F$  to be a symmetric polynomial of degree higher than 1, one can get slightly better lower bounds for  $M_{4,\varepsilon}$ ; for instance setting  $\varepsilon = 5/21$  and choosing  $F$  to be a cubic polynomial, we were able to obtain the bound  $M_{4,\varepsilon} \geq 2.05411$ . On the other hand, the best lower bound for  $M_{3,\varepsilon}$  that we were able to obtain was 1.91726 (taking  $\varepsilon = 56/113$  and optimizing over cubic polynomials). Again, see [www.dropbox.com/sh/0xb4xrsx4qmua7u/W0huo2Gx7f/Polymath8b](https://www.dropbox.com/sh/0xb4xrsx4qmua7u/W0huo2Gx7f/Polymath8b) for the relevant code and data.

#### 7.4 Three-dimensional cutoffs

In this section we establish Theorem 3.15. We relabel the variables  $(t_1, t_2, t_3)$  as  $(x, y, z)$ , thus our task is to locate a piecewise polynomial function  $F : [0, +\infty)^3 \rightarrow \mathbb{R}$  supported on the simplex

$$R := \left\{ (x, y, z) \in [0, +\infty)^3 : x + y + z \leq \frac{3}{2} \right\}$$

and symmetric in the  $x, y, z$  variables, obeying the vanishing marginal condition

$$\int_0^\infty F(x, y, z) dz = 0 \quad (129)$$

whenever  $x, y \geq 0$  with  $x + y > 1 + \varepsilon$ , and such that

$$J(F) > 2I(F) \quad (130)$$

where

$$J(F) := 3 \int_{x+y \leq 1-\varepsilon} \left( \int_0^\infty F(x, y, z) dz \right)^2 dx dy \quad (131)$$

and

$$I(F) := \int_R F(x, y, z)^2 dx dy dz \quad (132)$$

and

$$\varepsilon := 1/4.$$

Our strategy will be as follows. We will decompose the simplex  $R$  (up to null sets) into a carefully selected set of disjoint open polyhedra  $P_1, \dots, P_m$  (in fact  $m$  will be 60), and on each  $P_i$  we will take  $F(x, y, z)$  to be a low degree polynomial  $F_i(x, y, z)$  (indeed, the degree will never exceed 3). The left and right-hand sides of (130) become quadratic functions in the coefficients of the  $F_i$ . Meanwhile, the requirement of symmetry, as well as the marginal requirement (129), imposes some linear constraints on these coefficients. In principle, this creates a finite-dimensional quadratic program, which one can try to solve numerically. However, to make this strategy practical, one needs to keep the number of linear constraints imposed on the coefficients to be fairly small, as compared with the total number of coefficients. To achieve this, the following properties on the polyhedra  $P_i$  are desirable:

- (Symmetry) If  $P_i$  is a polytope in the partition, then every reflection of  $P_i$  formed by permuting the  $x, y, z$  coordinates should also lie in the partition.
- (Graph structure) Each polytope  $P_i$  should be of the form

$$\{(x, y, z) : (x, y) \in Q_i; a_i(x, y) < z < b_i(x, y)\}, \quad (133)$$

where  $a_i(x, y), b_i(x, y)$  are linear forms and  $Q_i$  is a polygon.

- (Epsilon splitting) Each  $Q_i$  is contained in one of the regions  $\{(x, y) : x + y < 1 - \varepsilon\}$ ,  $\{(x, y) : 1 - \varepsilon < x + y < 1 + \varepsilon\}$ , or  $\{(x, y) : 1 + \varepsilon < x + y < 3/2\}$ .

Observe that the vanishing marginal condition (129) now takes the form

$$\sum_{i: (x, y) \in Q_i} \int_{a_i(x, y)}^{b_i(x, y)} F_i(x, y, z) dz = 0 \quad (134)$$

for every  $x, y > 0$  with  $x + y > 1 + \varepsilon$ . If the set  $\{i : (x, y) \in Q_i\}$  is fixed, then the left-hand side of (134) is a polynomial in  $x, y$  whose coefficients depend linearly on the coefficients of the  $F_i$ , and thus (134) imposes a set of linear conditions on these coefficients for each possible set  $\{i : (x, y) \in Q_i\}$  with  $x + y > 1 + \varepsilon$ .

Now we describe the partition we will use. This partition can in fact be used for all  $\varepsilon$  in the interval  $[1/4, 1/3]$ , but the endpoint  $\varepsilon = 1/4$  has some simplifications which allowed for reasonably good numerical results. To obtain the symmetry property, it is natural to split  $R$  (modulo null sets) into six polyhedra  $R_{xyz}, R_{xzy}, R_{yxz}, R_{yzx}, R_{zxy}, R_{zyx}$ , where

$$\begin{aligned} R_{xyz} &:= \{(x, y, z) \in R : x + y < y + z < z + x\} \\ &= \{(x, y, z) : 0 < y < x < z; x + y + z \leq 3/2\} \end{aligned}$$

and the other polyhedra are obtained by permuting the indices  $x, y, z$ , thus for instance

$$\begin{aligned} R_{yxz} &:= \{(x, y, z) \in R : y + x < x + z < z + y\} \\ &= \{(x, y, z) : 0 < x < y < z; y + x + z \leq 3/2\}. \end{aligned}$$

To obtain the epsilon splitting property, we decompose  $R_{xyz}$  (modulo null sets) into eight sub-polytopes

$$\begin{aligned} A_{xyz} &= \{(x, y, z) \in R : x + y < y + z < z + x < 1 - \varepsilon\}, \\ B_{xyz} &= \{(x, y, z) \in R : x + y < y + z < 1 - \varepsilon < z + x < 1 + \varepsilon\}, \\ C_{xyz} &= \{(x, y, z) \in R : x + y < 1 - \varepsilon < y + z < z + x < 1 + \varepsilon\}, \\ D_{xyz} &= \{(x, y, z) \in R : 1 - \varepsilon < x + y < y + z < z + x < 1 + \varepsilon\}, \\ E_{xyz} &= \{(x, y, z) \in R : x + y < y + z < 1 - \varepsilon < 1 + \varepsilon < z + x\}, \\ F_{xyz} &= \{(x, y, z) \in R : x + y < 1 - \varepsilon < y + z < 1 + \varepsilon < z + x\}, \\ G_{xyz} &= \{(x, y, z) \in R : x + y < 1 - \varepsilon < 1 + \varepsilon < y + z < z + x\}, \\ H_{xyz} &= \{(x, y, z) \in R : 1 - \varepsilon < x + y < y + z < 1 + \varepsilon < z + x\}; \end{aligned}$$

the other five polytopes  $R_{xzy}, R_{yxz}, R_{yzx}, R_{zxy}, R_{zyx}$  are decomposed similarly, leading to a partition of  $R$  into  $6 \times 8 = 48$  polytopes. This is almost the partition we will use; however there is a technical difficulty arising from the fact that some of the permutations of  $F_{xyz}$  do not obey the graph structure property. So we will split  $F_{xyz}$  further, into the three pieces

$$\begin{aligned} S_{xyz} &= \{(x, y, z) \in F_{xyz} : z < 1/2 + \varepsilon\}, \\ T_{xyz} &= \{(x, y, z) \in F_{xyz} : z > 1/2 + \varepsilon; x > 1/2 - \varepsilon\}, \\ U_{xyz} &= \{(x, y, z) \in F_{xyz} : x < 1/2 - \varepsilon\}. \end{aligned}$$

Thus  $R_{xyz}$  is now partitioned into ten polytopes  $A_{xyz}, B_{xyz}, C_{xyz}, D_{xyz}, E_{xyz}, S_{xyz}, T_{xyz}, U_{xyz}, G_{xyz}, H_{xyz}$ , and similarly for permutations of  $R_{xyz}$ , leading to a decomposition of  $R$  into  $6 \times 10 = 60$  polytopes.

A symmetric piecewise polynomial function  $F$  supported on  $R$  can now be described (almost everywhere) by specifying a polynomial function  $F|_P: P \rightarrow \mathbb{R}$  for the ten polytopes  $P = A_{xyz}, B_{xyz}, C_{xyz}, D_{xyz}, E_{xyz}, S_{xyz}, T_{xyz}, U_{xyz}, G_{xyz}, H_{xyz}$ , and then extending by symmetry, thus for instance

$$F|_{A_{yxz}}(x, y, z) = F|_{A_{xyz}}(z, x, y).$$

As discussed earlier, the expressions  $I(F), J(F)$  can now be written as quadratic forms in the coefficients of the  $F|_P$ , and the vanishing marginal condition (129) imposes some linear constraints on these coefficients.



Observe that the polytope  $D_{xyz}$  and all of its permutations make no contribution to either the functional  $J(F)$  or to the marginal condition (129), and give a non-negative contribution to  $I(F)$ . Thus without loss of generality we may assume that

$$F \downarrow_{D_{xyz}} = 0.$$

However, the other nine polytopes  $A_{xyz}, B_{xyz}, C_{xyz}, E_{xyz}, S_{xyz}, T_{xyz}, U_{xyz}, G_{xyz}, H_{xyz}$  have at least one permutation which gives a non-trivial contribution to either  $J(F)$  or to (129), and cannot be easily eliminated.

Now we compute  $I(F)$ . By symmetry we have

$$I(F) = 3! I(F \downarrow_{R_{xyz}}) = 6 \sum_P I(F \downarrow_P)$$

where  $P$  ranges over the nine polytopes  $A_{xyz}, B_{xyz}, C_{xyz}, E_{xyz}, S_{xyz}, T_{xyz}, U_{xyz}, G_{xyz}, H_{xyz}$ . A tedious but straightforward computation shows that

$$\begin{aligned} I(F \downarrow_{A_{xyz}}) &= \int_{x=0}^{1/2-\varepsilon/2} \int_{y=0}^x \int_{z=x}^{1-\varepsilon-x} F \downarrow_{A_{xyz}}^2 dz dy dx \\ I(F \downarrow_{B_{xyz}}) &= \left( \int_{z=1/2-\varepsilon/2}^{1/2+\varepsilon/2} \int_{x=1-\varepsilon-z}^z + \int_{z=1/2+\varepsilon/2}^{1-\varepsilon} \int_{x=1-\varepsilon-z}^{1+\varepsilon-z} \right) \int_{y=0}^{1-\varepsilon-z} F \downarrow_{B_{xyz}}^2 dy dx dz \\ I(F \downarrow_{C_{xyz}}) &= \left( \int_{y=0}^{1/2-3\varepsilon/2} \int_{x=y}^{y+2\varepsilon} + \int_{y=1/2-3\varepsilon/2}^{1/2-\varepsilon} \int_{x=y}^{1-\varepsilon-y} \right) \int_{z=1-\varepsilon-y}^{1+\varepsilon-x} \\ &\quad + \int_{y=1/2-\varepsilon}^{1/2-\varepsilon/2} \int_{x=y}^{1-\varepsilon-y} \int_{z=1-\varepsilon-y}^{3/2-x-y} F \downarrow_{C_{xyz}}^2 dz dx dy \\ I(F \downarrow_{E_{xyz}}) &= \int_{z=1/2+\varepsilon/2}^{1-\varepsilon} \int_{x=1+\varepsilon-z}^z \int_{y=0}^{1-\varepsilon-z} F \downarrow_{E_{xyz}}^2 dy dx dz \\ I(F \downarrow_{S_{xyz}}) &= \left( \int_{y=0}^{1/2-3\varepsilon/2} \int_{z=1-\varepsilon-y}^{1/2+\varepsilon} + \int_{y=1/2-3\varepsilon/2}^{1/2-\varepsilon} \int_{z=y+2\varepsilon}^{1/2+\varepsilon} \right) \int_{x=1+\varepsilon-z}^{1-\varepsilon-y} F \downarrow_{S_{xyz}}^2 dx dz dy \\ I(F \downarrow_{T_{xyz}}) &= \left( \int_{z=1/2+\varepsilon}^{1/2+2\varepsilon} \int_{x=1+\varepsilon-z}^{3/2-z} + \int_{z=1/2+2\varepsilon}^{1+\varepsilon} \int_{x=1/2-\varepsilon}^{3/2-z} \right) \int_{y=0}^{3/2-x-z} F \downarrow_{T_{xyz}}^2 dy dz dx \\ I(F \downarrow_{U_{xyz}}) &= \int_{x=0}^{1/2-\varepsilon} \int_{y=0}^x \int_{z=1+\varepsilon-x}^{1+\varepsilon-y} F \downarrow_{U_{xyz}}^2 dz dy dx \\ I(F \downarrow_{G_{xyz}}) &= \int_{x=0}^{1/2-\varepsilon} \int_{y=0}^x \int_{z=1+\varepsilon-y}^{3/2-x-y} F \downarrow_{G_{xyz}}^2 dx dz dy \end{aligned}$$

and

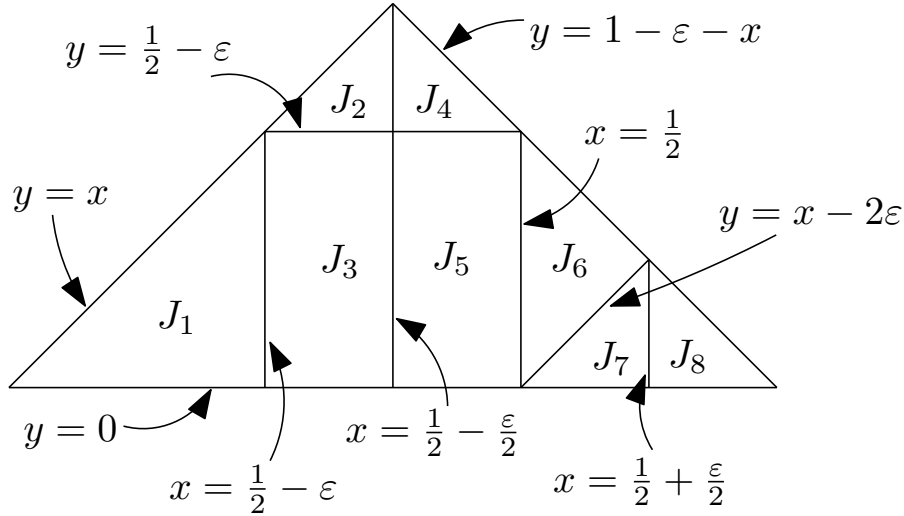
$$\begin{aligned} I(F \downarrow_{H_{xyz}}) &= \left( \int_{x=1/2+\varepsilon/2}^{1-\varepsilon} \int_{y=1-\varepsilon-x}^{3/2-2x} + \int_{x=1-\varepsilon}^{3/4} \int_{y=0}^{3/2-2x} \right) \int_{z=x}^{3/2-x-y} \\ &\quad + \int_{x=1/2}^{1/2+\varepsilon/2} \int_{y=1-\varepsilon-x}^{1/2-\varepsilon} \int_{z=1+\varepsilon-x}^{3/2-x-y} F \downarrow_{H_{xyz}}^2 dz dy dx. \end{aligned}$$

Now we consider the quantity  $J(F)$ . Here we only have the symmetry of swapping  $x$  and  $y$ , so that

$$J(F) = 6 \int_{0 < y < x; x+y < 1-\varepsilon} \left( \int_0^{3/2-x-y} F(x, y, z) dz \right)^2 dx dy.$$

The region of integration meets the polytopes  $A_{xyz}$ ,  $A_{yzx}$ ,  $A_{zyx}$ ,  $B_{xyz}$ ,  $B_{zyx}$ ,  $C_{xyz}$ ,  $E_{xyz}$ ,  $E_{zyx}$ ,  $S_{xyz}$ ,  $T_{xyz}$ ,  $U_{xyz}$ , and  $G_{xyz}$ .

Projecting these polytopes to the  $(x, y)$ -plane, we have the diagram:



This diagram is drawn to scale in the case when  $\varepsilon = 1/4$ , otherwise there is a separation between the  $J_5$  and  $J_7$  regions. For these eight regions there are eight corresponding integrals  $J_1, J_2, \dots, J_8$ , thus

$$J(F) = 6(J_1 + \dots + J_8).$$

We have

$$J_1 = \int_{x=0}^{1/2-\varepsilon} \int_{y=0}^x \left( \int_{z=0}^y F \downarrow_{A_{yzx}} + \int_{z=y}^x F \downarrow_{A_{zyx}} + \int_{z=x}^{1-\varepsilon-x} F \downarrow_{A_{xyz}} + \int_{z=1-\varepsilon-x}^{1-\varepsilon-y} F \downarrow_{B_{xyz}} \right. \\ \left. + \int_{z=1-\varepsilon-y}^{1+\varepsilon-x} F \downarrow_{C_{xyz}} + \int_{z=1+\varepsilon-x}^{1+\varepsilon-y} F \downarrow_{U_{xyz}} + \int_{z=1+\varepsilon-y}^{3/2-x-y} F \downarrow_{G_{xyz}} dz \right)^2 dy dx.$$

Next comes

$$J_2 = \int_{x=1/2-\varepsilon}^{1/2-\varepsilon/2} \int_{y=1/2-\varepsilon}^x \left( \int_{z=0}^y F \downarrow_{A_{yzx}} + \int_{z=y}^x F \downarrow_{A_{zyx}} + \int_{z=x}^{1-\varepsilon-x} F \downarrow_{A_{xyz}} + \int_{z=1-\varepsilon-x}^{1-\varepsilon-y} F \downarrow_{B_{xyz}} \right. \\ \left. + \int_{z=1-\varepsilon-y}^{3/2-x-y} F \downarrow_{C_{xyz}} dz \right)^2 dy dx.$$

Third is the piece

$$J_3 = \int_{x=1/2-\varepsilon}^{1/2-\varepsilon/2} \int_{y=0}^{1/2-\varepsilon} \left( \int_{z=0}^y F \downarrow_{A_{yzx}} + \int_{z=y}^x F \downarrow_{A_{zyx}} + \int_{z=x}^{1-\varepsilon-x} F \downarrow_{A_{xyz}} + \int_{z=1-\varepsilon-x}^{1-\varepsilon-y} F \downarrow_{B_{xyz}} \right. \\ \left. + \int_{z=1-\varepsilon-y}^{1+\varepsilon-x} F \downarrow_{C_{xyz}} + \int_{z=1+\varepsilon-x}^{3/2-x-y} F \downarrow_{T_{xyz}} dz \right)^2 dy dx.$$

We now have dealt with all integrals involving  $A_{xyz}$ , and all remaining integrals pass through  $B_{zyx}$ . Continuing, we have

$$J_4 = \int_{x=1/2-\varepsilon/2}^{1/2} \int_{y=1/2-\varepsilon}^{1-\varepsilon-x} \left( \int_{z=0}^y F \downarrow_{A_{yzx}} + \int_{z=y}^{1-\varepsilon-x} F \downarrow_{A_{zyx}} + \int_{z=1-\varepsilon-x}^x F \downarrow_{B_{zyx}} + \int_{z=x}^{1-\varepsilon-y} F \downarrow_{B_{xyz}} \right. \\ \left. + \int_{z=1-\varepsilon-y}^{3/2-x-y} F \downarrow_{C_{xyz}} dz \right)^2 dy dx.$$

Another component is

$$J_5 = \int_{x=1/2-\varepsilon/2}^{1/2} \int_{y=0}^{1/2-\varepsilon} \left( \int_{z=0}^y F \downarrow_{A_{yzx}} + \int_{z=y}^{1-\varepsilon-x} F \downarrow_{A_{zyx}} \right. \\ \left. + \int_{z=1-\varepsilon-x}^x F \downarrow_{B_{zyx}} + \int_{z=x}^{1-\varepsilon-y} F \downarrow_{B_{xyz}} + \int_{z=1-\varepsilon-y}^{1+\varepsilon-x} F \downarrow_{C_{xyz}} + \int_{z=1+\varepsilon-x}^{3/2-x-y} F \downarrow_{T_{xyz}} dz \right)^2 dy dx.$$

The most complicated piece is

$$J_6 = \left( \int_{x=1/2}^{2\varepsilon} \int_{y=0}^{1-\varepsilon-x} + \int_{x=2\varepsilon}^{1/2+\varepsilon/2} \int_{y=x-2\varepsilon}^{1-\varepsilon-x} \right) \left( \int_{z=0}^y F \downarrow_{A_{yzx}} + \int_{z=y}^{1-\varepsilon-x} F \downarrow_{A_{zyx}} + \int_{z=1-\varepsilon-x}^x F \downarrow_{B_{zyx}} \right. \\ \left. + \int_{z=x}^{1-\varepsilon-y} F \downarrow_{B_{xyz}} + \int_{z=1-\varepsilon-y}^{1+\varepsilon-x} F \downarrow_{C_{xyz}} + \int_{z=1+\varepsilon-x}^{1/2+\varepsilon} F \downarrow_{S_{xyz}} + \int_{z=1/2+\varepsilon}^{3/2-x-y} F \downarrow_{T_{xyz}} dz \right)^2 dy dx.$$

Here we use  $\left( \int_{x=1/2}^{2\varepsilon} \int_{y=0}^{1-\varepsilon-x} + \int_{x=2\varepsilon}^{1/2+\varepsilon/2} \int_{y=x-2\varepsilon}^{1-\varepsilon-x} \right) f(x, y) dy dx$  as an abbreviation for

$$\int_{x=1/2}^{2\varepsilon} \int_{y=0}^{1-\varepsilon-x} f(x, y) dy dx + \int_{x=2\varepsilon}^{1/2+\varepsilon/2} \int_{y=x-2\varepsilon}^{1-\varepsilon-x} f(x, y) dy dx.$$

We have now exhausted  $C_{xyz}$ . The seventh piece is

$$J_7 = \int_{x=2\varepsilon}^{1/2+\varepsilon/2} \int_{y=0}^{x-2\varepsilon} \left( \int_{z=0}^y F \downarrow_{A_{yzx}} + \int_{z=y}^{1-\varepsilon-x} F \downarrow_{A_{zyx}} + \int_{z=1-\varepsilon-x}^x F \downarrow_{B_{zyx}} \right. \\ \left. + \int_{z=x}^{1+\varepsilon-x} F \downarrow_{B_{xyz}} + \int_{z=1+\varepsilon-x}^{1-\varepsilon-y} F \downarrow_{E_{xyz}} + \int_{z=1-\varepsilon-y}^{1/2+\varepsilon} F \downarrow_{S_{xyz}} + \int_{z=1/2+\varepsilon}^{3/2-x-y} F \downarrow_{T_{xyz}} dz \right)^2 dy dx.$$

Finally, we have

$$J_8 = \int_{x=1/2+\varepsilon/2}^{1-\varepsilon} \int_{y=0}^{1-\varepsilon-x} \left( \int_{z=0}^y F \downarrow_{A_{yzx}} + \int_{z=y}^{1-\varepsilon-x} F \downarrow_{A_{zyx}} + \int_{z=1-\varepsilon-x}^{1+\varepsilon-x} F \downarrow_{B_{zyx}} \right. \\ \left. + \int_{z=1+\varepsilon-x}^x F \downarrow_{E_{zyx}} + \int_{z=x}^{1-\varepsilon-y} F \downarrow_{E_{xyz}} + \int_{z=1-\varepsilon-y}^{1/2+\varepsilon} F \downarrow_{S_{xyz}} + \int_{z=1/2+\varepsilon}^{3/2-x-y} F \downarrow_{T_{xyz}} dz \right)^2 dy dx.$$

In the case  $\varepsilon = 1/4$ , the marginal conditions (129) reduce to requiring

$$\int_{z=0}^{3/2-x-y} F \downarrow_{G_{yzx}} dz = 0 \quad (135)$$

$$\int_{z=0}^y F \downarrow_{G_{yzx}} + \int_{z=y}^{3/2-x-y} F \downarrow_{G_{zyx}} dz = 0 \quad (136)$$

$$\int_{z=0}^{1+\varepsilon-x} F \downarrow_{U_{yzx}} + \int_{z=1+\varepsilon-x}^y F \downarrow_{G_{yzx}} + \int_{z=y}^{3/2-x-y} F \downarrow_{G_{zyx}} dz = 0 \quad (137)$$

$$\int_{z=0}^{1+\varepsilon-x} F \downarrow_{U_{yzx}} + \int_{z=1+\varepsilon-x}^{3/2-x-y} F \downarrow_{G_{yzx}} dz = 0 \quad (138)$$

$$\int_{z=0}^{3/2-x-y} F \downarrow_{T_{yzx}} dz = 0 \quad (139)$$

$$\int_{z=0}^{1-\varepsilon-x} F \downarrow_{E_{yzx}} + \int_{z=1-\varepsilon-x}^{1-\varepsilon-y} F \downarrow_{S_{yzx}} + \int_{z=1-\varepsilon-y}^{3/2-x-y} F \downarrow_{H_{yzx}} dz = 0. \quad (140)$$

Each of these constraints is only required to hold for some portion of the parameter space  $\{(x, y) : 1 + \varepsilon \leq x + y \leq 3/2\}$ , but as the left-hand sides are all polynomial functions in  $x, y$  (using the signed definite integral  $\int_b^a = -\int_a^b$ ), it is equivalent to require that all coefficients of these polynomial functions vanish.

Now we specify  $F$ . After some numerical experimentation, we have found the simplest choice of  $F$  that still achieves the desired goal comes by taking  $F(x, y, z)$  to be a polynomial of degree 1 on each of  $E_{xyz}, S_{xyz}, H_{xyz}$ , degree 2 on  $T_{xyz}$ , vanishing on  $D_{xyz}$ , and degree 3 on the remaining five relevant components of  $R_{xyz}$ . After solving the quadratic program, rounding, and clearing denominators, we arrive at the choice

$$\begin{aligned} F \downarrow_{A_{xyz}} &:= -66 + 96x - 147x^2 + 125x^3 + 128y - 122xy + 104x^2y - 275y^2 + 394y^3 + 99z \\ &\quad - 58xz + 63x^2z - 98yz + 51xyz + 41y^2z - 112z^2 + 24xz^2 + 72yz^2 + 50z^3 \\ F \downarrow_{B_{xyz}} &:= -41 + 52x - 73x^2 + 25x^3 + 108y - 66xy + 71x^2y - 294y^2 + 56xy^2 + 363y^3 \\ &\quad + 33z + 15xz + 22x^2z - 40yz - 42xyz + 75y^2z - 36z^2 - 24xz^2 + 26yz^2 + 20z^3 \\ F \downarrow_{C_{xyz}} &:= -22 + 45x - 35x^2 + 63y - 99xy + 82x^2y - 140y^2 + 54xy^2 + 179y^3 \\ F \downarrow_{E_{xyz}} &:= -12 + 8x + 32y \\ F \downarrow_{S_{xyz}} &:= -6 + 8x + 16y \\ F \downarrow_{T_{xyz}} &:= 18 - 30x + 12x^2 + 42y - 20xy - 66y^2 - 45z + 34xz + 22z^2 \\ F \downarrow_{U_{xyz}} &:= 94 - 1823x + 5760x^2 - 5128x^3 + 54y - 168x^2y + 105y^2 + 1422xz - 2340x^2z \\ &\quad - 192y^2z - 128z^2 - 268xz^2 + 64z^3 \\ F \downarrow_{G_{xyz}} &:= 5274 - 19833x + 18570x^2 - 5128x^3 - 18024y + 44696xy - 20664x^2y + 16158y^2 \\ &\quad - 19056xy^2 - 4592y^3 - 10704z + 26860xz - 12588x^2z + 24448yz - 30352xyz \\ &\quad - 10980y^2z + 7240z^2 - 9092xz^2 - 8288yz^2 - 1632z^3 \\ F \downarrow_{H_{xyz}} &:= 8z. \end{aligned}$$

One may compute that

$$I(F) = \frac{62082439864241}{507343011840}$$

and

$$J(F) = \frac{9933190664926733}{40587440947200}$$

with all the marginal conditions (135)-(140) obeyed, thus

$$\frac{J(F)}{I(F)} = 2 + \frac{286648173}{4966595189139280}$$

and (130) follows.

## 8 The parity problem

In this section we argue why the “parity barrier” of Selberg [57] prohibits sieve-theoretic methods, such as the ones in this paper, from obtaining any bound on  $H_1$  that is stronger than  $H_1 \leq 6$ , even on the assumption of strong distributional conjectures such as the generalized Elliott-Halberstam conjecture  $\text{GEH}[\vartheta]$ , and even if one uses sieves other than the Selberg sieve. Our discussion will be somewhat informal and heuristic in nature.

We begin by briefly recalling how the bound  $H_1 \leq 6$  on GEH (i.e., Theorem 1.4(xii)) was proven. This was deduced from the claim  $\text{DHL}[3, 2]$ , or more specifically from the claim that the set

$$A := \{n \in \mathbb{N} : \text{at least two of } n, n+2, n+6 \text{ are prime}\} \quad (141)$$

was infinite.

To do this, we (implicitly) established a lower bound

$$\sum_n \nu(n) \mathbf{1}_A(n) > 0$$

for some non-negative weight  $\nu : \mathbb{N} \rightarrow \mathbb{R}^+$  supported on  $[x, 2x]$  for a sufficiently large  $x$ . This bound was in turn established (after a lengthy sieve-theoretic analysis, and with a carefully chosen weight  $\nu$ ) from upper bounds on various discrepancies. More precisely, one required good upper bounds (on average) for the expressions

$$\left| \sum_{x \leq n \leq 2x: n=a \pmod{q}} f(n+h) - \frac{1}{\varphi(q)} \sum_{x \leq n \leq 2x: (n+h, q)=1} f(n+h) \right| \quad (142)$$

for all  $h \in \{0, 2, 6\}$  and various residue classes  $a \pmod{q}$  with  $q \leq x^{1-\varepsilon}$  and arithmetic functions  $f$ , such as the constant function  $f = 1$ , the von Mangoldt function  $f = \Lambda$ , or Dirichlet convolutions  $f = \alpha \star \beta$  of the type considered in Claim 2.6. (In the presentation of this argument in previous sections, the shift by  $h$  was eliminated using the change of variables  $n' = n + h$ , but for the current discussion it is important that we do not use this shift.) One also required good asymptotic control on the main terms

$$\sum_{x \leq n \leq 2x: (n+h, q)=1} f(n+h). \quad (143)$$

Once one eliminates the shift by  $h$ , an inspection of these arguments reveals that they would be equally valid if one inserted a further non-negative weight  $\omega : \mathbb{N} \rightarrow \mathbb{R}^+$  in the summation over  $n$ . More precisely, the above sieve-theoretic argument would also deduce the lower bound

$$\sum_n \nu(n) \mathbf{1}_A(n) \omega(n) > 0$$

if one had control on the weighted discrepancies

$$\left| \sum_{x \leq n \leq 2x: n=a \pmod{q}} f(n+h) \omega(n) - \frac{1}{\varphi(q)} \sum_{x \leq n \leq 2x: (n+h, q)=1} f(n+h) \omega(n) \right| \quad (144)$$

and on the weighted main terms

$$\sum_{x \leq n \leq 2x: (n+h, q)=1} f(n+h)\omega(n) \quad (145)$$

that were of the same form as in the unweighted case  $\omega = 1$ .

Now suppose for instance that one was trying to prove the bound  $H_1 \leq 4$ . A natural way to proceed here would be to replace the set  $A$  in (141) with the smaller set

$$A' := \{n \in \mathbb{N} : n, n+2 \text{ are both prime}\} \cup \{n \in \mathbb{N} : n+2, n+6 \text{ are both prime}\} \quad (146)$$

and hope to establish a bound of the form

$$\sum_n \nu(n) \mathbf{1}_{A'}(n) > 0$$

for a well-chosen function  $\nu : \mathbb{N} \rightarrow \mathbb{R}^+$  supported on  $[x, 2x]$ , by deriving this bound from suitable (averaged) upper bounds on the discrepancies (142) and control on the main terms (143). If the arguments were sieve-theoretic in nature, then (as in the  $H_1 \leq 6$  case), one could then also deduce the lower bound

$$\sum_n \nu(n) \mathbf{1}_{A'}(n) \omega(n) > 0 \quad (147)$$

for any non-negative weight  $\omega : \mathbb{N} \rightarrow \mathbb{R}^+$ , provided that one had the same control on the weighted discrepancies (144) and weighted main terms (145) that one did on (142), (143).

We apply this observation to the weight

$$\begin{aligned} \omega(n) &:= (1 - \lambda(n)\lambda(n+2))(1 - \lambda(n+2)\lambda(n+6)) \\ &= 1 - \lambda(n)\lambda(n+2) - \lambda(n+2)\lambda(n+6) + \lambda(n)\lambda(n+6) \end{aligned}$$

where  $\lambda(n) := (-1)^{\Omega(n)}$  is the Liouville function. Observe that  $\omega$  vanishes for any  $n \in A'$ , and hence

$$\sum_n \nu(n) \mathbf{1}_{A'}(n) \omega(n) = 0 \quad (148)$$

for any  $\nu$ . On the other hand, the “Möbius randomness law” (see e.g. [34]) predicts a significant amount of cancellation for any non-trivial sum involving the Möbius function  $\mu$ , or the closely related Liouville function  $\lambda$ . For instance, the expression

$$\sum_{x \leq n \leq 2x: n \equiv a \pmod{q}} \lambda(n+h)$$

is expected to be very small (of size<sup>[7]</sup>  $O(\frac{x}{q} \log^{-A} x)$  for any fixed  $A$ ) for any residue class  $a \pmod{q}$  with  $q \leq x^{1-\varepsilon}$ , and any  $h \in \{0, 2, 6\}$ ; similarly for more complicated expressions such as

$$\sum_{x \leq n \leq 2x: n \equiv a \pmod{q}} \lambda(n+2)\lambda(n+6)$$

---

<sup>[7]</sup>Indeed, one might be even more ambitious and conjecture a square-root cancellation  $\ll \sqrt{x/q}$  for such sums (see [40] for some similar conjectures), although such stronger cancellations generally do not play an essential role in sieve-theoretic computations.

or

$$\sum_{x \leq n \leq 2x: n=a \pmod{q}} \Lambda(n) \lambda(n+2) \lambda(n+6)$$

or more generally

$$\sum_{x \leq n \leq 2x: n=a \pmod{q}} f(n) \lambda(n+2) \lambda(n+6)$$

where  $f$  is a Dirichlet convolution  $\alpha \star \beta$  of the form considered in Claim 2.6. Similarly for expressions such as

$$\sum_{x \leq n \leq 2x: n=a \pmod{q}} f(n) \lambda(n) \lambda(n+2);$$

note from the complete multiplicativity of  $\lambda$  that  $(\alpha \star \beta) \lambda = (\alpha \lambda) \star (\beta \lambda)$ , so if  $f$  is of the form in Claim 2.6, then  $f \lambda$  is also. In view of these observations (and similar observations arising from permutations of  $\{0, 2, 6\}$ ), we conclude (heuristically, at least) that all the bounds that are believed to hold for (142), (143) should also hold (up to minor changes in the implied constants) for (144), (145). Thus, if the bound  $H_1 \leq 4$  could be proven in a sieve-theoretic fashion, one should be able to conclude the bound (147), which is in direct contradiction to (148).

**Remark 8.1** Similar arguments work for any set of the form

$$A_H := \{n \in \mathbb{N} : \exists n \leq p_1 < p_2 \leq n + H; p_1, p_2 \text{ both prime, } p_2 - p_1 \leq 4\}$$

and any fixed  $H > 0$ , to prohibit any non-trivial lower bound on  $\sum_n \nu(n) \mathbf{1}_{A_H}(n)$  from sieve-theoretic methods. Indeed, one uses the weight

$$\omega(n) := \prod_{0 \leq i \leq i' \leq H; (n+i, 3) = (n+i', 3) = 1; i' - i \leq 4} (1 - \lambda(n+i) \lambda(n+i'));$$

we leave the details to the interested reader. This seems to block any attempt to use any argument based only on the distribution of the prime numbers and related expressions in arithmetic progressions to prove  $H_1 \leq 4$ .

The same arguments of course also prohibit a sieve-theoretic proof of the twin prime conjecture  $H_1 = 2$ . In this case one can use the simpler weight  $\omega(n) = 1 - \lambda(n) \lambda(n+2)$  to rule out such a proof, and the argument is essentially due to Selberg [57].

Of course, the parity barrier could be circumvented if one were able to introduce stronger sieve-theoretic axioms than the “linear” axioms currently available (which only control sums of the form (142) or (143)). For instance, if one were able to obtain non-trivial bounds for “bilinear” expressions such as

$$\sum_{x \leq n \leq 2x} f(n) \Lambda(n+2) = \sum_d \sum_m \alpha(d) \beta(m) \mathbf{1}_{[x, 2x]}(dm) \Lambda(dm+2)$$

for functions  $f = \alpha \star \beta$  of the form in Claim 2.6, then (by a modification of the proof of Proposition 2.7) one would very likely obtain non-trivial bounds on

$$\sum_{x \leq n \leq 2x} \Lambda(n) \Lambda(n+2)$$

which would soon lead to a proof of the twin prime conjecture. Unfortunately, we do not know of any plausible way to control such bilinear expressions. (Note however that there are some other situations in which bilinear sieve axioms may be established, for instance in the argument of Friedlander and Iwaniec [21] establishing an infinitude of primes of the form  $a^2 + b^4$ .)

## 9 Additional remarks

The proof of Theorem 3.2(xii) may be modified to establish the following variant:

**Proposition 9.1** *Assume the generalized Elliott-Halberstam conjecture  $\text{GEH}[\vartheta]$  for all  $0 < \vartheta < 1$ . Let  $0 < \varepsilon < 1/2$  be fixed. Then if  $x$  is a sufficiently large multiple of 6, there exists a natural number  $n$  with  $\varepsilon x \leq n \leq (1 - \varepsilon)x$  such that at least two of  $n, n - 2, x - n$  are prime. Similarly if  $n - 2$  is replaced by  $n + 2$ .*

Note that if at least two of  $n, n - 2, x - n$  are prime, then either  $n, n + 2$  are twin primes, or else at least one of  $x, x - 2$  is expressible as the sum of two primes, and Theorem 1.5 easily follows.

*Proof (Sketch)* We just discuss the case of  $n - 2$ , as the  $n + 2$  case is similar. Observe from the Chinese remainder theorem (and the hypothesis that  $x$  is divisible by 6) that one can find a residue class  $b \pmod{W}$  such that  $b, b - 2, x - b$  are all coprime to  $W$  (in particular, one has  $b \equiv 1 \pmod{6}$ ). By a routine modification of the proof of Lemma 3.4, it suffices to find a non-negative weight function  $\nu: \mathbb{N} \rightarrow \mathbb{R}^+$  and fixed quantities  $\alpha > 0$  and  $\beta_1, \beta_2, \beta_3 \geq 0$ , such that one has the asymptotic upper bound

$$\sum_{\substack{\varepsilon x \leq n \leq (1-\varepsilon)x \\ n \equiv b \pmod{W}}} \nu(n) \leq \mathfrak{S}(\alpha + o(1)) B^{-k} \frac{(1 - 2\varepsilon)x}{W},$$

the asymptotic lower bounds

$$\begin{aligned} \sum_{\substack{\varepsilon x \leq n \leq (1-\varepsilon)x \\ n \equiv b \pmod{W}}} \nu(n) \theta(n) &\geq \mathfrak{S}(\beta_1 - o(1)) B^{1-k} \frac{(1 - 2\varepsilon)x}{\varphi(W)} \\ \sum_{\substack{\varepsilon x \leq n \leq (1-\varepsilon)x \\ n \equiv b \pmod{W}}} \nu(n) \theta(n + 2) &\geq \mathfrak{S}(\beta_2 - o(1)) B^{1-k} \frac{(1 - 2\varepsilon)x}{\varphi(W)} \\ \sum_{\substack{\varepsilon x \leq n \leq (1-\varepsilon)x \\ n \equiv b \pmod{W}}} \nu(n) \theta(x - n) &\geq \mathfrak{S}(\beta_3 - o(1)) B^{1-k} \frac{(1 - 2\varepsilon)x}{\varphi(W)} \end{aligned}$$

and the inequality

$$\beta_1 + \beta_2 + \beta_3 > 2\alpha,$$

where  $\mathfrak{S}$  is the singular series

$$\mathfrak{S} := \prod_{p|x(x-2); p > w} \frac{p}{p-1}.$$

We select  $\nu$  to be of the form

$$\nu(n) = \left( \sum_{j=1}^J c_j \lambda_{F_{j,1}}(n) \lambda_{F_{j,2}}(n+2) \lambda_{F_{j,3}}(x-n) \right)^2$$



for various fixed coefficients  $c_1, \dots, c_J \in \mathbb{R}$  and fixed smooth compactly supported functions  $F_{j,i} : [0, +\infty) \rightarrow \mathbb{R}$  with  $j = 1, \dots, J$  and  $i = 1, \dots, 3$ . It is then routine<sup>[8]</sup> to verify that analogues of Theorem 3.5 and Theorem 3.6 hold for the various components of  $\nu$ , with the role of  $x$  in the right-hand side replaced by  $(1 - 2\varepsilon)x$ , and the claim then follows by a suitable modification of Theorem 3.14, taking advantage of the function  $F$  constructed in Theorem 3.15.  $\square$

It is likely that the bounds in Theorem 1.4 can be improved further by refining the sieve-theoretic methods employed in this paper, with the exception of part (xii) for which the parity problem prevents further improvement, as discussed in Section 8. We list some possible avenues to such improvements as follows:

- 1 In Theorem 3.13, the bound  $M_{k,\varepsilon} > 4$  was obtained for some  $\varepsilon > 0$  and  $k = 50$ . It is possible that  $k$  could be lowered slightly, for instance to  $k = 49$ , by further numerical computations, but we were only barely able to establish the  $k = 50$  bound after two weeks of computation. However, there may be a more efficient way to solve the required variational problem (e.g. by selecting a more efficient basis than the symmetric monomial basis) that would allow one to advance in this direction; this would improve the bound  $H_1 \leq 246$  slightly. Extrapolation of existing numerics also raises the possibility that  $M_{53}$  exceeds 4, in which case the bound of 270 in Theorem 1.4(vii) could be lowered to 264.
- 2 To reduce  $k$  (and thus  $H_1$ ) further, one could try to solve another variational problem, such as the one arising in Theorem 3.10 or in Theorem 3.14, rather than trying to lower bound  $M_k$  or  $M_{k,\varepsilon}$ . It is also possible to use the more complicated versions of  $\text{MPZ}[\varpi, \delta]$  established in [52] (in which the modulus  $q$  is assumed to be densely divisible rather than smooth) to replace the truncated simplex appearing in Theorem 3.10 with a more complicated region (such regions also appear implicitly in [52, §4.5]). However, in the medium-dimensional setting  $k \approx 50$ , we were not able to accurately and rapidly evaluate the various integrals associated to these variational problems when applied to a suitable basis of functions. One key difficulty here is that whereas polynomials appear to be an adequate choice of basis for the  $M_k$ , an analysis of the Euler-Lagrange equation reveals that one should use piecewise polynomial basis functions instead for more complicated variational problems such as the  $M_{k,\varepsilon}$  problem (as was done in the three-dimensional case in Section 7.4), and these are difficult to work with in medium dimensions. From our experience with the low  $k$  problems, it looks like one should allow these piecewise polynomials to have relatively high degree on some polytopes, low degree on other polytopes, and vanish completely on yet further polytopes<sup>[9]</sup>, but we do not have a systematic understanding of what the optimal placement of degrees should be.
- 3 In Theorem 3.14, the function  $F$  was required to be supported in the simplex  $\frac{k}{k-1} \cdot \mathcal{R}_k$ . However, one can consider functions  $F$  supported in other regions  $R$ , subject to the constraint that all elements of the sumset  $R + R$  lie in a region treatable by one of the cases of Theorem 3.6. This could potentially lead to other optimization problems that lead to superior numerology, although again it appears difficult to perform efficient numerics for such problems in the medium  $k$  regime  $k \approx 50$ . One possibility would be to adopt a “free boundary” perspective, in which the support of  $F$  is not fixed in advance, but is allowed to evolve by some iterative numerical scheme.

<sup>[8]</sup>One new technical difficulty here is that some of the various moduli  $[d_j, d'_j]$  arising in these arguments are not required to be coprime at primes  $p > w$  dividing  $x$  or  $x - 2$ ; this requires some modification to Lemma 4.1 that ultimately leads to the appearance of the singular series  $\mathfrak{S}$ . However, these modifications are quite standard, and we do not give the details here.

<sup>[9]</sup>In particular, the optimal choice  $F$  for  $M_{k,\varepsilon}$  should vanish on the polytope  $\{(t_1, \dots, t_k) \in (1 + \varepsilon) \cdot \mathcal{R}_k : \sum_{i \neq i_0} t_i \geq 1 - \varepsilon \text{ for all } i_0 = 1, \dots, k\}$ .

- 4 To improve the bounds on  $H_m$  for  $m = 2, 3, 4, 5$ , one could seek a better lower bound on  $M_k$  than the one provided by Theorem 6.7; one could also try to lower bound more complicated quantities such as  $M_{k,\varepsilon}$ .
- 5 One could attempt to improve the range of  $\varpi, \delta$  for which estimates of the form  $\text{MPZ}[\varpi, \delta]$  are known to hold, which would improve the results of Theorem 1.4(ii)-(vi). For instance, we believe that the condition  $600\varpi + 180\delta < 7$  in Theorem 2.5 could be improved slightly to  $1080\varpi + 330\delta < 13$  by refining the arguments in [52], but this requires a hypothesis of square root cancellation in a certain four-dimensional exponential sum over finite fields, which we have thus far been unable to establish rigorously. Another direction to pursue would be to improve the  $\delta$  parameter, or to otherwise relax the requirement of smoothness in the moduli, in order to reduce the need to pass to a truncation of the simplex  $\mathcal{R}_k$ , which is the primary reason why the  $m = 1$  results are currently unable to use the existing estimates of the form  $\text{MPZ}[\varpi, \delta]$ . Another speculative possibility is to seek  $\text{MPZ}[\varpi, \delta]$  type estimates which only control distribution for a positive proportion of smooth moduli, rather than for all moduli, and then to design a sieve  $\nu$  adapted to just that proportion of moduli (cf. [17]). Finally, there may be a way to combine the arguments currently used to prove  $\text{MPZ}[\varpi, \delta]$  with the automorphic forms (or “Kloostermania”) methods used to prove nontrivial equidistribution results with respect to a fixed modulus, although we do not have any ideas on how to actually achieve such a combination.
- 6 It is also possible that one could tighten the argument in Lemma 3.4, for instance by establishing a non-trivial lower bound on the portion of the sum  $\sum_n \nu(n)$  when  $n + h_1, \dots, n + h_k$  are all composite, or a sufficiently strong upper bound on the pair correlations  $\sum_n \theta(n + h_i)\theta(n + h_j)$  (see [2, §6] for a recent implementation of this latter idea). However, our preliminary attempts to exploit these adjustments suggested that the gain from the former idea would be exponentially small in  $k$ , whereas the gain from the latter would also be very slight (perhaps reducing  $k$  by  $O(1)$  in large  $k$  regimes, e.g.  $k \geq 5000$ ).
- 7 All of our sieves used are essentially of Selberg type, being the square of a divisor sum. We have experimented with a number of non-Selberg type sieves (for instance trying to exploit the obvious positivity of  $1 - \sum_{p \leq x; p|n} \frac{\log p}{\log x}$  when  $n \leq x$ ), however none of these variants offered a numerical improvement over the Selberg sieve. Indeed it appears that after optimizing the cutoff function  $F$ , the Selberg sieve is in some sense a “local maximum” in the space of non-negative sieve functions, and one would need a radically different sieve to obtain numerically superior results.
- 8 Our numerical bounds for the diameter  $H(k)$  of the narrowest admissible  $k$ -tuple are known to be exact for  $k \leq 342$ , but there is scope for some slight improvement for larger values of  $k$ , which would lead to some improvements in the bounds on  $H_m$  for  $m = 2, 3, 4, 5$ . However, we believe that our bounds on  $H_m$  are already fairly close (e.g. within 10%) of optimal, so there is only a limited amount of gain to be obtained solely from this component of the argument.

## 10 Narrow admissible tuples

In this section we outline the methods used to obtain the numerical bounds on  $H(k)$  given by Theorem 3.3, which are reproduced below:

- 1  $H(3) = 6$ ,
- 2  $H(50) = 246$ ,
- 3  $H(51) = 252$ ,
- 4  $H(54) = 270$ ,
- 5  $H(5511) \leq 52\,116$ ,
- 6  $H(35\,410) \leq 398\,130$ ,
- 7  $H(41\,588) \leq 474\,266$ ,
- 8  $H(309\,661) \leq 4\,137\,854$ ,

0, 4, 6, 16, 30, 34, 36, 46, 48, 58, 60, 64, 70, 78, 84, 88, 90, 94, 100, 106,  
 108, 114, 118, 126, 130, 136, 144, 148, 150, 156, 160, 168, 174, 178, 184,  
 190, 196, 198, 204, 210, 214, 216, 220, 226, 228, 234, 238, 240, 244, 246.

**Figure 1** Admissible 50-tuple realizing  $H(50) = 246$ .

0, 6, 10, 12, 22, 36, 40, 42, 52, 54, 64, 66, 70, 76, 84, 90, 94, 96, 100, 106,  
 112, 114, 120, 124, 132, 136, 142, 150, 154, 156, 162, 166, 174, 180, 184,  
 190, 196, 202, 204, 210, 216, 220, 222, 226, 232, 234, 240, 244, 246, 250, 252.

**Figure 2** Admissible 51-tuple realizing  $H(51) = 252$ .

- 9  $H(1\,649\,821) \leq 24\,797\,814$ ,
- 10  $H(75\,845\,707) \leq 1\,431\,556\,072$ ,
- 11  $H(3\,473\,955\,908) \leq 80\,550\,202\,480$ .

### 10.1 $H(k)$ values for small $k$

The equalities in the first four bounds (1)-(4) were previously known. The case  $H(3) = 6$  is obvious: the admissible 3-tuples  $(0, 2, 6)$  and  $(0, 4, 6)$  have diameter 6 and no 3-tuple of smaller diameter is admissible. The cases  $H(50) = 246$ ,  $H(51) = 252$ , and  $H(54) = 270$  follow from results of Clark and Jarvis [10]. They define  $\varrho^*(x)$  to be the largest integer  $k$  for which there exists an admissible  $k$ -tuple that lies in a half-open interval  $(y, y + x]$  of length  $x$ . For each integer  $k > 1$ , the largest  $x$  for which  $\varrho^*(x) = k$  is precisely  $H(k + 1)$ . Table 1 of [10] lists these largest  $x$  values for  $2 \leq k \leq 170$ , and we find that  $H(50) = 246$ ,  $H(51) = 252$ , and  $H(54) = 270$ . Admissible tuples that realize these bounds are shown in Figures 1, 2 and 3.

### 10.2 $H(k)$ bounds for mid-range $k$

As previously noted, exact values for  $H(k)$  are known only for  $k \leq 342$ . The upper bounds on  $H(k)$  for the five cases (5)-(9) were obtained by constructing admissible  $k$ -tuples using techniques developed during the first part of the Polymath8 project. These are described in detail in Section 3 of [53], but for the sake of completeness we summarize the most relevant methods here.

#### 10.2.1 Fast admissibility testing

A key component of all our constructions is the ability to efficiently determine whether a given  $k$ -tuple  $\mathcal{H} = (h_1, \dots, h_k)$  is admissible. We say that  $\mathcal{H}$  is *admissible modulo*  $p$  if its elements do not form a complete set of residues modulo  $p$ . Any  $k$ -tuple  $\mathcal{H}$  is automatically admissible modulo all primes  $p > k$ , since a  $k$ -tuple cannot occupy more than  $k$  residue classes; thus we only need to test admissibility modulo primes  $p < k$ .

A simple way to test admissibility modulo  $p$  is to enumerate the elements of  $\mathcal{H}$  modulo  $p$  and keep track of which residue classes have been encountered in a table with  $p$  boolean-valued entries. Assuming the elements of  $\mathcal{H}$  have absolute value bounded by  $O(k \log k)$  (true of all the tuples we consider), this approach yields a total bit-complexity of  $O(k^2 / \log k \cdot M(\log k))$ , where  $M(n)$  denotes the complexity of

0, 4, 10, 18, 24, 28, 30, 40, 54, 58, 60, 70, 72, 82, 84, 88, 94, 102, 108, 112, 114,  
 118, 124, 130, 132, 138, 142, 150, 154, 160, 168, 172, 174, 180, 184, 192, 198, 202,  
 208, 214, 220, 222, 228, 234, 238, 240, 244, 250, 252, 258, 262, 264, 268, 270.

**Figure 3** Admissible 54-tuple realizing  $H(54) = 270$ .

multiplying two  $n$ -bit integers, which, up to a constant factor, also bounds the complexity of division with remainder. Applying the Schönhage-Strassen bound  $M(n) = O(n \log n \log \log n)$  from [56], this is  $O(k^2 \log \log k \log \log \log k)$ , essentially quadratic in  $k$ .

This approach can be improved by observing that for most of the primes  $p < k$  there are likely to be many unoccupied residue classes modulo  $p$ . In order to verify admissibility at  $p$  it is enough to find one of them, and we typically do not need to check them all in order to do so. Using a heuristic model that assumes the elements of  $\mathcal{H}$  are approximately equidistributed modulo  $p$ , one can determine a bound  $m < p$  such that  $k$  random elements of  $\mathbb{Z}/p\mathbb{Z}$  are unlikely to occupy all of the residue classes in  $[0, m]$ . By representing the  $k$ -tuple  $\mathcal{H}$  as a boolean vector  $\mathcal{B} = (b_0, \dots, b_{h_k - h_1})$  in which  $b_i = 1$  if and only if  $i = h_j - h_1$  for some  $h_j \in \mathcal{H}$ , we can efficiently test whether  $\mathcal{H}$  occupies every residue class in  $[0, m]$  by examining the entries

$$b_0, \dots, b_m, b_p, \dots, b_{p+m}, b_{2p}, \dots, b_{2p+m}, \dots$$

of  $\mathcal{B}$ . The key point is that when  $p < k$  is large, say  $p > (1 + \epsilon)k / \log k$ , we can choose  $m$  so that we only need to examine a small subset of the entries in  $\mathcal{B}$ . Indeed, for primes  $p > k/c$  (for any constant  $c$ ), we can take  $m = O(1)$  and only need to examine  $O(\log k)$  elements of  $\mathcal{B}$  (assuming its total size is  $O(k \log k)$ , which applies to all the tuples we consider here).

Of course it may happen that  $\mathcal{H}$  occupies every residue class in  $[0, m]$  modulo  $p$ . In this case we revert to our original approach of enumerating the elements of  $\mathcal{H}$  modulo  $p$ , but we expect this to happen for only a small proportion of the primes  $p < k$ . Heuristically, this reduces the complexity of admissibility testing by a factor of  $O(\log k)$ , making it sub-quadratic. In practice we find this approach to be much more efficient than the straight-forward method when  $k$  is large. See [52, §3.1] for further details.

### 10.2.2 Sieving methods

Our techniques for constructing admissible  $k$ -tuples all involve sieving an integer interval  $[s, t]$  of residue classes modulo primes  $p < k$  and then selecting an admissible  $k$ -tuple from the survivors. There are various approaches one can take, depending on the choice of interval and the residue classes to sieve. We list four of these below, starting with the classical sieve of Eratosthenes and proceeding to more modern variations.

- **Sieve of Eratosthenes.** We sieve an interval  $[2, x]$  to obtain admissible  $k$ -tuples

$$p_{m+1}, \dots, p_{m+k}.$$

with  $m$  as small as possible. If we sieve the residue class  $0(p)$  for all primes  $p \leq k$  we have  $m = \pi(k)$  and  $p_{m+1} > k$ . In this case no admissibility testing is required, since the residue class

$0(p)$  is unoccupied for all  $p \leq k$ . Applying the Prime Number Theorem in the forms

$$p_k = k \log k + k \log \log k - k + O\left(k \frac{\log \log k}{\log k}\right),$$

$$\pi(x) = \frac{x}{\log x} + O\left(\frac{x}{\log^2 x}\right),$$

this construction yields the upper bound

$$H(k) \leq k \log k + k \log \log k - k + o(k). \quad (149)$$

As an optimization, rather than sieving modulo every prime  $p \leq k$  we instead sieve modulo increasing primes  $p$  and stop as soon as the first  $k$  survivors form an admissible tuple. This will typically happen for some  $p_m < k$ .

- **Hensley-Richards sieve.** The bound in (149) was improved by Hensley and Richards [32, 33, 54], who observed that rather than sieving  $[2, x]$  it is better to sieve the interval  $[-x/2, x/2]$  to obtain admissible  $k$ -tuples of the form

$$-p_{m+[k/2]-1}, \dots, p_{m+1}, \dots, -1, 1, \dots, p_{m+1}, \dots, p_{m+[(k+1)/2]-1},$$

where we again wish to make  $m$  as small as possible. It follows from Lemma 5 of [33] that one can take  $m = o(k/\log k)$ , leading to the improved upper bound

$$H(k) \leq k \log k + k \log \log k - (1 + \log 2)k + o(k). \quad (150)$$

- **Shifted Schinzel sieve.** As noted by Schinzel in [55], in the Hensley-Richards sieve it is slightly better to sieve  $1(2)$  rather than  $0(2)$ ; this leaves unsieved powers of 2 near the center of the interval  $[-x/2, x/2]$  that would otherwise be removed (more generally, one can sieve  $1(p)$  for many small primes  $p$ , but we did not). Additionally, we find that shifting the interval  $[-x/2, x/2]$  can yield significant improvements (one can also view this as changing the choices of residue classes).

This leads to the following approach: we sieve an interval  $[s, s+x]$  of odd integers and multiples of odd primes  $p \leq p_m$ , where  $x$  is large enough to ensure at least  $k$  survivors, and  $m$  is large enough to ensure that the survivors form an admissible tuple, with  $x$  and  $m$  minimal subject to these constraints. A tuple of exactly  $k$  survivors is then chosen to minimize the diameter. By varying  $s$  and comparing the results, we can choose a starting point  $s \in [-x/2, x/2]$  that yields the smallest final diameter. For large  $k$  we typically find  $s \approx k$  is optimal, as opposed to  $s \approx -(k/2) \log k$  in the Hensley-Richards sieve.

- **Shifted greedy sieve.** As a further optimization, we can allow greater freedom in the choice of residue class to sieve. We begin as in the shifted Schinzel sieve, but for primes  $p \leq p_m$  that exceed  $2\sqrt{k \log k}$ , rather than sieving  $0(p)$  we choose a minimally occupied residue class  $a(p)$ . As above we sieve the interval  $[s, s+x]$  for varying values of  $s \in [-x/2, x/2]$  and select the best result, but unlike the shifted Schinzel sieve, for large  $k$  we typically choose  $s \approx -(k/\log k - k)/2$ .

We remark that while one might suppose that it would be better to choose a minimally occupied residue class at all primes, not just the larger ones, we find that this is generally not the case. Fixing a structured choice of residue classes for the small primes avoids the erratic behavior that can result from making greedy choices too soon (see [28, Fig. 1] for an illustration of this).

Table 4 lists the bounds obtained by applying each of these techniques (in the online version of this paper, each table entry includes a link to the constructed tuple). To the admissible tuples obtained

**Table 4** Upper bounds on  $H(k)$  for selected values of  $k$ .

| $k$                            | 5511   | 35 410  | 41 588  | 309 661   | 1 649 821  |
|--------------------------------|--------|---------|---------|-----------|------------|
| $k$ primes past $k$            | 56 538 | 433 992 | 516 586 | 4 505 700 | 26 916 060 |
| Eratothenes                    | 55 160 | 424 636 | 505 734 | 4 430 212 | 26 540 720 |
| Hensley-Richards               | 54 480 | 415 642 | 494 866 | 4 312 612 | 25 841 884 |
| Shifted Schinzel               | 53 774 | 411 060 | 489 056 | 4 261 858 | 25 541 910 |
| Shifted greedy                 | 52 296 | 399 936 | 476 028 | 4 142 780 | 24 798 306 |
| Best known                     | 52 116 | 398 130 | 474 266 | 4 137 854 | 24 797 814 |
| $\lfloor k \log k + k \rfloor$ | 52 985 | 406 320 | 483 899 | 4 224 777 | 25 268 951 |

using the shifted greedy sieve we additionally applied various local optimizations that are detailed in [52, §3.6]. As can be seen in the table, the additional improvement due to these local optimizations is quite small compared to that gained by using better sieving algorithms, especially when  $k$  is large.

Table 4 also lists the value  $\lfloor k \log k + k \rfloor$  that we conjecture as an upper bound on  $H(k)$  for all sufficiently large  $k$ .

### 10.3 $H(k)$ bounds for large $k$

. The upper bounds on  $H(k)$  for the last two cases (10) and (11) were obtained using modified versions of the techniques described above that are better suited to handling very large values of  $k$ . These entail three types of optimizations that are summarized in the subsections below.

#### 10.3.1 Improved time complexity

As noted above, the complexity of admissibility testing is quasi-quadratic in  $k$ . Each of the techniques listed in §10.2 involves optimizing over a parameter space whose size is at least quasi-linear in  $k$ , leading to an overall quasi-cubic time complexity for constructing a narrow admissible  $k$ -tuple; this makes it impractical to handle  $k > 10^9$ . We can reduce this complexity in a number of ways.

First, we can combine parameter optimization and admissibility testing. In both the sieve of Eratosthenes and Hensley-Richards sieves, taking  $m = k$  guarantees an admissible  $k$ -tuple. For  $m < k$ , if the corresponding  $k$ -tuple is inadmissible, it is typically because it is inadmissible modulo the smallest prime  $p_{m+1}$  that appears in the tuple. This suggests a heuristic approach in which we start with  $m = k$ , and then iteratively reduce  $m$ , testing the admissibility of each  $k$ -tuple modulo  $p_{m+1}$  as we go, until we can proceed no further. We then verify that the last  $k$ -tuple that was admissible modulo  $p_{m+1}$  is also admissible modulo all primes  $p > p_{m+1}$  (we know it is admissible at all primes  $p \leq p_m$  because we have sieved a residue class for each of these primes). We expect this to be the case, but if not we can increase  $m$  as required. Heuristically this yields a quasi-quadratic running time, and in practice it takes less time to find the minimal  $m$  than it does to verify the admissibility of the resulting  $k$ -tuple.

Second, we can avoid a complete search of the parameter space. In the case of the shifted Schinzel sieve, for example, we find empirically that taking  $s = k$  typically yields an admissible  $k$ -tuple whose diameter is not much larger than that achieved by an optimal choice of  $s$ ; we can then simply focus on optimizing  $m$  using the strategy described above. Similar comments apply to the shifted greedy sieve.

#### 10.3.2 Improved space complexity

We expect a narrow admissible  $k$ -tuple to have diameter  $d = (1 + o(1))k \log k$ . Whether we encode this tuple as a sequence of  $k$  integers, or as a bitmap of  $d + 1$  bits, as in the fast admissibility testing algorithm, we will need approximately  $k \log k$  bits. For  $k > 10^9$  this may be too large to conveniently fit in memory. We can reduce the space to  $O(k \log \log k)$  bits by encoding the  $k$ -tuple as a sequence of  $k - 1$  gaps; the average gap between consecutive entries has size  $\log k$  and can be encoded in  $O(\log \log k)$  bits. In practical terms, for the sequences we constructed almost all gaps can be encoded using a single 8-bit byte for each gap.

**Table 5** Upper bounds on  $H(k)$  for selected values of  $k$ .

| $k$                            | 75 845 707    | 3 473 955 908  |
|--------------------------------|---------------|----------------|
| $k$ primes past $k$            | 1 541 858 666 | 84 449 123 072 |
| Eratosthenes                   | 1 526 698 470 | 83 833 839 848 |
| Hensley-Richards               | 1 488 227 220 | 81 912 638 914 |
| Shifted Schinzel               | 1 467 584 468 | 80 761 835 464 |
| Shifted Greedy                 | 1 431 556 072 | not available  |
| Best known                     | 1 431 556 072 | 80 550 202 480 |
| $\lfloor k \log k + k \rfloor$ | 1 452 006 268 | 79 791 764 059 |

One can further reduce space by partitioning the sieving interval into windows. For the construction of our largest tuples, we used windows of size  $O(\sqrt{d})$  and converted to a gap-sequence representation only after sieving at all primes up to an  $O(\sqrt{d})$  bound.

### 10.3.3 Parallelization

With the exception of the greedy sieve, all the techniques described above are easily parallelized. The greedy sieve is more difficult to parallelize because the choice of a minimally occupied residue class modulo  $p$  depends on the set of survivors obtained after sieving modulo primes less than  $p$ . To address this issue we modified the greedy approach to work with batches of consecutive primes of size  $n$ , where  $n$  is a multiple of the number of parallel threads of execution. After sieving fixed residue classes modulo all small primes  $p < 2\sqrt{k \log k}$ , we determine minimally occupied residue classes for the next  $n$  primes in parallel, sieve these residue classes, and then proceed to the next batch of  $n$  primes.

In addition to the techniques described above, we also considered a modified Schinzel sieve in which we check admissibility modulo each successive prime  $p$  before sieving multiples of  $p$ , in order to verify that sieving modulo  $p$  is actually necessary. For values of  $p$  close to but slightly less than  $p_m$  it will often be the case that the set of survivors is already admissible modulo  $p$ , even though it does contain multiples of  $p$  (because some other residue class is unoccupied). As with the greedy sieve, when using this approach we sieve residue classes in batches of size  $n$  to facilitate parallelization.

### 10.3.4 Results for large $k$

Table 5 lists the bounds obtained for the two largest values of  $k$ . For  $k = 75\,845\,707$  the best results were obtained with a shifted greedy sieve that was modified for parallel execution as described above, using the fixed shift parameter  $s = -(k \log k - k)/2$ . A list of the sieved residue classes is available at [math.mit.edu/~drew/greedy\\_75845707\\_1431556072.txt](http://math.mit.edu/~drew/greedy_75845707_1431556072.txt).

This file contains values of  $k$ ,  $s$ ,  $d$ , and  $m$ , along with a list of prime indices  $n_i > m$  and residue classes  $r_i$  such that sieving the interval  $[s, s + d]$  of odd integers, multiples of  $p_n$  for  $1 < n \leq m$ , and at  $r_i$  modulo  $p_{n_i}$  yields an admissible  $k$ -tuple.

For  $k = 3\,473\,955\,908$  we did not attempt any form of greedy sieving due to practical limits on the time and computational resources available. The best results were obtained using a modified Schinzel sieve that avoids unnecessary sieving, as described above, using the fixed shift parameter  $s = k0$ . A list of the sieved residue classes is available at

[math.mit.edu/~drew/schinzel\\_3473955908\\_80550202480.txt](http://math.mit.edu/~drew/schinzel_3473955908_80550202480.txt).

This file contains values of  $k$ ,  $s$ ,  $d$ , and  $m$ , along with a list of prime indices  $n_i > m$  such that sieving the interval  $[s, s + d]$  of odd integers, multiples of  $p_n$  for  $1 < n \leq m$ , and multiples of  $p_{n_i}$  yields an admissible  $k$ -tuple.

Source code for our implementation is available at [math.mit.edu/~drew/ompadm\\_v0.5.tar](http://math.mit.edu/~drew/ompadm_v0.5.tar); this code can be used to verify the admissibility of both the tuples listed above.



## Acknowledgements

This paper is part of the *Polymath project*, which was launched by Timothy Gowers in February 2009 as an experiment to see if research mathematics could be conducted by a massive online collaboration. The current project (which was administered by Terence Tao) is the eighth project in this series, and this is the second paper arising from that project, after [52]. Further information on the Polymath project can be found on the web site [michaelnielsen.org/polymath1](http://michaelnielsen.org/polymath1). Information about this specific project may be found at

[michaelnielsen.org/polymath1/index.php?title=Bounded\\_gaps\\_between\\_primes](http://michaelnielsen.org/polymath1/index.php?title=Bounded_gaps_between_primes)

and a full list of participants and their grant acknowledgments may be found at

[michaelnielsen.org/polymath1/index.php?title=Polymath8\\_grant\\_acknowledgments](http://michaelnielsen.org/polymath1/index.php?title=Polymath8_grant_acknowledgments)

We thank Thomas Engelsma for supplying us with his data on narrow admissible tuples, and Henryk Iwaniec for useful suggestions. We also thank the anonymous referees for some suggestions in improving the content and exposition of the paper.

## References

1. J. Andersson, *Bounded prime gaps in short intervals*, preprint.
2. W. D. Banks, T. Freiberg, J. Maynard, *On limit points of the sequence of normalized prime gaps*, preprint.
3. W. D. Banks, T. Freiberg, C. L. Turnage-Butterbaugh, *Consecutive primes in tuples*, preprint.
4. J. Benatar, *The existence of small prime gaps in subsets of the integers*, preprint.
5. E. Bombieri, *Le Grand Crible dans la Théorie Analytique des Nombres*, *Astérisque* **18** (1987), (Seconde ed.).
6. E. Bombieri, *The asymptotic sieve*, *Rend. Accad. Naz.* **XL** (5) 1/2 (1975/76), 243–269 (1977).
7. E. Bombieri, J. Friedlander, H. Iwaniec, *Primes in arithmetic progressions to large moduli*, *Acta Math.* **156** (1986), no. 3–4, 203–251.
8. A. Castillo, C. Hall, R. J. Lemke Oliver, P. Pollack, L. Thompson, *Bounded gaps between primes in number fields and function fields*, preprint.
9. L. Chua, S. Park, G. D. Smith, *Bounded gaps between primes in special sequences*, preprint.
10. D. Clark, N. Jarvis, *Dense admissible sequences*, *Math. Comp.* **70** (2001), no. 236, 1713–1718.
11. P. Deligne, *La conjecture de Weil. II*, *Publications Mathématiques de l’IHÉS* **52** (1980), 137–252.
12. L. E. Dickson, *A new extension of Dirichlet’s theorem on prime numbers*, *Messenger of Mathematics* **33** (1904), 155–161.
13. P. D. T. A. Elliott, H. Halberstam, *A conjecture in prime number theory*, *Symp. Math.* **4** (1968), 59–72.
14. B. Farkas, J. Pintz, S. Révész, *On the optimal weight function in the Goldston-Pintz-Yıldırım method for finding small gaps between consecutive primes*, to appear in: *Paul Turán Memorial Volume: Number Theory, Analysis and Combinatorics*, de Gruyter, Berlin, 2013.
15. K. Ford, *On Bombieri’s asymptotic sieve*, *Trans. Amer. Math. Soc.* **357** (2005), 1663–1674.
16. É. Fouvry, *Autour du théorème de Bombieri-Vinogradov*, *Acta Math.* **152** (1984), no. 3–4, 219–244.
17. E. Fouvry, *Théorie de Brun-Titchmarsh: application au théorème de Fermat*, *Invent. Math.* **79** (1985), no. 2, 383–407.
18. T. Freiberg, *A note on the theorem of Maynard and Tao*, preprint.
19. J. Friedlander, A. Granville, *Relevance of the residue class to the abundance of primes*, *Proceedings of the Amalfi Conference on Analytic Number Theory* (Maiori, 1989), 95–103, Univ. Salerno, Salerno, 1992.
20. J. Friedlander, A. Granville, A. Hildebrand, H. Maier, *Oscillation theorems for primes in arithmetic progressions and for sifting functions*, *J. Amer. Math. Soc.* **4** (1991), no. 1, 25–86.
21. J. Friedlander, H. Iwaniec, *The polynomial  $X^2 + Y^4$  captures its primes*, *Ann. of Math.* (2) **148** (1998), no. 3, 945–1040.
22. É. Fouvry, H. Iwaniec, *Primes in arithmetic progressions*, *Acta Arith.* **42** (1983), no. 2, 197–218.
23. J. Friedlander, H. Iwaniec, *Opera del Cribro*, American Mathematical Society Colloquium Publications vol. 57, 2010.
24. P. X. Gallagher, *Bombieri’s mean value theorem*, *Mathematika* **15** (1968), 1–6.
25. D. Goldston, J. Pintz, C. Yıldırım, *Primes in tuples. I*, *Ann. of Math.* **170** (2009), no. 2, 819–862.
26. D. Goldston, S. Graham, J. Pintz, C. Yıldırım, *Small gaps between primes or almost primes*, *Trans. Amer. Math. Soc.* **361** (2009), no. 10, 5285–5330.
27. D. Goldston, C. Yıldırım, *Higher correlations of divisor sums related to primes. I. Triple correlations*, *Integers* **3** (2003), A5, 66 pp.
28. D. Gordon, G. Rodemich, *Dense admissible sets*, *Algorithmic number theory* (Portland, OR, 1998), 216–225, *Lecture Notes in Comput. Sci.*, 1423, Springer, Berlin, 1998.
29. A. Granville, *Bounded gaps between primes*, preprint.
30. G. H. Hardy, 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.
31. D. R. Heath-Brown, *Prime numbers in short intervals and a generalized Vaughan identity*, *Canad. J. Math.* **34** (1982), no. 6, 1365–1377.
32. D. Hensley, I. Richards, *On the incompatibility of two conjectures concerning primes*, *Analytic number theory* (Proc. Sympos. Pure Math., Vol. XXIV, St. Louis Univ., St. Louis, Mo., 1972), pp. 123–127. Amer. Math. Soc., Providence, R.I., 1973.
33. D. Hensley, I. Richards, *Primes in intervals*, *Acta Arith.* **25** (1973/74), 375–391.
34. H. Iwaniec, E. Kowalski, *Analytic Number Theory*, American Mathematical Society Colloquium Publications Vol. 53, 2004.
35. H. Li, H. Pan, *Bounded gaps between primes of the special form*, preprint.
36. J. Maynard, *Bounded length intervals containing two primes and an almost-prime*, *Bull. Lond. Math. Soc.* **45** (2013), 753–764.
37. J. Maynard, *Bounded length intervals containing two primes and an almost-prime II*, preprint.
38. J. Maynard, *Small gaps between primes*, to appear, *Annals Math.*.
39. J. Maynard, *Dense clusters of primes in subsets*, preprint.
40. H. L. Montgomery, *Topics in Multiplicative Number Theory*, volume 227 of *Lecture Notes in Math.* Springer, New York, 1971.
41. H. L. Montgomery, R. C. Vaughan, *Multiplicative Number Theory I. Classical Theory*, *Cambridge studies in advanced mathematics*, 2007.
42. Y. Motohashi, *An induction principle for the generalization of Bombieri’s Prime Number Theorem*, *Proc. Japan Acad.* **52** (1976), 273–275.
43. Y. Motohashi, J. Pintz, *A smoothed GPY sieve*, *Bull. Lond. Math. Soc.* **40** (2008), no. 2, 298–310.
44. J. Pintz, *Are there arbitrarily long arithmetic progressions in the sequence of twin primes?*, *An irregular mind. Szemerédi is 70*. Bolyai Soc. Math. Stud. **21**, Springer, 2010, pp. 525–559.
45. J. Pintz, *The bounded gap conjecture and bounds between consecutive Goldbach numbers*, *Acta Arith.* **155** (2012), no. 4, 397–405.
46. J. Pintz, *Polignac Numbers, Conjectures of Erdős on Gaps between Primes, Arithmetic Progressions in Primes, and the Bounded Gap Conjecture*, preprint.
47. J. Pintz, *A note on bounded gaps between primes*, preprint.
48. J. Pintz, *On the ratio of consecutive gaps between primes*, preprint.
49. J. Pintz, *On the distribution of gaps between consecutive primes*, preprint.
50. P. Pollack, *Bounded gaps between primes with a given primitive root*, preprint.
51. P. Pollack, L. Thompson, *Arithmetic functions at consecutive shifted primes*, preprint.
52. D. H. J. Polymath, *New equidistribution estimates of Zhang type*, submitted.



53. D. H. J. Polymath, *New equidistribution estimates of Zhang type, and bounded gaps between primes*, unpublished at [arxiv.org/abs/1402.0811v2](https://arxiv.org/abs/1402.0811v2).
54. I. Richards, *On the incompatibility of two conjectures concerning primes; a discussion of the use of computers in attacking a theoretical problem*, Bull. Amer. Math. Soc. 80 (1974), 419–438.
55. A. Schinzel, *Remarks on the paper “Sur certaines hypothèses concernant les nombres premiers”*, Acta Arith. 7 (1961/1962) 1–8.
56. A. Schönhage, V. Strassen, *Schnelle Multiplikation großer Zahlen*, Computing (Arch. Elektron. Rechnen) 7 (1971), 281–292.
57. A. Selberg, *On elementary methods in prime number-theory and their limitations*, in Proc. 11th Scand. Math. Cong. Trondheim (1949), Collected Works, Vol. I, 388–397, Springer-Verlag, Berlin-Göttingen-Heidelberg, 1989.
58. H. Siebert, *Einige Analoga zum Satz von Siegel-Walfisz*, in: Zahlentheorie (Tagung, Math. Forschungsinst., Oberwolfach, 1970), Bibliographisches Inst., Mannheim, 1971, 173–184.
59. K. Soundararajan, *Small gaps between prime numbers: the work of Goldston-Pintz-Yıldırım*, Bull. Amer. Math. Soc. (N.S.) 44 (2007), no. 1, 1–18.
60. J. Thorner, *Bounded Gaps Between Primes in Chebotarev Sets*, preprint.
61. T. S. Trudgian, *A poor man’s improvement on Zhang’s result: there are infinitely many prime gaps less than 60 million*, preprint.
62. R. C. Vaughan, *Sommes trigonométriques sur les nombres premiers*, C. R. Acad. Sci. Paris Sér. A 285 (1977), 981–983.
63. A. I. Vinogradov, *The density hypothesis for Dirichlet L-series*, Izv. Akad. Nauk SSSR Ser. Mat. (in Russian) 29 (1956), 903–934.
64. A. Weil, *Sur les courbes algébriques et les variétés qui s’en déduisent*, Actualités Sci. Ind. 1041, Hermann, 1948.
65. Y. Zhang, *Bounded gaps between primes*, Annals Math 179 (2014), 1121–1174.