Lean formalization of bounded gaps between primes

A unified blueprint of the works of James Maynard and Polymath8b

Evan Chen, Sidharth Hariharan, Kenny Lau, Bhavik Mehta,
Ashvin A. Swaminathan, Jesse Thorner, Edison Xie

Abstract

Let p_n denote the n-th prime. We blueprint the formalization by AxiomProver of the bounds

$$\liminf_{n\to\infty}(p_{n+1}-p_n)\le 600, \qquad \liminf_{n\to\infty}(p_{n+1}-p_n)\le 246$$

proved by James Maynard and Polymath8b, respectively. Both results follow from a single uniform treatment of the Polymath8b generalization of James Maynard's modification of the sieve-theoretic work of Goldston, Pintz, and Yıldırım. Our formalization assumes the Bombieri–Vinogradov theorem and no other results.

Contents

1 Introduction

1.1 The main results

The long-standing **twin prime conjecture** asserts that, if p_n is the n-th prime, then

$$\liminf_{n\to\infty}(p_{n+1} - p_n) = 2.$$

One of many reasons for the difficulty of the conjecture is the fact that it imposes additive structure on primes, which are defined multiplicatively. The twin prime conjecture is a special case of a more general conjecture, the **Hardy–Littlewood k-tuples conjecture**, which we now state. Suppose that $\mathcal{H}_k = \{h_1, \dots, h_k\}$ is an *admissible set* (Definition 3.15). The conjecture then states that there are infinitely many positive integers n such that each of the numbers $n + h_1, \dots, n + h_k$ is prime. Since $\{0, 2\}$ is an admissible set, the Hardy–Littlewood k-tuples conjecture implies the twin prime conjecture.

The prime number theorem implies that as $n \to \infty$, the average value of $p_{n+1} - p_n$ is asymptotic to $\log p_n$. Foundational work of Goldston, Pintz, and Yıldırım developed the "GPY method," which led to the proof that

$$\liminf_{n\to\infty} \frac{p_{n+1} - p_n}{\log p_n} = 0.$$

Their method crucially relies on the distribution of primes in arithmetic progressions given by the **Bombieri–Vinogradov theorem**: If $0 < \varepsilon < \frac{1}{2}$ and $A > 0$ are fixed, then for all $x \geq 2$,

$$\sum_{q \leq x^{1/2-\varepsilon}} \max_{\gcd(a,q)=1} \left|\pi(x; q, a) - \frac{\pi(x)}{\varphi(q)}\right| \ll_A \frac{x}{(\log x)^A}.$$

The **Elliott–Halberstam conjecture** asserts that $x^{1/2-\varepsilon}$ can be replaced with $x^{1-\varepsilon}$. If there exists a fixed $\delta > 0$ such that $1/2 - \varepsilon$ can be improved to $1/2 + \delta$, then the GPY method suffices to establish the infinitude of *bounded gaps between primes*, namely

$$\liminf_{n\to\infty}(p_{n+1} - p_n) < \infty.$$

In May 2013, Zhang established deep new results on the distribution of primes in arithmetic progressions, a variant of the Bombieri–Vinogradov theorem. When combined with a modification of the GPY method, it yields for the first time the existence of infinitely many bounded gaps between primes:

$$\liminf_{n\to\infty}(p_{n+1} - p_n) \leq 70\,000\,000.$$

The work of Zhang was meticulously refined by the Polymath 8a project, the net result of which was the bound

$$\liminf_{n\to\infty}(p_{n+1}-p_n)\leq 4\,680.$$

In November 2013, the breakthrough work of Maynard[1] established a substantially more robust version of the GPY method that leads to significant theoretical and numerical improvements over the work of Zhang using only the original Bombieri–Vinogradov theorem. Writing

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

Maynard proved that $H_m \ll m^3 e^{4m}$, and in particular that $H_1 \leq 600$. Shortly thereafter, the Polymath8b project theoretically and numerically refined these ideas, obtaining $H_m \ll me^{(4-\frac{28}{157})m}$ and in particular $H_1 \leq 246$.

This blueprint develops both explicit bounds from a single multidimensional sieve. The only result that is assumed is the Bombieri–Vinogradov theorem.

Theorem (First main result)**.** *Assume the Bombieri–Vinogradov theorem. Then $H_1 \leq 600$; that is,*

$$\liminf_{n\to\infty}(p_{n+1}-p_n)\leq 600.$$

Theorem (Second main result)**.** *Assume the Bombieri–Vinogradov theorem. Then $H_1 \leq 246$; that is,*

$$\liminf_{n\to\infty}(p_{n+1}-p_n)\leq 246.$$

1.2 Blueprinting bounded gaps

This blueprint is designed to facilitate a formalisation of the bounds $H_1 \leq 600$ and $H_1 \leq 246$. It exhibits several key differences from a typical research article, some of which we now outline. For convenience, we call lemmas, sublemmas, propositions, and theorems **provable statements**; a definition or notation is a **declaration**; and an **entity** is a provable statement or a declaration. Given an entity, all other entities that are referenced in the statement (resp. proof) of the entity are catalogued in a `uses-notation` (resp. `uses-proof`) environment. A **uses statement** is a `uses-notation` or `uses-proof` environment.

1. Each declaration defines exactly one object. Within each declaration environment, only one declaration is introduced. No declaration environment contains any provable statement. Conversely, a provable statement only contains statements to be proved; no declarations are introduced in any provable statement. When recalling a declaration, only declarations are referenced, not provable statements.

2. Each provable statement has a minimal set of hypotheses under which exactly one stated conclusion holds. Multiple conclusions that might be bundled together must be separated into smaller provable statements.

3. Provable statements are decomposed into smaller, "atomic" provable statements, each of which will ideally correspond with a single formal declaration.

4. The dependency graph of our proof is directed and absent of any cycles. The edges of this graph are indicated by our uses statements. An entity may only cite entities that come before; no "forward-referencing" is permitted. A uses statement associated to a declaration cannot cite a provable statement.

[1] Around the same time, Tao arrived at similar, slightly weaker conclusions.

5. Apart from the main target theorems, every entity must be cited by at least one uses statement.

6. Every reference inside a provable statement or its proof must appear in the uses statement of said provable statement.

7. Superfluous or duplicated entities are deleted. In particular, there are no `remark` or `corollary` environments.

8. Imported or external results are stated once in a dedicated "external inputs" section and explicitly cited at least once in a uses statement.

9. The exposition is as "linear" as feasible, which we loosely describe as follows.

   (a) Declarations come first.

   (b) Lemmas and sublemmas come next; they incorporate the declarations that came before. Lemmas are not proved before the entities in their uses statements are given (and proved, if necessary).

   (c) Propositions are then proved using the lemmas and declarations. Propositions are not proved before the entities in their uses statements are given (and proved, if necessary).

   (d) The main theorems are then proved using all pertinent entities that precede them. A main theorem is not proved before the entities in its uses statement are given (and proved, if necessary).

2 Overview

We give a blueprint of the proofs of the two bounds of Section 1, in a form meant to be formalized in Lean. This section outlines the method; the rest of the paper carries it out. Throughout, the only analytic input is the Bombieri–Vinogradov theorem.

2.1 The GPY method and the sum $S_2(\lambda) - \rho S_1(\lambda)$

Fix an admissible k-tuple $\mathcal{H} = \{h_1, \dots, h_k\}$ (Definition 3.15); we seek infinitely many n for which at least $\lfloor \rho \rfloor + 1$ of $n + h_1, \dots, n + h_k$ are prime. Following Goldston–Pintz–Yıldırım, introduce weights $w_n \geq 0$ and set

$$S(N, \rho) \;:=\; \sum_{N < n \leq 2N} \Big(\sum_{i=1}^{k} \chi_{\mathbb{P}}(n + h_i) - \rho \Big) w_n,$$

with $\chi_{\mathbb{P}}$ the characteristic function of the primes. If $S(N, \rho) > 0$ for all large N, then infinitely many n have at least $\lfloor \rho \rfloor + 1$ of the $n + h_i$ prime: some summand is positive, and since $w_n \geq 0$ that n satisfies $\sum_i \chi_{\mathbb{P}}(n + h_i) > \rho$. Writing

$$S_1(\lambda) = \sum_{N < n \leq 2N} w_n, \qquad S_2(\lambda) = \sum_{N < n \leq 2N} \Big(\sum_{i} \chi_{\mathbb{P}}(n + h_i) \Big) w_n,$$

we have

$$S(N, \rho) = S_2(\lambda) - \rho S_1(\lambda).$$

One chooses w_n to make $S_2(\lambda)/S_1(\lambda)$ large.

2.2 The sieve weights and the W-trick

Maynard's weights are

$$w_n \;=\; \Big(\sum_{d_i \mid n+h_i \;\forall i} \lambda(d_1,\dots,d_k)\Big)^2, \qquad \lambda : \mathbb{N}^k \to \mathbb{R},$$

supported on $\prod_i d_i \le R$, where R is a truncation parameter below the level of distribution ϑ. The one-dimensional choice $\lambda(d_1,\dots,d_k) = \mu(d)\,F\big(\log(R/d)\big)$ with $d = d_1\cdots d_k$ recovers the Selberg weights of Goldston–Pintz–Yıldırım; for these the level $\vartheta < 1/2$ does not suffice to make $S_2(\lambda) - \rho S_1(\lambda) > 0$ for any $\rho \ge 1$, and it is the extra freedom of the k-dimensional λ that removes this obstruction. The $\lambda(d_1,\dots,d_k)$ are built from a fixed profile F on a subset of $\mathbb{R}^k$; the subset distinguishes the two bounds.

To handle small primes, we restrict w_n to be nonzero only for n in a single residue class $v_0 \pmod W$, where $W = \prod_{p \le D_0} p$ and $D_0 = \log\log\log N$. By admissibility of $\mathcal{H}$ and the Chinese remainder theorem there is a v_0 with $v_0 + h_i$ coprime to W for every i, so no prime $p \le D_0$ divides any $n + h_i$.

2.3 The variational problem and the two cases

Write $\mathcal{R}_k(s) := \{t \in \mathbb{R}^k_{\ge 0} : t_1 + \cdots + t_k \le s\}$ for the standard simplex scaled to total mass s, and $\mathcal{R}_k := \mathcal{R}_k(1)$. Fix $\varepsilon \ge 0$ and take F supported on $\mathcal{R}_k(1+\varepsilon)$. With W the primorial of the W-trick, the sieve sums satisfy

$$\begin{aligned} S_1(\lambda) \;&=\; \big(1+o(1)\big)\,\frac{\varphi(W)^k}{W^{k+1}}\,N\,(\log R)^k\,I_k(F), \\ S_2(\lambda) \;&=\; \big(1+o(1)\big)\,\frac{\log R}{\log N}\,\frac{\varphi(W)^k}{W^{k+1}}\,N\,(\log R)^k \sum_{m=1}^{k} J^{(m)}_{k,\varepsilon}(F), \end{aligned}$$

where $I_k(F) = \int_{\mathcal{R}_k(1+\varepsilon)} F^2$ and $J^{(m)}_{k,\varepsilon}(F)$ is the L^2 mass of the mth marginal $\int F\,dt_m$ over the slice $\mathcal{R}_{k-1}(1-\varepsilon)$ (Section 5 gives the precise definitions). Since $\log R = (\vartheta/2 - \delta)\log N$,

$$\frac{S_2(\lambda)}{S_1(\lambda)} \;\sim\; \Big(\frac{\vartheta}{2} - \delta\Big) M_{k,\varepsilon}, \qquad M_{k,\varepsilon} := \sup_F \frac{\sum_{m=1}^{k} J^{(m)}_{k,\varepsilon}(F)}{I_k(F)}.$$

Provided ϑ and ε satisfy a compatibility condition, $S(N,\rho)$ is eventually positive when $\rho < (\vartheta/2)\,M_{k,\varepsilon}$; two of the $n + h_i$ are then prime for infinitely many n, giving a gap at most the diameter of $\mathcal{H}$, once $M_{k,\varepsilon} > 2/\vartheta$.

The two bounds are the cases $\varepsilon = 0$ and $\varepsilon = 1/25$.

I. $\varepsilon = 0$**,** $k = 105$**.** Here $\mathcal{R}_k(1+\varepsilon) = \mathcal{R}_k$ and $M_{105,0} > 4$. Since Bombieri–Vinogradov gives every $\vartheta < 1/2$ and $4 > 2/\vartheta$ for ϑ near $1/2$, the criterion holds; with an admissible 105-tuple of diameter 600 this gives $H_1 \le 600$.

II. $\varepsilon = 1/25$**,** $k = 50$**.** Enlarging the support to $\mathcal{R}_k(1+\varepsilon)$ raises the supremum to $M_{50,1/25} > 4.0043$, exhibited by an explicit F from a finite basis of symmetric polynomials, for which the ratio is a quotient of rational quadratic forms verified by the certificate of Section 13. With an admissible 50-tuple of diameter 246 this gives $H_1 \le 246$.

The first case is unconditional given Bombieri–Vinogradov. The second also requires the finite numerical inequality $M_{50,1/25} > 4.0043$ (Section 13), which we prove by exact computation.

3 Definitions and preliminaries

This section collects the objects and standard facts used throughout: the classical arithmetic functions, admissible sets, the level of distribution and the two external inputs (Bombieri–Vinogradov and the prime number theorem), the W-trick, and the truncation parameter R with the support condition it imposes on the sieve weights.

3.1 Arithmetic functions and standard notation

Notation 3.1 (Standard symbols)**.** We write $\mathbb{N} = \{1, 2, 3, \dots\}$ for the positive integers and $\mathbb{P} = \{2, 3, 5, \dots\}$ for the primes. For integers a, b we write $(a, b) = \gcd(a, b)$ and $[a, b] = \operatorname{lcm}(a, b)$, and we call a and b *coprime* when $(a, b) = 1$. For real x we write $\lfloor x \rfloor$ and $\lceil x \rceil$ for the floor and the ceiling. Finally, $\chi_{\mathbb{P}} : \mathbb{N} \to \{0, 1\}$ denotes the indicator function of the primes, so that $\chi_{\mathbb{P}}(n) = 1$ if $n \in \mathbb{P}$ and $\chi_{\mathbb{P}}(n) = 0$ otherwise.

Lean:
- `PrimeGaps.primeIndicator`

Uses(notation): $\mathbb{N}$, $\mathbb{P}$, (a, b), $[a, b]$, $\lfloor x \rfloor$, $\lceil x \rceil$, $\chi_{\mathbb{P}}$ (index)

Notation 3.2 (Asymptotic notation)**.** For real-valued functions f, g, the statement $f \ll g$ means $|f| \le C|g|$ for some constant C, which may depend on the tuple length k and on $\mathcal{H}$ (both fixed throughout) but never on N. A subscript, as in $f \ll_A g$, records an additional permitted dependence of C. We write $f = O(g)$ for $f \ll g$, and $f = o(g)$ when $f/g \to 0$ as $N \to \infty$.

Lean:
- `Asymptotics.IsLittleO`
- `Asymptotics.IsBigO`

Uses(notation): $f \ll g$, $O(g)$, $o(g)$, k, $\mathcal{H}$, N (index)

Definition 3.3 (Möbius function μ)**.** The *Möbius function* $\mu : \mathbb{N} \to \{-1, 0, 1\}$ is defined by

$$\mu(n) := \begin{cases} 1 & \text{if } n = 1, \\ (-1)^r & \text{if } n = p_1 p_2 \cdots p_r \text{ is a product of } r \text{ distinct primes}, \\ 0 & \text{if } p^2 \mid n \text{ for some prime } p. \end{cases}$$

Lean:
- `ArithmeticFunction.moebius`

Uses(notation): μ, $\mathbb{N}$, $\mathbb{P}$ (index)

Definition 3.4 (Euler totient φ)**.** The *Euler totient function* $\varphi : \mathbb{N} \to \mathbb{N}$ is defined by

$$\varphi(n) := \#\{a \in \{1, 2, \dots, n\} : (a, n) = 1\}.$$

Lean:
- `ArithmeticFunction.totient`

Uses(notation): φ, $\mathbb{N}$, (a, b) (index)

Definition 3.5 (r-fold divisor function τ_r)**.** For an integer $r \ge 0$, the *r-fold divisor function* $\tau_r : \mathbb{N} \to \mathbb{N}$ counts ordered factorisations into r positive-integer factors:

$$\tau_r(n) := \#\{(d_1, d_2, \dots, d_r) \in \mathbb{N}^r : d_1 d_2 \cdots d_r = n\}.$$

We abbreviate $\tau := \tau_2$, the ordinary divisor-counting function $\tau(n) = \#\{d \ge 1 : d \mid n\}$.

Lean:

- `ArithmeticFunction.termτ_`
- `ArithmeticFunction.tau_apply_eq_card_finMulAntidiag`

Uses(notation): τ_r, τ, $\mathbb{N}$ (index)

Definition 3.6 (The function g)**.** Let $g : \mathbb{N} \to \mathbb{N}$ be the multiplicative function determined on prime powers by

$$g(p) \;:=\; p-2, \qquad g(p^a) \;:=\; p^{a-2}(p-1)^2 \quad (a \geq 2),$$

for every prime p, together with $g(1) = 1$.

Lean:

- `ArithmeticFunction.detotient`

Uses(notation): g, $\mathbb{N}$, $\mathbb{P}$ (index)

Lemma 3.7 (Möbius divisor-sum identity)**.** *For every positive integer s,*

$$\sum_{t|s} \mu(t) \;=\; \mathbf{1}[s=1].$$

Lean:

- `ArithmeticFunction.sum_divisors_moebius`
- `ArithmeticFunction.sum_divisors_coe_moebius`

Uses (statement): 3.3

Uses(notation): μ, $\mathbb{P}$ (index)

Proof. If $s = 1$, the only divisor is $t = 1$ and $\mu(1) = 1$ by Definition 3.3, so the sum equals 1. Suppose $s > 1$ and let $p_1, \ldots, p_r$ $(r \geq 1)$ be the distinct primes dividing s. By Definition 3.3, $\mu(t) = 0$ unless t is squarefree, so only the 2^r divisors of $p_1 \cdots p_r$ contribute, and such a divisor $t = \prod_{i \in I} p_i$ contributes $(-1)^{|I|}$. Grouping by $|I|$ gives

$$\sum_{t|s} \mu(t) \;=\; \sum_{j=0}^{r} \binom{r}{j} (-1)^j \;=\; (1-1)^r \;=\; 0$$

by the binomial theorem, since $r \geq 1$. □

Lemma 3.8 (Submultiplicativity of τ_r)**.** *For every integer $r \geq 0$ and all positive integers m, n,*

$$\tau_r(mn) \;\leq\; \tau_r(m)\,\tau_r(n),$$

with equality when $(m, n) = 1$.

Lean:

- `ArithmeticFunction.isMultiplicative_tau`
- `ArithmeticFunction.isSubmultiplicative_tau`

Uses (statement): 3.5

Uses(notation): τ_r, (a, b), $\mathbb{N}$ (index)

Proof. By Definition 3.5, τ_r is the r-fold Dirichlet convolution of the constant function $\mathbf{1}$ with itself. Convolution preserves multiplicativity, whence $\tau_r(mn) = \tau_r(m)\tau_r(n)$ whenever $(m, n) = 1$; this gives the asserted equality.

For the inequality in general, it suffices to produce an injection from the ordered r-fold factorisations of mn into the pairs consisting of an ordered r-fold factorisation of m and one of n. Fix a factorisation $mn = d_1 \cdots d_r$. For each prime p distribute the exponent $v_p(d_i)$ between m and n by the greedy rule: assign to the m-part the first $\min(v_p(d_i), \text{remaining exponent of } p \text{ in } m)$ of the copies of p occurring in d_i, processing $i = 1, \ldots, r$ in order, and assign the rest to the n-part. This produces factorisations $m = a_1 \cdots a_r$ and $n = b_1 \cdots b_r$ with $d_i = a_i b_i$, and the original factorisation is recovered from the pair, so the assignment is injective. Hence $\tau_r(mn) \le \tau_r(m)\tau_r(n)$. □

Lemma 3.9 (Equality $\tau_r(n)^2 = \tau_{r^2}(n)$ on squarefree integers)**.** *For every integer $r \ge 0$ and every squarefree positive integer n,*

$$\tau_r(n)^2 \;=\; \tau_{r^2}(n).$$

Lean:

- `ArithmeticFunction.tau_sq_eq_tau_sq`

Uses (statement): 3.5

Uses(notation): τ_r, $\mathbb{P}$ (index)

Proof. Both sides are multiplicative in n by Lemma 3.8, so for squarefree $n = p_1 \cdots p_s$ with distinct primes p_i it suffices to verify the identity at a single prime argument. An ordered r-fold factorisation of a prime p has exactly one factor equal to p and all others equal to 1, so $\tau_r(p) = r$; likewise $\tau_{r^2}(p) = r^2$. Hence $\tau_r(p)^2 = r^2 = \tau_{r^2}(p)$, and multiplying over $p_1, \ldots, p_s$ gives $\tau_r(n)^2 = \tau_{r^2}(n)$. (The hypothesis that n be squarefree is essential: for $n = p^2$ one has $\tau_r(p^2)^2 = \binom{r+1}{2}^2 < \binom{r^2+1}{2} = \tau_{r^2}(p^2)$ in general.) □

Uses (proof): 3.8

Lemma 3.10 (Dirichlet convolution identity $\varphi = g * \mathbf{1}$)**.** *For every positive integer n,*

$$\varphi(n) \;=\; \sum_{u \mid n} g(u).$$

Lean:

- `Nat.totient_eq_sum_divisors_g`

Uses (statement): 3.6, 3.4

Uses(notation): φ, g, $\mathbb{N}$ (index)

Proof. Both sides are multiplicative functions of n: this is standard for φ, and the divisor sum of the multiplicative function g (Definition 3.6) is again multiplicative. It therefore suffices to compare the two sides at a prime power $n = p^a$ with $a \ge 1$.

For $a = 1$,

$$\sum_{u \mid p} g(u) \;=\; g(1) + g(p) \;=\; 1 + (p - 2) \;=\; p - 1 \;=\; \varphi(p).$$

For $a \geq 2$, using $g(p^j) = p^{j-2}(p-1)^2$ for $2 \leq j \leq a$,

$$\sum_{u \mid p^a} g(u) \;=\; 1 + (p-2) + (p-1)^2 \sum_{j=2}^{a} p^{j-2} \;=\; (p-1) + (p-1)^2 \cdot \frac{p^{a-1}-1}{p-1} \;=\; (p-1)\,p^{a-1},$$

which is $\varphi(p^a)$. Multiplicativity now gives the identity for every n. □

Lemma 3.11 (Euler product identity for $\sum_{d \mid n} \mu(d)/d$)**.** *For every positive integer n,*

$$\sum_{d \mid n} \frac{\mu(d)}{d} \;=\; \frac{\varphi(n)}{n} \;=\; \prod_{p \mid n} \Bigl(1 - \frac{1}{p}\Bigr).$$

Lean:

- `Nat.totient_div_self_eq_prod`
- `ArithmeticFunction.sum_moebius_div_self`

Uses (statement): 3.3, 3.4

Uses(notation): μ, φ, $\mathbb{P}$ (index)

Proof. The function $n \mapsto \sum_{d \mid n} \mu(d)/d$ is the Dirichlet convolution of the multiplicative function $d \mapsto \mu(d)/d$ with the constant function $\mathbf{1}$, hence multiplicative. Evaluating at a prime power p^a with $a \geq 1$ and using $\mu(p^j) = 0$ for $j \geq 2$ (Definition 3.3),

$$\sum_{d \mid p^a} \frac{\mu(d)}{d} \;=\; \frac{\mu(1)}{1} + \frac{\mu(p)}{p} \;=\; 1 - \frac{1}{p}.$$

Hence, writing $n = \prod_{p \mid n} p^{a_p}$, the divisor sum equals $\prod_{p \mid n}(1 - 1/p)$. On the other hand φ is multiplicative with $\varphi(p^a) = p^a(1 - 1/p)$ (Definition 3.4), so $\varphi(n)/n = \prod_{p \mid n}(1 - 1/p)$ as well. The two displayed equalities follow. □

Lemma 3.12 (Dirichlet expansion of $n/\varphi(n)$)**.** *For every positive integer n,*

$$\frac{n}{\varphi(n)} \;=\; \sum_{e \mid n} \frac{\mu(e)^2}{\varphi(e)}.$$

Lean:

- `ArithmeticFunction.sum_totient_inv`
- `ArithmeticFunction.sum_squarefree_totient_inv`
- `ArithmeticFunction.sum_moebius_sq_div_totient`

Uses (statement): 3.3, 3.4

Uses(notation): μ, φ, $\mathbb{N}$ (index)

Proof. Both sides are multiplicative in n; for the right-hand side this is because μ^2 and $1/\varphi$ are multiplicative and the divisor sum of a multiplicative function is multiplicative. It therefore suffices to compare at a prime power $n = p^a$ with $a \geq 1$. Since $\mu(p^j)^2 = 0$ for $j \geq 2$,

$$\sum_{e \mid p^a} \frac{\mu(e)^2}{\varphi(e)} \;=\; \frac{1}{\varphi(1)} + \frac{1}{\varphi(p)} \;=\; 1 + \frac{1}{p-1} \;=\; \frac{p}{p-1} \;=\; \frac{p^a}{\varphi(p^a)},$$

the last equality because $\varphi(p^a) = p^{a-1}(p-1)$. Multiplicativity gives the identity in general. □

Lemma 3.13 (Lower bound for φ in terms of τ)**.** *For every positive integer n,*

$$n \;\le\; \tau(n)\,\varphi(n), \qquad \textit{equivalently} \qquad \frac{n}{\tau(n)} \;\le\; \varphi(n).$$

Lean:

- `Nat.inv_d_totient_le`
- `ArithmeticFunction.le_tau₂_mul_totient`
- `ArithmeticFunction.div_tau₂_le_totient`

Uses (statement): 3.4, 3.5

Uses(notation): φ, τ, $\omega(n)$, $\mathbb{P}$ (index)

Proof. Write $n = \prod_{i=1}^{r} p_i^{a_i}$ with distinct primes p_i, so $r = \omega(n)$ and $\tau(n) = \prod_i (a_i + 1) \ge 2^r$. For every prime $p \ge 2$ we have $1 - 1/p \ge 1/2$, hence by Lemma 3.11

$$\frac{\varphi(n)}{n} \;=\; \prod_{i=1}^{r}\Big(1 - \frac{1}{p_i}\Big) \;\ge\; 2^{-r} \;\ge\; \frac{1}{\tau(n)}.$$

Multiplying through by $n\,\tau(n)$ gives $n \le \tau(n)\varphi(n)$. □

Uses (proof): 3.11

Lemma 3.14 (Union bound for finite sums)**.** *Let X be a finite index set, let $\{A_x\}_{x\in X}$ be finite subsets of an ambient set, and write $\mathbf{1}_A$ for the indicator function of A. Then, for every point y,*

$$\mathbf{1}_{\bigcup_{x\in X} A_x}(y) \;\le\; \sum_{x\in X} \mathbf{1}_{A_x}(y).$$

Proof. If $y \notin \bigcup_x A_x$ both sides are 0. If $y \in \bigcup_x A_x$, then $y \in A_{x_0}$ for at least one $x_0 \in X$, so the left-hand side equals 1 while the right-hand side is a sum of non-negative terms one of which equals 1; the inequality follows. □

Throughout the blueprint, Lemma 3.14 is applied in the following form. Given a sum $\Sigma = \sum_{t\in\mathcal{T}} c(t)$ of non-negative terms and a covering $\mathcal{T} = \bigcup_{x\in X} \mathcal{T}_x$ by finitely many sub-events, multiplying the displayed indicator inequality by $c(t) \ge 0$ and summing yields

$$\Sigma \;\le\; \sum_{x\in X} \sum_{t\in\mathcal{T}_x} c(t).$$

The phrase "apply the union bound over the choices of x" denotes this passage.

3.2 Admissible sets

Definition 3.15 (Admissible set)**.** Let $k \in \mathbb{N}$. A finite set $\mathcal{H} = \{h_1, \dots, h_k\} \subset \mathbb{Z}_{\ge 0}$ of k distinct non-negative integers is *admissible* if, for every prime p, there exists an integer a_p with

$$a_p \not\equiv h_i \pmod{p} \qquad \text{for every } i \in \{1, \dots, k\}.$$

Lean:

- `Finset.Admissible`

Uses(notation): $\mathcal{H}$, k, h_i, $\mathbb{N}$, $\mathbb{P}$ (index)

Lemma 3.16 (Cardinality characterisation of admissibility)**.** *A finite set $\mathcal{H} \subset \mathbb{Z}_{\geq 0}$ is admissible (Definition 3.15) if and only if*

$$\#\{h \bmod p : h \in \mathcal{H}\} \ < \ p \qquad \text{for every prime } p.$$

Lean:
- `PrimeGaps.lem_admissible_cardinality_char`

Uses (statement): 3.15

Uses(notation): $\mathcal{H}$, k, h_i, $\mathbb{P}$ (index)

Proof. Fix a prime p and set $S_p := \{h \bmod p : h \in \mathcal{H}\} \subseteq \mathbb{Z}/p\mathbb{Z}$. There exists a_p with $a_p \not\equiv h_i \pmod p$ for every i if and only if $S_p \neq \mathbb{Z}/p\mathbb{Z}$: given such an a_p, its residue class is missing from S_p; conversely, if $S_p \subsetneq \mathbb{Z}/p\mathbb{Z}$, any representative of a class not in S_p will do. Since $\mathbb{Z}/p\mathbb{Z}$ is finite of cardinality p, the condition $S_p \neq \mathbb{Z}/p\mathbb{Z}$ is equivalent to $\#S_p < p$. Applying this for every prime p gives the claim. □

Lemma 3.17 (Admissibility is a finite check)**.** *A finite set $\mathcal{H} \subset \mathbb{Z}_{\geq 0}$ is admissible (Definition 3.15) if and only if, for every prime $p \leq \#\mathcal{H}$, there exists $a < p$ with $h \not\equiv a \pmod p$ for every $h \in \mathcal{H}$.*

Lean:
- `Finset.admissible_iff_le_card`

Uses (statement): 3.15

Uses(notation): $\mathcal{H}$, k, h_i, $\mathbb{P}$ (index)

Proof. Admissibility trivially implies the stated condition, since the latter only restricts the primes $p \leq \#\mathcal{H}$ and, as in the proof of Lemma 3.16, one may always replace a missed residue by its representative in $\{0, 1, \ldots, p-1\}$.

Conversely, assume the condition and let p be an arbitrary prime. If $p \leq \#\mathcal{H}$ there is nothing to prove. If $p > \#\mathcal{H}$, then the image of $\mathcal{H}$ in $\mathbb{Z}/p\mathbb{Z}$ has at most $\#\mathcal{H} < p$ elements, so it is a proper subset of $\mathbb{Z}/p\mathbb{Z}$ and some residue class is missed; by Lemma 3.16 this is exactly what admissibility requires at p. Hence $\mathcal{H}$ is admissible. □

Uses (proof): 3.16

Lemma 3.17 reduces admissibility — a condition on all primes — to a finite verification, so the admissibility of an explicit tuple is decidable.

Definition 3.18 (Diameter of $\mathcal{H}$)**.** Let $\mathcal{H} = \{h_1, \ldots, h_k\}$ be a finite non-empty set of non-negative integers. The *diameter* of $\mathcal{H}$ is

$$\operatorname{diam}(\mathcal{H}) \ := \ \max_{1 \leq i,j \leq k} (h_i - h_j).$$

Lean:
- `Finset.diameter`

Uses(notation): $\operatorname{diam}(\mathcal{H})$, $\mathcal{H}$, h_i, k (index)

Lemma 3.19 (Diameter as largest minus smallest)**.** *Let $\mathcal{H} = \{h_1, \ldots, h_k\}$ be a finite non-empty set of non-negative integers, labelled so that $h_1 < h_2 < \cdots < h_k$. Then*

$$\operatorname{diam}(\mathcal{H}) \ = \ h_k - h_1.$$

Lean:
- `Finset.diameter_eq_max_sub_min`

Uses (statement): 3.18

Uses(notation): $\operatorname{diam}(\mathcal{H})$, $\mathcal{H}$, h_i, k (index)

Proof. The pair $(i,j)=(k,1)$ is admissible in the maximum of Definition 3.18 and contributes $h_k - h_1$, so $\operatorname{diam}(\mathcal{H}) \geq h_k - h_1$. Conversely, for any i,j we have $h_i \leq h_k$ and $h_j \geq h_1$, whence $h_i - h_j \leq h_k - h_1$; taking the maximum over i,j gives $\operatorname{diam}(\mathcal{H}) \leq h_k - h_1$. □

3.3 Prime counting and the level of distribution

Notation 3.20 (Prime-counting functions)**.** For real x, a modulus $q \in \mathbb{N}$, a residue a modulo q, and a subset $S \subseteq \mathbb{N}$, set

$$\pi_S(x) := \#\{p \leq x : p \in \mathbb{P},\ p \in S\}, \qquad \pi_S(x;q,a) := \#\{p \leq x : p \in \mathbb{P},\ p \in S,\ p \equiv a \pmod q\}.$$

When $S = \mathbb{N}$ we drop it from the notation and write $\pi(x)$ and $\pi(x;q,a)$; these are the classical prime-counting function and its analogue in an arithmetic progression. A real argument is always interpreted through its floor, so $\pi_S(x) = \pi_S(\lfloor x \rfloor)$.

Uses(notation): $\pi(x)$, $\pi(x;q,a)$, $\mathbb{P}$, $\mathbb{N}$, $\lfloor x \rfloor$ (index)

Definition 3.21 (Level of distribution)**.** Let $S \subseteq \mathbb{N}$, let $\vartheta \in \mathbb{R}$, and let $\Xi \in \mathbb{N}$ be an auxiliary modulus. We say that S *has level of distribution* ϑ *relative to* Ξ if for every $A \geq 1$ there is a constant $c = c(A) > 0$ such that, for every real $x \geq 3$,

$$\sum_{\substack{1 \leq q \leq x^{\vartheta} \\ (q,\Xi)=1}} \max_{\substack{a \bmod q \\ (a,q)=1}} \left| \pi_S(x;q,a) - \frac{\pi_S(x)}{\varphi(q)} \right| \leq \frac{c\,x}{(\log x)^A}.$$

The unadorned phrase "the primes have level of distribution ϑ" means that $\mathbb{P}$ has level of distribution ϑ relative to $\Xi = 1$.

Lean:
- `Nat.HasLevelOfDistribution`

Uses (statement): 3.4, 3.20

Uses(notation): $\pi(x;q,a)$, $\pi(x)$, φ, ϑ, (a,b), $\mathbb{N}$, $\mathbb{P}$ (index)

By Definition 3.21, the level of distribution depends on S only through its primes, so the hypothesis may be stated for a set S that is not itself a set of primes.

Lemma 3.22 (Insensitivity to intersecting with the primes)**.** *Let $S \subseteq \mathbb{N}$, let $\vartheta \in \mathbb{R}$ and let $\Xi \in \mathbb{N}$. Then $S \cap \mathbb{P}$ has level of distribution ϑ relative to Ξ if and only if S does.*

Uses (statement): 3.21

Uses(notation): ϑ, $\mathbb{P}$, $\mathbb{N}$ (index)

Proof. By Notation 3.20, both π_S and $\pi_S(\cdot\,;q,a)$ count only elements of S that are prime, so

$$\pi_{S\cap\mathbb{P}}(x) = \pi_S(x) \qquad \text{and} \qquad \pi_{S\cap\mathbb{P}}(x;q,a) = \pi_S(x;q,a)$$

for all x,q,a. The two instances of the inequality in Definition 3.21 are therefore literally the same, so each implies the other. □

Uses (proof): 3.20

Theorem 1 (Bombieri–Vinogradov)**.** *For every $\theta < 1/2$, the set $\mathbb{N}$ has level of distribution θ relative to $\Xi = 1$.*

Lean:
- `hasLevelOfDistribution_Prime`
- `BombieriVinogradov`

Uses (statement): 3.21

Uses(notation): ϑ, $\mathbb{N}$ (index)

Proof. This is the classical theorem of Bombieri and Vinogradov, taken as an external input; see Bombieri (1965), Vinogradov (1965), and the textbook treatment in Davenport's *Multiplicative Number Theory*. No proof is given here. □

Lemma 3.23 (Bombieri–Vinogradov for the primes)**.** *Assume Bombieri–Vinogradov (Theorem 1). Then for every $\theta < 1/2$ the set $\mathbb{P}$ of primes has level of distribution θ relative to $\Xi = 1$.*

Uses (statement): 1, 3.21

Uses(notation): ϑ, $\mathbb{P}$, $\mathbb{N}$ (index)

Proof. Fix $\theta < 1/2$. By Theorem 1 the set $\mathbb{N}$ has level of distribution θ relative to 1. Since $\mathbb{P} = \mathbb{N} \cap \mathbb{P}$, Lemma 3.22 applied with $S = \mathbb{N}$ transfers the property from $\mathbb{N}$ to $\mathbb{P}$. □

Uses (proof): 3.22

3.4 External analytic inputs

Two analytic results are quoted from the literature: Bombieri–Vinogradov (Theorem 1, above) and the prime number theorem, stated below. Bombieri–Vinogradov is a hypothesis of both main theorems and is never discharged. The prime number theorem is used once, to count the primes in the dyadic interval, and is quoted with a secondary term. The size bound for the primorial W uses Chebyshev-type bounds.

In a proof, "external input" means that the assertion is quoted from the literature and no proof is given here.

Theorem 2 (Prime number theorem with secondary term)**.** *There exist a constant $C > 0$ and an integer N_0 such that, for every integer $N \geq N_0$,*

$$\left|\pi(N) - \frac{N}{\log N}\right| \leq \frac{C\,N}{(\log N)^2}.$$

Uses (statement): 3.20

Uses(notation): $\pi(x)$, N, $O(g)$ (index)

Proof. This is the prime number theorem in the form $\pi(x) = x/\log x + O\big(x/(\log x)^2\big)$, equivalently $\psi(x) = x + O\big(x/\log x\big)$ after partial summation and the passage from ψ to θ to π. It is taken as an external input; no proof is given here. □

3.5 Prime counts in a dyadic interval

The sieve runs over the dyadic interval $N < n \le 2N$. The two quantities below count the primes in this window and record their discrepancy in arithmetic progressions.

Notation 3.24 (Prime-counting in $(N, 2N]$)**.**

$$X_N \;:=\; \sum_{N<n\le 2N} \chi_{\mathbb{P}}(n) \;=\; \#\{p \in \mathbb{P} : N < p \le 2N\}.$$

Lean:
- `Nat.primeCountingIoc`

Uses (statement): 3.1

Uses(notation): X_N, $\chi_{\mathbb{P}}$, N, $\mathbb{P}$ (index)

Lemma 3.25 (X_N as a difference of prime-counting functions)**.** *For every $N \ge 0$,*

$$X_N \;=\; \pi(2N) - \pi(N).$$

Lean:
- `PrimeGaps.primeCountingIoc_eq_primeCounting_diff`

Uses (statement): 3.24, 3.20

Uses(notation): X_N, $\chi_{\mathbb{P}}$, N, $\mathbb{P}$, $\pi(x)$ (index)

Proof. By Notation 3.20, $\pi(x) = \sum_{n\le x} \chi_{\mathbb{P}}(n)$, the primes $\le x$ being counted once each by the indicator $\chi_{\mathbb{P}}$. The set of primes $\le N$ is contained in the set of primes $\le 2N$, so the two counts differ by exactly the primes in $(N, 2N]$:

$$\pi(2N) - \pi(N) \;=\; \sum_{n\le 2N} \chi_{\mathbb{P}}(n) \;-\; \sum_{n\le N} \chi_{\mathbb{P}}(n) \;=\; \sum_{N<n\le 2N} \chi_{\mathbb{P}}(n) \;=\; X_N,$$

the last equality being Notation 3.24. □

Theorem 3 (Prime number theorem for $(N, 2N]$)**.** *There exist a constant $C > 0$ and an integer N_0 such that, for every integer $N \ge N_0$,*

$$\left| X_N \;-\; \frac{N}{\log N} \right| \;\le\; \frac{C\,N}{(\log N)^2}.$$

Lean:
- `PNT.primeCountingIoc_self_two_mul`

Uses (statement): 3.24

Uses(notation): X_N, N, $O(g)$, $\pi(x)$ (index)

Proof. By Lemma 3.25, $X_N = \pi(2N) - \pi(N)$. Let C_0 and N_0' be as furnished by Theorem 2, and take $N \ge \max(N_0', 2)$. Applying Theorem 2 at $2N$ and at N, and using $\log(2N) \ge \log N$ to weaken the error term at $2N$,

$$\left|\pi(2N) - \frac{2N}{\log(2N)}\right| \le \frac{2C_0N}{(\log N)^2}, \qquad \left|\pi(N) - \frac{N}{\log N}\right| \le \frac{C_0N}{(\log N)^2}.$$

It remains to compare the main terms. Since $\log(2N) = \log 2 + \log N$,

$$\frac{2N}{\log(2N)} - \frac{2N}{\log N} \;=\; -\frac{2N\log 2}{(\log 2 + \log N)\log N},$$

whose absolute value is at most $2N\log 2/(\log N)^2$ because $(\log 2 + \log N)\log N \geq (\log N)^2$. Combining, and writing

$$X_N - \frac{N}{\log N} = \Big(\pi(2N) - \frac{2N}{\log 2N}\Big) - \Big(\pi(N) - \frac{N}{\log N}\Big) + \Big(\frac{2N}{\log 2N} - \frac{2N}{\log N}\Big),$$

the triangle inequality gives the claim with $C = 3C_0 + 2\log 2$. □

Uses (proof): 3.25, 2

Notation 3.26 (Local error $E(N,q)$)**.** For $q \in \mathbb{N}$,

$$E(N,q) \;:=\; 1 \;+\; \sup_{(a,q)=1} \Big| \sum_{\substack{N<n\leq 2N \\ n\equiv a \pmod q}} \chi_{\mathbb{P}}(n) \;-\; \frac{1}{\varphi(q)} \sum_{N<n\leq 2N} \chi_{\mathbb{P}}(n) \Big|,$$

the supremum being over the residue classes a modulo q that are coprime to q. The additive 1 is a convenience: it makes $E(N,q) \geq 1$, so that $E(N,q)$ may be used as a multiplicative majorant without a separate case for a vanishing discrepancy.

Lean:
- `Nat.primeCountingIocError`

Uses (statement): 3.4, 3.24

Uses(notation): $E(N,q)$, N, φ, $\chi_{\mathbb{P}}$, $\mathbb{N}$, (a,b), X_N (index)

Lemma 3.27. *Assume Bombieri–Vinogradov (Theorem 1). Let $A > 0$ and $\varepsilon \in (0, 1/2)$ be fixed. Then there is a constant $C = C(A,\varepsilon) > 0$ such that, for every $N \geq 1$,*

$$\sum_{\substack{1\leq q\leq N \\ q<N^{1/2-\varepsilon}}} \mu(q)^2\, E(N,q) \;\leq\; \frac{C\,N}{(\log N)^A}.$$

Lean:
- `PrimeGaps.lem_BV_restated`

Uses (statement): 1, 3.26, 3.3

Uses(notation): N, μ, $E(N,q)$, φ, $\chi_{\mathbb{P}}$, (a,b), ϑ, X_N, $\pi(x)$, $\pi(x;q,a)$ (index)

Proof. Write $Q := \{q : 1 \leq q \leq N,\ q < N^{1/2-\varepsilon}\}$ for the range of moduli, and put $\theta := 1/2 - \varepsilon$, so that $0 < \theta < 1/2$ and $Q \subseteq \{q : 1 \leq q \leq N^{\theta}\}$.

Step 1: bound $E(N,q)$ by two cumulative discrepancies. Fix q and a residue a coprime to q. The window count $\sum_{N<n\leq 2N,\, n\equiv a} \chi_{\mathbb{P}}(n)$ equals $\pi(2N;q,a) - \pi(N;q,a)$, and $\sum_{N<n\leq 2N} \chi_{\mathbb{P}}(n) = X_N = \pi(2N) - \pi(N)$ by Lemma 3.25. Hence the quantity inside the supremum in Notation 3.26 equals

$$\Big(\pi(2N;q,a) - \frac{\pi(2N)}{\varphi(q)}\Big) - \Big(\pi(N;q,a) - \frac{\pi(N)}{\varphi(q)}\Big),$$

so by the triangle inequality

$$E(N,q) \;\le\; 1 \;+\; D(2N,q) \;+\; D(N,q), \qquad D(x,q) := \max_{(a,q)=1}\Big|\pi(x;q,a) - \frac{\pi(x)}{\varphi(q)}\Big|.$$

Step 2: sum over $q \in Q$. Since $\mu(q)^2 \le 1$, summing Step 1 over $q \in Q$ gives

$$\sum_{q\in Q}\mu(q)^2 E(N,q) \;\le\; \#Q \;+\; \sum_{q\in Q} D(2N,q) \;+\; \sum_{q\in Q} D(N,q),$$

and $\#Q \le N^{1/2-\varepsilon}$ by the definition of Q.

Step 3: apply the level-of-distribution bound. Set $A' := \max(A,1)$. Since $\theta < 1/2$, Lemma 3.23 says the primes have level of distribution θ relative to $\Xi = 1$, so Definition 3.21 supplies $c = c(A') > 0$ with $\sum_{q\le x^\theta} D(x,q) \le c\,x(\log x)^{-A'}$ for all $x \ge 3$. All summands $D(x,q)$ are non-negative, so restricting to $q \in Q \subseteq \{q \le x^\theta\}$ (valid at $x = N$ by the choice of θ, and at $x = 2N$ since $N^\theta \le (2N)^\theta$ for $\theta \ge 0$) yields, for $N \ge 3$,

$$\sum_{q\in Q} D(N,q) \le \frac{c\,N}{(\log N)^{A'}} \qquad \text{and} \qquad \sum_{q\in Q} D(2N,q) \le \frac{2c\,N}{(\log 2N)^{A'}} \le \frac{2c\,N}{(\log N)^{A'}}.$$

Moreover $\log N \ge 1$ for $N \ge 3$, so $(\log N)^{-A'} \le (\log N)^{-A}$.

Step 4: absorb the trivial term and the small N. The term $\#Q \le N^{1/2-\varepsilon}$ satisfies $N^{1/2-\varepsilon} \le KN(\log N)^{-A}$ for a constant $K = K(A,\varepsilon)$ and all $N \ge 2$: indeed $(\log N)^A \ll_{A,\varepsilon} N^{1/2+\varepsilon}$, and $N^{1/2-\varepsilon}\cdot N^{1/2+\varepsilon} = N$. Combining Steps 2–4 gives the asserted bound with a suitable constant for all N beyond an explicit threshold; enlarging the constant to accommodate the finitely many remaining $N \ge 1$ (for which the left-hand side is a finite sum, and is empty when $N = 1$) gives the lemma. □

Uses (proof): 3.25, 3.23, 3.21, 3.24

3.6 The W-trick

The primes are not equidistributed among the residue classes of a small modulus, and a sieve over $N < n \le 2N$ inherits this defect in every local factor. The W-trick removes the obstruction by restricting n to a single residue class v_0 modulo the product W of all small primes, chosen so that each shifted value $n + h_i$ is coprime to W. The cutoff defining "small" is a triple logarithm, large enough for the local factors to converge and small enough that W is negligible against any power of N.

Definition 3.28 (The primorial cutoff D_0)**.** For real N, set

$$D_0 \;:=\; \log\log\log N.$$

Lean:

- `PrimeGaps.D₀`

Uses(notation): D_0, N (index)

Definition 3.29 (The primorial W)**.** Let D_0 be as in Definition 3.28. Set

$$W \;:=\; \prod_{\substack{p\le\lfloor D_0\rfloor\\ p\in\mathbb{P}}} p,$$

the product of all primes not exceeding $\lfloor D_0\rfloor$ (an empty product, equal to 1, when $D_0 < 2$).

Lean:
- `Finset.CompatibleWith`
- `PrimeGaps.sieveModulus`

Uses (statement): 3.28

Uses(notation): W, D_0, N, $\mathbb{P}$, $\lfloor x \rfloor$ (index)

Lemma 3.30 (Prime divisors of W)**.** *For every prime p,*

$$p \mid W \iff p \le D_0.$$

Uses (statement): 3.29, 3.28

Uses(notation): W, D_0, $\mathbb{P}$, $\lfloor x \rfloor$ (index)

Proof. By Definition 3.29, W is a product of distinct primes, so a prime p divides W if and only if p is one of the factors, that is, if and only if $p \le \lfloor D_0 \rfloor$. Since p is an integer, $p \le \lfloor D_0 \rfloor$ is equivalent to $p \le D_0$. □

Lemma 3.31 (W is squarefree)**.** *W is squarefree.*

Uses (statement): 3.29

Uses(notation): W, $\mathbb{P}$, μ (index)

Proof. By Definition 3.29, W is a product of *distinct* primes, each occurring to the first power. Hence no prime square divides W, i.e. W is squarefree. □

Lemma 3.32 (Size of W)**.** *For all sufficiently large N,*

$$W \le (\log\log N)^2.$$

Lean:
- `PrimeGaps.lem_W_size`

Uses (statement): 3.29, 3.28

Uses(notation): W, N, D_0, $\mathbb{P}$, $\lfloor x \rfloor$ (index)

Proof. We use the elementary Chebyshev-type bound $\prod_{p \le n} p \le 4^n$, valid for every integer $n \ge 0$; it is proved by induction on n using the divisibility of $\binom{2m+1}{m}$ by all primes in $(m+1, 2m+1]$ together with $\binom{2m+1}{m} \le 4^m$. No appeal to the prime number theorem is required.

By Definition 3.29 and this bound,

$$W = \prod_{p \le \lfloor D_0 \rfloor} p \le 4^{\lfloor D_0 \rfloor} \le 4^{D_0}$$

whenever $D_0 \ge 0$. Now $D_0 = \log\log\log N$ by Definition 3.28, so $e^{D_0} = \log\log N$ and hence

$$4^{D_0} = e^{D_0 \log 4} = \left(e^{D_0}\right)^{\log 4} = (\log\log N)^{\log 4}.$$

Finally $\log 4 \le 2$, so once N is large enough that $\log\log N \ge 1$ — equivalently $N \ge e^e$, which also forces $D_0 \ge 0$ — we may raise the exponent from $\log 4$ to 2 and conclude $W \le (\log\log N)^2$. □

Definition 3.33 (Compatible residue class v_0)**.** Let $\mathcal{H} = \{h_1, \dots, h_k\}$ be a finite set of non-negative integers and let W be as in Definition 3.29. An integer v_0 is *compatible with $\mathcal{H}$ modulo W* if

$$(v_0 + h_i,\, W) \;=\; 1 \qquad \text{for every } i \in \{1, \dots, k\}.$$

Uses (statement): 3.29

Uses(notation): v_0, W, $\mathcal{H}$, h_i, k, (a, b) (index)

Lemma 3.34 (Existence of v_0)**.** *Let $\mathcal{H}$ be an admissible set (Definition 3.15) and let N be arbitrary. Then there exists v_0 with $0 \le v_0 < W$ that is compatible with $\mathcal{H}$ modulo W (Definition 3.33).*

Lean:

- `Finset.Admissible.exists_compatibleWith`

Uses (statement): 3.15, 3.29, 3.33

Uses(notation): v_0, $\mathcal{H}$, k, W, D_0, N, h_i, $\mathbb{P}$, (a, b) (index)

Proof. By Definition 3.29, W is a product of distinct primes $p \le \lfloor D_0 \rfloor$. For each such prime p, admissibility of $\mathcal{H}$ (Definition 3.15) supplies a residue a_p with $a_p \not\equiv h_i \pmod p$ for every i; equivalently, setting $b_p := -a_p$, we have $b_p + h_i \not\equiv 0 \pmod p$ for every i.

The primes dividing W are pairwise coprime, so by the Chinese remainder theorem there is an integer v with $v \equiv b_p \pmod p$ for every prime $p \mid W$. Put $v_0 := v \bmod W$, so $0 \le v_0 < W$ and $v_0 \equiv b_p \pmod p$ for each such p.

It remains to check compatibility. Fix i. Since W is squarefree (Lemma 3.31), $(v_0 + h_i, W) = 1$ holds if and only if no prime divisor of W divides $v_0 + h_i$; and by Lemma 3.30 the prime divisors of W are exactly the primes $p \le D_0$, which are the primes handled above. For such a p we have $v_0 + h_i \equiv b_p + h_i \not\equiv 0 \pmod p$ by the choice of b_p. Hence $(v_0 + h_i, W) = 1$ for every i, i.e. v_0 is compatible with $\mathcal{H}$ modulo W. □

Uses (proof): 3.30, 3.31

Notation 3.35 (Standing notation)**.** We fix throughout: an integer $k \ge 2$; an admissible set $\mathcal{H} = \{h_1, \dots, h_k\}$ with $h_1 < \cdots < h_k$; the parameters D_0 and W of Definitions 3.28 and 3.29; and a residue v_0 compatible with $\mathcal{H}$ modulo W (Definition 3.33). All constants implicit in $O(\cdot)$, $\ll$ and $o(\cdot)$ may depend on k and on $\mathcal{H}$, but never on N.

Uses (statement): 3.15, 3.28, 3.29, 3.33, 3.2

Uses(notation): k, $\mathcal{H}$, h_i, D_0, W, v_0, $O(g)$, $f \ll g$, $o(g)$, N (index)

3.7 The truncation parameter R and the sieve support

Definition 3.36 (The auxiliary parameter δ)**.** Let ϑ be a level of distribution (Definition 3.21). Throughout the blueprint, δ denotes a fixed real number in the open interval

$$\delta \;\in\; \big(0,\, \vartheta/2\big),$$

which may be chosen arbitrarily small. The proofs treat δ as a fixed positive constant, and the dependence of error terms on δ is suppressed in the asymptotic notation of Notation 3.2.

Uses (statement): 3.21, 3.2

Uses(notation): ϑ, δ (index)

Definition 3.37 (The truncation R)**.** Let δ be as in Definition 3.36. The *sieve truncation parameter* is

$$R \;:=\; N^{\vartheta/2-\delta}.$$

Lean:

- `PrimeGaps.sieveTruncation`
- `PrimeGaps.sieveTruncation.termR`

Uses (statement): 3.36

Uses(notation): R, N, ϑ, δ (index)

Lemma 3.38 (R^2W lies below $N^{1/2}$)**.** *Let ϑ and δ be as in Definitions 3.21 and 3.36, let R be as in Definition 3.37, and let η be a real number with*

$$\eta \;<\; \tfrac{1}{2} - (\vartheta - 2\delta).$$

Then, for all sufficiently large N,

$$R^2\, W \;\le\; N^{1/2-\eta}.$$

Lean:

- `PrimeGaps.sieve_support_le`

Uses (statement): 3.37, 3.29, 3.36, 3.21

Uses(notation): ϑ, δ, R, W, N, $o(g)$ (index)

Proof. Set $\gamma := \frac{1}{2} - \eta - (\vartheta - 2\delta)$, which is positive by hypothesis. By Definition 3.37,

$$R^2 \;=\; N^{\vartheta-2\delta},$$

so it suffices to show that $W \le N^{\gamma}$ for all large N, since then $R^2W \le N^{\vartheta-2\delta+\gamma} = N^{1/2-\eta}$.

By Lemma 3.32, $W \le (\log\log N)^2$ for all large N, and $\log\log N \le \log N$ for $N \ge e$. Hence $W \le (\log N)^2$ for all large N. Since $(\log N)^2 = o(N^{\gamma})$ for every $\gamma > 0$ — powers of the logarithm are dominated by every positive power of N — we obtain $W \le N^{\gamma}$ for all large N, as required. □

Uses (proof): 3.32

The support condition below is imposed on every sieve weight. It makes the divisor sums finite, keeps the moduli below the Bombieri–Vinogradov range through Lemma 3.38, and decouples the sieve from the small primes absorbed into W.

Definition 3.39 (Permissible support at a truncation level)**.** Let T be a positive integer, the *truncation level*. A function $\lambda : \mathbb{N}^k \to \mathbb{R}$ has *permissible support at level T* if $\lambda(d_1, \dots, d_k) = 0$ whenever one of the following three conditions fails:

(i) $\prod_{i=1}^{k} d_i \le T$;

(ii) $\left(\prod_{i=1}^{k} d_i,\, W\right) = 1$;

(iii) $\prod_{i=1}^{k} d_i$ is squarefree.

Two levels occur below. The development on the standard simplex takes $T = \lfloor R \rfloor$, and the development on the enlarged simplex takes $T = \lfloor R^{1+\varepsilon} \rfloor$; the enlargement buys exactly the longer divisors that the larger level admits. Where the level is clear from context we say simply *permissible support*.

Lean:
- `Finset.mem_permissibleSupport_iff`
- `Finset.permissibleSupport`
- `Finsupp.HasPermissibleSupport`

Uses (statement): 3.37, 3.29

Uses(notation): $\lambda(d_1, \dots, d_k)$, R, W, k, (a, b), $\mathbf{d}$, $\mathbb{N}$ (index)

Lemma 3.40 (Permissibility passes to divisors)**.** *Let* $\mathbf{r} = (r_1, \dots, r_k)$ *satisfy the three conditions (i)–(iii) of Definition 3.39, and let* $\mathbf{d} = (d_1, \dots, d_k)$ *satisfy* $d_i \mid r_i$ *for every* i*. Then* $\mathbf{d}$ *satisfies (i)–(iii) as well.*

Uses (statement): 3.39

Uses(notation): $\mathbf{d}$, $\mathbf{r}$, R, W, k, (a, b), $\mathbb{N}$ (index)

Proof. Since $d_i \mid r_i$ for every i, we have $\prod_i d_i \mid \prod_i r_i$. Because $\prod_i r_i$ is squarefree it is in particular non-zero, so $\prod_i d_i \le \prod_i r_i \le R$, giving (i). A divisor of an integer coprime to W is itself coprime to W, giving (ii), and a divisor of a squarefree integer is squarefree, giving (iii). □

Notation 3.41 (Singular series)**.** For a density $\gamma : \mathbb{N} \to \mathbb{R}$ the associated *singular series* is

$$\mathfrak{S}(\gamma) \;:=\; \prod_{p \in \mathbb{P}} \frac{1 - 1/p}{1 - \gamma(p)/p},$$

the product being taken over all primes.

Lean:
- `PrimeGaps.singularSeries`
- `PrimeGaps.term𝔖`
- `PrimeGaps.singularSeries_primorial`

Uses(notation): $\mathfrak{S}(\gamma)$, $\mathbb{P}$, γ_g (index)

4 The sieve datum, the sieve weights, and the auxiliary y-variables

4.1 The sieve datum

The sieve estimates below — the Mertens-type partial summations, the evaluation of the local densities, the singular products — depend on the arithmetic function γ_g only through a short list of properties: multiplicativity, a uniform gap between $\gamma(p)$ and p, closeness of $\gamma(p)$ to 1 away from a fixed squarefree modulus, and a two-sided Mertens bound on the weighted prime log-sums. We collect this list as a single object, the *sieve datum*, consisting of the density γ, the modulus V, and the four constants A_1, A_2, A_3, L. Every statement below is proved for an arbitrary sieve datum and applied twice: once to the W-tricked density γ_g, and once, through the same lemmas with a different modulus, in the r_i-coordinate analysis of $S_2^{(m)}$.

Definition 4.1 (Mertens deviation of a weight)**.** Let $f : \mathbb{N} \to \mathbb{R}$ and let $2 \le w \le z$ be real. The *Mertens deviation* of f on $[w, z]$ is

$$\Delta(f; w, z) \;:=\; \sum_{\substack{w \le p \le z \\ p \in \mathbb{P}}} \frac{f(p)\,\log p}{p} \;-\; \log(z/w).$$

Uses(notation): $\mathbb{P}$, $T(z)$, $\mathbb{N}$ (index)

Definition 4.2 (Density)**.** A *density* is a function $\gamma : \mathbb{N} \to \mathbb{R}$ such that

(i) $\gamma(n) \ge 0$ for every n;

(ii) $\gamma(1) = 1$;

(iii) $\gamma(mn) = \gamma(m)\,\gamma(n)$ whenever $(m, n) = 1$;

(iv) $\gamma(p) < p$ for every prime p.

Uses(notation): γ, $\mathbb{N}$, $\mathbb{P}$, (a, b) (index)

Definition 4.3 (Sieve datum)**.** A *sieve datum* is a tuple $\mathcal{S} = (\gamma, V, A_1, A_2, A_3, L)$ consisting of a function $\gamma : \mathbb{N} \to \mathbb{R}$, an integer $V \ge 1$, and real numbers A_1, A_2, A_3, L, subject to the following axioms.

(D1) *Density.* γ is a density in the sense of Definition 4.2.

(D2) *Growth.* $0 < A_1 < 1$ and $\gamma(p)/p \le 1 - A_1$ for every prime p.

(D3) *Modulus.* V is squarefree, and $\gamma(p) = 0$ for every prime $p \mid V$.

(D4) *Unit density off the modulus.* $A_3 \ge 0$ and $|\gamma(p) - 1| \le A_3/p$ for every prime $p \nmid V$.

(D5) *Mertens.* $A_2 > 0$, $L \ge 0$, and $-L \le \Delta(\gamma; w, z) \le A_2$ for all real $2 \le w \le z$, with Δ as in Definition 4.1.

We call γ the *density*, V the *modulus*, and A_1, A_2, A_3, L the *constants* of $\mathcal{S}$.

Uses (statement): 4.2, 4.1

Uses(notation): γ, V, A_1, A_2, A_3, L, $\mathbb{P}$, $\mathbb{N}$ (index)

Axiom (D1) is stated separately from the sieve datum because the same notion of density is also imposed level by level on the N-indexed family of W-tricked densities from which the constants A_2 and L are extracted; the two uses share one definition rather than duplicating four axioms. Note also that (D2) forces $\gamma(p) \le (1 - A_1)p$ *uniformly* in p, which is strictly stronger than the pointwise inequality $\gamma(p) < p$ of (D1)(iv); both are recorded because the weaker one is what the construction of g_* needs, and the stronger one is what the analytic estimates need.

Definition 4.4 (The derived multiplicative function g_*)**.** Let $\mathcal{S} = (\gamma, V, A_1, A_2, A_3, L)$ be a sieve datum (Definition 4.3). The *derived function* $g_* : \mathbb{N} \to \mathbb{R}$ is the totally multiplicative function determined by its prime values

$$g_*(p) \;:=\; \frac{\gamma(p)}{p - \gamma(p)};$$

explicitly, $g_*(n) := \prod_p g_*(p)^{v_p(n)}$, where $v_p(n)$ is the exponent of p in n, so that $g_*(1) = 1$.

Lean:
- `PrimeGaps.SieveDatum.gStar`

Uses (statement): 4.3

Uses(notation): g_*, γ, $\mathbb{P}$, $\mathbb{N}$ (index)

Lemma 4.5 (g_* is well defined and non-negative)**.** *Let $\mathcal{S}$ be a sieve datum and p a prime. Then $p - \gamma(p) > 0$, and $g_*(p) \geq 0$.*

Lean:
- `PrimeGaps.lem_g_star_well_defined`

Uses (statement): 4.3, 4.4

Uses(notation): g_*, γ, A_1, $\mathbb{P}$ (index)

Proof. Axiom (D1)(iv) of Definition 4.3 gives $\gamma(p) < p$, i.e. $p - \gamma(p) > 0$; hence the quotient defining $g_*(p)$ makes sense. (Axiom (D2) sharpens this to $p - \gamma(p) \geq A_1 p$, a bound uniform in p, which is what the analytic applications use.) Since $\gamma(p) \geq 0$ by (D1)(i) and the denominator is positive, $g_*(p) = \gamma(p)/(p - \gamma(p)) \geq 0$. □

Because g_* is totally multiplicative and non-negative at every prime, $g_*(n) \geq 0$ for every n, and on a squarefree n one has $g_*(n) = \prod_{p \mid n} g_*(p)$. We use both facts freely below.

Definition 4.6 (The convolution summand h)**.** Let $\mathcal{S}$ be a sieve datum with derived function g_* (Definition 4.4). Define $h : \mathbb{N} \to \mathbb{R}$ by

$$h(d) \;:=\; \mu(d)^2\, g_*(d).$$

Lean:
- `PrimeGaps.SieveDatum.h`

Uses (statement): 4.4, 3.3

Uses(notation): $h(d)$, g_*, μ, $\mathbb{N}$ (index)

Lemma 4.7 (h is supported on integers coprime to the modulus)**.** *Let $\mathcal{S} = (\gamma, V, A_1, A_2, A_3, L)$ be a sieve datum. If $d \geq 1$ satisfies $(d, V) > 1$, then $h(d) = 0$.*

Uses (statement): 4.3, 4.6, 4.4

Uses(notation): $h(d)$, g_*, V, (a, b), $\mathbb{N}$ (index)

Proof. Since $(d, V) > 1$, the integer (d, V) has a prime divisor p; then $p \mid d$ and $p \mid V$. By axiom (D3) of Definition 4.3, $\gamma(p) = 0$; by Lemma 4.5 the denominator $p - \gamma(p)$ is strictly positive, so Definition 4.4 gives $g_*(p) = 0/(p - \gamma(p)) = 0$. Write $d = p\,c$ with $c \geq 1$. Total multiplicativity of g_* gives $g_*(d) = g_*(p)\, g_*(c) = 0$, whence $h(d) = \mu(d)^2\, g_*(d) = 0$. □

Uses (proof): 4.5

Thus h is a non-negative arithmetic function supported on the squarefree integers coprime to V, and its partial sums are the input to the partial-summation lemma below. The next function records the deviation of g_* from the model density $1/(p-1)$ at each prime not dividing the modulus.

Definition 4.8 (The convolution defect b)**.** Let $\mathcal{S} = (\gamma, V, A_1, A_2, A_3, L)$ be a sieve datum with derived function g_*. Define $b : \mathbb{N} \to \mathbb{R}$ by

$$b(n) := \begin{cases} \displaystyle\prod_{p \mid n} \Big(g_*(p) - \frac{1}{p-1} \Big), & \text{if } n \text{ is squarefree and } (n, V) = 1, \\ 0, & \text{otherwise,} \end{cases}$$

the product being over the distinct primes dividing n (so $b(1) = 1$).

Lean:

- `PrimeGaps.SieveDatum.bDefect`

Uses (statement): 4.3, 4.4

Uses(notation): $b(p)$, g_*, V, (a, b), $\mathbb{N}$, $\mathbb{P}$ (index)

By construction b is multiplicative, vanishes off the squarefree integers coprime to V, and takes the value

$$b(p) \;=\; g_*(p) - \frac{1}{p-1} \;=\; \frac{\gamma(p)}{p - \gamma(p)} - \frac{1}{p-1} \;=\; \frac{p\,(\gamma(p) - 1)}{(p - \gamma(p))(p-1)}$$

at a prime $p \nmid V$; axiom (D4) therefore makes $b(p)$ of size $O(1/p^2)$.

We now record the density to which the construction is applied.

Definition 4.9 (The W-tricked density γ_g)**.** Let $V \geq 1$. Define $\gamma_g^{(V)} : \mathbb{N} \to \mathbb{R}$ to be the totally multiplicative function with prime values

$$\gamma_g^{(V)}(p) := \begin{cases} 0, & p \mid V, \\ \dfrac{p}{p-1}, & p \nmid V, \end{cases}$$

that is, $\gamma_g^{(V)}(n) := \prod_p \gamma_g^{(V)}(p)^{v_p(n)}$, with $v_p(n)$ the exponent of p in n.

Uses(notation): γ_g, V, $\mathbb{P}$, $\mathbb{N}$ (index)

Lemma 4.10 (The Maynard sieve datum)**.** *Let $V \geq 1$ be squarefree with $2 \mid V$, and suppose $A_2 > 0$ and $L \geq 0$ are such that $-L \leq \Delta\big(\gamma_g^{(V)}; w, z\big) \leq A_2$ for all real $2 \leq w \leq z$. Let $\mathcal{S}_V$ be the six-tuple*

$$\mathcal{S}_V = \Big(\gamma_g^{(V)},\ V,\ A_1 = \tfrac{1}{2},\ A_2,\ A_3 = 2,\ L \Big).$$

Then $\mathcal{S}_V$ is a sieve datum in the sense of Definition 4.3.

Uses (statement): 4.3, 4.9, 4.1

Uses(notation): γ_g, V, A_1, A_2, A_3, L, $\mathbb{P}$, $\mathbb{N}$ (index)

Proof. Write $\gamma := \gamma_g^{(V)}$. Axiom (D5) is the hypothesis, and the squarefreeness of V together with $\gamma(p) = 0$ for $p \mid V$ (immediate from Definition 4.9) is axiom (D3). We check the remaining axioms.

(D1) Density. Every prime value $\gamma(p)$ is non-negative — it is either 0 or $p/(p-1) > 0$ — so the product over the prime factorisation defining $\gamma(n)$ is non-negative, and $\gamma(1) = 1$ as an empty product. Total multiplicativity of γ over prime factorisations gives $\gamma(mn) = \gamma(m)\gamma(n)$ for all coprime m, n (indeed for all nonzero m, n). For (iv), if $p \mid V$ then $\gamma(p) = 0 < p$. If $p \nmid V$, then

$p \neq 2$ because $2 \mid V$, so $p \geq 3$; and $p/(p-1) < p$ is equivalent to $1 < p-1$, i.e. to $p > 2$, which holds.

(D2) Growth with $A_1 = 1/2$. Certainly $0 < 1/2 < 1$. If $p \mid V$ then $\gamma(p)/p = 0 \leq 1/2$. If $p \nmid V$ then $p \geq 3$ as above, and $\gamma(p)/p = 1/(p-1) \leq 1/2$.

(D4) Unit density with $A_3 = 2$. Clearly $2 \geq 0$. Let $p \nmid V$. Then

$$|\gamma(p) - 1| \;=\; \Big|\frac{p}{p-1} - 1\Big| \;=\; \frac{1}{p-1},$$

and $1/(p-1) \leq 2/p$ is equivalent to $p \leq 2(p-1)$, i.e. to $p \geq 2$, which holds for every prime. □

In the application V is the primorial modulus W of the W-trick (Definition 3.29), which is squarefree and even as soon as $D_0(N) \geq 2$; the constants A_2 and L are then supplied, uniformly in N, by the Mertens estimates for the family γ_g.

4.2 Tensor sieve coefficients

Notation 4.11 (Boldface k-tuple convention)**.** A boldface symbol such as $\mathbf{d}$ denotes the k-tuple of positive integers $(d_1, \ldots, d_k)$, and we use the following conventions throughout.

- $\lambda(\mathbf{d})$ denotes the value $\lambda(d_1, \ldots, d_k)$. The bare symbol λ denotes the function $\mathbb{N}^k \to \mathbb{R}$ itself (as when it is passed as an argument to w_n, S_1, S_2 or $S_2^{(m)}$); we never write a subscripted form such as $\lambda_{\mathbf{d}}$.
- $\sum_{\mathbf{d}}$ denotes the iterated sum $\sum_{d_1} \sum_{d_2} \cdots \sum_{d_k}$, each variable ranging over $\mathbb{N}$, and $\prod_i d_i$ the product $d_1 d_2 \cdots d_k$.
- $\sum_{d_i \mid n+h_i\, \forall i}$ abbreviates $\sum_{d_1 \mid n+h_1} \sum_{d_2 \mid n+h_2} \cdots \sum_{d_k \mid n+h_k}$.
- Analogous conventions apply to $\mathbf{e}$, $\mathbf{r}$, $\mathbf{u}$, $\mathbf{a}$, $\mathbf{b}$.

All the sieve coefficient functions occurring below are *finitely supported*: the set $\{\mathbf{d} : \lambda(\mathbf{d}) \neq 0\}$ is finite. Every $\mathbf{d}$-sum written without further qualification is therefore a finite sum, and no convergence question arises.

Uses(notation): $\mathbf{d}$, $\mathbf{e}$, $\mathbf{r}$, $\mathbf{u}$, $\mathbf{a}$, $\mathbf{b}$, $\lambda(d_1, \ldots, d_k)$, $\sum_{\mathbf{d}}$, k, h_i, $\mathbb{N}$ (index)

The sieve coefficients are indexed by k-tuples of divisors, one per shift h_i, subject to three conditions: a truncation on the product of the divisors, coprimality to the W-trick modulus, and squarefreeness of the product. The truncation level is carried as a parameter rather than fixed to R: Instantiation I runs at level R and Instantiation II at level $R^{1+\varepsilon}$, and every lemma below is stated once, at an arbitrary level.

Two consequences of (iii) are used constantly and without comment: each individual d_i is squarefree (being a divisor of the squarefree $\prod_i d_i$), and the d_i are pairwise coprime. Both instantiations take λ with permissible support: Instantiation I at level $T = \lfloor R \rfloor$, Instantiation II at level $T = \lfloor R^{1+\varepsilon} \rfloor$.

4.3 The sieve weights and the sums $S_1(\lambda)$, $S_2(\lambda)$

Definition 4.12 (Sieve weight)**.** Let $\lambda : \mathbb{N}^k \to \mathbb{R}$ be finitely supported and let $n \geq 1$ be an integer. The *sieve weight* at n is

$$w_n \;:=\; \Big(\sum_{d_i \mid n+h_i \;\forall i} \lambda(d_1, \ldots, d_k) \Big)^2.$$

Lean:

- `PrimeGaps.weight`
- `PrimeGaps.termW`

Uses (statement): 4.11

Uses(notation): w_n, $\lambda(d_1, \dots, d_k)$, h_i, k, $\mathbb{N}$ (index)

Lemma 4.13 (Non-negativity of w_n)**.** *For every finitely supported $\lambda : \mathbb{N}^k \to \mathbb{R}$ and every integer $n \geq 1$, we have $w_n \geq 0$.*

Lean:
- `PrimeGaps.w_nonneg`

Uses (statement): 4.12

Uses(notation): w_n, $\lambda(d_1, \dots, d_k)$, h_i, k (index)

Proof. By Definition 4.12, w_n is the square of the real number $\sum_{d_i \mid n+h_i \ \forall i} \lambda(\mathbf{d})$, and squares of real numbers are non-negative. □

Non-negativity of w_n is the only property of the weight used in the passage from positivity of the sieve sums to primes; everything else about w_n enters through the asymptotic evaluation of S_1 and S_2.

Definition 4.14 (The sum $S_1(\lambda)$)**.** Let $\lambda : \mathbb{N}^k \to \mathbb{R}$ be finitely supported, with sieve weight w_n as in Definition 4.12. Set

$$S_1(\lambda) := \sum_{\substack{N < n \leq 2N \\ n \equiv v_0 \pmod W}} w_n.$$

Lean:
- `PrimeGaps.S₁`

Uses (statement): 4.12, 3.29, 3.33

Uses(notation): $S_1(\lambda)$, N, v_0, W, w_n, k, $\lambda(d_1, \dots, d_k)$ (index)

Definition 4.15 (The sum $S_2(\lambda)$)**.** Let $\lambda : \mathbb{N}^k \to \mathbb{R}$ be finitely supported, with sieve weight w_n as in Definition 4.12. Set

$$S_2(\lambda) := \sum_{\substack{N < n \leq 2N \\ n \equiv v_0 \pmod W}} \Big(\sum_{i=1}^{k} \chi_{\mathbb{P}}(n + h_i) \Big) w_n,$$

so that each n is weighted by the number of indices i for which $n + h_i$ is prime.

Lean:
- `PrimeGaps.S₂`

Uses (statement): 4.12, 3.29, 3.33

Uses(notation): $S_2(\lambda)$, N, v_0, W, $\chi_{\mathbb{P}}$, h_i, w_n, k, $\lambda(d_1, \dots, d_k)$ (index)

Definition 4.16 (The sum $S_2^{(m)}(\lambda)$)**.** Let $m \in \{1, \dots, k\}$ and let $\lambda : \mathbb{N}^k \to \mathbb{R}$ be finitely supported, with sieve weight w_n as in Definition 4.12. Set

$$S_2^{(m)}(\lambda) \;:=\; \sum_{\substack{N<n\le 2N \\ n\equiv v_0 \pmod W}} \chi_{\mathbb{P}}(n+h_m)\, w_n.$$

Lean:

- `PrimeGaps.S₂m`

Uses (statement): 4.12, 3.29, 3.33

Uses(notation): $S_2^{(m)}(\lambda)$, N, v_0, W, $\chi_{\mathbb{P}}$, h_i, w_n, k, $\lambda(d_1, \dots, d_k)$ (index)

Lemma 4.17 (Decomposition of $S_2(\lambda)$)**.** *For every finitely supported $\lambda : \mathbb{N}^k \to \mathbb{R}$,*

$$S_2(\lambda) \;=\; \sum_{m=1}^{k} S_2^{(m)}(\lambda).$$

Lean:

- `PrimeGaps.sum_S₂m_eq_S₂`

Uses (statement): 4.15, 4.16

Uses(notation): $S_2(\lambda)$, $S_2^{(m)}(\lambda)$, k, N, v_0, W, $\chi_{\mathbb{P}}$, h_i, w_n (index)

Proof. Fix n in the index set of Definition 4.15. Distributing the finite inner sum across the product with w_n gives

$$\Big(\sum_{i=1}^{k} \chi_{\mathbb{P}}(n+h_i)\Big) w_n \;=\; \sum_{i=1}^{k} \chi_{\mathbb{P}}(n+h_i)\, w_n.$$

Substituting this into Definition 4.15 and interchanging the two finite summations — over $n \in (N, 2N]$ with $n \equiv v_0 \pmod W$, and over $i \in \{1, \dots, k\}$ — yields

$$S_2(\lambda) \;=\; \sum_{i=1}^{k} \sum_{\substack{N<n\le 2N \\ n\equiv v_0 \pmod W}} \chi_{\mathbb{P}}(n+h_i)\, w_n.$$

By Definition 4.16 the inner sum is exactly $S_2^{(i)}(\lambda)$; renaming the index i to m gives the claim. □

Definition 4.18 (The sieve difference $S(\lambda, \rho)$)**.** Let $\lambda : \mathbb{N}^k \to \mathbb{R}$ be finitely supported and let $\rho > 0$ be real. Set

$$S(\lambda, \rho) \;:=\; S_2(\lambda) \;-\; \rho\, S_1(\lambda).$$

Uses (statement): 4.14, 4.15

Uses(notation): $S(\lambda)$, $S_1(\lambda)$, $S_2(\lambda)$, ρ, $\lambda(d_1, \dots, d_k)$, k (index)

Unwinding Definitions 4.14 and 4.15, the sieve difference is the single sum

$$S(\lambda, \rho) \;=\; \sum_{\substack{N<n\le 2N \\ n\equiv v_0 \pmod W}} \Big(\sum_{i=1}^{k} \chi_{\mathbb{P}}(n+h_i) \;-\; \rho\Big) w_n,$$

whose summands are the weights $w_n \ge 0$ multiplied by the excess of the prime count over ρ. Positivity of $S(\lambda, \rho)$ therefore cannot come from the weights alone; it forces the excess to be positive somewhere. This is the GPY reduction, and it is the sole point at which the sieve produces primes.

Lemma 4.19. *Let $\lambda : \mathbb{N}^k \to \mathbb{R}$ be finitely supported and let $\rho > 0$. If $S(\lambda, \rho) > 0$, then there exists an integer n with $N < n \le 2N$ and $n \equiv v_0 \pmod W$ such that at least $\lfloor \rho + 1 \rfloor$ of $n + h_1, \dots, n + h_k$ are prime.*

Lean:
- `PrimeGaps.lem_S_positive_implies_primes`

Uses (statement): 4.18, 4.14, 4.15, 4.12, 3.29, 3.34

Uses(notation): $S(\lambda)$, $S_1(\lambda)$, $S_2(\lambda)$, ρ, N, v_0, W, $\lfloor x \rfloor$, h_i, k, $\chi_{\mathbb{P}}$, w_n, $\lambda(d_1, \dots, d_k)$ (index)

Proof. Write $c(n) := \#\{i \in \{1, \dots, k\} : n + h_i \in \mathbb{P}\} = \sum_{i=1}^{k} \chi_{\mathbb{P}}(n + h_i)$, so that, as computed above,

$$S(\lambda, \rho) \;=\; \sum_{\substack{N < n \le 2N \\ n \equiv v_0 \pmod W}} \big(c(n) - \rho\big) w_n.$$

Suppose, for contradiction, that $c(n) \le \rho$ for every n in the index set. By Lemma 4.13 each $w_n \ge 0$, so each summand $(c(n) - \rho) w_n$ is non-positive, and therefore $S(\lambda, \rho) \le 0$, contradicting the hypothesis. Hence there is an n in the index set with $c(n) > \rho$.

It remains to convert $c(n) > \rho$ into $c(n) \ge \lfloor \rho + 1 \rfloor$. Since $c(n)$ is an integer and $c(n) > \rho$, we have $c(n) \ge \lfloor \rho \rfloor + 1$ when $\rho \notin \mathbb{Z}$, and $c(n) \ge \rho + 1$ when $\rho \in \mathbb{Z}$; in either case $c(n) \ge \lfloor \rho + 1 \rfloor$, because $\lfloor \rho + 1 \rfloor = \lfloor \rho \rfloor + 1$. □

Uses (proof): 4.13

4.4 The auxiliary y-variables

The coefficients $\lambda(\mathbf{d})$ are natural for the divisor-sum manipulations of S_1 and S_2, but ill-suited as variational unknowns: the divisibility relations couple them, and the diagonal form of the sieve sums is not visible in them. The change of variables $\lambda \leftrightarrow y$ below is a linear, involutive Möbius-type substitution which diagonalises S_1 and, after a further substitution $y \rightsquigarrow y^{(m)}$, also $S_2^{(m)}$. We give here the definitions and the inversion identities; the asymptotic evaluation of S_1 and $S_2^{(m)}$ in terms of y and $y^{(m)}$ is carried out later.

Definition 4.20 (The y-transform y_λ of λ)**.** Let $\lambda : \mathbb{N}^k \to \mathbb{R}$ be finitely supported. Define $y_\lambda : \mathbb{N}^k \to \mathbb{R}$ by

$$y_\lambda(r_1, \dots, r_k) \;:=\; \Big(\prod_{i=1}^{k} \mu(r_i)\,\varphi(r_i)\Big) \sum_{\substack{\mathbf{d} \in \mathbb{N}^k \\ r_i \mid d_i \text{ and } d_i \text{ squarefree } \forall i}} \frac{\lambda(d_1, \dots, d_k)}{\prod_{i=1}^{k} d_i}.$$

The $\mathbf{d}$-sum is finite because λ is finitely supported, and y_λ is again finitely supported. We write $y := y_\lambda$ and call its values the *y-variables* of λ.

Lean:
- `PrimeGaps.lToY`
- `PrimeGaps.lToY_apply`

Uses (statement): 3.3, 3.4, 4.11

Uses(notation): $y_\lambda(r_1, \dots, r_k)$, $y(r_1, \dots, r_k)$, $\lambda(d_1, \dots, d_k)$, μ, φ, k, $\mathbf{d}$, $\mathbf{r}$, $\mathbb{N}$ (index)

The squarefreeness restriction on the summation variables is vacuous in every application: if λ has permissible support at some level, then each d_i in its support is squarefree by Definition 3.39(iii), and the sum reduces to the unrestricted $\sum_{r_i \mid d_i\, \forall i}$. Carrying the restriction in the definition lets the inversion identities below hold for arbitrary finitely supported λ, with no support hypothesis; the map $\lambda \mapsto y_\lambda$ is $\mathbb{R}$-linear.

Lemma 4.21 (The y-variables inherit permissible support)**.** *Let $T \geq 1$ and let $\lambda : \mathbb{N}^k \to \mathbb{R}$ have permissible support at level T (Definition 3.39). Then y_λ has permissible support at level T.*

Lean:
- `Finsupp.HasPermissibleSupport.lToY`

Uses (statement): 3.39, 4.20

Uses(notation): $\lambda(d_1, \dots, d_k)$, $y(r_1, \dots, r_k)$, R, W, μ, k, (a, b), $\mathbf{r}$, $\mathbb{N}$ (index)

Proof. Suppose $y_\lambda(\mathbf{r}) \neq 0$. Then some summand of the defining sum in Definition 4.20 is nonzero, so there is a $\mathbf{d}$ with $\lambda(\mathbf{d}) \neq 0$ and $r_i \mid d_i$ for every i. It therefore suffices to show that the three conditions of Definition 3.39 are inherited from $\mathbf{d}$ to any coordinatewise divisor $\mathbf{r}$ of $\mathbf{d}$. That is exactly Lemma 3.40, applied at level T.

In detail: since $r_i \mid d_i$ for each i, we have $\prod_i r_i \mid \prod_i d_i$. As all d_i are nonzero, $\prod_i d_i \geq 1$ and hence $\prod_i r_i \leq \prod_i d_i \leq T$, which is (i). A divisor of an integer coprime to W is coprime to W, giving (ii); and a divisor of a squarefree integer is squarefree, giving (iii). □

Uses (proof): 3.40

Definition 4.22 (The Möbius transform λ_y of y)**.** Let $y : \mathbb{N}^k \to \mathbb{R}$ be finitely supported. Define $\lambda_y : \mathbb{N}^k \to \mathbb{R}$ by

$$\lambda_y(d_1, \dots, d_k) \;:=\; \Big(\prod_{i=1}^{k} \mu(d_i)\, d_i\Big) \sum_{\substack{\mathbf{r} \in \mathbb{N}^k \\ d_i \mid r_i \text{ and } r_i \text{ squarefree } \forall i}} \frac{y(r_1, \dots, r_k)}{\prod_{i=1}^{k} \varphi(r_i)}.$$

The $\mathbf{r}$-sum is finite because y is finitely supported, and λ_y is again finitely supported.

Lean:
- `PrimeGaps.yToL_apply`
- `PrimeGaps.yToL`

Uses (statement): 3.3, 3.4, 4.11

Uses(notation): $\lambda_y(d_1, \dots, d_k)$, $y(r_1, \dots, r_k)$, μ, φ, k, $\mathbf{d}$, $\mathbf{r}$, $\mathbb{N}$ (index)

The transform $y \mapsto \lambda_y$ is $\mathbb{R}$-linear, and the argument of Lemma 4.21 applies verbatim (with the roles of $\mathbf{d}$ and $\mathbf{r}$ exchanged) to show that it too preserves permissible support at any level T. The next two lemmas identify the two transforms as inverses.

Lemma 4.23 (Möbius inversion: λ from y)**.** *Let $\lambda : \mathbb{N}^k \to \mathbb{R}$ be finitely supported, and let $y : \mathbb{N}^k \to \mathbb{R}$ satisfy $y = y_\lambda$ (Definition 4.20). Then for every $\mathbf{d} \in \mathbb{N}^k$,*

$$\lambda_y(d_1, \dots, d_k) \;=\; \begin{cases} \lambda(d_1, \dots, d_k), & \text{if every } d_i \text{ is squarefree,} \\ 0, & \text{otherwise.} \end{cases}$$

Lean:

- `PrimeGaps.yToL_lToY`

Uses (statement): 4.20, 4.22, 3.3, 3.4

Uses(notation): $\lambda(d_1,\dots,d_k)$, $\lambda_y(d_1,\dots,d_k)$, $y(r_1,\dots,r_k)$, μ, φ, k, $\mathbf{d}$, $\mathbf{r}$, $\mathbb{N}$ (index)

Proof. If some d_i is not squarefree, then every $\mathbf{r}$ contributing to Definition 4.22 has $d_i \mid r_i$ with r_i squarefree, forcing d_i squarefree — a contradiction. So the sum is empty and $\lambda_y(\mathbf{d}) = 0$. Assume henceforth that every d_i is squarefree.

Both transforms are $\mathbb{R}$-linear in their argument, and every finitely supported function is a finite linear combination of point masses; so it suffices to prove the identity when $\lambda = c \cdot \mathbf{1}_{\mathbf{d}^{(0)}}$ is a point mass at some $\mathbf{d}^{(0)}$ with weight c. For such λ,

$$y(\mathbf{r}) \;=\; \Big(\prod_i \mu(r_i)\varphi(r_i)\Big)\frac{c}{\prod_i d_i^{(0)}} \quad \text{if } r_i \mid d_i^{(0)} \text{ and } d_i^{(0)} \text{ is squarefree for all } i,$$

and $y(\mathbf{r}) = 0$ otherwise; in particular y is supported on the coordinatewise divisors of $\mathbf{d}^{(0)}$. If some $d_i^{(0)}$ fails to be squarefree, or if $d_i \nmid d_i^{(0)}$ for some i, then no $\mathbf{r}$ contributes to $\lambda_y(\mathbf{d})$ and both sides vanish (on the right, $\lambda(\mathbf{d}) = 0$ unless $\mathbf{d} = \mathbf{d}^{(0)}$, and the excluded cases exclude that). Otherwise substitute the displayed formula for y into Definition 4.22: the factors $\varphi(r_i)$ cancel, and

$$\lambda_y(\mathbf{d}) \;=\; \frac{c\prod_i \mu(d_i)\,d_i}{\prod_i d_i^{(0)}} \sum_{\substack{\mathbf{r}\\ d_i\mid r_i\mid d_i^{(0)}\ \forall i}} \prod_{i=1}^{k}\mu(r_i) \;=\; \frac{c\prod_i \mu(d_i)\,d_i}{\prod_i d_i^{(0)}} \prod_{i=1}^{k}\Big(\sum_{d_i\mid r_i\mid d_i^{(0)}}\mu(r_i)\Big),$$

the sum having factorised across coordinates. For squarefree $d_i^{(0)}$ the inner sum is $\mu(d_i)\,\mathbf{1}[d_i = d_i^{(0)}]$, by Möbius inversion applied to the divisor chain $d_i \mid r_i \mid d_i^{(0)}$ (writing $r_i = d_i s_i$ with $s_i \mid d_i^{(0)}/d_i$ and using $\mu(d_i s_i) = \mu(d_i)\mu(s_i)$, valid since $d_i^{(0)}$ is squarefree, together with $\sum_{s\mid t}\mu(s) = \mathbf{1}[t=1]$). Hence the whole expression vanishes unless $\mathbf{d} = \mathbf{d}^{(0)}$, in which case it equals

$$\frac{c\prod_i \mu(d_i)\,d_i}{\prod_i d_i}\prod_i \mu(d_i) \;=\; c\prod_i \mu(d_i)^2 \;=\; c,$$

because each d_i is squarefree. This is $\lambda(\mathbf{d})$, as required. □

Lemma 4.24 (Möbius inversion: y from λ)**.** *Let $y : \mathbb{N}^k \to \mathbb{R}$ be finitely supported and let λ_y be as in Definition 4.22. Then for every $\mathbf{r} \in \mathbb{N}^k$,*

$$y_{\lambda_y}(r_1,\dots,r_k) \;=\; \begin{cases} y(r_1,\dots,r_k), & \text{if every } r_i \text{ is squarefree,}\\ 0, & \text{otherwise.}\end{cases}$$

Lean:

- `PrimeGaps.lToY_yToL`

Uses (statement): 4.20, 4.22, 3.3, 3.4

Uses(notation): $y_\lambda(r_1,\dots,r_k)$, $y(r_1,\dots,r_k)$, $\lambda_y(d_1,\dots,d_k)$, μ, φ, k, $\mathbf{d}$, $\mathbf{r}$, $\mathbb{N}$ (index)

Proof. The argument is that of Lemma 4.23 with the roles of the two transforms exchanged. If some r_i is not squarefree, every $\mathbf{d}$ contributing to Definition 4.20 at $\mathbf{r}$ has $r_i \mid d_i$ with d_i squarefree, which is impossible; so $y_{\lambda_y}(\mathbf{r}) = 0$. Assume every r_i squarefree. By linearity reduce

to $y = c \cdot \mathbf{1}_{\mathbf{r}^{(0)}}$ a point mass. Then λ_y is supported on the coordinatewise divisors of $\mathbf{r}^{(0)}$, and substituting Definition 4.22 into Definition 4.20 the factors d_i cancel against $1/\prod_i d_i$, leaving

$$y_{\lambda_y}(\mathbf{r}) \;=\; \frac{c\prod_i \mu(r_i)\varphi(r_i)}{\prod_i \varphi(r_i^{(0)})} \prod_{i=1}^{k} \Big(\sum_{r_i \mid d_i \mid r_i^{(0)}} \mu(d_i) \Big).$$

As before each inner sum equals $\mu(r_i)\,\mathbf{1}[r_i = r_i^{(0)}]$, so the expression vanishes unless $\mathbf{r} = \mathbf{r}^{(0)}$, and then equals $c\prod_i \mu(r_i)^2 = c = y(\mathbf{r})$. □

Uses (proof): 4.23

Together, Lemmas 4.23 and 4.24 say that the two transforms are mutually inverse $\mathbb{R}$-linear isomorphisms of the space of finitely supported functions $\mathbb{N}^k \to \mathbb{R}$ that vanish at every tuple with some non-squarefree entry. In particular one may specify a sieve either by giving λ or by giving y; the second is the form used in the variational problem.

The analysis of $S_2^{(m)}$ requires a second, closely related transform, in which the m-th coordinate is frozen at 1 and the totient φ is replaced by the function g defined next.

On squarefree arguments — the only ones at which g is ever evaluated below, since it always occurs multiplied by μ — one has $g(n) = \prod_{p \mid n}(p-2)$.

Definition 4.25 (The $y^{(m)}$-variables)**.** Let $m \in \{1, \dots, k\}$ and let $\lambda : \mathbb{N}^k \to \mathbb{R}$ be finitely supported. Define $y^{(m)} : \mathbb{N}^k \to \mathbb{R}$ by

$$y^{(m)}(r_1, \dots, r_k) \;:=\; \Big(\prod_{i=1}^{k} \mu(r_i)\, g(r_i) \Big) \sum_{\substack{\mathbf{d} \in \mathbb{N}^k,\ d_m = 1 \\ r_i \mid d_i \text{ and } d_i \text{ squarefree } \forall i}} \frac{\lambda(d_1, \dots, d_k)}{\prod_{i=1}^{k} \varphi(d_i)}.$$

Lean:

- `PrimeGaps.ym`
- `PrimeGaps.ym_apply`

Uses (statement): 3.3, 3.6, 3.4, 4.11

Uses(notation): $y^{(m)}(r_1, \dots, r_k)$, μ, g, $\lambda(d_1, \dots, d_k)$, φ, k, $\mathbf{d}$, $\mathbf{r}$, $\mathbb{N}$ (index)

Because $\mu(1) = g(1) = \varphi(1) = 1$, the factors carrying the index m contribute trivially: the prefactor may equally be written $\prod_{i \neq m} \mu(r_i) g(r_i)$ and the denominator $\prod_{i \neq m} \varphi(d_i)$, which is the shape in which $y^{(m)}$ appears in the evaluation of $S_2^{(m)}$.

Lemma 4.26 (Support of $y^{(m)}$)**.** *Let $m \in \{1, \dots, k\}$ and $\lambda : \mathbb{N}^k \to \mathbb{R}$ finitely supported. If $y^{(m)}(r_1, \dots, r_k) \neq 0$, then $r_m = 1$.*

Lean:

- `PrimeGaps.eq_one_of_ym_ne_zero`

Uses (statement): 4.25

Uses(notation): $y^{(m)}(r_1, \dots, r_k)$, k, $\mathbf{r}$ (index)

Proof. If $y^{(m)}(\mathbf{r}) \neq 0$ then the defining sum of Definition 4.25 has a nonzero summand, so there is a $\mathbf{d}$ with $d_m = 1$ and $r_i \mid d_i$ for every i. Taking $i = m$ gives $r_m \mid 1$, hence $r_m = 1$. □

Finally we record the coefficient function that the two instantiations use in S_1 and S_2. It is specified in the y-variables — a rescaled sample of a fixed function F on the truncation scale — and the corresponding λ is produced by the Möbius transform of Definition 4.22. The analytic input is F, and λ is the derived object.

By construction y_0 has permissible support at level T, hence is finitely supported, so Definition 4.22 applies to it.

Since the Möbius transform preserves permissible support, λ_0 has permissible support at level T, so all of Definitions 4.12–4.18 apply to it. Instantiation I takes $T = \lfloor R \rfloor$ with F supported in the standard simplex $\mathcal{R}_k$; Instantiation II takes $T = \lfloor R^{1+\varepsilon} \rfloor$ with F supported in the enlarged simplex $\mathcal{T}_\varepsilon$. Everything above is common to both.

Definition 4.27 (Maximal modulus of a finitely supported system)**.** Let A be a set and let $f : A \to \mathbb{R}$ be finitely supported. Its *maximal modulus* is

$$\|f\|_{\max} := \max_{a \in \operatorname{supp} f} |f(a)|,$$

with the convention $\|f\|_{\max} := 0$ when f vanishes identically.

Lean:

- `Finsupp.le_maxRealAbs`
- `Finsupp.maxRealAbs`
- `Finsupp.maxRealAbs_le_iff`

Uses(notation): $y_{\max}$, $y^{(m)}_{\max}$, $y(r_1, \dots, r_k)$ (index)

Definition 4.28 (The truncated weight sum $H(z)$)**.** Let S be a sieve datum with associated weight h (Definition 4.6). For a real $z > 0$ set

$$H(z) := \sum_{0<d<z} h(d),$$

the sum taken over positive integers d strictly below z.

Lean:

- `PrimeGaps.SieveDatum.H`

Uses (statement): 4.6

Uses(notation): $H(z)$, $h(d)$ (index)

Definition 4.29 (The Maynard sieve datum)**.** Let N be a positive integer with $2 \le D_0(N)$. The *Maynard sieve datum* at N is the sieve datum (Definition 4.3) whose density is the W-tricked density γ_g of Definition 4.9, whose modulus is $V := W$, and whose growth and structure constants are $A_1 := 1/2$ and $A_3 := 2$. Its remaining constants A_2 and L are the deviation bounds carried by the datum; that they may be taken uniformly in N is established in Section 6.

Lean:

- `PrimeGaps.maynardSieveDatum`

Uses (statement): 4.3, 3.29, 4.9

Uses(notation): γ_g, V, A_1, A_2, A_3, L, W, D_0, N (index)

5 The variational problem

The sieve estimates of the preceding sections reduce the problem to one in the calculus of variations. A *profile* F supported on a simplex determines the sieve coefficients λ; the leading term of $S_1(\lambda)$ is the L^2 mass $I_k(F)$ of the profile, and the leading term of $S_2^{(m)}(\lambda)$ is the L^2 mass $J_{k,\varepsilon}^{(m)}(F)$ of its m-th marginal. The supremum of the ratio $\big(\sum_m J_{k,\varepsilon}^{(m)}(F)\big)/I_k(F)$ over profiles F is the constant $M_{k,\varepsilon}$.

The construction is carried out for an enlargement parameter $\varepsilon \geq 0$. The profile is supported on the enlarged simplex $\mathcal{T}_\varepsilon$ of total mass $1+\varepsilon$, while each marginal is taken over the shrunken slice $\mathcal{S}_\varepsilon$ of total mass $1-\varepsilon$; at $\varepsilon = 0$ both reduce to the standard simplex and its faces, and every statement below specialises to the classical one. The denominator $I_k(F) = \int F^2$ does not depend on ε; the enlargement enters only through the wider class of admissible competitors.

We introduce $\mathcal{T}_\varepsilon$, F, I_k, $J_{k,\varepsilon}^{(m)}$, $M_{k,\varepsilon}$ and r_k, record the basic properties of $M_{k,\varepsilon}$ (finiteness, positivity, L^2-continuity of the two functionals, density of smooth profiles), and prove the variational criterion (Lemma 5.22), which passes from a lower bound on $M_{k,\varepsilon}$ to the strict inequality used by the sieve estimates. Both cases invoke Lemma 5.22, one at $\varepsilon = 0$, the other at $\varepsilon = 1/25$. The remaining lemmas bound $M_{k,\varepsilon}$ from below by evaluating the two functionals on explicit polynomial profiles.

5.1 The simplex and the profile

Notation 5.1 (Power sums)**.** For $j \in \mathbb{Z}_{\geq 1}$ we write P_j for the j-th power sum in the coordinates $t_1, \dots, t_k$,

$$P_j(t_1, \dots, t_k) \;:=\; \sum_{i=1}^{k} t_i^{\,j},$$

so that $P_1 = t_1 + \cdots + t_k$ and $P_2 = t_1^2 + \cdots + t_k^2$. When one coordinate is distinguished we write P_j' for the corresponding power sum in the remaining $k-1$ coordinates.

Lean:

- `PrimeGaps.termP₁_`
- `PrimeGaps.termP₂_`

Uses(notation): P_j, P_j', k (index)

Definition 5.2 (Scaled standard simplex)**.** Let $k \in \mathbb{Z}_{\geq 0}$ and $s \in \mathbb{R}$. The *scaled standard simplex* of dimension k and scale s is

$$\mathcal{R}_k(s) \;:=\; \Big\{ t = (t_1, \dots, t_k) \in \mathbb{R}^k \;:\; t_i \geq 0 \text{ for every } i,\; \sum_{i=1}^{k} t_i \leq s \Big\}.$$

Three instances recur, and we abbreviate them. For $\varepsilon \geq 0$:

$$\mathcal{R}_k \;:=\; \mathcal{R}_k(1), \qquad \mathcal{T}_\varepsilon \;:=\; \mathcal{R}_k(1+\varepsilon), \qquad \mathcal{S}_\varepsilon \;:=\; \mathcal{R}_{k-1}(1-\varepsilon),$$

called respectively the *standard simplex*, the *ε-enlarged simplex* and the *ε-shrunken slice*. Thus $\mathcal{T}_0 = \mathcal{R}_k$ and $\mathcal{S}_0 = \mathcal{R}_{k-1}$; we also write $\mathcal{R}_k^{(m)}$ for the copy of $\mathcal{R}_{k-1}$ obtained by deleting the m-th coordinate of $\mathbb{R}^k$, so that $\mathcal{S}_0 = \mathcal{R}_k^{(m)}$ under that identification.

Lean:

- `PrimeGaps.termℛ_`
- `Affine.stdAffineSimplex.closedInterior_eq`
- `EuclideanSpace.mem_scaledStdSimplex_iff`

- `Affine.stdAffineSimplex`

Uses(notation): $\mathcal{R}_k$, $\mathcal{R}_k^{(m)}$, $\mathcal{R}_k$, $\mathcal{T}_\varepsilon$, $\mathcal{S}_\varepsilon$, ε, k (index)

Definition 5.3 (Maynard's combinatorial polynomial $G_{b,j}$)**.** For $b, j, x \in \mathbb{Z}_{\geq 0}$ set

$$G_{b,j}(x) \;:=\; b! \sum_{\substack{(d_1,\dots,d_x)\in\mathbb{Z}_{\geq 0}^x \\ d_1+\cdots+d_x=b}} \prod_{i=1}^{x} \frac{(j\,d_i)!}{d_i!},$$

the sum being over all x-tuples of non-negative integers with total b. Each quotient $(jd_i)!/d_i!$ is an integer, so $G_{b,j}(x) \in \mathbb{Z}_{\geq 0}$. Discarding the zero entries of a tuple and recording only the r positive ones — each reduced tuple being counted $\binom{x}{r}$ times, since the weight takes the value 1 at $d_i = 0$ — gives the equivalent form

$$G_{b,j}(x) \;=\; b! \sum_{r=1}^{b} \binom{x}{r} \sum_{\substack{d_1+\cdots+d_r=b \\ d_i\geq 1}} \prod_{i=1}^{r} \frac{(jd_i)!}{d_i!} \qquad (b \geq 1),$$

in which x occurs only inside the binomial coefficients; this exhibits $G_{b,j}$ as a polynomial in x of degree b.

Lean:
- `PrimeGaps.maynardG`
- `PrimeGaps.<termG[_,_]>`

Uses(notation): $G_{b,j}(x)$ (index)

Definition 5.4 (Extension by zero from the simplex)**.** Let $k \in \mathbb{N}$ and let $P : \mathbb{R}^k \to \mathbb{R}$. The *profile attached to* P is

$$F_P \;:=\; P \cdot \mathbf{1}_{\mathcal{R}_k}, \qquad \text{that is,} \qquad F_P(t) = \begin{cases} P(t), & t \in \mathcal{R}_k, \\ 0, & t \notin \mathcal{R}_k. \end{cases}$$

Lean:
- `PrimeGaps.F_P`

Uses (statement): 5.2

Uses(notation): F_P, $\mathcal{R}_k$, k, F (index)

Definition 5.5 (The monomial index set $\mathcal{B}_d$)**.** For $d \in \mathbb{Z}_{\geq 0}$ set

$$\mathcal{B}_d \;:=\; \{\, (b,c) \in \mathbb{Z}_{\geq 0}^2 \;:\; b + 2c \leq d \,\}.$$

A pair $(b, c) \in \mathcal{B}_d$ indexes the symmetric polynomial $(1-P_1)^b\, P_2^c$ in $t_1, \dots, t_k$; these polynomials span the space of profiles used to bound $M_{k,0}$ from below.

Lean:
- `PrimeGaps.basisB`

Uses (statement): 5.1

Uses(notation): $\mathcal{B}_d$, P_j, k (index)

5.2 The functionals I_k, $J_{k,\varepsilon}^{(m)}$ and the constant $M_{k,\varepsilon}$

Definition 5.6 (The functional I_k)**.** For $k \in \mathbb{N}$ and $F \in L^2(\mathbb{R}^k)$,

$$I_k(F) \;:=\; \int_{\mathbb{R}^k} F(t_1,\dots,t_k)^2\, dt_1 \cdots dt_k.$$

If F vanishes outside a measurable set S then $I_k(F) = \int_S F^2 = \|F\|^2_{L^2(S)}$; in particular $I_k(F) = \|F\|^2_{L^2(\mathcal{T}_\varepsilon)}$ for a profile supported in $\mathcal{T}_\varepsilon$. The functional carries no dependence on ε.

Uses (statement): 5.2

Uses(notation): $I_k(F)$, $I_k(F)$, F, k, $\mathcal{R}_k$, $\mathcal{T}_\varepsilon$, $\|f\|_{L^2(X)}$ (index)

Definition 5.7 (The marginal functional $J_{k,\varepsilon}^{(m)}$)**.** Let $k \ge 1$, let $\varepsilon \ge 0$, let $m \in \{1,\dots,k\}$, and let $F : \mathbb{R}^k \to \mathbb{R}$ vanish outside $\mathcal{T}_\varepsilon$. Write T_m for the *m-th marginal operator*

$$(T_m F)(t_1,\dots,\widehat{t_m},\dots,t_k) \;:=\; \int_0^\infty F(t_1,\dots,t_k)\, dt_m,$$

defined for almost every $(t_i)_{i\neq m} \in \mathbb{R}^{k-1}$, and set

$$J_{k,\varepsilon}^{(m)}(F) \;:=\; \int_{\mathcal{S}_\varepsilon} (T_m F)^2 \prod_{i\neq m} dt_i \;=\; \|T_m F\|^2_{L^2(\mathcal{S}_\varepsilon)}.$$

We abbreviate $J_k^{(m)} := J_{k,0}^{(m)}$, the marginal functional of the standard problem, whose outer domain $\mathcal{S}_0$ is $\mathcal{R}_k^{(m)}$.

Lean:
- `PrimeGaps.JBilinCLM`
- `PrimeGaps.J`

Uses (statement): 5.2

Uses(notation): $J_k^{(m)}(F)$, $J_\varepsilon^{(i)}(F)$, T_m, F, k, $\mathcal{T}_\varepsilon$, $\mathcal{S}_\varepsilon$, ε, $\mathcal{R}_k^{(m)}$, $\|f\|_{L^2(X)}$ (index)

Definition 5.8 (The variational constant $M_{k,\varepsilon}$)**.** For $k \ge 1$ and $\varepsilon \ge 0$,

$$M_{k,\varepsilon} \;:=\; \sup\left\{ \frac{\sum_{m=1}^k J_{k,\varepsilon}^{(m)}(F)}{I_k(F)} \;\middle|\; F \in L^2(\mathcal{T}_\varepsilon),\; I_k(F) \neq 0 \right\}.$$

The supremum is taken over the whole of $L^2(\mathcal{T}_\varepsilon)$, each competitor being extended by zero outside $\mathcal{T}_\varepsilon$: no smoothness, no symmetry and no sign condition is imposed on F. We abbreviate $M_k := M_{k,0}$.

Lean:
- `PrimeGaps.M`

Uses (statement): 5.6, 5.7

Uses(notation): M_k, $\mathcal{Q}_\varepsilon(F)$, $I_k(F)$, $J_k^{(m)}(F)$, F, k, $\mathcal{T}_\varepsilon$, ε, $\|f\|_{L^2(X)}$ (index)

Definition 5.9 (The prime count $r_k(\vartheta)$)**.** For $\vartheta \in (0,1)$ and $k \geq 1$,

$$r_k(\vartheta) \;:=\; \left\lceil \frac{\vartheta\, M_k}{2} \right\rceil \in \mathbb{Z}_{\geq 0},$$

where $M_k = M_{k,0}$. This is the number of primes the sieve delivers among $n+h_1, \dots, n+h_k$ at level of distribution ϑ.

Lean:
- `PrimeGaps.r`

Uses (statement): 5.8, 3.21

Uses(notation): r_k, M_k, ϑ, $\lceil x \rceil$, k (index)

5.3 From a profile to the sieve coefficients

Definition 5.10 (The y-variables attached to a profile)**.** Let $k \in \mathbb{N}$, let $R > 1$ and $W \in \mathbb{N}$, and let $F : \mathbb{R}^k \to \mathbb{R}$. Define $y_0 : \mathbb{N}^k \to \mathbb{R}$ by

$$y_0(r_1, \dots, r_k) \;:=\; \begin{cases} F\left(\dfrac{\log r_1}{\log R}, \dots, \dfrac{\log r_k}{\log R}\right), & (r_1, \dots, r_k) \text{ admissible for } (R, W), \\ 0, & \text{otherwise,} \end{cases}$$

where a tuple is *admissible for* (R,W) in the sense of Definition 3.39, i.e. $\prod_i r_i \leq R$, $\gcd(\prod_i r_i, W) = 1$ and $\prod_i r_i$ is squarefree.

Lean:
- `PrimeGaps.y₀`
- `PrimeGaps.y₀_apply`

Uses (statement): 3.39

Uses(notation): $y(r_1, \dots, r_k)$, F, R, W, k, $\mathbb{N}$, (a,b) (index)

Definition 5.11 (The sieve coefficients attached to a profile)**.** With k, R, W and F as in Definition 5.10, the coefficients $\lambda_0 : \mathbb{N}^k \to \mathbb{R}$ attached to F are the image of y_0 under the $y \mapsto \lambda$ correspondence of Definition 4.22:

$$\lambda_0(d_1, \dots, d_k) \;=\; \Big(\prod_{i=1}^{k} \mu(d_i)\, d_i\Big) \sum_{\substack{(r_1,\dots,r_k) \text{ admissible for } (R,W) \\ d_i \mid r_i \text{ for every } i}} \frac{y_0(r_1, \dots, r_k)}{\prod_{i=1}^{k} \varphi(r_i)}.$$

Taking F supported in $\mathcal{T}_\varepsilon$ and truncation level $R^{1+\varepsilon}$ produces the enlarged weight; taking F supported in $\mathcal{R}_k$ and level R produces the standard one.

Lean:
- `PrimeGaps.l₀_apply`

Uses (statement): 5.10, 4.22, 3.39, 5.2

Uses(notation): $\lambda_0(d_1, \dots, d_k)$, λ_ε, $\lambda_y(d_1, \dots, d_k)$, $y(r_1, \dots, r_k)$, μ, φ, $\mathbf{d}$, $\mathbf{r}$, k, R, W, $\mathcal{T}_\varepsilon$, ε (index)

5.4 Basic properties of I_k, $J_{k,\varepsilon}^{(m)}$ and $M_{k,\varepsilon}$

Lemma 5.12 (Each marginal is dominated by the L^2 mass)**.** *Let $k \geq 1$, let $\varepsilon \geq 0$, let $F \in L^2(\mathcal{T}_\varepsilon)$ and let $m \in \{1, \dots, k\}$. Then*

$$J_{k,\varepsilon}^{(m)}(F) \;\leq\; (1+\varepsilon)\, I_k(F).$$

Uses (statement): 5.6, 5.7, 5.2

Uses(notation): $I_k(F)$, $J_k^{(m)}(F)$, T_m, F, k, $\mathcal{T}_\varepsilon$, $\mathcal{S}_\varepsilon$, ε, $\|f\|_{L^2(X)}$ (index)

Proof. Fix $(t_i)_{i\neq m} \in \mathcal{S}_\varepsilon$ and consider the fibre

$$E \;:=\; \{\, s \geq 0 : (t_1, \dots, t_{m-1}, s, t_{m+1}, \dots, t_k) \in \mathcal{T}_\varepsilon \,\}.$$

Since every $t_i \geq 0$, the fibre is the interval $[0,\, 1+\varepsilon-\sum_{i\neq m} t_i]$ (empty if the right endpoint is negative), whose length is at most $1+\varepsilon$. Since F vanishes off $\mathcal{T}_\varepsilon$, the Cauchy–Schwarz inequality on E gives

$$\big|(T_m F)\big((t_i)_{i\neq m}\big)\big|^2 \;=\; \Big|\int_E F\, ds\Big|^2 \;\leq\; |E| \int_E F^2\, ds \;\leq\; (1+\varepsilon) \int_E F^2\, ds.$$

Integrating over $(t_i)_{i\neq m} \in \mathcal{S}_\varepsilon$ and applying Tonelli's theorem to the right-hand side, the iterated integral is bounded by the integral of F^2 over all of $\mathbb{R}^k$, which is $I_k(F)$. Hence $J_{k,\varepsilon}^{(m)}(F) \leq (1+\varepsilon) I_k(F)$.

Equivalently: T_m is a bounded linear operator $L^2(\mathcal{T}_\varepsilon) \to L^2(\mathcal{S}_\varepsilon)$ whose norm is at most the square root of the essential supremum of the fibre lengths, and that essential supremum is at most $1+\varepsilon$. □

Lemma 5.13 (Iterated form of the standard marginal functional)**.** *Let $k \geq 1$, let $m \in \{1, \dots, k\}$, and let $F \in L^2(\mathcal{R}_k)$ vanish outside $\mathcal{R}_k$. Then*

$$J_k^{(m)}(F) \;=\; \int_{\mathcal{R}_k^{(m)}} \Big(\int_0^{1-\sum_{i\neq m} t_i} F(t_1, \dots, t_{m-1}, s, t_{m+1}, \dots, t_k)\, ds \Big)^2 \prod_{i\neq m} dt_i.$$

Uses (statement): 5.7, 5.2

Uses(notation): $J_k^{(m)}(F)$, $\mathcal{R}_k^{(m)}$, $\mathcal{R}_k$, F, k, T_m (index)

Proof. At $\varepsilon = 0$ the outer domain $\mathcal{S}_0$ is $\mathcal{R}_k^{(m)}$, so only the inner integral has to be identified. Fix $(t_i)_{i\neq m} \in \mathcal{R}_k^{(m)}$. Then every $t_i \geq 0$ and $\sum_{i\neq m} t_i \leq 1$, and the fibre of $\mathcal{R}_k$ over this point is exactly $[0,\, 1-\sum_{i\neq m} t_i]$. For s beyond that endpoint the inserted point lies outside $\mathcal{R}_k$, so F vanishes there by hypothesis; hence the integral $\int_0^\infty F\, dt_m$ defining $T_m F$ agrees with the truncated integral $\int_0^{1-\sum_{i\neq m} t_i} F\, ds$. Squaring and integrating over $\mathcal{R}_k^{(m)}$ gives the claim.

The hypothesis that F vanish *pointwise* outside $\mathcal{R}_k$, and not merely almost everywhere on $\mathcal{R}_k$, is essential: T_m samples F along the whole half-line, so a modification of F on a set that is null in $\mathcal{R}_k$ but not in $\mathbb{R}^k$ would change the left-hand side. □

Uses (proof): 5.12

Lemma 5.14 (L^2-continuity of I_k)**.** *For every $\varepsilon \geq 0$ the functional $I_k : L^2(\mathcal{T}_\varepsilon) \to \mathbb{R}$ is continuous.*

Lean:
- `continuous_norm`

Uses (statement): 5.6, 5.2

Uses(notation): $I_k(F)$, F, k, $\mathcal{T}_\varepsilon$, ε, $\|f\|_{L^2(X)}$ (index)

Proof. On $L^2(\mathcal{T}_\varepsilon)$ we have $I_k(F) = \|F\|^2_{L^2(\mathcal{T}_\varepsilon)}$ by Definition 5.6, so I_k is the square of the norm, and the norm on a normed space is continuous. Quantitatively, for $F, G \in L^2(\mathcal{T}_\varepsilon)$ the factorisation $F^2 - G^2 = (F-G)(F+G)$ together with Cauchy–Schwarz and the triangle inequality gives

$$|I_k(F) - I_k(G)| \;\le\; \|F-G\|\,\|F+G\| \;\le\; \big(\|F\| + \|G\|\big)\,\|F-G\|,$$

all norms being those of $L^2(\mathcal{T}_\varepsilon)$. □

Lemma 5.15 (L^2-continuity of $J^{(m)}_{k,\varepsilon}$)**.** *For every $k \ge 1$, every $\varepsilon \ge 0$ and every $m \in \{1, \dots, k\}$, the functional $J^{(m)}_{k,\varepsilon} : L^2(\mathcal{T}_\varepsilon) \to \mathbb{R}$ is continuous.*

Lean:
- `PrimeGaps.continuous_JEps`
- `PrimeGaps.continuous_J`

Uses (statement): 5.7, 5.2

Uses(notation): $J^{(m)}_k(F)$, T_m, F, k, $\mathcal{T}_\varepsilon$, $\mathcal{S}_\varepsilon$, ε, $\|f\|_{L^2(X)}$ (index)

Proof. The marginal operator T_m is linear, and by the proof of Lemma 5.12 it is bounded from $L^2(\mathcal{T}_\varepsilon)$ to $L^2(\mathcal{S}_\varepsilon)$, with norm at most $(1+\varepsilon)^{1/2}$; hence it is continuous. The map $H \mapsto \|H\|^2_{L^2(\mathcal{S}_\varepsilon)}$ is continuous, being the square of a norm. By Definition 5.7, $J^{(m)}_{k,\varepsilon}$ is the composite of these two continuous maps. Quantitatively, the argument of Lemma 5.14 applied in $L^2(\mathcal{S}_\varepsilon)$ together with the operator bound gives

$$\big|J^{(m)}_{k,\varepsilon}(F) - J^{(m)}_{k,\varepsilon}(G)\big| \;\le\; (1+\varepsilon)\big(\|F\| + \|G\|\big)\,\|F-G\|$$

with the norms of $L^2(\mathcal{T}_\varepsilon)$. □

Uses (proof): 5.12, 5.14

Lemma 5.16 (Upper bound for $M_{k,\varepsilon}$)**.** *For every $k \ge 1$ and every $\varepsilon \ge 0$ the supremum defining $M_{k,\varepsilon}$ is finite, and*

$$M_{k,\varepsilon} \;\le\; (1+\varepsilon)\,k.$$

In particular $M_k \le k$.

Uses (statement): 5.8, 5.2

Uses(notation): M_k, $I_k(F)$, $J^{(m)}_k(F)$, k, $\mathcal{T}_\varepsilon$, ε (index)

Proof. Let $F \in L^2(\mathcal{T}_\varepsilon)$ with $I_k(F) \neq 0$. Summing the bound of Lemma 5.12 over the k values of m gives $\sum_{m=1}^{k} J^{(m)}_{k,\varepsilon}(F) \le (1+\varepsilon)k\, I_k(F)$, so the Rayleigh quotient at F is at most $(1+\varepsilon)k$. The family of quotients is therefore bounded above, its supremum exists in $\mathbb{R}$, and it is at most $(1+\varepsilon)k$. □

Uses (proof): 5.12

Lemma 5.17 (Positivity of $M_{k,\varepsilon}$)**.** *For every $k \geq 1$ and every $\varepsilon \in [0,1)$ we have $M_{k,\varepsilon} > 0$.*

Uses (statement): 5.8, 5.2

Uses(notation): M_k, $I_k(F)$, $J_k^{(m)}(F)$, k, $\mathcal{R}_k$, $\mathcal{T}_\varepsilon$, $\mathcal{S}_\varepsilon$, ε, $\mathrm{vol}(A)$ (index)

Proof. Take $F := \mathbf{1}_{\mathcal{R}_k}$, which vanishes outside $\mathcal{R}_k \subseteq \mathcal{T}_\varepsilon$ and so is an element of $L^2(\mathcal{T}_\varepsilon)$. The simplex $\mathcal{R}_k$ contains the non-empty open set $\{t : t_i > 0\ \forall i,\ \sum_i t_i < 1\}$, hence has positive volume, and $I_k(F) = \mathrm{vol}(\mathcal{R}_k) > 0$; so F is an admissible competitor.

Fix m. For $(t_i)_{i\neq m}$ with $t_i \geq 0$ and $\sum_{i\neq m} t_i \leq 1$ the fibre of $\mathcal{R}_k$ over that point is the interval $[0,\, 1-\sum_{i\neq m} t_i]$, so, exactly as in the proof of Lemma 5.13, $(T_m F)((t_i)_{i\neq m}) = 1 - \sum_{i\neq m} t_i$. The shrunken slice $\mathcal{S}_\varepsilon$ is contained in $\mathcal{R}_{k-1}$ because $\varepsilon \geq 0$, and it contains the non-empty open set $V := \{t : t_i > 0\ \forall i,\ \sum_i t_i < (1-\varepsilon)/2\}$ because $\varepsilon < 1$; on V the integrand $(1-\sum_{i\neq m} t_i)^2$ exceeds $1/4$. Hence

$$J_{k,\varepsilon}^{(m)}(F) \;\geq\; \tfrac{1}{4}\,\mathrm{vol}(V) \;>\; 0.$$

Therefore $\sum_m J_{k,\varepsilon}^{(m)}(F) > 0$, the Rayleigh quotient at F is strictly positive, and $M_{k,\varepsilon}$, being at least that quotient, is strictly positive. □

Uses (proof): 5.13

5.5 Smooth profiles and the variational criterion

The sieve asymptotics for S_1 and $S_2^{(m)}$ require a *smooth* profile supported in $\mathcal{T}_\varepsilon$, whereas $M_{k,\varepsilon}$ is a supremum over all of $L^2(\mathcal{T}_\varepsilon)$. Smooth functions supported in the simplex are L^2-dense there, hence (by the continuity of I_k and $J_{k,\varepsilon}^{(m)}$) they realise the supremum to within any prescribed error, and the resulting near-maximiser satisfies the strict inequality required by the sieve estimates. The density statement is proved on the standard simplex and transported to $\mathcal{T}_\varepsilon$ by the dilation $t \mapsto (1+\varepsilon)^{-1}t$, which carries $\mathcal{T}_\varepsilon$ onto $\mathcal{R}_k$.

Lemma 5.18 (Smooth cut-off to an inset of $\mathcal{R}_k$)**.** *Let $k \geq 2$ and let η satisfy $0 < \eta < 1/(2k)$. There exists a smooth function $\chi : \mathbb{R}^k \to [0,1]$ such that*

1. *$\chi(t) = 1$ for every $t \in \mathcal{R}_k$ satisfying $t_i \geq \eta$ for all i and $\sum_i t_i \leq 1-\eta$;*
2. *$\chi(t) = 0$ for every $t \notin \mathcal{R}_k$;*
3. *for every measurable $G : \mathbb{R}^k \to \mathbb{R}$ and every $M \geq 0$ with $|G| \leq M$ on $\mathcal{R}_k$,*

$$\int_{\mathcal{R}_k} \big(G - G\chi\big)^2 \;\leq\; M^2\,\mathrm{vol}\big(\{t \in \mathcal{R}_k : t_i < \eta \text{ for some } i \text{ or } \sum_i t_i > 1-\eta\}\big) \;\leq\; M^2 \cdot 2k\eta.$$

Lean:

- `PrimeGaps.exists_smooth_simplex_cutoff`

Uses (statement): 5.2

Uses(notation): $\mathcal{R}_k$, $\mathcal{R}_k^{(\varepsilon)}$, k, $C^\infty(X)$, $\mathrm{vol}(A)$ (index)

Proof. Fix a smooth non-decreasing $\psi : \mathbb{R} \to [0,1]$ with $\psi(u) = 0$ for $u \le 0$ and $\psi(u) = 1$ for $u \ge 1$, and put $\psi_\eta(u) := \psi(u/\eta)$. Set

$$\chi(t_1, \dots, t_k) \;:=\; \Big(\prod_{i=1}^{k} \psi_\eta(t_i)\Big) \cdot \psi_\eta\Big(1 - \sum_{i=1}^{k} t_i\Big).$$

This is a finite product of smooth functions, hence smooth, with values in $[0,1]$.

(i) If $t \in \mathcal{R}_k^{(\eta)}$ then $t_i \ge \eta$ for each i and $\sum_i t_i \le 1 - \eta$, so every factor equals 1.

(ii) If $t \notin \mathcal{R}_k$ then either $t_i < 0$ for some i, or $\sum_i t_i > 1$; in the first case $\psi_\eta(t_i) = 0$ and in the second $\psi_\eta(1 - \sum_i t_i) = 0$.

(iii) On $\mathcal{R}_k^{(\eta)}$ we have $G - G\chi = 0$ by (i), while on all of $\mathcal{R}_k$ we have $|G - G\chi| = |G|\,|1 - \chi| \le M$. Hence

$$\int_{\mathcal{R}_k} (G - G\chi)^2 = \int_{\mathcal{R}_k \setminus \mathcal{R}_k^{(\eta)}} (G - G\chi)^2 \le M^2 \operatorname{vol}\big(\mathcal{R}_k \setminus \mathcal{R}_k^{(\eta)}\big).$$

For the second inequality, $\mathcal{R}_k \setminus \mathcal{R}_k^{(\eta)}$ is covered by the k slabs $\{t \in \mathcal{R}_k : t_i < \eta\}$ together with the slab $\{t \in \mathcal{R}_k : \sum_i t_i > 1 - \eta\}$. Each of the first k slabs has volume at most $\eta/(k-1)!$, since its cross-section at fixed t_i lies in a $(k-1)$-dimensional simplex of volume $1/(k-1)!$; the last slab, being the difference of two simplices, has volume $\big(1 - (1-\eta)^k\big)/k! \le \eta/(k-1)!$. Summing gives $\operatorname{vol}(\mathcal{R}_k \setminus \mathcal{R}_k^{(\eta)}) \le (k+1)\eta/(k-1)! \le 2k\eta$ for $k \ge 2$. □

Lemma 5.19 (L^2-density of smooth functions supported in $\mathcal{R}_k$)**.** *Let $k \ge 2$, let $F_0 \in L^2(\mathcal{R}_k)$ and let $\delta > 0$. Then there exists a smooth $F_1 : \mathbb{R}^k \to \mathbb{R}$ whose closed support is contained in $\mathcal{R}_k$ and which satisfies*

$$\|F_0 - F_1\|_{L^2(\mathcal{R}_k)} \;<\; \delta.$$

Lean:

- `PrimeGaps.smooth_L2_approx_on_simplex`

Uses (statement): 5.2

Uses(notation): $\mathcal{R}_k$, k, F, δ, $\|f\|_{L^2(X)}$, $C^\infty(X)$, $C_c^\infty(X)$ (index)

Proof. Extend F_0 by zero to $\mathbb{R}^k$; the extension lies in $L^2(\mathbb{R}^k)$ with the same norm. By the density of smooth compactly supported functions in $L^2(\mathbb{R}^k)$ there is a smooth compactly supported $\widetilde{F}$ with $\|F_0 - \widetilde{F}\|_{L^2(\mathbb{R}^k)} < \delta/3$; restricting the integration to $\mathcal{R}_k$ only decreases the left-hand side. Being continuous with compact support, $\widetilde{F}$ is bounded on $\mathcal{R}_k$; let $M \ge 0$ be a bound for $|\widetilde{F}|$ there.

Choose $\eta \in (0, 1/(2k))$ so small that $M\sqrt{2k\eta} < \delta/3$, and let χ be the cut-off supplied by Lemma 5.18 for this η. Put $F_1 := \widetilde{F} \cdot \chi$. Then F_1 is smooth, and its closed support is contained in that of χ, hence in the closed set $\mathcal{R}_k$. By part (iii) of Lemma 5.18 applied with $G = \widetilde{F}$,

$$\|\widetilde{F} - F_1\|_{L^2(\mathcal{R}_k)} \;\le\; M\sqrt{2k\eta} \;<\; \delta/3,$$

and the triangle inequality gives $\|F_0 - F_1\|_{L^2(\mathcal{R}_k)} < 2\delta/3 < \delta$. □

Uses (proof): 5.18

Lemma 5.20 (Smooth approximation of the variational quotient)**.** *Let $k \ge 2$, let $f \in L^2(\mathcal{R}_k)$ with $I_k(f) > 0$, and let $\eta > 0$. Then there is $F \in C^\infty(\mathbb{R}^k)$ with $\operatorname{supp} F \subset \mathcal{R}_k$, $I_k(F) > 0$ and*

$$\left| \frac{\sum_{m=1}^{k} J_k^{(m)}(F)}{I_k(F)} \;-\; \frac{\sum_{m=1}^{k} J_k^{(m)}(f)}{I_k(f)} \right| \;<\; \eta.$$

Lean:

- `PrimeGaps.exists_smooth_ratio_close`

Uses (statement): 5.2, 5.6, 5.7

Uses(notation): $\mathcal{R}_k$, k, F, $I_k(F)$, $J_k^{(m)}(F)$, $\|f\|_{L^2(X)}$, $C^\infty(X)$ (index)

Proof. Write $\mathcal{Q}(g) := \big(\sum_m J_k^{(m)}(g)\big)/I_k(g)$ for the variational quotient, defined on the set where $I_k \neq 0$. Its numerator is a finite sum of the functionals $J_k^{(m)}$, continuous by Lemma 5.15, and its denominator I_k is continuous by Lemma 5.14 and is non-zero at f; hence $\mathcal{Q}$ is continuous at f.

Choose $\delta_1 > 0$ such that $\|g - f\|_{L^2(\mathcal{R}_k)} < \delta_1$ implies $|\mathcal{Q}(g) - \mathcal{Q}(f)| < \eta$, and, using continuity of I_k at f together with $I_k(f) > 0$, choose $\delta_2 > 0$ such that $\|g - f\|_{L^2(\mathcal{R}_k)} < \delta_2$ implies $I_k(g) > 0$. Apply Lemma 5.19 with $\delta := \min(\delta_1, \delta_2)$ to obtain a smooth F supported on $\mathcal{R}_k$ with $\|f - F\|_{L^2(\mathcal{R}_k)} < \delta$. Both conclusions hold for this F. □

Uses (proof): 5.14, 5.15, 5.19

5.6 The main inequality on S

Lemma 5.21 (Smooth near-maximisers for $M_{k,\varepsilon}$)**.** *Let $k \geq 2$, let $\varepsilon \in [0, 1)$ and let $\delta > 0$. Then there exists a smooth $F : \mathbb{R}^k \to \mathbb{R}$ whose closed support is contained in $\mathcal{T}_\varepsilon$, with $I_k(F) > 0$ and*

$$(M_{k,\varepsilon} - \delta)\, I_k(F) \;<\; \sum_{m=1}^{k} J_{k,\varepsilon}^{(m)}(F).$$

Lean:

- `PrimeGaps.M_pos_and_le_and_exists_smooth_lt_sum_J`

Uses (statement): 5.8, 5.6, 5.7, 5.2

Uses(notation): M_k, $I_k(F)$, $J_k^{(m)}(F)$, F, k, $\mathcal{R}_k$, $\mathcal{T}_\varepsilon$, ε, δ, $C^\infty(X)$ (index)

Proof. Step 1: smooth functions supported in $\mathcal{T}_\varepsilon$ are L^2-dense there. Let $\sigma := 1 + \varepsilon > 0$ and let Λ_ε be the dilation

$$(\Lambda_\varepsilon G)(t) \;:=\; \sigma^{-k/2}\, G(\sigma^{-1} t).$$

Since $\sigma^{-1}\mathcal{T}_\varepsilon = \mathcal{R}_k$, the map Λ_ε carries functions supported in $\mathcal{R}_k$ to functions supported in $\mathcal{T}_\varepsilon$, preserves smoothness, and is an isometry of $L^2(\mathcal{R}_k)$ onto $L^2(\mathcal{T}_\varepsilon)$ by the change of variables $t = \sigma u$. Composing Lemma 5.19 with Λ_ε shows: for every $H_0 \in L^2(\mathcal{T}_\varepsilon)$ and every $\delta' > 0$ there is a smooth H_1 with closed support in $\mathcal{T}_\varepsilon$ and $\|H_0 - H_1\|_{L^2(\mathcal{T}_\varepsilon)} < \delta'$.

Step 2: transfer along the Rayleigh quotient. By Lemma 5.17 we have $M_{k,\varepsilon} > 0$, and by Lemma 5.16 the supremum is finite. Write $Q(G) := \big(\sum_m J_{k,\varepsilon}^{(m)}(G)\big)/I_k(G)$. Since $M_{k,\varepsilon}$ is the supremum of Q, there is $G_0 \in L^2(\mathcal{T}_\varepsilon)$ with

$$\max\Big(M_{k,\varepsilon} - \tfrac{\delta}{2},\; \tfrac{M_{k,\varepsilon}}{2}\Big) \;<\; Q(G_0),$$

and in particular $I_k(G_0) > 0$, since a competitor must have non-zero mass. Both $\sum_m J_{k,\varepsilon}^{(m)}$ and I_k are continuous on $L^2(\mathcal{T}_\varepsilon)$ (Lemmas 5.14 and 5.15), so Q is continuous at G_0 and I_k stays positive near G_0: there is $\delta' > 0$ such that every H with $\|G_0 - H\| < \delta'$ satisfies $I_k(H) > 0$ and $|Q(H) - Q(G_0)| < \delta/2$.

By Step 1 choose such an $H = F$ smooth with closed support in $\mathcal{T}_\varepsilon$. Then $I_k(F) > 0$ and $Q(F) > Q(G_0) - \delta/2 > M_{k,\varepsilon} - \delta$; multiplying by $I_k(F) > 0$ gives the asserted inequality. □

Uses (proof): 5.17, 5.16, 5.14, 5.15, 5.19, 5.20

The following lemma is used by both cases of the argument. It converts a lower bound on $M_{k,\varepsilon}$ into the strict inequality

$$\rho\, I_k(F) \;<\; \Big(\frac{\vartheta}{2} - \delta\Big) \sum_{m=1}^{k} J_{k,\varepsilon}^{(m)}(F)$$

for a *smooth* profile F — exactly the hypothesis under which the S_1 and $S_2^{(m)}$ asymptotics force $S_2(\lambda_0) - \rho\, S_1(\lambda_0) > 0$ for all large N. Taking $\varepsilon = 0$ gives the standard criterion; taking $\varepsilon > 0$ gives the enlarged one.

Lemma 5.22 (The variational criterion)**.** *Let $k \geq 2$, let $\varepsilon \in [0, 1)$, let $\vartheta \in (0, 1)$ and let $\rho > 0$. Suppose that*

$$M_{k,\varepsilon} \;>\; \frac{2\rho}{\vartheta}.$$

Then there exist a smooth $F : \mathbb{R}^k \to \mathbb{R}$ whose closed support is contained in $\mathcal{T}_\varepsilon$ and which satisfies $I_k(F) > 0$, and a real $\delta \in (0, \vartheta/2)$, such that

$$\rho\, I_k(F) \;<\; \Big(\frac{\vartheta}{2} - \delta\Big) \sum_{m=1}^{k} J_{k,\varepsilon}^{(m)}(F).$$

Lean:

- `PrimeGaps.maynard_smooth_witness_ineq`

Uses (statement): 5.8, 5.6, 5.7, 5.2

Uses(notation): M_k, $\mathcal{Q}_\varepsilon(F)$, $I_k(F)$, $J_k^{(m)}(F)$, F, k, $\mathcal{T}_\varepsilon$, ε, ϑ, ρ, δ, $C^\infty(X)$ (index)

Proof. Write $M := M_{k,\varepsilon}$, so that $2\rho/\vartheta < M$ by hypothesis, and set

$$\eta \;:=\; \frac{\vartheta}{2}\Big(M - \frac{2\rho}{\vartheta}\Big) \;>\; 0, \qquad \text{so that} \qquad \rho \;=\; \frac{\vartheta M}{2} - \eta.$$

Apply Lemma 5.21 with tolerance $\delta_0 := \min\big(\eta/\vartheta,\ M/2\big) > 0$: it produces a smooth F with closed support in $\mathcal{T}_\varepsilon$, $I_k(F) > 0$, and

$$Q \;:=\; \frac{\sum_m J_{k,\varepsilon}^{(m)}(F)}{I_k(F)} \;>\; M - \delta_0.$$

Since $\delta_0 \leq M/2$ we get $Q > M/2 > 0$.

Put $\delta := \eta/(2Q)$, which is positive. From $Q > M - \delta_0 \geq M - \eta/\vartheta$ we get $\vartheta(M - Q) < \eta$, hence

$$\frac{\vartheta M}{2} - \eta \;<\; \frac{\vartheta Q}{2} - \frac{\eta}{2} \;=\; \Big(\frac{\vartheta}{2} - \frac{\eta}{2Q}\Big) Q \;=\; \Big(\frac{\vartheta}{2} - \delta\Big) Q.$$

Multiplying by $I_k(F) > 0$ and using $Q\, I_k(F) = \sum_m J_{k,\varepsilon}^{(m)}(F)$ gives

$$\rho\, I_k(F) \;<\; \Big(\frac{\vartheta}{2} - \delta\Big) \sum_{m=1}^{k} J_{k,\varepsilon}^{(m)}(F).$$

Finally $\delta < \vartheta/2$: the displayed inequality has strictly positive left-hand side (as $\rho > 0$ and $I_k(F) > 0$), and $\sum_m J_{k,\varepsilon}^{(m)}(F) \geq 0$ since each term is an integral of a square, so the factor $\vartheta/2 - \delta$ must be strictly positive. □

Uses (proof): 5.21, 5.20

Because $M_{k,\varepsilon}$ is defined as a supremum, the hypothesis of Lemma 5.22 is equivalent to the existence of a single competitor $F_0 \in L^2(\mathcal{T}_\varepsilon)$ with $I_k(F_0) \neq 0$ and $\sum_m J^{(m)}_{k,\varepsilon}(F_0) > (2\rho/\vartheta)\, I_k(F_0)$. Both cases use it: one supplies a polynomial competitor on $\mathcal{R}_k$ at $k = 105$, $\rho = 2$, the other a certified competitor on $\mathcal{T}_{1/25}$ at $k = 50$, $\rho = 1$.

5.7 Lower bounds for $M_{k,\varepsilon}$: polynomial profiles

A lower bound $M_{k,\varepsilon} > c$ is obtained by exhibiting a single competitor. On the standard simplex the competitors used are the profiles F_P of Definition 5.4 attached to symmetric polynomials

$$P \;=\; \sum_{i=1}^{d} a_i\,(1-P_1)^{b_i} P_2^{c_i}, \qquad (b_i, c_i) \in \mathcal{B}_{d'},\ a_i \in \mathbb{R},$$

in the monomial family of Definition 5.5. For such profiles both $I_k(F_P)$ and $\sum_m J^{(m)}_k(F_P)$ are quadratic forms in the coefficient vector with explicit rational entries; the ratio is then a generalised Rayleigh quotient, whose supremum over the coefficients is the largest generalised eigenvalue. Verifying $M_k > c$ therefore reduces to a finite rational computation.

Lemma 5.23 (Dirichlet's integral over the simplex)**.** *For all $k \in \mathbb{N}$ and all $a, a_1, \dots, a_k \in \mathbb{Z}_{\geq 0}$,*

$$\int_{\mathcal{R}_k} \Big(1 - \sum_{i=1}^{k} t_i\Big)^{a} \prod_{i=1}^{k} t_i^{a_i}\, dt_1 \cdots dt_k \;=\; \frac{a!\,\prod_{i=1}^{k} a_i!}{\big(k + a + \sum_{i=1}^{k} a_i\big)!}.$$

Lean:

- `EuclideanSpace.dirichlet_integral`

Uses (statement): 5.2

Uses(notation): $\mathcal{R}_k$, k, $\Gamma(s)$, $B(a,b)$ (index)

Proof. One proves the scaled form: for every real $s \geq 0$,

$$\int_{\mathcal{R}_k(s)} \Big(s - \sum_{i} t_i\Big)^{a} \prod_{i} t_i^{a_i}\, dt \;=\; s^{\,k+a+\sum_i a_i}\, \frac{a!\,\prod_i a_i!}{\big(k + a + \sum_i a_i\big)!},$$

by induction on k; the statement of the lemma is the case $s = 1$. (The scaled form is also what is needed to integrate over $\mathcal{T}_\varepsilon$ and $\mathcal{S}_\varepsilon$.)

For $k = 0$ both sides equal $s^a\, a!/a! = s^a$. For the inductive step, the integrand is continuous on the compact set $\mathcal{R}_{k+1}(s)$, hence integrable, and Fubini's theorem lets us peel off the first coordinate:

$$\int_{\mathcal{R}_{k+1}(s)} \cdots \;=\; \int_0^{s} t_1^{a_1} \Bigg(\int_{\mathcal{R}_k(s-t_1)} \Big((s-t_1) - \sum_{i\geq 2} t_i\Big)^{a} \prod_{i\geq 2} t_i^{a_i}\, dt_2 \cdots dt_{k+1} \Bigg) dt_1.$$

By the inductive hypothesis applied with scale $s - t_1$, the inner integral equals $(s - t_1)^{\,k+a+\sum_{i\geq 2} a_i}\, a! \prod_{i\geq 2} a_i!/(k + a + \sum_{i\geq 2} a_i)!$. The remaining one-dimensional integral is a Beta integral: for $A, e \in \mathbb{Z}_{\geq 0}$ and $s \geq 0$,

$$\int_0^{s} (s-t)^{A}\, t^{e}\, dt \;=\; s^{\,A+e+1}\, \frac{A!\,e!}{(A+e+1)!},$$

which follows from $B(A+1, e+1) = A!\,e!/(A+e+1)!$ after the substitution $t = su$. Combining the two factorial expressions and simplifying gives the scaled formula in dimension $k+1$. □

Lemma 5.24 (Power-sum moments over the simplex)**.** *For all $k \in \mathbb{N}$ and all $a, b \in \mathbb{Z}_{\geq 0}$,*

$$\int_{\mathcal{R}_k} (1 - P_1)^a \, P_2^b \, dt_1 \cdots dt_k \;=\; \frac{a!\, G_{b,2}(k)}{(k + a + 2b)!},$$

where P_1 and P_2 are the power sums of Notation 5.1 and $G_{b,j}$ is the polynomial of Definition 5.3.

Lean:

- `PrimeGaps.lem_integration_formula`

Uses (statement): 5.2, 5.1, 5.3

Uses(notation): $\mathcal{R}_k$, k, P_j, $G_{b,j}(x)$ (index)

Proof. Expand $P_2^b = \left(\sum_i t_i^2\right)^b$ by the multinomial theorem:

$$\left(\sum_{i=1}^{k} t_i^2\right)^b \;=\; \sum_{\substack{v \in \mathbb{Z}^k_{\geq 0} \\ v_1 + \cdots + v_k = b}} \frac{b!}{\prod_i v_i!} \prod_{i=1}^{k} t_i^{2v_i}.$$

Integrating term by term over $\mathcal{R}_k$ and applying Lemma 5.23 with exponents $a_i = 2v_i$ (so $\sum_i a_i = 2b$) gives

$$\int_{\mathcal{R}_k} (1 - P_1)^a P_2^b \;=\; \frac{a!}{(k + a + 2b)!} \sum_{\substack{v \in \mathbb{Z}^k_{\geq 0} \\ \sum_i v_i = b}} \frac{b!}{\prod_i v_i!} \prod_{i=1}^{k} (2v_i)!.$$

By Definition 5.3, $G_{b,2}(k) = b! \sum_{\sum_i v_i = b} \prod_i (2v_i)!/v_i!$, which is literally the displayed sum, since $b!/\prod_i v_i!$ times $\prod_i (2v_i)!$ equals $b! \prod_i (2v_i)!/v_i!$. □

Uses (proof): 5.23

Lemma 5.25 (The L^2 mass of a polynomial profile)**.** *Let $k \in \mathbb{N}$, let $d \in \mathbb{N}$, let $a_1, \ldots, a_d \in \mathbb{R}$ and let $(b_1, c_1), \ldots, (b_d, c_d) \in \mathbb{Z}^2_{\geq 0}$, and let P satisfy $P = \sum_{i=1}^{d} a_i (1 - P_1)^{b_i} P_2^{c_i}$. Let F_P be the profile of Definition 5.4. Then*

$$I_k(F_P) \;=\; \sum_{i=1}^{d} \sum_{j=1}^{d} a_i a_j \, \frac{(b_i + b_j)!\, G_{c_i + c_j, 2}(k)}{(k + b_i + b_j + 2c_i + 2c_j)!}.$$

Lean:

- `PrimeGaps.lem_Ik_quadratic_I`

Uses (statement): 5.6, 5.4, 5.5, 5.3, 5.1

Uses(notation): $I_k(F)$, F_P, $\mathcal{R}_k$, k, P_j, $G_{b,j}(x)$, $\mathcal{B}_d$ (index)

Proof. Since F_P vanishes off $\mathcal{R}_k$ and agrees with P on it, $I_k(F_P) = \int_{\mathcal{R}_k} P^2$. Expanding the square,

$$P^2 \;=\; \sum_i \sum_j a_i a_j \, (1 - P_1)^{b_i + b_j} \, P_2^{c_i + c_j},$$

using $(1 - P_1)^{b_i} (1 - P_1)^{b_j} = (1 - P_1)^{b_i + b_j}$ and likewise for P_2. The integrand is continuous on the compact set $\mathcal{R}_k$, so the finite sum may be integrated term by term, and each term is evaluated by Lemma 5.24 with $a = b_i + b_j$ and $b = c_i + c_j$. □

Uses (proof): 5.24

Lemma 5.26 (The marginal mass of a polynomial profile)**.** *Let $k, d \geq 1$, let $a_1, \dots, a_d \in \mathbb{R}$, let $(b_1, c_1), \dots, (b_d, c_d) \in \mathbb{Z}^2_{\geq 0}$, let $P = \sum_{i=1}^{d} a_i (1 - P_1)^{b_i} P_2^{c_i}$, and let F_P be the profile of Definition 5.4. Then, for every $m \in \{1, \dots, k\}$,*

$$J_k^{(m)}(F_P) \;=\; \sum_{i=1}^{d} \sum_{j=1}^{d} a_i a_j \sum_{c'=0}^{c_i} \sum_{c''=0}^{c_j} \binom{c_i}{c'} \binom{c_j}{c''} \frac{\gamma(b_i, b_j, c_i, c_j, c', c'')\; G_{c'+c'',\,2}(k-1)}{\big(k + b_i + b_j + 2c_i + 2c_j + 1\big)!},$$

where γ is the function given by

$$\gamma(b_i, b_j, c_i, c_j, c', c'') \;=\; \frac{b_i!\; b_j!\; (2c_i - 2c')!\; (2c_j - 2c'')!\; \big(b_i + b_j + 2c_i + 2c_j - 2c' - 2c'' + 2\big)!}{\big(b_i + 2c_i - 2c' + 1\big)!\; \big(b_j + 2c_j - 2c'' + 1\big)!}.$$

In particular $J_k^{(m)}(F_P)$ does not depend on m.

Lean:

- `PrimeGaps.lem_quad_forms`

Uses (statement): 5.7, 5.4, 5.5, 5.2, 5.3, 5.1

Uses(notation): $J_k^{(m)}(F)$, F_P, $\mathcal{R}_k$, $\mathcal{R}_k^{(m)}$, k, P_j, P_j', $G_{b,j}(x)$, $\mathcal{B}_d$ (index)

Proof. By Lemma 5.13,

$$J_k^{(m)}(F_P) \;=\; \int_{\mathcal{R}_k^{(m)}} \Big(\int_0^{1-P_1'} P(\dots, s, \dots)\, ds \Big)^2 \prod_{i \neq m} dt_i,$$

where $P_1' = \sum_{i \neq m} t_i$ and s occupies the m-th slot. Since P is symmetric, the value of the inner integral is independent of which slot is singled out, and the outer domain $\mathcal{R}_k^{(m)}$ is a copy of $\mathcal{R}_{k-1}$; this gives the asserted independence of m, and we may take $m = 1$.

The inner integral. Writing $P_2' = \sum_{i \neq m} t_i^2$, the i-th monomial of P evaluated at $(s, (t_i)_{i \neq m})$ is $a_i (1 - P_1' - s)^{b_i} (s^2 + P_2')^{c_i}$. Binomially expanding $(s^2 + P_2')^{c_i} = \sum_{c' \leq c_i} \binom{c_i}{c'} (P_2')^{c'} s^{2c_i - 2c'}$ and evaluating each one-dimensional integral by the Beta formula $\int_0^L (L - s)^B s^E \, ds = L^{B+E+1} B!\, E! / (B + E + 1)!$ with $L = 1 - P_1'$ gives

$$\int_0^{1-P_1'} P \, ds \;=\; \sum_{i=1}^{d} a_i \sum_{c'=0}^{c_i} \binom{c_i}{c'} (P_2')^{c'} \, (1 - P_1')^{\,b_i + 2c_i - 2c' + 1} \, \frac{b_i!\, (2c_i - 2c')!}{(b_i + 2c_i - 2c' + 1)!}.$$

Squaring. Multiplying this finite sum by itself produces the fourfold sum over (i, j, c', c''), with the (P_2')-powers combining to $(P_2')^{c'+c''}$, the $(1 - P_1')$-powers combining to $(1 - P_1')^{\,b_i + b_j + 2c_i + 2c_j - 2c' - 2c'' + 2}$, and the factorial prefactors combining to $\gamma(b_i, b_j, c_i, c_j, c', c'')$ divided by $\big(b_i + b_j + 2c_i + 2c_j - 2c' - 2c'' + 2\big)!$.

The outer integral. Integrating term by term over $\mathcal{R}_k^{(m)} \cong \mathcal{R}_{k-1}$ and applying Lemma 5.24 in dimension $k - 1$ with $a = b_i + b_j + 2c_i + 2c_j - 2c' - 2c'' + 2$ and $b = c' + c''$ replaces each term by

$$\frac{\big(b_i + b_j + 2c_i + 2c_j - 2c' - 2c'' + 2\big)!\; G_{c'+c'',\,2}(k-1)}{\big(k + b_i + b_j + 2c_i + 2c_j + 1\big)!},$$

the denominator being $(k - 1) + a + 2b$ factorial. The factorial in the numerator cancels the one introduced in the squaring step, leaving exactly the stated expression. □

Uses (proof): 5.13, 5.24

Lemma 5.27 (Positive definiteness of the Gram matrices)**.** *Let (X, μ) be a measure space, let $f_1, \dots, f_d : X \to \mathbb{R}$ be measurable with f_i^2 integrable, and suppose the family is linearly independent modulo null sets: whenever $\alpha \in \mathbb{R}^d$ satisfies $\sum_i \alpha_i f_i = 0$ μ-almost everywhere, then $\alpha = 0$. Let $c > 0$, and let A be the symmetric matrix satisfying*

$$A \;=\; (A_{ij})_{1 \le i,j \le d}, \qquad A_{ij} \;=\; c \int_X f_i f_j \, d\mu.$$

Then A is positive definite.

Lean:

- `MeasureTheory.lem_A1_posdef`

Uses(notation): $\|f\|_{L^2(X)}$, $\mathcal{R}_k$, $\mathcal{R}_k^{(m)}$, $I_k(F)$, $J_k^{(m)}(F)$, T_m, F_P, P_j, k (index)

Proof. Symmetry is clear. For $\alpha \in \mathbb{R}^d$, expanding the square and using integrability of each product $f_i f_j$ (a consequence of Cauchy–Schwarz and the integrability of the f_i^2),

$$\alpha^{\top} A \alpha \;=\; c \int_X \Big(\sum_{i=1}^{d} \alpha_i f_i \Big)^2 d\mu \;\ge\; 0.$$

If $\alpha^{\top} A \alpha = 0$ then, as $c > 0$ and the integrand is non-negative, $\sum_i \alpha_i f_i = 0$ almost everywhere, whence $\alpha = 0$ by hypothesis. So the form is strictly positive on $\mathbb{R}^d \setminus \{0\}$.

This is applied twice. Taking $X = \mathcal{R}_k$ with Lebesgue measure and f_i the monomials $(1-P_1)^{b_i} P_2^{c_i}$ shows that the matrix A_I whose quadratic form is $I_k(F_P)$ (Lemma 5.25) is positive definite. Taking $X = \mathcal{R}_k^{(m)}$ and f_i the marginals $T_m f_i$ for one fixed m shows that the matrix with entries $\int (T_m f_i)(T_m f_j)$ is positive definite; the remaining $k-1$ matrices in the sum $A_J := \big(\sum_m \int (T_m f_i)(T_m f_j)\big)_{ij}$, whose quadratic form is $\sum_m J_k^{(m)}(F_P)$ (Lemma 5.26), are positive semidefinite by the same computation, so A_J is positive definite. Only the independence of the marginals for a *single* m is needed. □

Uses (proof): 5.25, 5.26

Lemma 5.28 (The generalised Rayleigh quotient)**.** *Let $d \ge 1$ and let A_1, A_2 be real symmetric positive definite $d \times d$ matrices. Then there exists $\lambda \in \mathbb{R}$ with all of the following properties simultaneously:*

1. *$\alpha^{\top} A_2 \alpha \le \lambda \, \alpha^{\top} A_1 \alpha$ for every $\alpha \in \mathbb{R}^d$, with equality for some $\alpha_0 \ne 0$;*
2. *$\lambda = \sup\big\{ (\alpha^{\top} A_2 \alpha)/(\alpha^{\top} A_1 \alpha) : \alpha \ne 0 \big\}$;*
3. *$A_2 \alpha = \lambda A_1 \alpha$ for some $\alpha \ne 0$, and $\nu \le \lambda$ whenever $A_2 \alpha = \nu A_1 \alpha$ for some $\alpha \ne 0$;*
4. *λ is the largest element of the spectrum of $A_1^{-1} A_2$.*

Lean:

- `PrimeGaps.lem_rayleigh`

Uses(notation): $\|f\|_{L^2(X)}$ (index)

Proof. Since A_1 is positive definite it factors as $A_1 = LL^\top$ with L invertible. Put $B := L^{-1}A_2(L^{-1})^\top$; then B is symmetric, and it is positive definite because $\beta^\top B\beta = \alpha^\top A_2\alpha$ for $\alpha := (L^{-1})^\top\beta$, and $\beta \mapsto \alpha$ is a linear bijection. Substituting $\beta := L^\top\alpha$ turns the generalised quotient into an ordinary one:

$$\frac{\alpha^\top A_2\alpha}{\alpha^\top A_1\alpha} = \frac{\beta^\top B\beta}{\beta^\top\beta}, \qquad \alpha \neq 0 \iff \beta \neq 0.$$

By the spectral theorem B has an orthonormal eigenbasis with real eigenvalues; let λ be the largest of them and β_0 a corresponding unit eigenvector. Expanding β in the eigenbasis gives $\beta^\top B\beta \leq \lambda\,\beta^\top\beta$ for all β, with equality at β_0; this is (i), and (ii) follows since the quotient is bounded above by λ and attains it.

For (iii), $B\beta = \nu\beta$ with $\beta \neq 0$ is equivalent, under $\beta = L^\top\alpha$, to $A_2\alpha = \nu A_1\alpha$ with $\alpha \neq 0$; so the generalised eigenvalues of the pair (A_1, A_2) are exactly the eigenvalues of B, and λ is the largest. For (iv), $A_2\alpha = \nu A_1\alpha$ is equivalent to $(A_1^{-1}A_2)\alpha = \nu\alpha$ because A_1 is invertible, and for a square real matrix the spectrum consists precisely of its eigenvalues; so the spectrum of $A_1^{-1}A_2$ is the set of generalised eigenvalues, whose largest element is λ. □

The three preceding lemmas combine into the criterion by which every lower bound for $M_k = M_{k,0}$ is certified.

Proposition 5.29 (Polynomial certificates for lower bounds on M_k)**.** *Let $k \geq 1$ and $d \geq 1$, let $(b_1, c_1), \ldots, (b_d, c_d) \in \mathbb{Z}^2_{\geq 0}$ be pairwise distinct, and let $\alpha = (a_1, \ldots, a_d) \in \mathbb{R}^d$ be non-zero. Write $A(\alpha)$ and $B(\alpha)$ for the two quadratic forms of Lemmas 5.25 and 5.26, that is,*

$$A(\alpha) = \sum_{i,j} a_i a_j \frac{(b_i + b_j)!\; G_{c_i+c_j,\,2}(k)}{(k + b_i + b_j + 2c_i + 2c_j)!},$$

$$B(\alpha) = \sum_{i,j} a_i a_j \sum_{c'=0}^{c_i} \sum_{c''=0}^{c_j} \binom{c_i}{c'} \binom{c_j}{c''} \frac{\gamma(b_i, b_j, c_i, c_j, c', c'')\; G_{c'+c'',\,2}(k-1)}{(k + b_i + b_j + 2c_i + 2c_j + 1)!},$$

with γ the explicit rational constant of Lemma 5.26. If $c \in \mathbb{R}$ satisfies $c\,A(\alpha) < k\,B(\alpha)$, then $M_k > c$.

Uses (statement): 5.8, 5.4, 5.5, 5.3, 5.1, 5.25, 5.26

Uses(notation): M_k, $I_k(F)$, $J_k^{(m)}(F)$, F_P, $\mathcal{B}_d$, P_j, $G_{b,j}(x)$, k (index)

Proof. Let $P := \sum_i a_i(1 - P_1)^{b_i}P_2^{c_i}$ and let F_P be the associated profile, which vanishes outside $\mathcal{R}_k = \mathcal{T}_0$. The monomials $(1 - P_1)^{b_i}P_2^{c_i}$, for pairwise distinct index pairs, are linearly independent as functions on $\mathcal{R}_k$ modulo null sets; hence by Lemma 5.27 the Gram matrix A_I whose quadratic form is $A(\alpha)$ is positive definite, and therefore $A(\alpha) > 0$ because $\alpha \neq 0$. By Lemma 5.25, $I_k(F_P) = A(\alpha) > 0$, so F_P is an admissible competitor in Definition 5.8 at $\varepsilon = 0$. By Lemma 5.26 each of the k marginal masses equals $B(\alpha)$, so $\sum_{m=1}^{k} J_k^{(m)}(F_P) = k\,B(\alpha)$. Definition 5.8 therefore gives

$$M_k \geq \frac{k\,B(\alpha)}{A(\alpha)} > c,$$

the last step being the hypothesis divided by $A(\alpha) > 0$.

The hypothesis is an inequality between two rational numbers whenever α is rational, so it can be certified by an exact finite computation. The optimal constant obtainable this way from a fixed monomial family is, by Lemma 5.28 applied to the pair (A_I, A_J) — both positive definite by Lemma 5.27 — equal to k times the largest element of the spectrum of $A_I^{-1}A_J$. □

Uses (proof): 5.27, 5.28

The two explicit instances of this computation are carried out in the two later sections: on the standard simplex at $k = 105$, with the 42 monomials indexed by $\mathcal{B}_{11}$ and $c = 4$, giving $M_{105} > 4$; and on the ε-enlarged simplex at $k = 50$, $\varepsilon = 1/25$, with a larger basis, giving $M_{50,\,1/25} > 4$. In each case the bound is used in Lemma 5.22.

Lemma 5.30 (The Beta integral)**.** *Let $a, b \geq 1$ be integers. Then*

$$\int_0^1 t^{a-1}(1-t)^{b-1}\,dt \;=\; \frac{(a-1)!\,(b-1)!}{(a+b-1)!}.$$

Lean:

- `Nat.betaIntegral_eq_factorial_mul_div`
- `Real.lem_beta_eval`

Uses(notation): $B(a,b)$ (index)

Proof. The integral is the Euler Beta function $B(a,b)$ evaluated at positive integer arguments. For integers it satisfies $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$, and $\Gamma(n) = (n-1)!$ for $n \geq 1$; substituting gives the stated quotient of factorials. □

6 Auxiliary arithmetic estimates

This section collects the arithmetic estimates used in the sieve analysis: the elementary totient and Möbius identities, the Mertens-type sums over integers coprime to a modulus, the evaluation of the truncated Mertens sum $\sum_{u \leq R,\,(u,W)=1} \mu(u)^2/\varphi(u)$, the verification of the sieve-datum hypotheses for the two densities, the asymptotic for the summatory function $H(z)$ of the sieve weight, and the partial-summation lemma that converts H into a G-weighted sum.

Throughout, $\mathcal{W}$, V, m and M denote generic moduli; W and R are the primorial and the truncation of Definitions 3.29 and 3.37, and $\mathrm{e} = \exp(1)$ denotes the base of the natural logarithm (never a summation variable).

6.1 Elementary totient and Möbius identities

Lemma 6.1 (Multiplicativity formula for φ on a least common multiple)**.** *Let d and e be squarefree positive integers (each squarefree; the product de need not be). Then*

$$\varphi\big([d,e]\big)\,\varphi\big((d,e)\big) \;=\; \varphi(d)\,\varphi(e).$$

Lean:

- `PrimeGaps.lem_phi_lcm_product`

Uses (statement): 3.4

Uses(notation): φ, (a,b), $[a,b]$, $\mathbb{N}$ (index)

Proof. Since d and e are squarefree, so are $[d,e]$ and (d,e), and

$$\varphi(n) \;=\; \prod_{p \mid n}(p-1)$$

for every squarefree n. The prime divisors of $[d,e]$ are exactly those of d together with those of e, while the prime divisors of (d,e) are exactly those dividing both. Hence, counting each prime with the multiplicity with which it occurs on either side,

$$\prod_{p\mid[d,e]}(p-1)\cdot\prod_{p\mid(d,e)}(p-1)=\prod_{p\mid d}(p-1)\cdot\prod_{p\mid e}(p-1),$$

which is the assertion. □

Uses (proof): 3.4

Lemma 6.2 (Reciprocal LCM splitting for φ)**.** *Let d and e be squarefree positive integers. Then*

$$\frac{1}{\varphi([d,e])}=\frac{1}{\varphi(d)\,\varphi(e)}\sum_{u\mid(d,e)}g(u),$$

with g as in Definition 3.6.

Lean:
- `PrimeGaps.lem_lcm_phi_split`

Uses (statement): 3.4, 3.6, 6.1, 3.10

Uses(notation): φ, g, (a,b), $[a,b]$ (index)

Proof. By Lemma 6.1, $\varphi([d,e])=\varphi(d)\varphi(e)/\varphi((d,e))$, all the factors being positive. The integer (d,e) is squarefree, so Lemma 3.10 applies to it and gives $\varphi((d,e))=\sum_{u\mid(d,e)}g(u)$. Substituting and inverting yields the claim. □

Uses (proof): 6.1, 3.10

Lemma 6.3 (Reciprocal LCM splitting)**.** *For all positive integers d,e (no squarefreeness required),*

$$\frac{1}{[d,e]}=\frac{1}{de}\sum_{u\mid(d,e)}\varphi(u).$$

Lean:
- `PrimeGaps.lem_lcm_split`

Uses (statement): 3.4

Uses(notation): φ, (a,b), $[a,b]$ (index)

Proof. The classical identity $\sum_{u\mid n}\varphi(u)=n$, valid for every positive integer n, applied at $n=(d,e)$ turns the right-hand side into $(d,e)/(de)$. The claim is then the identity $[d,e]\,(d,e)=de$. □

Sublemma 6.4 (Divisor-sum form of $1/\varphi$)**.** *For every squarefree integer $u\geq 1$,*

$$\frac{1}{\varphi(u)}=\frac{1}{u}\sum_{d\mid u}\frac{\mu(d)^2}{\varphi(d)}.$$

Lean:
- `PrimeGaps.one_div_totient_eq_one_div_mul_sum_moebius_sq_div_totient`

Uses (statement): 3.4, 3.3

Uses(notation): μ, φ, $\mathbb{P}$ (index)

Proof. For squarefree u,

$$\frac{u}{\varphi(u)} = \prod_{p\mid u}\frac{p}{p-1} = \prod_{p\mid u}\Big(1+\frac{1}{p-1}\Big) = \sum_{d\mid u}\prod_{p\mid d}\frac{1}{p-1} = \sum_{d\mid u}\frac{\mu(d)^2}{\varphi(d)}.$$

The third equality expands the product over $p \mid u$ as a sum over the subsets of the prime divisors of u, i.e. over the divisors $d \mid u$; the fourth uses $\prod_{p\mid d} 1/(p-1) = \mu(d)^2/\varphi(d)$ on squarefree d. Dividing by u gives the claim. □

Sublemma 6.5 (Per-coordinate lcm-decomposition bijection)**.** *Let m be a squarefree positive integer, and let $\mathcal{D}_m$ and $\mathcal{C}_m$ be the sets*

$$\mathcal{D}_m = \{(d,e) : d,e \geq 1 \text{ squarefree}, [d,e] = m\}, \qquad \mathcal{C}_m = \{(c,a,b) : c,a,b \geq 1 \text{ pairwise coprime}, cab = m\}.$$

Then $(d,e) \mapsto \big((d,e), d/(d,e), e/(d,e)\big)$ maps $\mathcal{D}_m$ into $\mathcal{C}_m$ and is a bijection, with inverse $(c,a,b) \mapsto (ca, cb)$.

Lean:

- `PrimeGaps.phi_psi_id`

Uses(notation): (a,b), $[a,b]$, $\mathbb{N}$ (index)

Proof. Let $(d,e) \in \mathcal{D}_m$ and put $c := (d,e)$, $a := d/c$, $b := e/c$. Then $cab = (d,e)\,[d,e]/(d,e)\cdots$; more directly, $ca = d$, $cb = e$ and $cab = d\,b = d\,e/(d,e) = [d,e] = m$. The integers a and b are coprime because any common prime divisor would divide both d/c and e/c and hence, m being squarefree, already lie in c; the same squarefreeness of $m = cab$ forces c to be coprime to both a and b. So the map lands in $\mathcal{C}_m$.

Conversely let $(c,a,b) \in \mathcal{C}_m$ and put $d := ca$, $e := cb$. Both are divisors of the squarefree m, hence squarefree, and $(d,e) = c$ and $[d,e] = cab = m$ by pairwise coprimality; so $(d,e) \in \mathcal{D}_m$. The two constructions are mutually inverse: starting from (d,e) one recovers $\big((d,e)\big)\cdot\big(d/(d,e)\big) = d$ and likewise for e; starting from (c,a,b) one recovers $(ca,cb) \mapsto (c,a,b)$ because $(ca,cb) = c$ when $(a,b) = 1$. □

6.2 Divisor sums, Mertens-type bounds and convergent series

Definition 6.6 (The von Mangoldt function Λ)**.** Define $\Lambda : \mathbb{N} \to \mathbb{R}_{\geq 0}$ by

$$\Lambda(n) := \begin{cases} \log p, & n = p^a \text{ for some prime } p \text{ and some } a \geq 1, \\ 0, & \text{otherwise.} \end{cases}$$

Uses(notation): $\mathbb{P}$, $\mathbb{N}$ (index)

Lemma 6.7 (Mertens' first theorem, von Mangoldt form)**.** *For every integer $N \geq 1$,*

$$\Big|\sum_{d\leq N}\frac{\Lambda(d)}{d} - \log N\Big| \leq \log 4 + 5.$$

Uses (statement): 6.6

Uses(notation): $\mathbb{P}$, $\mathbb{N}$, $\vartheta(x)$ (index)

Proof. Step 1: a Chebyshev upper bound. For every $N \geq 0$,

$$\sum_{d \leq N} \Lambda(d) \;\leq\; (\log 4 + 4)\, N.$$

This is the classical elementary bound obtained from the central binomial coefficient: $\prod_{N<p\leq 2N} p$ divides $\binom{2N}{N} \leq 4^N$, so $\sum_{N<p\leq 2N} \log p \leq N \log 4$, and summing the resulting dyadic estimate together with the $O(\sqrt{N}\log N)$ contribution of the proper prime powers gives the stated linear bound with the explicit constant $\log 4 + 4$.

Step 2: the factorial identity. Since every $n \geq 1$ satisfies $\log n = \sum_{d \mid n} \Lambda(d)$, grouping by d gives

$$\sum_{n \leq N} \log n \;=\; \sum_{d \leq N} \Lambda(d) \Big\lfloor \frac{N}{d} \Big\rfloor.$$

Replacing $\lfloor N/d \rfloor$ by N/d costs at most $\sum_{d \leq N} \Lambda(d)$, which by Step 1 is at most $(\log 4 + 4)N$. Hence

$$\Big| \sum_{n \leq N} \log n \;-\; N \sum_{d \leq N} \frac{\Lambda(d)}{d} \Big| \;\leq\; (\log 4 + 4)\, N.$$

Step 3: Stirling-free evaluation of $\sum_{n \leq N} \log n$. Comparing the sum with $\int_1^N \log t \, dt = N \log N - N + 1$ by monotonicity of log gives $\big|\sum_{n \leq N} \log n - N \log N\big| \leq N$. Combining with Step 2 and dividing by N yields

$$\Big| \sum_{d \leq N} \frac{\Lambda(d)}{d} - \log N \Big| \;\leq\; \log 4 + 5. \qquad \square$$

Uses (proof): 6.6

Lemma 6.8 (Mertens-type estimate for $\sum \mu(u)^2 \tau_k(u)/\varphi(u)$)**.** *For every integer $k \geq 1$ there is a constant $C = C(k) > 0$ such that, for every real $z \geq 2$,*

$$\sum_{u \leq z} \frac{\mu(u)^2\, \tau_k(u)}{\varphi(u)} \;\leq\; C\, (\log z)^k.$$

Lean:

- `ArithmeticFunction.sum_moebius_sq_mul_tau_div_totient_le`
- `WeightedDivisorSum.weightedSum_le`

Uses (statement): 3.3, 3.4, 3.5

Uses(notation): μ, φ, τ_r, k, $\mathbb{P}$, $\mathbb{N}$ (index)

Proof. Rankin majorisation. For any non-negative $a : \mathbb{N} \to \mathbb{R}_{\geq 0}$, any real $Z \geq 1$ and any $\kappa \geq 0$ one has $(Z/u)^\kappa \geq 1$ for $u \leq Z$, so $\sum_{u \leq Z} a(u) \leq Z^\kappa \sum_{u \geq 1} a(u) u^{-\kappa}$. Apply this with $a(u) := \mu(u)^2 \tau_k(u)/\varphi(u)$ and $\kappa := 1/\log(\mathrm{e}z) \in (0, 1)$.

Euler factorisation. The majorant $u \mapsto \mu(u)^2 \tau_k(u)/(\varphi(u) u^\kappa)$ is non-negative, multiplicative and supported on squarefree integers; since $\tau_k(p) = k$ and $\varphi(p) = p - 1$, and all prime-power terms with exponent ≥ 2 vanish,

$$\sum_{u \geq 1} \frac{\mu(u)^2 \tau_k(u)}{\varphi(u) u^\kappa} \;=\; \prod_p \Big(1 + \frac{k}{(p-1)p^\kappa}\Big).$$

Logarithmic bound. Using $\log(1+x) \le x$ and splitting $1/(p-1) = 1/p + 1/(p(p-1))$,

$$\log \prod_p \Big(1 + \frac{k}{(p-1)p^\kappa}\Big) \;\le\; k \sum_p \frac{1}{p^{1+\kappa}} \;+\; k \sum_p \frac{1}{p(p-1)},$$

and the second sum converges absolutely. For the first, start from Lemma 6.7: discarding the proper prime powers (whose total contribution to $\sum_{d \le t} \Lambda(d)/d$ is $O(1)$, being bounded by $\sum_p \sum_{a \ge 2} (\log p)/p^a \ll \sum_p (\log p)/p^2$) gives the one-sided prime form $\sum_{p \le t} (\log p)/p \le \log t + O(1)$ for $t \ge 2$. Abel summation of this against the weight $1/\log t$ yields the elementary one-sided Mertens second bound $\sum_{p \le t} 1/p \le \log\log t + B$ with B absolute. A second Abel summation, now against the weight $t^{-\kappa}$, gives $\sum_p p^{-1-\kappa} = \log(1/\kappa) + O(1)$. With $\kappa = 1/\log(ez)$ this is $\log\log(ez) + O(1) \le \log\log z + O(1)$ for $z \ge 3$ (the range $2 \le z < 3$ being trivial). Only upper bounds on prime sums are used, so no Tauberian or Dirichlet-series input enters.

Conclusion. Exponentiating gives $\prod_p (1 + k(p-1)^{-1}p^{-\kappa}) \le C_1(k)(\log z)^k$, while $Z^\kappa = z^{1/\log(ez)} \le \mathrm{e}$. Combining with the Rankin majorisation proves the lemma. □

Uses (proof): 6.7

Lemma 6.9 (Truncated coprime divisor sum)**.** *For every integer $k \ge 1$ there is a constant $C = C(k) > 0$ such that, for every positive integer D and every real $Z \ge 2$,*

$$\sum_{\substack{r \le Z \\ (r,D)=1}} \frac{\mu(r)^2\, \tau_k(r)}{\varphi(r)} \;\le\; C\,(\log Z)^k.$$

The constant is independent of D.

Lean:

- `PrimeGaps.lem_truncated_divisor_sum`

Uses (statement): 3.3, 3.4, 3.5, 6.8

Uses(notation): μ, φ, τ_r, k, (a,b), $\mathbb{N}$ (index)

Proof. The summand is non-negative, so discarding the coprimality condition only increases the sum; the claim is then Lemma 6.8. (The point of the statement is the uniformity in D: the same constant serves for every modulus, which is what the downstream applications, where D varies with the sieve variables, require.) □

Uses (proof): 6.8

Lemma 6.10 (k-fold restricted reciprocal-totient sum)**.** *For every integer $k \ge 1$ there is a constant $C = C(k) > 0$ with the following property. For every real $R \ge 2$, every modulus $W \ge 1$ and every $\mathbf{d} = (d_1, \dots, d_k) \in \mathbb{Z}_{\ge 1}^k$,*

$$\Big(\prod_{i=1}^k d_i\Big) \sum_{\substack{r_1,\dots,r_k \ge 1 \\ d_i \mid r_i \ \forall i \\ \prod_i r_i \le R,\ \mu(\prod_i r_i)^2 = 1,\ (\prod_i r_i, W) = 1}} \prod_{i=1}^k \frac{1}{\varphi(r_i)} \;\le\; C\,(\log R)^k.$$

Uses (statement): 3.3, 3.4, 6.8

Uses(notation): μ, φ, k, R, W, $\mathbf{d}$, $\mathbf{r}$, (a,b), $\mathbb{N}$ (index)

Proof. Write $r_i = d_i s_i$ and $s' := \prod_i s_i$. Because $\prod_i r_i$ is squarefree, the d_i are pairwise coprime and coprime to s', and $\prod_i \varphi(r_i) = \varphi(\prod_i d_i)\varphi(s')$ by multiplicativity of φ on coprime factors. Hence the left-hand side is at most

$$\frac{\prod_i d_i}{\varphi(\prod_i d_i)} \sum_{\substack{s' \le R/\prod_i d_i \\ \mu(s')^2=1,\ (s', W\prod_i d_i)=1}} \frac{\tau_k(s')}{\varphi(s')},$$

since a squarefree s' arises from at most $\tau_k(s')$ tuples $(s_1, \dots, s_k)$ with $\prod_i s_i = s'$. For the prefactor, the Dirichlet expansion of Lemma 3.12 gives $\prod_i d_i/\varphi(\prod_i d_i) = \sum_{D \mid \prod_i d_i} \mu(D)^2/\varphi(D)$, which is bounded by $\sum_{D \mid \prod_i d_i} \mu(D)^2\tau_k(D)/\varphi(D)$; and every prime dividing $\prod_i d_i$ is at most R. The remaining s'-sum is a truncated coprime divisor sum at modulus $\big(W\prod_i d_i\big)$, so Lemma 6.9 bounds it by $C(k)(\log R)^k$ with a constant that does not depend on that modulus — the uniformity there is exactly what is needed, since the modulus varies with $\mathbf{d}$. Both factors are thus dominated by the truncated sum $\sum_{u \le R} \mu(u)^2\tau_k(u)/\varphi(u)$ of Lemma 6.8 — more precisely, the whole expression is bounded by the Euler product $\prod_{p \le R}\big(1 + k/(p-1)\big)$, and the argument of Lemma 6.8 bounds this by $C(k)(\log R)^k$ for $R \ge 2$. □

Uses (proof): 6.8, 3.12, 6.9

Lemma 6.11 (Convergent sum bound for φ)**.** $\displaystyle\sum_{s \ge 1} \frac{\mu(s)^2}{\varphi(s)^2} < \infty.$

Lean:

- `PrimeGaps.lem_convergent_sum_phi`

Uses (statement): 3.3, 3.4

Uses(notation): μ, φ, τ, $\mathbb{N}$ (index)

Proof. The summand is non-negative, so it suffices to majorise it by a convergent series. The inequality $s \le \varphi(s)\tau(s)$ established in the proof of Lemma 3.13 gives $1/\varphi(s) \le \tau(s)/s$, whence

$$\frac{\mu(s)^2}{\varphi(s)^2} \le \frac{\tau(s)^2}{s^2} = \frac{\tau(s)}{\sqrt{s}} \cdot \frac{\tau(s)}{s^{3/2}} \le \frac{2\,\tau(s)}{s^{3/2}},$$

the last step by the elementary bound $\tau(s) \le 2\sqrt{s}$ (divisors pair up as $u \leftrightarrow s/u$ about $\sqrt{s}$). The Dirichlet series $\sum_{s \ge 1} \tau(s) s^{-3/2} = \big(\sum_{s \ge 1} s^{-3/2}\big)^2$ converges since $3/2 > 1$, so the partial sums of the original series are uniformly bounded, and a series of non-negative terms with bounded partial sums converges. □

Uses (proof): 3.13

Lemma 6.12 (Convergent sum bound for g)**.** *With g the multiplicative function of Definition 3.6 ($g(p) = p - 2$),*

$$\sum_{\substack{s \ge 1 \\ (s,2)=1}} \frac{\mu(s)^2}{g(s)^2} < \infty,$$

the summand being read as 0 whenever $\mu(s)^2 = 0$; the sum is thus supported on odd squarefree s, for which $g(s) = \prod_{p \mid s}(p-2) > 0$.

Lean:

- `PrimeGaps.convergent_sum_g`

Uses (statement): 3.3, 3.6

Uses(notation): μ, g, (a, b), $\mathbb{P}$, $\mathbb{N}$ (index)

Proof. The summand is non-negative and multiplicative, supported on odd squarefree integers. Its local factor at an odd prime p is $1 + (p-2)^{-2}$, and at $p = 2$ it is 1. Since $(p-2)^{-2} \le 9p^{-2}$ for every odd prime p (indeed $p - 2 \ge p/3$ for $p \ge 3$), the series $\sum_p (p-2)^{-2}$ converges, so $\prod_{p \text{ odd}} (1 + (p-2)^{-2}) < \infty$. Every partial sum over a finite set of integers is bounded by this Euler product, whence convergence. □

6.3 The constants $A(M)$ and $B(M)$

Sublemma 6.13 (Euler product for $\zeta(2)$)**.** *With $\zeta(2) = \sum_{n \ge 1} n^{-2}$,*

$$\zeta(2) \;=\; \prod_{p} \Bigl(1 - \frac{1}{p^2}\Bigr)^{-1},$$

both the series and the product being convergent.

Lean:
- `PrimeGaps.eulerProduct_zetaTwo`
- `PrimeGaps.summable_zetaTwo`
- `PrimeGaps.multipliable_eulerProduct_zetaTwo`

Uses(notation): $\zeta(2)$, $\mathbb{P}$ (index)

Proof. By unique factorisation, every $n \ge 1$ has a unique representation $n = \prod_p p^{a_p}$ with $a_p \ge 0$ and only finitely many a_p non-zero, so $n^{-2} = \prod_p p^{-2a_p}$. The series $\sum_n n^{-2}$ converges absolutely (it is bounded by $1 + \int_1^\infty t^{-2}\,dt = 2$), which licenses reorganising it as a product over primes of the geometric series $\sum_{a \ge 0} p^{-2a} = (1 - p^{-2})^{-1}$; the rearrangement is justified by letting $P \to \infty$ in the partial product over $p \le P$, whose expansion exhausts all $n \ge 1$ in the limit. □

Definition 6.14 (The auxiliary Dirichlet sum $A(M)$)**.** For every positive integer M, set

$$A(M) \;:=\; \sum_{\substack{e \ge 1 \\ (e, M) = 1}} \frac{\mu(e)}{e^2}.$$

Lean:
- `PrimeGaps.A`

Uses (statement): 3.3

Uses(notation): $A(M)$, μ, (a, b), $\mathbb{P}$ (index)

Sublemma 6.15 (Absolute convergence and the bound $|A(M)| \le \zeta(2)$)**.** *For every positive integer M the series of Definition 6.14 converges absolutely, and $|A(M)| \le \zeta(2)$.*

Lean:
- `PrimeGaps.A_abs_le_zetaTwo`

Uses (statement): 6.14

Uses(notation): $A(M)$, μ, $\zeta(2)$, (a,b) (index)

Proof. Each summand satisfies $|\mu(e)/e^2| \le 1/e^2$, so

$$\sum_{\substack{e\ge 1\\(e,M)=1}} \left|\frac{\mu(e)}{e^2}\right| \le \sum_{e\ge 1}\frac{1}{e^2} = \zeta(2) < \infty.$$

This is absolute convergence, and the triangle inequality applied to the (absolutely convergent) sum gives $|A(M)| \le \zeta(2)$. □

Sublemma 6.16 (Closed form for $A(M)$ in terms of $\zeta(2)$)**.** *For every positive integer M (squarefreeness is not required),*

$$A(M) = \frac{1/\zeta(2)}{\prod_{p\mid M}(1-1/p^2)}.$$

Lean:
- `PrimeGaps.A_closed_form`

Uses (statement): 6.14, 6.13, 6.15

Uses(notation): $A(M)$, μ, $\zeta(2)$, (a,b), $\mathbb{P}$ (index)

Proof. By Sublemma 6.15 the defining series converges absolutely, so it may be expanded as an Euler product. Its local factor at a prime $p \mid M$ is 1 (only $e = 1$ contributes at that prime), and at a prime $p \nmid M$ it is $\sum_{a\ge 0}\mu(p^a)p^{-2a} = 1 - p^{-2}$, since $\mu(p^a) = 0$ for $a \ge 2$. Hence

$$A(M) = \prod_{p\nmid M}\Big(1-\frac{1}{p^2}\Big).$$

Splitting the product over all primes as $\prod_p = \prod_{p\mid M}\cdot\prod_{p\nmid M}$ and invoking $\prod_p(1-p^{-2}) = 1/\zeta(2)$ (Sublemma 6.13) gives the stated closed form. Note that only the set of prime divisors of M enters, which is why no squarefreeness hypothesis is needed. □

Uses (proof): 6.13, 6.15

Definition 6.17 (The auxiliary Dirichlet sum $B(M)$)**.** For every positive integer M, set

$$B(M) := \sum_{\substack{e\ge 1\\(e,M)=1}} \frac{\mu(e)\log e}{e^2}.$$

The series converges absolutely, since $|\mu(e)\log e/e^2| \le (\log e)/e^2$ and $\sum_{e\ge 1}(\log e)/e^2 < \infty$.

Lean:
- `PrimeGaps.B`

Uses (statement): 3.3

Uses(notation): $B(M)$, μ, (a,b), $\mathbb{P}$ (index)

6.4 Coprime harmonic and coprime-density sums

Definition 6.18 (The prime-cost quantity ℓ_m)**.** For a positive integer m, set

$$\ell_m \;:=\; \sum_{p\mid m} \frac{\log p}{p-1},$$

the sum running over the distinct prime divisors of m.

Lean:

- `PrimeGaps.ellV`

Uses(notation): ℓ_m, $\mathbb{P}$ (index)

Sublemma 6.19 (Coprime harmonic sum)**.** *There is an absolute constant $C>0$ such that, for every squarefree integer $m\ge 1$ and every real $Z\ge 1$,*

$$\Big|\sum_{\substack{n\le Z\\(n,m)=1}} \frac{1}{n} \;-\; \frac{\varphi(m)}{m}\big(\log Z+\gamma_0+\ell_m\big)\Big| \;\le\; C\,\frac{\tau(m)}{Z},$$

where γ_0 is the Euler–Mascheroni constant.

Lean:

- `PrimeGaps.coprime_reciprocal_sum_asymptotic`

Uses (statement): 6.18, 3.3, 3.4

Uses(notation): μ, φ, τ, (a,b), ℓ_m, γ_0 (index)

Proof. Detect the coprimality condition by $\mathbf{1}[(n,m)=1]=\sum_{d\mid(n,m)}\mu(d)$ and swap the two finite sums:

$$\sum_{\substack{n\le Z\\(n,m)=1}}\frac{1}{n} \;=\; \sum_{d\mid m}\mu(d)\sum_{\substack{n\le Z\\ d\mid n}}\frac{1}{n} \;=\; \sum_{d\mid m}\frac{\mu(d)}{d}\sum_{n'\le Z/d}\frac{1}{n'}.$$

The classical harmonic estimate $\sum_{n'\le Y}1/n'=\log Y+\gamma_0+O(1/Y)$ gives

$$\sum_{\substack{n\le Z\\(n,m)=1}}\frac{1}{n} \;=\; (\log Z+\gamma_0)\sum_{d\mid m}\frac{\mu(d)}{d} \;-\; \sum_{d\mid m}\frac{\mu(d)\log d}{d} \;+\; O\Big(\frac{\tau(m)}{Z}\Big),$$

the error using $\sum_{d\mid m}|\mu(d)|=\tau(m)$ for squarefree m. Now $\sum_{d\mid m}\mu(d)/d=\prod_{p\mid m}(1-1/p)=\varphi(m)/m$, and differentiating the identity $\sum_{d\mid m}\mu(d)d^{-s}=\prod_{p\mid m}(1-p^{-s})$ at $s=1$ gives

$$-\sum_{d\mid m}\frac{\mu(d)\log d}{d} \;=\; \prod_{p\mid m}\Big(1-\frac{1}{p}\Big)\sum_{p\mid m}\frac{p^{-1}\log p}{1-p^{-1}} \;=\; \frac{\varphi(m)}{m}\,\ell_m.$$

Combining the three displays proves the claim, with an absolute implied constant. □

Lemma 6.20 (Mertens sum $\sum 1/n$ restricted to $(n,\mathcal{W})=1$)**.** *There is an absolute constant $C>0$ with the following property. Let $\mathcal{W}\ge 1$ be a squarefree integer and let x be a real number with*

$$x \;\ge\; \frac{\mathcal{W}\,\tau(\mathcal{W})}{\varphi(\mathcal{W})}.$$

Then

$$\Big|\sum_{\substack{n\le x\\(n,\mathcal{W})=1}}\frac{1}{n} \;-\; \frac{\varphi(\mathcal{W})}{\mathcal{W}}\log x\Big| \;\le\; C\,\frac{\varphi(\mathcal{W})}{\mathcal{W}}\big(1+\log\log(e\mathcal{W})\big).$$

Lean:
- `PrimeGaps.mertens_reciprocal_general`

Uses (statement): 3.3, 3.4, 6.19, 6.18

Uses(notation): μ, φ, τ, (a,b), $\mathbb{P}$, γ_0, ℓ_m (index)

Proof. Sublemma 6.19 applied with $m = \mathcal{W}$ and $Z = x$ gives

$$\sum_{\substack{n\le x\\(n,\mathcal{W})=1}} \frac{1}{n} \;=\; \frac{\varphi(\mathcal{W})}{\mathcal{W}}\big(\log x+\gamma_0+\ell_{\mathcal{W}}\big) \;+\; O\Big(\frac{\tau(\mathcal{W})}{x}\Big).$$

The hypothesis on x is exactly $\tau(\mathcal{W})/x \le \varphi(\mathcal{W})/\mathcal{W}$, so the error term is $O(\varphi(\mathcal{W})/\mathcal{W})$; the same bound covers the constant $\gamma_0\,\varphi(\mathcal{W})/\mathcal{W}$. It remains to bound $\ell_{\mathcal{W}} = \sum_{p\mid\mathcal{W}}(\log p)/(p-1)$. Split the prime divisors of $\mathcal{W}$ at $1+\log\mathcal{W}$. For $p > 1+\log\mathcal{W}$, use $\sum_{p\mid\mathcal{W}}\log p = \log\mathcal{W}$ (valid since $\mathcal{W}$ is squarefree) and $p-1 > \log\mathcal{W}$ to bound that part by 1; more crudely, the contribution is $O(1)$. For $p \le 1+\log\mathcal{W}$, use $(\log p)/(p-1) \le 2(\log p)/p$ and the elementary Mertens bound $\sum_{p\le t}(\log p)/p \le \log t + O(1)$ of Lemma 6.7 with $t = 1+\log\mathcal{W}$, which gives $O\big(\log(1+\log\mathcal{W})\big) = O\big(1+\log\log(\mathrm{e}\mathcal{W})\big)$. Hence $\ell_{\mathcal{W}} \ll 1+\log\log(\mathrm{e}\mathcal{W})$, and collecting all error contributions proves the lemma. The normalisation $1+\log\log(\mathrm{e}\mathcal{W})$ rather than $\log\log\mathcal{W}$ is used so that the error term is defined and non-negative for every $\mathcal{W}\ge 1$; for large $\mathcal{W}$ the two agree up to an additive constant. □

Uses (proof): 6.19, 6.18, 6.7

Definition 6.21 (The coprime density partial sum $A_m(x)$)**.** For a positive integer m and a real number x, set

$$A_m(x) \;:=\; \sum_{\substack{f\le x\\(f,m)=1}} \frac{\mu(f)^2}{\varphi(f)},$$

the sum running over the positive integers $f \le x$ coprime to m; thus $A_m(x) = 0$ for $x < 1$.

Lean:
- `PrimeGaps.MaynardOffDiagonal.sumA`

Uses (statement): 3.3, 3.4

Uses(notation): $A_m(x)$, μ, φ, (a,b), $\lfloor x\rfloor$ (index)

Sublemma 6.22 (Coprime density sum)**.** *There is an absolute constant $C > 0$ such that, for every squarefree integer $m \ge 1$ and every real $x \ge 2$,*

$$\Big|A_m(x) \;-\; \frac{\varphi(m)}{m}\big(\log x+\ell_m\big)\Big| \;\le\; C\Big(\frac{\varphi(m)}{m} \;+\; \frac{\tau(m)\log(2mx)}{\sqrt{x}}\Big).$$

Lean:
- `PrimeGaps.exists_abs_sumA_sub_le`

Uses (statement): 6.21, 6.18, 6.19

Uses(notation): $A_m(x)$, μ, φ, τ, (a,b), ℓ_m, γ_0, $\zeta(2)$ (index)

Proof. Step 1: a Dirichlet cofactor. Let κ be the multiplicative function defined by the Dirichlet convolution identity $\mu^2/\varphi = \kappa * (1/\mathrm{id})$, i.e.

$$\frac{\mu(f)^2}{\varphi(f)} \;=\; \sum_{de=f} \kappa(d)\,\frac{1}{e}.$$

Comparing local Euler factors shows $\kappa(1)=1$, $\kappa(p)=1/(p(p-1))$, $\kappa(p^2)=-1/(p(p-1))$ and $\kappa(p^j)=0$ for $j\ge 3$; in particular κ is supported on cube-free integers, its local factor $1+\kappa(p)+\kappa(p^2)$ equals 1 at every prime, and

$$\sum_{d\ge 1} |\kappa(d)|\,\sqrt{d} \;<\; \infty, \qquad \sum_{d\ge 1} |\kappa(d)|\log(2d) \;<\; \infty.$$

Both convergence statements follow from the Euler-product bound $\sum_{j\ge 0}|\kappa(p^j)|p^{j/2} \le \exp\bigl(2p^{-3/2}\bigr)$, together with the elementary partial-summation bound $\sum_{d\le X}|\kappa(d)|\,d \ll \sqrt{X}\log(2X)$, which in turn comes from writing a cube-free d uniquely as b^2a with a squarefree, $(a,b)=1$, and using $\sum_{b\le\sqrt{X}} 1/\varphi(b) \ll 1+\log X$. Consequently, for every $Y\ge 1$,

$$\sum_{d\ge Y} |\kappa(d)| \;\ll\; \frac{\log(2Y)}{\sqrt{Y}}, \qquad \sum_{d\ge Y} |\kappa(d)|\log(d/Y) \;\ll\; \frac{\log(2Y)}{\sqrt{Y}}$$

(a dyadic decomposition of the range $d\ge Y$).

Step 2: the hyperbola identity. Summing $\mu^2/\varphi = \kappa*(1/\mathrm{id})$ over $f\le x$ with $(f,m)=1$ and noting that $(f,m)=1$ is equivalent to $(d,m)=(e,m)=1$ for $f=de$,

$$A_m(x) \;=\; \sum_{\substack{d\le x\\(d,m)=1}} \kappa(d) \sum_{\substack{e\le x/d\\(e,m)=1}} \frac{1}{e}.$$

Step 3: insert the harmonic asymptotic. Apply Sublemma 6.19 to the inner sum for each d, obtaining the main term $\frac{\varphi(m)}{m}\bigl(\log(x/d)+\gamma_0+\ell_m\bigr)$ and an error $O(\tau(m)d/x)$; the errors sum, against $\sum_{d\le x}|\kappa(d)|d \ll \sqrt{x}\log(2x)$, to $O\bigl(\tau(m)\log(2x)/\sqrt{x}\bigr)$. For the main term, complete the d-sum to all $d\ge 1$ coprime to m; by Step 1 the discarded tail $d>x$ contributes $O\bigl(\tau(m)\log(2mx)/\sqrt{x}\bigr)$ (using $\ell_m \le \log m$ to absorb the m into the logarithm). Since κ has local mass 1 at every prime, $\sum_{(d,m)=1}\kappa(d)=1$, and the completed main term is

$$\frac{\varphi(m)}{m}\bigl(\log x+\gamma_0+\ell_m\bigr) \;-\; \frac{\varphi(m)}{m}\sum_{\substack{d\ge 1\\(d,m)=1}} \kappa(d)\log d.$$

The last sum is $O(1)$ by the second convergence statement of Step 1, and $\gamma_0\varphi(m)/m = O(\varphi(m)/m)$. Collecting all contributions gives the stated bound, with an absolute constant. □

Uses (proof): 6.19, 6.18, 6.21

6.5 The sum T_d and the Mertens sum over integers coprime to W

Definition 6.23 (The auxiliary sum T_d)**.** Let $R\ge 1$ be real and let $W\ge 1$ and $d\ge 1$ be integers. Set

$$T_d \;:=\; \sum_{\substack{v\le R/d\\(v,dW)=1}} \frac{\mu(v)^2}{v}.$$

The sum is empty, and hence $T_d=0$, once $d>R$.

Lean:
- `PrimeGaps.TInner`

Uses (statement): 3.3

Uses(notation): T_d, R, W, μ, (a,b), $\lfloor x \rfloor$ (index)

Sublemma 6.24 (Outer-sum decomposition for $\sum \mu^2/\varphi$)**.** *For every integer $R \geq 1$ and every integer $W \geq 1$,*

$$\sum_{\substack{u \leq R \\ (u,W)=1}} \frac{\mu(u)^2}{\varphi(u)} = \sum_{\substack{d \leq R \\ (d,W)=1}} \frac{\mu(d)^2}{d\,\varphi(d)}\, T_d,$$

with T_d as in Definition 6.23.

Lean:
- `PrimeGaps.sum_moebius_sq_div_totient_eq_sum_mul_TInner`

Uses (statement): 6.23, 3.3, 3.4

Uses(notation): μ, φ, T_d, R, W, (a,b) (index)

Proof. Substitute the identity $1/\varphi(u) = (1/u)\sum_{d \mid u} \mu(d)^2/\varphi(d)$ of Sublemma 6.4 — legitimate because $\mu(u)^2 \neq 0$ restricts the outer sum to squarefree u — and swap the two finite summations:

$$\sum_{\substack{u \leq R \\ (u,W)=1}} \frac{\mu(u)^2}{\varphi(u)} = \sum_{\substack{d \leq R \\ (d,W)=1}} \frac{\mu(d)^2}{\varphi(d)} \sum_{\substack{u \leq R,\ d \mid u \\ (u,W)=1}} \frac{\mu(u)^2}{u}.$$

Fix a squarefree d with $(d,W)=1$ and write $u = dv$ with $v \leq R/d$. On the support of $\mu(u)^2$ the integer u is squarefree, which for $u = dv$ is equivalent to v squarefree with $(v,d)=1$; and $(u,W)=1$ together with $(d,W)=1$ is equivalent to $(v,W)=1$. So the conditions on u become "v squarefree and $(v,dW)=1$". Since $\mu(u)^2/u = \mu(v)^2/(dv)$ on that support, the inner sum equals T_d/d, which is the claim. □

Uses (proof): 6.4

Sublemma 6.25 (Möbius squarefree-inversion step for T_d)**.** *For every real $R \geq 1$, every integer $W \geq 1$ and every integer $d \geq 1$ (no squarefreeness or coprimality is required),*

$$T_d = \sum_{\substack{e \leq \sqrt{R/d} \\ (e,dW)=1}} \frac{\mu(e)}{e^2} \sum_{\substack{w \leq R/(de^2) \\ (w,dW)=1}} \frac{1}{w}.$$

Lean:
- `PrimeGaps.T_decomposition`

Uses (statement): 6.23, 3.3

Uses(notation): T_d, R, W, μ, (a,b) (index)

Proof. Start from the squarefree-detection identity $\mu(v)^2 = \sum_{e^2 \mid v} \mu(e)$, valid for every positive integer v: both sides are multiplicative, and at a prime power p^a the right-hand side is 1 for $a \in \{0,1\}$ and $1+\mu(p) = 0$ for $a \geq 2$, matching $\mu(p^a)^2$. Insert it into $T_d = \sum_{v \leq R/d,\,(v,dW)=1} \mu(v)^2/v$ and swap the two finite sums. Writing $v = e^2 w$ turns the constraint $e^2 \mid v$ into a free w-sum with $w \leq R/(de^2)$, and the constraint $(v, dW) = 1$ into $(e, dW) = (w, dW) = 1$; the e-range is $e \leq \sqrt{R/d}$ because $e^2 \leq v \leq R/d$. Since $1/v = (1/e^2)(1/w)$, the displayed identity follows. (When $(e, dW) > 1$ the inner sum is empty, so restricting the e-sum to $(e, dW) = 1$ changes nothing.) □

Sublemma 6.26 (Inner w-sum via Mertens for $e \leq R^{1/8}$)**.** *There is a constant $C > 0$ such that, for all sufficiently large N, the following holds for every squarefree d with $(d, W) = 1$ and $d \leq R^{1/2}$ and every integer e with $1 \leq e \leq R^{1/8}$ and $(e, dW) = 1$. Let $\mathcal{W} = dW$. Then*

$$\Big| \sum_{\substack{w \leq R/(de^2) \\ (w,\mathcal{W})=1}} \frac{1}{w} \; - \; \frac{\varphi(\mathcal{W})}{\mathcal{W}} \big(\log R - \log d - 2\log e\big) \Big| \;\leq\; C\, \frac{\varphi(\mathcal{W})}{\mathcal{W}} \big(1 + \log\log(\mathrm{e}\mathcal{W})\big).$$

Lean:

- `PrimeGaps.slem_T_d_inner_w_sum`

Uses (statement): 3.37, 3.29, 6.20, 3.32

Uses(notation): R, W, φ, (a,b), N, $\lfloor x \rfloor$ (index)

Proof. Since d is squarefree, W is squarefree (Definition 3.29) and $(d, W) = 1$, the modulus $\mathcal{W} = dW$ is a positive squarefree integer. We apply Lemma 6.20 with this modulus and with $x := R/(de^2)$, and must verify $x \geq \mathcal{W}\tau(\mathcal{W})/\varphi(\mathcal{W})$.

Lower bound on x. From $d \leq R^{1/2}$ and $e \leq R^{1/8}$ we get $de^2 \leq R^{1/2} \cdot R^{1/4} = R^{3/4}$, hence $x = R/(de^2) \geq R^{1/4}$.

Upper bound on $\mathcal{W}\tau(\mathcal{W})/\varphi(\mathcal{W})$. For squarefree n the quantity $n\tau(n)/\varphi(n)$ factors over the primes dividing n, each contributing $2p/(p-1) \leq 4$; hence $\mathcal{W}\tau(\mathcal{W})/\varphi(\mathcal{W}) \leq 4^{\omega(\mathcal{W})}$. Next, for any fixed integer $B \geq 2$, splitting the prime divisors of n at B gives $\omega(n) \leq \pi(B) + \log n/\log B$, hence $4^{\omega(n)} \leq 4^{\pi(B)}\, n^{\log 4/\log B}$. Choose $B := 4^8 = 65536$, so that $\log 4/\log B = 1/8$ and, with $K := 4^{\pi(B)}$ an absolute constant,

$$\frac{\mathcal{W}\tau(\mathcal{W})}{\varphi(\mathcal{W})} \;\leq\; K\, \mathcal{W}^{1/8}.$$

By Lemma 3.32 we have $W \ll (\log\log N)^2$, and $d \leq R^{1/2}$, so $\mathcal{W} = dW \leq R^{1/2}(\log\log N)^2$ and therefore

$$\frac{\mathcal{W}\tau(\mathcal{W})}{\varphi(\mathcal{W})} \;\leq\; K\, R^{1/16}\, (\log\log N)^{1/4}.$$

Since $R = N^{\vartheta/2-\delta}$ is a fixed positive power of N, the right-hand side is $\leq R^{1/4} \leq x$ for all sufficiently large N, the factor $(\log\log N)^{1/4}$ being of lower order than $R^{3/16}$. This verifies the hypothesis.

With the hypothesis verified, Lemma 6.20 gives

$$\Big| \sum_{\substack{w \leq R/(de^2) \\ (w,\mathcal{W})=1}} \frac{1}{w} - \frac{\varphi(\mathcal{W})}{\mathcal{W}} \log \frac{R}{de^2} \Big| \leq C\, \frac{\varphi(\mathcal{W})}{\mathcal{W}} \big(1 + \log\log(\mathrm{e}\mathcal{W})\big),$$

and $\log\big(R/(de^2)\big) = \log R - \log d - 2\log e$. □

Uses (proof): 6.20, 3.32

Sublemma 6.27 (Evaluation of T_d)**.** *There is a constant $C > 0$ such that, for all sufficiently large N and every squarefree d with $(d, W) = 1$ and $d \le R^{1/2}$. Let $\mathcal{W} = dW$. Then*

$$\Big|T_d \;-\; \frac{\varphi(\mathcal{W})}{\mathcal{W}}\big[A(\mathcal{W})\log(R/d) \;-\; 2B(\mathcal{W})\big]\Big| \;\le\; C\,\frac{\varphi(\mathcal{W})}{\mathcal{W}}\big(\log(2d) + \log\log W\big) \;+\; C\,\frac{\log R}{R^{1/8}}.$$

Lean:

- `PrimeGaps.slem_T_d_eval`

Uses (statement): 6.23, 3.37, 3.29, 6.14, 6.17

Uses(notation): T_d, R, W, $A(M)$, $B(M)$, φ, μ, (a, b), N (index)

Proof. Sublemma 6.25 rewrites T_d as an e-sum weighted by $\mu(e)/e^2$ times an inner w-sum. Split the e-range at $e = R^{1/8}$.

Range $e \le R^{1/8}$. Sublemma 6.26 evaluates the inner sum for each such e, with error $O\big(\frac{\varphi(\mathcal{W})}{\mathcal{W}}(1 + \log\log(\mathrm{e}\mathcal{W}))\big)$. Summing the main term against $\mu(e)/e^2$ over $e \le R^{1/8}$ coprime to $\mathcal{W}$ produces

$$\frac{\varphi(\mathcal{W})}{\mathcal{W}}\Big[\log(R/d)\sum_{\substack{e\le R^{1/8}\\(e,\mathcal{W})=1}}\frac{\mu(e)}{e^2} \;-\; 2\sum_{\substack{e\le R^{1/8}\\(e,\mathcal{W})=1}}\frac{\mu(e)\log e}{e^2}\Big].$$

Completing both e-sums to infinity replaces them by $A(\mathcal{W})$ and $B(\mathcal{W})$ of Definitions 6.14 and 6.17; by the tail bounds $\big|A(M) - \sum_{e\le m}\big| \le 1/m$ and $\big|B(M) - \sum_{e\le m}\big| \le (\log m + 1)/m$ (each a comparison with $\sum_{e>m} e^{-2}$ and $\sum_{e>m}(\log e)e^{-2}$ respectively), applied at $m = \lfloor R^{1/8}\rfloor$, the cost is $O(\log R \cdot R^{-1/8})$. The accumulated error terms contribute $O\big(\frac{\varphi(\mathcal{W})}{\mathcal{W}}(1+\log\log(\mathrm{e}\mathcal{W}))\big)$ after summation against $\sum_e e^{-2} \le 2$.

Range $e > R^{1/8}$. Here use the trivial bounds $|\mu(e)/e^2| \le 1/e^2$ and $\sum_{w\le R/(de^2)} 1/w \le 1 + \log R$, so this part is

$$\ll \log R \sum_{e>R^{1/8}}\frac{1}{e^2} \;\ll\; \frac{\log R}{R^{1/8}}.$$

Normalising the error. Finally $\log\log(\mathrm{e}\mathcal{W}) = \log\log(\mathrm{e}dW) \ll \log(2d) + \log\log W$ for $W \ge \mathrm{e}$, which converts the middle error into the stated shape. Adding the three contributions gives the sublemma. □

Uses (proof): 6.25, 6.26, 6.14, 6.17

Sublemma 6.28 (Tail bound for $d > R^{1/2}$)**.** *There is a constant $C > 0$ such that, for all sufficiently large N,*

$$\Big|\sum_{\substack{d>\sqrt{R},\ d\ \text{squarefree}\\(d,W)=1}}\frac{\mu(d)^2}{d\,\varphi(d)}\,T_d\Big| \;\le\; C\,\frac{(\log R)^3}{\sqrt{R}}.$$

Lean:

- `PrimeGaps.tailSum_ll`

Uses (statement): 6.23, 3.37, 3.29

Uses(notation): T_d, μ, φ, R, W, (a, b), τ (index)

Proof. From $\mu(v)^2/v \le 1/v$ and $\sum_{v\le R/d} 1/v \le 1+\log R$ we get $0 \le T_d \le 1+\log R \ll \log R$ for $R \ge 2$, uniformly in d and W. By Lemma 3.13, $1/(d\varphi(d)) \le \tau(d)/d^2$. Lemma 6.8 with $k=2$ gives $\sum_{d\le X} \mu(d)^2\tau(d)/\varphi(d) \ll (\log X)^2$, hence also $\sum_{d\le X} \mu(d)^2\tau(d)/d \ll (\log X)^2$; partial summation converts this into the tail estimate

$$\sum_{d>Y} \frac{\mu(d)^2\tau(d)}{d^2} \ll \frac{(\log Y)^2+1}{Y}.$$

Applying it with $Y=\sqrt{R}$ bounds $\sum_{d>\sqrt{R}} \mu(d)^2/(d\varphi(d)) \ll (\log R)^2/\sqrt{R}$, and multiplying by the uniform bound $T_d \ll \log R$ gives the claim. □

Uses (proof): 3.13, 6.8

Sublemma 6.29 (Leading-coefficient telescoping identity)**.** *For every integer $W \ge 1$,*

$$\sum_{\substack{d\ge 1\\(d,W)=1}} \frac{\mu(d)^2}{d^2} A(dW) \;=\; 1.$$

Lean:

- `PrimeGaps.main_thm`

Uses (statement): 6.14, 6.16, 6.13, 6.15

Uses(notation): μ, $A(M)$, W, $\zeta(2)$, $\mathbb{P}$, (a,b) (index)

Proof. Write $f(d) := \mu(d)^2 A(dW)/d^2$. By Sublemma 6.16, $A(dW) = (1/\zeta(2))/\prod_{p\mid dW}(1-p^{-2})$, so for squarefree d coprime to W

$$f(d) \;=\; A(W)\prod_{p\mid d} \frac{1/p^2}{1-1/p^2}.$$

In particular $f(1)=A(W)$, and f is multiplicative up to the normalising constant $A(W)$: for coprime squarefree d_1,d_2 both coprime to W one has $f(d_1d_2)A(W) = f(d_1)f(d_2)$; and at a prime $p\nmid W$, $f(p) = p^{-2}A(W)/(1-p^{-2})$. Since the values of f at distinct primes multiply in this normalised sense, and $\mu(d)^2$ restricts the sum to squarefree d, each prime $p\nmid W$ contributes independently either 1 or $p^{-2}(1-p^{-2})^{-1}$, giving

$$\sum_{\substack{d\ge 1\\(d,W)=1}} \frac{\mu(d)^2}{d^2} A(dW) \;=\; A(W)\prod_{p\nmid W}\Big(1+\frac{1}{p^2}\cdot\frac{1}{1-1/p^2}\Big).$$

(The rearrangement is licensed by absolute convergence: $|A(dW)| \le \zeta(2)$ by Sublemma 6.15 and $\sum_d d^{-2} < \infty$.) The elementary identity $1+x/(1-x) = 1/(1-x)$ with $x=p^{-2}$ turns the local factor into $(1-p^{-2})^{-1}$, so the product equals $\prod_{p\nmid W}(1-p^{-2})^{-1}$. Combining with $A(W) = (1/\zeta(2))/\prod_{p\mid W}(1-p^{-2})$,

$$\sum_{\substack{d\ge 1\\(d,W)=1}} \frac{\mu(d)^2}{d^2} A(dW) \;=\; \frac{1/\zeta(2)}{\prod_{p\mid W}(1-p^{-2})\prod_{p\nmid W}(1-p^{-2})} \;=\; \frac{1/\zeta(2)}{\prod_p(1-p^{-2})} \;=\; 1,$$

the last equality by Sublemma 6.13. □

Uses (proof): 6.16, 6.15, 6.13

Lemma 6.30 (Mertens sum over u coprime to W)**.** *Let R and W be as in Definitions 3.37 and 3.29. There is a constant $C > 0$ such that, for all sufficiently large N,*

$$\Big| \sum_{\substack{u \le R \\ (u,W)=1}} \frac{\mu(u)^2}{\varphi(u)} - \frac{\varphi(W)}{W} \log R \Big| \le C \frac{\varphi(W)}{W} \log\log W.$$

Lean:

- `PrimeGaps.lem_mertens_W`

Uses (statement): 3.37, 3.29, 3.3, 3.4

Uses(notation): R, W, μ, φ, N, (a, b) (index)

Proof. By Sublemma 6.24 the left-hand sum equals $\sum_d \mu(d)^2 T_d/(d\varphi(d))$. Sublemma 6.28 bounds the contribution of $d > \sqrt{R}$ by $O\big((\log R)^3/\sqrt{R}\big)$, which is $o(\varphi(W)/W)$ for large N since $\varphi(W)/W \gg 1/\log\log N$ by Lemma 3.32. Restrict to $d \le \sqrt{R}$.

Insert the expansion of Sublemma 6.27 and use $\varphi(dW)/(dW) = (\varphi(d)/d)(\varphi(W)/W)$, valid because $(d, W) = 1$. This gives

$$\sum_{\substack{u \le R \\ (u,W)=1}} \frac{\mu(u)^2}{\varphi(u)} = \frac{\varphi(W)}{W}\Big[(\log R)\,\mathcal{S}_1 - \mathcal{S}_2 - 2\mathcal{S}_3\Big] + O\Big(\frac{\varphi(W)}{W}\log\log W\Big),$$

where, with all sums over squarefree d coprime to W,

$$\mathcal{S}_1 := \sum_d \frac{\mu(d)^2 A(dW)}{d^2}, \qquad \mathcal{S}_2 := \sum_d \frac{\mu(d)^2 (\log d) A(dW)}{d^2}, \qquad \mathcal{S}_3 := \sum_d \frac{\mu(d)^2 B(dW)}{d^2},$$

and the error term also absorbs the $\log(2d)$ contributions of Sublemma 6.27 (which sum, against $\mu(d)^2\varphi(d)/(d^2\varphi(d)) \le d^{-2}$, to $O(1)$) and the $O(\log R/R^{1/8})$ terms. By Sublemma 6.29, $\mathcal{S}_1 = 1$ once the d-sum is completed, and completing it costs $O(R^{-1/2})$ by Sublemma 6.15. The sums $\mathcal{S}_2$ and $\mathcal{S}_3$ are $O(1)$ in absolute value: each is bounded by $\zeta(2)\sum_{d\ge1}(1+\log d)/d^2 \ll 1$, again using $|A(M)|, |B(M)| \ll 1$ from Sublemma 6.15 and the corresponding trivial bound for B. Assembling and absorbing the $O(1)$ and $O(R^{-1/2}\log R)$ contributions into $O\big(\frac{\varphi(W)}{W}\log\log W\big)$ — legitimate since $\log\log W \ge 1$ for N large — gives the lemma. □

Uses (proof): 6.24, 6.28, 6.27, 6.29, 6.15, 3.32, 6.14, 6.17

6.6 The two sieve densities

Throughout this subsection (γ, V) denotes a sieve datum in the sense of Definition 4.3, with parameters A_1, A_2, A_3, L, and g_* is the derived totally multiplicative function of Definition 4.4. We write

$$\Delta(\gamma; w, z) := \sum_{w \le p \le z} \frac{\gamma(p)\log p}{p} - \log(z/w) \qquad (2 \le w \le z)$$

for the Mertens deviation of a density γ.

Sublemma 6.31 (g_* for the Maynard datum)**.** *For the Maynard sieve datum (Definition 4.29) and every squarefree $r \geq 1$,*

$$g_*(r) \;=\; \begin{cases} \dfrac{\mu(r)^2}{g(r)}, & (r, W) = 1, \\ 0, & (r, W) > 1, \end{cases}$$

with g the multiplicative function of Definition 3.6.

Lean:

- `PrimeGaps.maynardSieveDatum_gStar_squarefree_coprime`
- `PrimeGaps.maynardSieveDatum_gStar_not_coprime`

Uses (statement): 4.29, 4.4, 3.6, 3.29

Uses(notation): μ, W, g, g_*, γ_g, (a, b) (index)

Proof. Both sides are totally multiplicative in r, so it suffices to compare them at primes. If $p \mid W$ then $\gamma_g(p) = 0$, so $g_*(p) = 0$; total multiplicativity then forces $g_*(r) = 0$ whenever some prime divisor of r divides W, i.e. whenever $(r, W) > 1$. If $p \nmid W$ then $\gamma_g(p) = p/(p-1)$ and

$$g_*(p) \;=\; \frac{\gamma_g(p)}{p - \gamma_g(p)} \;=\; \frac{p/(p-1)}{p - p/(p-1)} \;=\; \frac{1}{p-2} \;=\; \frac{1}{g(p)}.$$

For squarefree r coprime to W we have $\mu(r)^2 = 1$, and multiplying the prime-level identities over the (distinct) primes dividing r gives $g_*(r) = 1/g(r)$. □

Uses (proof): 4.5

A family of densities $(\gamma_N)_{N \geq 1}$ is called *W-tricked* if each γ_N is multiplicative and non-negative with $\gamma_N(1) = 1$, if $\gamma_N(p) = 0$ for every prime $p \mid W$, and if there is a single constant $c > 0$, independent of N, with $|\gamma_N(p) - 1| \leq c/p$ for every prime $p \nmid W$. Both γ_g and $\gamma^{(m)}$ are W-tricked, with c absolute. The next two sublemmas supply the two Mertens-deviation constants that a sieve datum (Definition 4.3) requires.

Sublemma 6.32 (Uniform upper Mertens deviation A_2)**.** *Let $(\gamma_N)_{N \geq 1}$ be a W-tricked family with constant c. Then there is $A_2 > 0$, depending only on c and in particular independent of N, such that for every N with $D_0 \geq 2$ and all reals $2 \leq w \leq z$,*

$$\Delta(\gamma_N; w, z) \;\leq\; A_2.$$

Uses (statement): 3.29, 3.28, 6.7

Uses(notation): γ, γ_g, W, D_0, A_2, $T(z)$, $\mathbb{P}$ (index)

Proof. Step 1: compare γ_N with the constant density. For a prime $p \nmid W$ write $\gamma_N(p)(\log p)/p = (\log p)/p + (\gamma_N(p) - 1)(\log p)/p$; the second term is at most $c(\log p)/p^2$ in absolute value, and $\sum_p (\log p)/p^2$ converges. For $p \mid W$ the term vanishes altogether, and $p \mid W$ holds exactly when $p \leq D_0$. Hence there is $E = E(c)$ with

$$\Big| \sum_{w \leq p \leq z} \frac{\gamma_N(p) \log p}{p} \;-\; \sum_{\max(w,\, D_0) \leq p \leq z} \frac{\log p}{p} \Big| \;\leq\; E$$

for all $2 \leq w \leq z$.

Step 2: Mertens' first theorem. By Lemma 6.7, discarding the proper prime powers as in the proof of Lemma 6.8, there is an absolute M with $\left|\sum_{p\le t}(\log p)/p - \log t\right| \le M$ for $t \ge 2$; hence $\sum_{a\le p\le z}(\log p)/p \le \log(z/a) + 2M$ for all $2 \le a \le z$.

Step 3. Apply Step 2 with $a := \max(w, D_0) \ge w$. Since $a \ge w$ we have $\log(z/a) \le \log(z/w)$, so

$$\Delta(\gamma_N; w, z) \;\le\; E + \log(z/a) + 2M - \log(z/w) \;\le\; E + 2M.$$

Thus $A_2 := E(c) + 2M$ works, and it does not depend on N, w or z. (The gain over the two-sided estimate below is that truncating the prime range *upward* to $\max(w, D_0)$ can only decrease the sum, so the $\log D_0$ discrepancy appears with a favourable sign.) □

Uses (proof): 6.7, 6.8

Sublemma 6.33 (Lower Mertens deviation $L \ll \log D_0$)**.** *Let $(\gamma_N)_{N\ge1}$ be a W-tricked family with constant c. Then there is $C > 0$, depending only on c, such that for every N with $D_0 \ge 2$ there exists a real L with*

$$0 \;\le\; L \;\le\; C \log D_0 \qquad \text{and} \qquad -L \;\le\; \Delta(\gamma_N; w, z) \quad \text{for all } 2 \le w \le z.$$

Uses (statement): 3.29, 3.28, 6.7, 6.32

Uses(notation): γ, γ_g, W, D_0, L, $T(z)$, $\mathbb{P}$ (index)

Proof. Retain the notation of the previous proof. Steps 1 and 2 there give, for $a = \max(w, D_0)$,

$$\Delta(\gamma_N; w, z) \;\ge\; -E + \log(z/a) - 2M - \log(z/w) \;=\; -E - 2M - \log\frac{a}{w}.$$

Now $w \le a = \max(w, D_0) \le w\, D_0$ (using $w \ge 2$ and $D_0 \ge 2$), so $0 \le \log(a/w) \le \log D_0$. Hence $\Delta(\gamma_N; w, z) \ge -\big(\log D_0 + E + 2M\big)$. Since $D_0 \ge 2$ we have $\log D_0 \ge \log 2 > 0$, so $L := \log D_0 + E + 2M \le C \log D_0$ with $C := 1 + (E + 2M)/\log 2$, which depends only on c. Non-negativity of L is clear. □

Uses (proof): 6.32, 6.7

Sublemma 6.34 (Singular product for γ_g)**.** *There are constants $C > 0$ and N_0 such that, for every $N \ge N_0$ with $D_0 \ge 2$,*

$$\left|\mathfrak{S}(\gamma_g) \;-\; \frac{\varphi(W)}{W}\right| \;\le\; C\,\frac{\varphi(W)}{W}\cdot\frac{1}{D_0},$$

with $\mathfrak{S}$ as in Notation 3.41. Equivalently, $\mathfrak{S}(\gamma_g) = \frac{\varphi(W)}{W}\big(1 + O(1/D_0)\big)$.

Lean:

- `PrimeGaps.singularSeries_maynardSieveDatum_γ_asymptotic`

Uses (statement): 3.41, 4.29, 3.29, 3.28

Uses(notation): $\mathfrak{S}(\gamma)$, γ_g, φ, D_0, W, $\mathbb{P}$ (index)

Proof. Split the defining product $\mathfrak{S}(\gamma_g) = \prod_p (1-\gamma_g(p)/p)^{-1}(1-1/p)$ over $p \mid W$ and $p \nmid W$. For $p \mid W$ one has $\gamma_g(p) = 0$ and the local factor is $1 - 1/p$; the product of these over $p \mid W$ is $\varphi(W)/W$. For $p \nmid W$ one has $\gamma_g(p) = p/(p-1)$, so

$$\Big(1 - \frac{\gamma_g(p)}{p}\Big)^{-1}\Big(1 - \frac{1}{p}\Big) \;=\; \Big(1 - \frac{1}{p-1}\Big)^{-1}\frac{p-1}{p} \;=\; \frac{(p-1)^2}{p(p-2)} \;=\; 1 + \frac{1}{p(p-2)}.$$

Every such p exceeds D_0, so the tail product satisfies $1 \le \prod_{p>D_0}\big(1 + \frac{1}{p(p-2)}\big) \le \exp\big(\sum_{p>D_0}\frac{1}{p(p-2)}\big) \le \exp(2/D_0)$, using $\sum_{n>t} n^{-2} \le 2/t$ and $p(p-2) \ge p^2/3$ for $p \ge 3$. Since $\mathrm{e}^{2/D_0} - 1 \le 4/D_0$ for $D_0 \ge 4$, the tail product is $1 + O(1/D_0)$, and multiplying by $\varphi(W)/W$ gives the claim. □

6.7 The convolution defect and the asymptotic for H

Sublemma 6.35 (Size of the convolution defect)**.** *Let b be the convolution defect of Definition 4.8. For every prime $p \nmid V$,*

$$|b(p)| \;\le\; \frac{2A_3}{A_1}\cdot\frac{1}{p^2}.$$

Lean:

- `PrimeGaps.SieveDatum.bDefect_bound`

Uses (statement): 4.3, 4.8, 4.4

Uses(notation): γ, A_1, A_3, V, $b(p)$, g_*, $\mathbb{P}$ (index)

Proof. Fix a prime $p \nmid V$. The structural hypothesis of Definition 4.3 gives $|\gamma(p) - 1| \le A_3/p$, and Lemma 4.5 gives $p - \gamma(p) \ge A_1 p$. Then

$$b(p) \;=\; \frac{\gamma(p)}{p-\gamma(p)} - \frac{1}{p-1} \;=\; \frac{p\big(\gamma(p)-1\big)}{\big(p-\gamma(p)\big)(p-1)},$$

so, using $p - 1 \ge p/2$,

$$|b(p)| \;\le\; \frac{p\cdot A_3/p}{A_1 p\cdot p/2} \;=\; \frac{2A_3}{A_1 p^2}.$$ □

Uses (proof): 4.5

Definition 6.36 (The normalised defect $\tilde{b}$)**.** With b as in Definition 4.8, set

$$\tilde{b}(e) \;:=\; b(e)\,\frac{\varphi(e)}{e} \qquad (e \ge 1).$$

Like b, the function $\tilde{b}$ is multiplicative and supported on squarefree integers coprime to V, and $|\tilde{b}(e)| \le |b(e)|$ because $\varphi(e)/e \in [0,1]$.

Uses (statement): 4.8, 3.4

Uses(notation): $b(p)$, φ, V, $\mathbb{N}$ (index)

Sublemma 6.37 (Uniform summability of the defect)**.** *There is a function $K = K(A_1, A_3) > 0$, depending only on A_1 and A_3, such that for every sieve datum with those parameters*

$$\sum_{e\ge 1} |b(e)|\,\tau(e)\,\sqrt{e} \;\le\; K(A_1, A_3).$$

Uses (statement): 4.3, 4.8, 6.35

Uses(notation): $b(p)$, τ, A_1, A_3, V, $\mathbb{P}$, $\mathbb{N}$ (index)

Proof. The function $e \mapsto |b(e)|\tau(e)\sqrt{e}$ is non-negative, multiplicative, supported on squarefree e, and takes the value 1 at $e = 1$. Hence its sum over all e equals the Euler product $\prod_p\bigl(1 + |b(p)|\tau(p)\sqrt{p}\bigr)$, the primes dividing V contributing the factor 1. For $p \nmid V$, Sublemma 6.35 and $\tau(p) = 2$ give

$$|b(p)|\,\tau(p)\sqrt{p} \;\le\; \frac{4A_3}{A_1}\cdot\frac{1}{p^{3/2}}.$$

Since $\sum_p p^{-3/2}$ converges, $\log\prod_p(1+\cdot) \le \sum_p |b(p)|\tau(p)\sqrt{p} \le (4A_3/A_1)\sum_p p^{-3/2}$, and exponentiating gives a bound depending only on A_1 and A_3. (In particular the sum is finite, so the Euler factorisation used above is legitimate.) □

Uses (proof): 6.35

Sublemma 6.38 (Convolution identity for h)**.** *Let h be as in Definition 4.6, and let $a : \mathbb{N} \to \mathbb{R}$ satisfy $a(n) = \mu(n)^2/\varphi(n)$ for every n. For every squarefree $d \ge 1$ with $(d, V) = 1$,*

$$h(d) \;=\; \sum_{e \mid d} b(e)\, a(d/e).$$

Lean:

- `PrimeGaps.SieveDatum.h_eq_sum_bDefect_mul_a`

Uses (statement): 4.6, 4.8, 4.4, 3.4, 3.3

Uses(notation): $h(d)$, $a(d)$, $b(p)$, g_*, μ, φ, (a, b), V (index)

Proof. For squarefree d, each divisor $e \mid d$ is squarefree and coprime to d/e, so the right-hand side is the value at d of the Dirichlet convolution of two multiplicative functions restricted to squarefree arguments; it therefore factors as $\prod_{p\mid d}\bigl(b(p)+a(p)\bigr)$, each prime going wholly into e or into d/e. Since $(d, V) = 1$, every $p \mid d$ satisfies $p \nmid V$, so $a(p) = 1/(p-1)$ and $b(p) + a(p) = g_*(p)$ by Definition 4.8. Hence the product equals $\prod_{p\mid d} g_*(p) = g_*(d) = \mu(d)^2 g_*(d) = h(d)$, using total multiplicativity of g_* and $\mu(d)^2 = 1$. □

Sublemma 6.39 (Convolution form of H)**.** *Let H be as in Definition 4.28. For every real z,*

$$H(z) \;=\; \sum_{e \ge 1} b(e) \sum_{\substack{0 < f < z/e \\ (f, Ve) = 1}} \frac{\mu(f)^2}{\varphi(f)},$$

the e-sum having only finitely many non-zero terms (those with $e < z$).

Uses (statement): 4.28, 4.6, 4.8, 6.38, 6.21

Uses(notation): $H(z)$, $h(d)$, $b(p)$, $a(d)$, $A_m(x)$, μ, φ, (a, b), V (index)

Proof. First, h is supported on integers coprime to V: this is Lemma 4.7, whose proof is that for $p \mid V$ the structural hypothesis gives $\gamma(p) = 0$, hence $g_*(p) = 0$, and total multiplicativity of g_* forces $h(d) = \mu(d)^2 g_*(d) = 0$ whenever $(d, V) > 1$. Therefore $H(z) = \sum_{0<d<z,\ (d,V)=1} h(d)$, and only squarefree d contribute.

Now insert Sublemma 6.38 and exchange the order of the two finite sums, grouping by the divisor e:

$$H(z) \;=\; \sum_{\substack{0<d<z\\(d,V)=1,\ d\text{ squarefree}}} \;\sum_{e\mid d} b(e)\, a(d/e) \;=\; \sum_{e\ge 1} b(e) \sum_{\substack{0<f<z/e\\(f,Ve)=1}} a(f),$$

where $f = d/e$: the constraints "$d < z$, $e \mid d$" become "$f < z/e$", while "d squarefree, $(d, V) = 1$" becomes "e squarefree coprime to V" (already enforced by the support of b) together with "f squarefree, $(f, Ve) = 1$" (the condition $(f, e) = 1$ being enforced by $\mu(f)^2$ inside $a(f)$ when combined with the squarefreeness of $d = ef$). Terms with $e \ge z$ have empty inner sum. □

Uses (proof): 6.38, 4.7, 4.4, 4.3

Sublemma 6.40 (Euler-product identification of $\mathfrak{S}(\gamma)$)**.** *For every sieve datum,*

$$\mathfrak{S}(\gamma) \;=\; \frac{\varphi(V)}{V} \sum_{e\ge 1} \tilde{b}(e),$$

the series converging absolutely. In particular $\mathfrak{S}(\gamma) > 0$.

Uses (statement): 3.41, 6.36, 4.3, 6.37

Uses(notation): $\mathfrak{S}(\gamma)$, γ, $b(p)$, φ, V, $\mathbb{P}$ (index)

Proof. Absolute convergence follows from $|\tilde{b}(e)| \le |b(e)| \le |b(e)|\tau(e)\sqrt{e}$ and Sublemma 6.37. Since $\tilde{b}$ is multiplicative, supported on squarefree integers coprime to V, the sum factors as $\sum_{e\ge1} \tilde{b}(e) = \prod_{p\nmid V}\big(1 + \tilde{b}(p)\big)$. For $p \nmid V$,

$$1 + \tilde{b}(p) \;=\; 1 + \frac{p-1}{p}\Big(g_*(p) - \frac{1}{p-1}\Big) \;=\; \frac{p-1}{p}\big(1 + g_*(p)\big) \;=\; \Big(1 - \frac{1}{p}\Big)\Big(1 - \frac{\gamma(p)}{p}\Big)^{-1},$$

using $1 + g_*(p) = 1 + \gamma(p)/(p - \gamma(p)) = (1 - \gamma(p)/p)^{-1}$. These are exactly the local factors of $\mathfrak{S}(\gamma)$ (Notation 3.41) at the primes $p \nmid V$. At the primes $p \mid V$ one has $\gamma(p) = 0$, so the local factor of $\mathfrak{S}(\gamma)$ is $1 - 1/p$, and $\prod_{p\mid V}(1 - 1/p) = \varphi(V)/V$. Multiplying the two contributions gives the identity. Positivity holds because each local factor is positive and the product converges. □

Uses (proof): 6.37, 4.4, 6.36

Sublemma 6.41 (Asymptotic for H)**.** *There are functions $C_1 = C_1(A_1, A_3) > 0$ and $C_2 = C_2(A_1, A_3) > 0$ such that, for every sieve datum with parameters A_1, A_2, A_3, L and every real $z \ge 2$,*

$$\big|H(z) - \mathfrak{S}(\gamma)\log z\big| \;\le\; C_1\,\mathfrak{S}(\gamma)\,(1 + \ell_V) \;+\; C_2\,\tau(V)\,z^{-1/8}\log(2Vz).$$

Both constants depend on A_1 and A_3 only; neither depends on A_2, L, γ, V or z.

Lean:

- `PrimeGaps.slem_H_error_assembly`

Uses (statement): 4.28, 4.3, 3.41, 6.18, 6.39, 6.22, 6.37, 6.40, 6.36

Uses(notation): $H(z)$, $\mathfrak{S}(\gamma)$, γ, φ, τ, ℓ_m, A_1, A_3, V, $b(p)$, $a(d)$, $A_m(x)$, (a, b) (index)

Proof. Throughout, all implied constants depend on A_1, A_3 only. Write $\mathcal{A}(x;m) := \sum_{0<f<x,\ (f,m)=1} \mu(f)^2/\varphi(f)$ for the strict truncation appearing in Sublemma 6.39; it differs from $A_m(x)$ of Definition 6.21 by at most one term $\mu(f)^2/\varphi(f) \ll x^{-1/2}$, so Sublemma 6.22 applies to $\mathcal{A}$ with the same error shape.

Preliminaries on $\tilde{b}$. From $|\tilde{b}(e)| \le |b(e)|$ and Sublemma 6.37,

$$\sum_e |\tilde{b}(e)|\,\tau(e)\,e^{1/2} = O(1), \quad \sum_e |\tilde{b}(e)|\,e^{1/4} = O(1), \quad \sum_e |\tilde{b}(e)|\,(1+\log e) = O(1),$$

the second and third following from the first by $e^{1/4} \le \tau(e)e^{1/2}$ and $1+\log e \le 3\tau(e)e^{1/2}$.

Step 1: main range $e \le \sqrt{z}$. For such e we have $z/e \ge \sqrt{z} \ge 2$, so Sublemma 6.22 applies with $m = Ve$. Using $(e,V)=1$, so that $\varphi(Ve)/(Ve) = (\varphi(V)/V)(\varphi(e)/e)$ and $\ell_{Ve} = \ell_V + \ell_e$, and $b(e)\varphi(e)/e = \tilde{b}(e)$,

$$\sum_{\substack{e\le\sqrt{z}\\(e,V)=1}} b(e)\,\mathcal{A}(z/e;Ve) = \frac{\varphi(V)}{V} \sum_{\substack{e\le\sqrt{z}\\(e,V)=1}} \tilde{b}(e)\big(\log z - \log e + \ell_V + \ell_e\big) + O\Big(\frac{\varphi(V)}{V}\Big) + \mathcal{R}(z),$$

where the $O(\varphi(V)/V)$ collects the $O(\varphi(Ve)/(Ve))$ errors against $\sum_e |\tilde{b}(e)| = O(1)$, and

$$\mathcal{R}(z) \;:=\; \sum_{\substack{e\le\sqrt{z}\\(e,V)=1}} b(e)\,O\Big(\frac{\tau(Ve)\log(2Vez)}{\sqrt{z/e}}\Big) \;=\; O\Big(\tau(V)\,z^{-1/2}\log(2Vz)\sum_e |b(e)|\tau(e)e^{1/2}\Big),$$

which is $O\big(\tau(V)z^{-1/2}\log(2Vz)\big)$ by Sublemma 6.37 (using $\tau(Ve) = \tau(V)\tau(e)$ for $(e,V)=1$).

Step 2: complete the e-sum. Extend the sum to all e coprime to V. The discarded range $e > \sqrt{z}$ is controlled by $|\log z - \log e + \ell_V + \ell_e| \ll \log(2z) + \ell_V + \log e$ together with $\sum_{e>\sqrt{z}} |\tilde{b}(e)| \le z^{-1/8}\sum_e |\tilde{b}(e)|e^{1/4} = O(z^{-1/8})$ and the same estimate with an extra $\log e$; this gives an error $O\big(\frac{\varphi(V)}{V}z^{-1/8}(\log 2z + \ell_V)\big)$. Similarly the tail $e > \sqrt{z}$ of the true sum of Sublemma 6.39 is $O\big(\frac{\varphi(V)}{V}z^{-1/8}\log 2z\big)$, since $\mathcal{A}(z/e;Ve) \ll \frac{\varphi(V)}{V}\log(2z)$ trivially.

For the completed sum, split $\log z - \log e + \ell_V + \ell_e$. By Sublemma 6.40, $\frac{\varphi(V)}{V}\sum_{(e,V)=1}\tilde{b}(e) = \mathfrak{S}(\gamma)$, so the $\log z$ and ℓ_V parts produce $\mathfrak{S}(\gamma)(\log z + \ell_V)$. The remaining part is

$$\frac{\varphi(V)}{V}\Big|\sum_{(e,V)=1}\tilde{b}(e)\big(\ell_e - \log e\big)\Big| \;\ll\; \frac{\varphi(V)}{V}\sum_e |\tilde{b}(e)|(1+\log e) \;=\; O\Big(\frac{\varphi(V)}{V}\Big),$$

using $0 \le \ell_e \le \log e$ for $e \ge 1$.

Step 3: comparison of $\varphi(V)/V$ with $\mathfrak{S}(\gamma)$. By Sublemma 6.40 and $\sum_e |\tilde{b}(e)| = O(1)$ we get $\mathfrak{S}(\gamma) \ll \varphi(V)/V$; conversely, the local factors at $p \nmid V$ satisfy $\log\big(1+\tilde{b}(p)\big) \ge -2A_3/p^2$ by Sublemma 6.35, so $\prod_{p\nmid V}(1+\tilde{b}(p)) \ge \exp\big(-2A_3\sum_p p^{-2}\big) \gg 1$ and hence $\varphi(V)/V \ll \mathfrak{S}(\gamma)$. The two quantities are therefore comparable, with constants depending on A_1, A_3 only, and every $O(\varphi(V)/V)$ above may be rewritten as $O(\mathfrak{S}(\gamma))$, hence absorbed into $O(\mathfrak{S}(\gamma)(1+\ell_V))$.

Collecting Steps 1–3 and noting $z^{-1/2} \le z^{-1/8}$ and $\log(2Vz) \ge 1$ for $z \ge 2$, $V \ge 1$, the two error shapes of the statement are obtained. □

Uses (proof): 6.39, 6.22, 6.37, 6.40, 6.35, 6.21, 6.18

6.8 Partial summation

Sublemma 6.42 (Cost of the removed primes)**.** *There is a function $C = C(A_3) > 0$ such that, for every sieve datum with parameters A_1, A_2, A_3, L,*

$$\ell_V \;=\; \sum_{p\mid V}\frac{\log p}{p-1} \;\le\; 2L \;+\; C(A_3).$$

Lean:

- `PrimeGaps.ellV_le_two_L_add_const`

Uses (statement): 4.3, 6.18, 6.7

Uses(notation): γ, L, A_3, V, ℓ_m, $T(z)$, $\mathbb{P}$ (index)

Proof. Fix $z \geq \max\{p : p \mid V\}$ (and $z \geq 2$). Since $\gamma(p) = 0$ for $p \mid V$,

$$\sum_{2\leq p\leq z} \frac{\gamma(p)\log p}{p} = \sum_{\substack{p\leq z\\ p\nmid V}} \frac{\log p}{p} + \sum_{\substack{p\leq z\\ p\nmid V}} \frac{(\gamma(p)-1)\log p}{p}.$$

By the structural hypothesis the second sum is at most $A_3\sum_p(\log p)/p^2 = O_{A_3}(1)$ in absolute value, the series converging. For the first, Mertens' first theorem $\sum_{p\leq z}(\log p)/p = \log z + O(1)$ (Lemma 6.7, after discarding the proper prime powers) gives

$$\sum_{2\leq p\leq z} \frac{\gamma(p)\log p}{p} = \log z - \sum_{p\mid V} \frac{\log p}{p} + O_{A_3}(1).$$

The lower half of the log-sum hypothesis at $w = 2$ reads $\sum_{2\leq p\leq z}\gamma(p)(\log p)/p \geq \log(z/2) - L$. Subtracting yields $\sum_{p\mid V}(\log p)/p \leq L + O_{A_3}(1)$. Finally $p/(p-1) \leq 2$ for every prime p, so $\ell_V \leq 2\sum_{p\mid V}(\log p)/p \leq 2L + O_{A_3}(1)$. □

Lemma 6.43 (Partial summation)**.** *There are functions $C_1 = C_1(A_1, A_2, A_3) > 0$ and $C_2 = C_2(A_1, A_3) > 0$ such that the following holds for every sieve datum with parameters A_1, A_2, A_3, L, every continuously differentiable $G : \mathbb{R} \to \mathbb{R}$, and every real $z \geq 2$:*

$$\Big|\sum_{0<d<z} \mu(d)^2 g_*(d)\, G\Big(\frac{\log d}{\log z}\Big) - \mathfrak{S}(\gamma)\,(\log z)\int_0^1 G(x)\,dx\Big| \leq C_1\,\mathfrak{S}(\gamma)\,L\,G_{\max} + C_2\,\frac{\tau(V)\log(2Vz)}{\log z}\,G_{\max},$$

where $G_{\max} = \sup_{t\in[0,1]}\big(|G(t)| + |G'(t)|\big)$.

Lean:

- `PrimeGaps.lem_partial_sum`

Uses (statement): 4.3, 4.4, 4.6, 4.28, 3.41, 6.41, 6.42

Uses(notation): μ, g_*, $h(d)$, $H(z)$, $\mathfrak{S}(\gamma)$, γ, τ, A_1, A_2, A_3, L, G, $G_{\max}$, V, ℓ_m, $C^\infty(X)$(index)

Proof. Write $S(z) := \sum_{0<d<z} h(d)G(\log d/\log z)$ with $h(d) = \mu(d)^2 g_*(d)$.

Step 1: Abel summation. The function H of Definition 4.28 is the summatory function of h; taking $\psi(t) := G(\log t/\log z)$, so that $\psi'(t) = G'(\log t/\log z)/(t\log z)$ and $\psi(z) = G(1)$, Abel summation gives the exact identity

$$S(z) = G(1)\,H(z) - \int_1^z \frac{G'(\log t/\log z)}{t\log z}\,H(t)\,dt.$$

(The integrand is continuous off the integers and H is a bounded monotone step function on $[1, z]$, so the integral exists.)

Step 2: separate main term and error. Write $H(t) = \mathfrak{S}(\gamma)\log t + r(t)$. For the main term, substitute $x := \log t/\log z$, so $dx = dt/(t\log z)$ and $\log t = x\log z$:

$$\mathfrak{S}(\gamma)\int_1^z \frac{G'(\log t/\log z)}{t\log z}\log t\,dt = \mathfrak{S}(\gamma)(\log z)\int_0^1 x\,G'(x)\,dx.$$

Integration by parts gives $\int_0^1 xG'(x)\,dx = G(1) - \int_0^1 G(x)\,dx$, so together with the boundary contribution $G(1)\mathfrak{S}(\gamma)\log z$ the main terms combine to exactly $\mathfrak{S}(\gamma)(\log z)\int_0^1 G(x)\,dx$. Consequently

$$\Big|S(z) - \mathfrak{S}(\gamma)(\log z)\int_0^1 G\Big| \;\le\; G_{\max}\Big(|r(z)| + \frac{1}{\log z}\int_1^z \frac{|r(t)|}{t}\,dt\Big),$$

using $|G(1)| \le G_{\max}$ and $|G'| \le G_{\max}$ on $[0,1]$.

Step 3: bound the error functional. By Sublemma 6.41,

$$|r(t)| \;\le\; C_1'\,\mathfrak{S}(\gamma)(1+\ell_V) \;+\; C_2'\,\tau(V)t^{-1/8}\log(2Vt) \qquad (2 \le t \le z),$$

with C_1', C_2' depending on A_1, A_3; on $[1,2]$ one has $|H(t)| \le 1$ and $\mathfrak{S}(\gamma)\log t \le \mathfrak{S}(\gamma)\log 2$, so the same shape holds after enlarging the constants. The first piece integrates to $O\big(\mathfrak{S}(\gamma)(1+\ell_V)\big)$ (since $\frac{1}{\log z}\int_1^z dt/t = 1$), and the second to $O\big(\tau(V)\log(2V)/\log z\big)$, because $\int_1^\infty t^{-9/8}\log(2Vt)\,dt = O(\log 2V)$ is a convergent integral. The boundary term $|r(z)|$ is dominated by the same two shapes, using $z^{-1/8} \le 1/\log z$ for $z \ge 2$.

Step 4: collapse ℓ_V to L. By Sublemma 6.42, $1 + \ell_V \le 1 + 2L + C(A_3)$. Absorbing the additive constants (the A_2-dependence entering through the parameter L supplied by the log-sum hypothesis) turns $\mathfrak{S}(\gamma)(1+\ell_V)$ into $C_1(A_1,A_2,A_3)\mathfrak{S}(\gamma)L$ plus a term of the second shape. Multiplying through by $G_{\max}$ gives the stated bound. □

Uses (proof): 6.41, 6.42, 6.18, 4.28

7 Relating the transformed weight systems

The sieve analysis uses four systems of variables: the profile F, the divisor weights $\lambda(d_1,\dots,d_k)$, the y-variables $y(r_1,\dots,r_k)$ (Definition 4.20), and, for each $m \in \{1,\dots,k\}$, the m-th marginal variables $y^{(m)}(r_1,\dots,r_k)$ (Definition 4.25). The $S_1(\lambda)$ analysis is carried out in the y-variables and the $S_2^{(m)}(\lambda)$ analysis in the $y^{(m)}$-variables; both are evaluated in terms of F. This section relates them: the y-variables of the F-derived sieve weight are evaluated in closed form, and $y^{(m)}$ is expressed in terms of y up to an error of relative size $1/D_0$.

Throughout, $k \ge 2$ is fixed, $\lambda : \mathbb{N}^k \to \mathbb{R}$ is a function of finite support, y denotes the y-transform of λ (Definition 4.20) and $y^{(m)}$ the m-th marginal transform (Definition 4.25). The parameters D_0, W and R are those of Definitions 3.28, 3.29 and 3.37, and "permissible support" always means Definition 3.39.

7.1 Finiteness of the marginal supremum

Lemma 7.1 ($y^{(m)}_{\max}$ is a finite maximum)**.** *Let $m \in \{1,\dots,k\}$ and let $\lambda : \mathbb{N}^k \to \mathbb{R}$ be a function of finite support. Then $y^{(m)}_{\max} < \infty$; in fact the supremum is attained as a maximum on a finite subset of $\mathbb{N}^k$.*

Lean:

- `PrimeGaps.ymMax_finite`

Uses (statement): 4.25, 4.27

Uses(notation): $y^{(m)}(r_1,\dots,r_k)$, $y^{(m)}_{\max}$, $\lambda(d_1,\dots,d_k)$, k, $\mathbb{N}$, $\mathbf{r}$ (index)

Proof. Suppose $y^{(m)}(r_1,\dots,r_k) \ne 0$. By Definition 4.25 the defining $\mathbf{d}$-sum has a nonzero summand, so there is a tuple $\mathbf{d} = (d_1,\dots,d_k) \in \mathrm{supp}(\lambda)$ with $r_i \mid d_i$ for every i; in particular $r_i \le d_i$ for every i. Since λ has finite support, the integer $M := \max\{d_i : \mathbf{d} \in \mathrm{supp}(\lambda),\ 1 \le i \le k\}$ is well defined and finite, and every $\mathbf{r}$ with $y^{(m)}(\mathbf{r}) \ne 0$ lies in the finite set $\{1,\dots,M\}^k$. Hence $y^{(m)}_{\max} = \max\{|y^{(m)}(\mathbf{r})| : \mathbf{r} \in \{1,\dots,M\}^k\}$ is a maximum over a finite set, and is finite. □

7.2 The y-variables of the F-derived weight

Lemma 7.2 (The transform of the F-derived weight is exactly y_0)**.** *Let $R > 0$, let $W \geq 1$, and let $F : \mathbb{R}^k \to \mathbb{R}$ be arbitrary. Let y_0 be the F-derived y-weight of Definition 5.10 and let λ_0 be the associated sieve weight of Definition 5.11. Then the y-transform of λ_0 (Definition 4.20) is y_0:*

$$y_{\lambda_0}(r_1, \dots, r_k) \;=\; y_0(r_1, \dots, r_k) \qquad \text{for every } (r_1, \dots, r_k) \in \mathbb{N}^k.$$

Lean:

- `PrimeGaps.lem_y_from_lambda_0`

Uses (statement): 5.10, 5.11, 4.20

Uses(notation): k, R, W, F, $y(r_1, \dots, r_k)$, $y_\lambda(r_1, \dots, r_k)$, $\lambda(d_1, \dots, d_k)$, $\lambda_y(d_1, \dots, d_k)$, $\mathbb{N}$, $\mathbf{r}$, μ (index)

Proof. By Definition 5.11 the weight λ_0 is the Möbius transform λ_{y_0} of y_0 (Definition 4.22). The round-trip identity Lemma 4.24 therefore gives

$$y_{\lambda_0}(\mathbf{r}) \;=\; y_{\lambda_{y_0}}(\mathbf{r}) \;=\; \begin{cases} y_0(\mathbf{r}) & \text{if } r_i \text{ is squarefree for every } i, \\ 0 & \text{otherwise.} \end{cases}$$

It remains to check that $y_0(\mathbf{r}) = 0$ whenever some r_i fails to be squarefree. In that case $\prod_j r_j$ is not squarefree, so $\mathbf{r}$ does not lie in the permissible support of Definition 3.39; since y_0 is by construction supported there (Definition 5.10), $y_0(\mathbf{r}) = 0$. Both sides therefore agree at every $\mathbf{r}$. □

Uses (proof): 5.11, 4.24, 4.22, 3.39

Lemma 7.3 (Piecewise closed form of y for the F-derived sieve weight)**.** *Let $k \geq 2$, let $R > 0$, let $W \geq 1$, and let $F : \mathbb{R}^k \to \mathbb{R}$ be a smooth function with $\operatorname{supp} F \subseteq \mathcal{R}_k$ (Definition 5.2). Let λ_0 be the F-derived sieve weight of Definition 5.11 and let y be its y-transform (Definition 4.20). Then, for every $(r_1, \dots, r_k) \in \mathbb{N}^k$,*

$$y(r_1, \dots, r_k) \;=\; \begin{cases} F\big(\frac{\log r_1}{\log R}, \dots, \frac{\log r_k}{\log R}\big) & \text{if } \prod_{i=1}^k r_i \text{ is squarefree and } \gcd\big(\prod_{i=1}^k r_i,\, W\big) = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Lean:

- `PrimeGaps.y_from_lambda`

Uses (statement): 5.11, 4.20, 5.2, 3.37, 3.29

Uses(notation): k, R, W, F, $\mathcal{R}_k$, $y(r_1, \dots, r_k)$, μ, (a, b), $\mathbb{N}$, $\mathbf{r}$, $C^\infty(X)$ (index)

Proof. By Lemma 7.2, $y = y_0$, and by Definition 5.10,

$$y_0(\mathbf{r}) \;=\; F\Big(\tfrac{\log r_1}{\log R}, \dots, \tfrac{\log r_k}{\log R}\Big) \cdot \mathbf{1}\big[\mathbf{r} \text{ lies in the permissible support}\big].$$

The permissible support consists of the tuples satisfying the three conditions $\prod_i r_i \leq R$, $\gcd(\prod_i r_i, W) = 1$, and $\prod_i r_i$ squarefree (Definition 3.39). Two of these are exactly the conditions in the assertion, so it suffices to prove that the truncation condition is automatic: precisely, that

$$F\Big(\tfrac{\log r_1}{\log R}, \dots, \tfrac{\log r_k}{\log R}\Big) \neq 0 \quad \text{and} \quad \textstyle\prod_i r_i \text{ squarefree} \qquad \Longrightarrow \qquad \prod_i r_i \leq R.$$

Assume the left-hand side and write $x_i := \log r_i / \log R$. Squarefreeness of $\prod_i r_i$ forces $r_i \geq 1$ for every i, so $\log r_i \geq 0$. Since F is nonzero at $\mathbf{x} = (x_1, \ldots, x_k)$ and $\operatorname{supp} F \subseteq \mathcal{R}_k$, we have $\mathbf{x} \in \mathcal{R}_k$, i.e. $x_i \geq 0$ for every i and $\sum_i x_i \leq 1$.

If $\log R > 0$, then $\sum_i x_i \leq 1$ reads $\sum_i \log r_i \leq \log R$, that is, $\log(\prod_i r_i) \leq \log R$; exponentiating gives $\prod_i r_i \leq R$, as required.

If instead $\log R \leq 0$ (that is, $0 < R \leq 1$), then each $x_i = \log r_i / \log R$ is a quotient of a non-negative number by a non-positive one, hence $x_i \leq 0$; combined with $x_i \geq 0$ this gives $\mathbf{x} = \mathbf{0}$. But $F(\mathbf{0}) = 0$: since $k \geq 1$, the origin is a limit of points outside $\mathcal{R}_k$ (move along a single coordinate ray into negative values), F vanishes identically off $\mathcal{R}_k$, and F is continuous, so $F(\mathbf{0}) = 0$ by continuity. This contradicts $F(\mathbf{x}) \neq 0$, so the case is vacuous. In either case the implication holds, and the piecewise formula follows. □

Uses (proof): 7.2, 5.10, 3.39

Definition 7.4 (C^1-norm $F_{\max}$ on $[0,1]^k$)**.** For a C^1 function $F : \mathbb{R}^k \to \mathbb{R}$, set

$$F_{\max} := \sup_{t \in [0,1]^k} \Big(|F(t)| + \sum_{i=1}^{k} \Big| \frac{\partial F}{\partial t_i}(t) \Big| \Big).$$

The supremum is over the compact unit cube, on which the displayed function is continuous, so it is finite; it is non-negative because the quantity inside the supremum is non-negative.

Uses(notation): F, $F_{\max}$, k, $C^\infty(X)$ (index)

7.3 Expressing $y^{(m)}$ through y

Lemma 7.5 (Substitute the inversion formula for λ into $y^{(m)}$)**.** *Let λ have permissible support, let $m \in \{1, \ldots, k\}$, and let $\mathbf{r} \in \mathbb{N}^k$ satisfy $r_m = 1$. Then*

$$y^{(m)}(r_1, \ldots, r_k) = \Big(\prod_{i=1}^{k} \mu(r_i)\, g(r_i) \Big) \sum_{\substack{\mathbf{d} \in \mathbb{N}^k \\ r_i \mid d_i \ \forall i,\ d_m = 1}} \Big(\prod_{i=1}^{k} \frac{\mu(d_i)\, d_i}{\varphi(d_i)} \Big) \sum_{\substack{\mathbf{a} \in \mathbb{N}^k \\ d_i \mid a_i \ \forall i}} \frac{y(a_1, \ldots, a_k)}{\prod_{i=1}^{k} \varphi(a_i)},$$

both sums running over tuples of positive *integers.*

Lean:

- `PrimeGaps.MaynardSubstitute.lem_ym_substitute_y`

Uses (statement): 4.25, 4.20, 3.39, 3.3, 3.6, 3.4

Uses(notation): k, $\lambda(d_1, \ldots, d_k)$, μ, φ, g, $\mathbf{d}$, $\mathbf{r}$, $y(r_1, \ldots, r_k)$, $y^{(m)}(r_1, \ldots, r_k)$, $\mathbb{N}$ (index)

Proof. By Definition 4.25,

$$y^{(m)}(r_1, \ldots, r_k) = \Big(\prod_{i=1}^{k} \mu(r_i)\, g(r_i) \Big) \sum_{\substack{\mathbf{d} \\ r_i \mid d_i \ \forall i,\ d_m = 1}} \frac{\lambda(d_1, \ldots, d_k)}{\prod_{i=1}^{k} \varphi(d_i)},$$

a finite sum because λ has finite support; moreover every contributing $\mathbf{d}$ has $\prod_i d_i$ squarefree, hence positive squarefree coordinates, by Definition 3.39. On such tuples Lemma 4.23 identifies λ with the Möbius transform of its own y-variables (Definition 4.22):

$$\lambda(\mathbf{d}) = \Big(\prod_{i=1}^{k} \mu(d_i)\, d_i \Big) \sum_{\substack{\mathbf{a} \\ d_i \mid a_i \ \forall i}} \frac{y(a_1, \ldots, a_k)}{\prod_{i=1}^{k} \varphi(a_i)}.$$

Dividing by $\prod_i \varphi(d_i)$ gives

$$\frac{\lambda(\mathbf{d})}{\prod_{i=1}^k \varphi(d_i)} \;=\; \Big(\prod_{i=1}^k \frac{\mu(d_i)\, d_i}{\varphi(d_i)}\Big) \sum_{\substack{\mathbf{a} \\ d_i \mid a_i \ \forall i}} \frac{y(a_1,\dots,a_k)}{\prod_{i=1}^k \varphi(a_i)},$$

and inserting this into the previous display yields the asserted formula. (The hypothesis $r_m = 1$ is not used in the derivation; it is recorded because $y^{(m)}$ vanishes otherwise, by Lemma 4.26.) □

Uses (proof): 4.23, 4.22, 4.26

Lemma 7.6 (Interchange the **d**- and **a**-summations)**.** *Let λ have permissible support, let $m \in \{1,\dots,k\}$, and let $\mathbf{r} \in \mathbb{N}^k$ satisfy $r_i \geq 1$ for every i and $r_m = 1$. Then*

$$y^{(m)}(r_1,\dots,r_k) \;=\; \Big(\prod_{i=1}^k \mu(r_i)\, g(r_i)\Big) \sum_{\substack{\mathbf{a}\in\mathbb{N}^k \\ r_i \mid a_i \ \forall i}} \frac{y(a_1,\dots,a_k)}{\prod_{i=1}^k \varphi(a_i)} \sum_{\substack{\mathbf{d}\in\mathbb{N}^k \\ r_i \mid d_i \mid a_i \ \forall i,\ d_m = 1}} \prod_{i=1}^k \frac{\mu(d_i)\, d_i}{\varphi(d_i)}.$$

Lean:

- `PrimeGaps.lem_ym_swap`

Uses (statement): 4.25, 4.20, 3.39, 3.3, 3.6, 3.4

Uses(notation): k, μ, φ, g, $\mathbf{d}$, $\mathbf{r}$, $y(r_1,\dots,r_k)$, $y^{(m)}(r_1,\dots,r_k)$, $\mathbb{N}$ (index)

Proof. The iterated sum of Lemma 7.5 is supported on the set of pairs $(\mathbf{d},\mathbf{a})$ with $r_i \mid d_i \mid a_i$ for every i and $d_m = 1$: the outer summation imposes $r_i \mid d_i$ and $d_m = 1$, and the inner one imposes $d_i \mid a_i$, which together force $r_i \mid a_i$. Both summations are finite, since y has finite support (Definition 4.20) and, for fixed $\mathbf{a}$, only the finitely many divisor tuples $\mathbf{d}$ of $\mathbf{a}$ contribute. We may therefore interchange the order of summation: grouping first by $\mathbf{a}$, the inner $\mathbf{d}$-sum runs over $\{\mathbf{d} : r_i \mid d_i \mid a_i \ \forall i,\ d_m = 1\}$. Multiplying by the outer prefactor $\prod_i \mu(r_i)g(r_i)$, which depends only on $\mathbf{r}$, gives the asserted identity. □

Uses (proof): 7.5

Lemma 7.7 (Evaluation of the inner divisor sum)**.** *Let a be a positive squarefree integer and let r be a positive divisor of a. Then*

$$\sum_{\substack{d \mid a \\ r \mid d}} \frac{\mu(d)\, d}{\varphi(d)} \;=\; \frac{\mu(a)\, r}{\varphi(a)}.$$

Lean:

- `PrimeGaps.lem_ym_eval_d_sum`

Uses (statement): 3.3, 3.4

Uses(notation): μ, φ, $\omega(n)$, $\mathbb{P}$, (a,b) (index)

Proof. Write $a = re$ with $e := a/r$. Since a is squarefree, so are r and e, and $(r, e) = 1$.

Every d with $r \mid d \mid a$ is uniquely of the form $d = rf$ with $f \mid e$, and then $(r, f) = 1$, so by multiplicativity of μ and of φ on coprime products, $\mu(d) = \mu(r)\mu(f)$ and $\varphi(d) = \varphi(r)\varphi(f)$. Substituting,

$$\sum_{\substack{d \mid a \\ r \mid d}} \frac{\mu(d)\, d}{\varphi(d)} \;=\; \frac{\mu(r)\, r}{\varphi(r)} \sum_{f \mid e} \frac{\mu(f)\, f}{\varphi(f)}.$$

The summand $f \mapsto \mu(f)f/\varphi(f)$ is multiplicative, so for squarefree e the divisor sum factors over the prime divisors of e:

$$\sum_{f \mid e} \frac{\mu(f)\, f}{\varphi(f)} \;=\; \prod_{p \mid e} \Big(1 + \frac{\mu(p)\, p}{\varphi(p)}\Big) \;=\; \prod_{p \mid e} \Big(1 - \frac{p}{p-1}\Big) \;=\; \prod_{p \mid e} \frac{-1}{p-1} \;=\; \frac{(-1)^{\omega(e)}}{\prod_{p \mid e}(p-1)} \;=\; \frac{\mu(e)}{\varphi(e)},$$

using $\mu(e) = (-1)^{\omega(e)}$ and $\varphi(e) = \prod_{p \mid e}(p-1)$ for squarefree e. Combining the two displays,

$$\sum_{\substack{d \mid a \\ r \mid d}} \frac{\mu(d)\, d}{\varphi(d)} \;=\; \frac{\mu(r)\, \mu(e)\, r}{\varphi(r)\, \varphi(e)} \;=\; \frac{\mu(a)\, r}{\varphi(a)},$$

the last step again by multiplicativity of μ and φ on the coprime product $a = re$. □

Lemma 7.8 ($y^{(m)}$ in terms of y, intermediate form)**.** *Let λ have permissible support, let $m \in \{1, \dots, k\}$, and let $\mathbf{r} \in \mathbb{N}^k$ satisfy $r_i \geq 1$ for every i and $r_m = 1$. Then*

$$y^{(m)}(r_1, \dots, r_k) \;=\; \Big(\prod_{i=1}^{k} \mu(r_i)\, g(r_i)\Big) \sum_{\substack{\mathbf{a} \in \mathbb{N}^k \\ r_i \mid a_i \;\forall i}} \frac{y(a_1, \dots, a_k)}{\prod_{i=1}^{k} \varphi(a_i)} \prod_{i \neq m} \frac{\mu(a_i)\, r_i}{\varphi(a_i)}.$$

Lean:

- `PrimeGaps.lem_ym_intermediate`

Uses (statement): 4.25, 4.20, 3.39, 3.3, 3.6, 3.4

Uses(notation): k, μ, φ, g, $\mathbf{d}$, $\mathbf{r}$, $y(r_1, \dots, r_k)$, $y^{(m)}(r_1, \dots, r_k)$, $\mathbb{N}$ (index)

Proof. Start from Lemma 7.6 and evaluate, for each fixed $\mathbf{a}$, the inner sum

$$\Sigma(\mathbf{a}) \;:=\; \sum_{\substack{\mathbf{d} \\ r_i \mid d_i \mid a_i \;\forall i,\; d_m = 1}} \prod_{i=1}^{k} \frac{\mu(d_i)\, d_i}{\varphi(d_i)}.$$

It suffices to treat those $\mathbf{a}$ with $y(\mathbf{a}) \neq 0$; by Lemma 4.21 such an $\mathbf{a}$ has $\prod_i a_i$ squarefree, so every a_i is positive and squarefree. The constraints on $\mathbf{d}$ are coordinatewise, so $\Sigma(\mathbf{a})$ factors as a product over i. The m-th factor is 1: the constraint $d_m = 1$ leaves the single term $\mu(1)\cdot 1/\varphi(1) = 1$, and $r_m = 1$ is compatible with it. For $i \neq m$ the i-th factor is $\sum_{r_i \mid d_i \mid a_i} \mu(d_i) d_i/\varphi(d_i)$, which equals $\mu(a_i) r_i/\varphi(a_i)$ by Lemma 7.7 applied with $a = a_i$ and $r = r_i$ (legitimate since a_i is squarefree and $r_i \mid a_i$). Hence $\Sigma(\mathbf{a}) = \prod_{i \neq m} \mu(a_i) r_i/\varphi(a_i)$, and substituting this into Lemma 7.6 gives the assertion. □

Uses (proof): 7.6, 7.7, 4.21

7.4 The diagonal decomposition of $y^{(m)}$

Definition 7.9 (The $y^{(m)}$-summand $T_{m,\mathbf{r}}$)**.** Let $m \in \{1,\dots,k\}$ and $\mathbf{r} \in \mathbb{N}^k$. For $\mathbf{a} \in \mathbb{N}^k$ set

$$T_{m,\mathbf{r}}(\mathbf{a}) \;:=\; \Big(\prod_{i=1}^{k} \mu(r_i)\, g(r_i)\Big) \cdot \frac{y(a_1,\dots,a_k)}{\prod_{i=1}^{k} \varphi(a_i)} \cdot \prod_{i \neq m} \frac{\mu(a_i)\, r_i}{\varphi(a_i)},$$

with the convention that the whole expression is read as 0 when some $a_i = 0$.

Uses (statement): 4.20, 3.3, 3.6, 3.4

Uses(notation): k, μ, φ, g, $\mathbf{r}$, $y(r_1,\dots,r_k)$, $\mathbb{N}$ (index)

Definition 7.10 (The free-coordinate marginal $Y^{(m)}$)**.** Let $m \in \{1,\dots,k\}$ and $\mathbf{r} \in \mathbb{N}^k$. Set

$$Y^{(m)}(r_1,\dots,r_k) \;:=\; \sum_{a \geq 1} \frac{y(r_1,\dots,r_{m-1},\, a,\, r_{m+1},\dots,r_k)}{\varphi(a)},$$

the sum running over the single free coordinate in the m-th slot. Only finitely many terms are nonzero, because y has finite support.

Uses (statement): 4.20, 3.4

Uses(notation): k, φ, $\mathbf{r}$, $y(r_1,\dots,r_k)$, $\mathbb{N}$ (index)

Lemma 7.11 (Diagonal and off-diagonal decomposition of $y^{(m)}$)**.** *Let λ have permissible support, let $m \in \{1,\dots,k\}$, and let $\mathbf{r} \in \mathbb{N}^k$ be such that every r_i is positive and squarefree and $r_m = 1$. Then*

$$y^{(m)}(r_1,\dots,r_k) \;=\; \Big(\prod_{i \neq m} \frac{g(r_i)\, r_i}{\varphi(r_i)^2}\Big)\, Y^{(m)}(r_1,\dots,r_k) \;+\; \sum_{\substack{\mathbf{a} \in \mathbb{N}^k:\ r_i \mid a_i\ \forall i \\ \exists\, j \neq m,\ a_j \neq r_j}} T_{m,\mathbf{r}}(\mathbf{a}).$$

Uses (statement): 4.25, 7.9, 7.10, 3.39, 3.6, 3.4

Uses(notation): k, μ, φ, g, $\mathbf{r}$, $y(r_1,\dots,r_k)$, $y^{(m)}(r_1,\dots,r_k)$, $\mathbb{N}$ (index)

Proof. By Lemma 7.8 and Definition 7.9,

$$y^{(m)}(\mathbf{r}) \;=\; \sum_{\substack{\mathbf{a} \\ r_i \mid a_i\ \forall i}} T_{m,\mathbf{r}}(\mathbf{a}),$$

the sum being finite because y has finite support and $T_{m,\mathbf{r}}$ vanishes wherever y does. Split the index set $\{\mathbf{a} : r_i \mid a_i\ \forall i\}$ into the *diagonal* part $\mathcal{D} := \{\mathbf{a} : r_i \mid a_i\ \forall i,\ a_i = r_i\ \forall i \neq m\}$ and its complement, the *off-diagonal* part $\mathcal{O} := \{\mathbf{a} : r_i \mid a_i\ \forall i,\ \exists j \neq m,\ a_j \neq r_j\}$. These are disjoint with union the whole index set, so the sum splits accordingly, and the second piece is the displayed off-diagonal sum.

It remains to evaluate the diagonal piece. Since $r_m = 1$ divides every integer, the tuples in $\mathcal{D}$ are exactly those obtained from $\mathbf{r}$ by replacing the m-th coordinate by an arbitrary $a \geq 0$, and distinct a give distinct tuples. For such a tuple, using $\varphi(a_i) = \varphi(r_i)$ and $\mu(a_i) = \mu(r_i)$ for $i \neq m$,

$$T_{m,\mathbf{r}}(\mathbf{a}) \;=\; \Big(\mu(r_m) g(r_m) \prod_{i \neq m} \mu(r_i) g(r_i)\Big) \cdot \frac{y(\mathbf{a})}{\varphi(a) \prod_{i \neq m} \varphi(r_i)} \cdot \prod_{i \neq m} \frac{\mu(r_i)\, r_i}{\varphi(r_i)}.$$

The factor $\mu(r_m)g(r_m)$ equals 1 because $r_m = 1$, and $\mu(r_i)^2 = 1$ for each $i \neq m$ because r_i is squarefree. Collecting the $i \neq m$ factors,

$$T_{m,\mathbf{r}}(\mathbf{a}) \;=\; \Big(\prod_{i\neq m} \frac{g(r_i)\, r_i}{\varphi(r_i)^2}\Big) \cdot \frac{y(\mathbf{a})}{\varphi(a)}.$$

Summing over $a \geq 1$ — the term $a = 0$ contributes 0 by the convention of Definition 7.9 — and recognising Definition 7.10 gives the diagonal contribution $\big(\prod_{i\neq m} g(r_i)r_i/\varphi(r_i)^2\big)\, Y^{(m)}(\mathbf{r})$, as asserted. □

Uses (proof): 7.8, 3.3

Lemma 7.12 (Approximation of the diagonal prefactor)**.** *There is a constant $C = C(k) \geq 0$ with the following property. Let $D_0 \geq 2$ be real and let $\mathbf{r} \in \mathbb{N}^k$ be such that every r_i is positive and squarefree and every prime factor of every r_i exceeds D_0. Then*

$$\Big|\prod_{i=1}^{k} \frac{g(r_i)\, r_i}{\varphi(r_i)^2} \;-\; 1\Big| \;\leq\; \frac{C}{D_0}.$$

Lean:

- `PrimeGaps.lem_ym_prefactor`

Uses (statement): 3.6, 3.4, 3.28

Uses(notation): k, φ, g, D_0, $\mathbb{P}$, $\mathbb{N}$, $\mathbf{r}$ (index)

Proof. Fix i. Since r_i is squarefree, both g and φ factor over the distinct prime divisors of r_i, and

$$\frac{g(r_i)\, r_i}{\varphi(r_i)^2} \;=\; \prod_{p\mid r_i} \frac{(p-2)\, p}{(p-1)^2} \;=\; \prod_{p\mid r_i}\Big(1 - \frac{1}{(p-1)^2}\Big),$$

because $(p-2)p = (p-1)^2 - 1$. Hence

$$\prod_{i=1}^{k} \frac{g(r_i)\, r_i}{\varphi(r_i)^2} \;=\; \prod_{j}(1 - x_j), \qquad x_j := \frac{1}{(p_j - 1)^2},$$

where $p_1, p_2, \ldots$ enumerates, with multiplicity, the prime divisors of $r_1, \ldots, r_k$. Every such prime exceeds $D_0 \geq 2$, so $p_j \geq 3$ and $0 \leq x_j \leq 1/4 \leq 1$. For numbers $x_j \in [0,1]$ one has the elementary bound $\big|\prod_j (1 - x_j) - 1\big| \leq \sum_j x_j$, proved by induction on the number of factors: if P is a partial product then $0 \leq P \leq 1$, so $|P(1-x) - 1| \leq |P - 1| + Px \leq |P - 1| + x$.

It therefore suffices to bound $\sum_j x_j$. For a single index i the primes $p \mid r_i$ are distinct and all exceed D_0, and $p - 1 \geq p/2$ for every prime p, so

$$\sum_{p\mid r_i} \frac{1}{(p-1)^2} \;\leq\; 4 \sum_{n > D_0} \frac{1}{n^2} \;\leq\; \frac{4}{\lfloor D_0 \rfloor} \;\leq\; \frac{8}{D_0},$$

using the integral comparison $\sum_{n>M} n^{-2} \leq \int_M^\infty t^{-2}\, dt = 1/M$ at the integer $M = \lfloor D_0 \rfloor$, and $\lfloor D_0 \rfloor \geq D_0/2$ for $D_0 \geq 1$. Summing over the k coordinates gives $\sum_j x_j \leq 8k/D_0$, so the lemma holds with $C := 8k$. □

Lemma 7.13 (Bound for the free-coordinate marginal)**.** *There is a constant $C = C(k) > 0$ with the following property. Let $W \geq 1$ and let $R \geq 1$ be real with $p \leq R$ for every prime $p \mid W$. Let λ have permissible support, with y-variables y. Then, for every $m \in \{1, \ldots, k\}$ and every $\mathbf{r} \in \mathbb{N}^k$,*

$$\big|Y^{(m)}(r_1, \ldots, r_k)\big| \;\leq\; C\, \frac{y_{\max}\, \varphi(W)\, (\log R + 1)}{W}.$$

Uses (statement): 7.10, 4.27, 3.39, 3.37, 3.29, 3.4

Uses(notation): k, R, W, μ, φ, $y(r_1,\dots,r_k)$, $y_{\max}$, (a,b), $\mathbb{N}$, $\mathbf{r}$, $f \ll g$ (index)

Proof. Write $\mathbf{r}^{(a)}$ for the tuple obtained from $\mathbf{r}$ by putting a in the m-th slot. By Lemma 4.21, $y(\mathbf{r}^{(a)}) = 0$ unless $\prod_i r_i^{(a)}$ is squarefree, coprime to W, and at most R; in particular a must lie in

$$\mathcal{S} \;:=\; \{a \in \mathbb{N} : a \le R,\ a \text{ squarefree},\ (a,W)=1\}.$$

Hence, by the triangle inequality and $|y| \le y_{\max}$,

$$\bigl|Y^{(m)}(\mathbf{r})\bigr| \;\le\; y_{\max} \sum_{a\in\mathcal{S}} \frac{1}{\varphi(a)}.$$

It remains to bound $A := \sum_{a\in\mathcal{S}} 1/\varphi(a)$. For squarefree a one has the divisor-sum identity $a/\varphi(a) = \sum_{d\mid a} \mu(d)^2/\varphi(d)$, so $1/\varphi(a) = a^{-1}\sum_{d\mid a}\mu(d)^2/\varphi(d)$. Writing $a = de$ and dropping the constraints linking d and e (all terms being non-negative),

$$A \;\le\; \Bigl(\sum_{\substack{d\ge 1\\ d \text{ squarefree},\ (d,W)=1}} \frac{\mu(d)^2}{\varphi(d)\,d}\Bigr)\cdot\Bigl(\sum_{\substack{e\le R\\ (e,W)=1}} \frac{1}{e}\Bigr).$$

The first factor is bounded by an absolute constant K: extending it to all squarefree $d \ge 1$ and factoring the resulting Euler product gives $\prod_p\bigl(1+1/(p(p-1))\bigr) < \infty$. The second factor is the Mertens-type coprime harmonic sum, which by Lemma 6.20 is $O\bigl((\varphi(W)/W)(\log R+1)\bigr)$ under the stated hypothesis that every prime factor of W is at most R. Multiplying the two bounds gives $A \le C\,\varphi(W)(\log R+1)/W$ for a constant $C = C(k) > 0$, and the lemma follows. □

Uses (proof): 4.21, 6.20, 3.3

Lemma 7.14 (Off-diagonal contribution: $a_j \ne r_j$ for some $j \ne m$)**.** *There is a constant $C = C(k) > 0$ with the following property. Let $\vartheta \in (0,1)$ and $\delta \in (0, \vartheta/2)$. Then there is N_0 such that for every $N \ge N_0$, every y of permissible support, every $m \in \{1,\dots,k\}$, and every $\mathbf{r} \in \mathbb{N}^k$ with $r_i \ge 1$ for every i and $r_m = 1$,*

$$\Bigl|\sum_{\substack{\mathbf{a}\in\mathbb{N}^k:\ r_i\mid a_i\ \forall i\\ \exists\, j\ne m,\ a_j\ne r_j}} T_{m,\mathbf{r}}(\mathbf{a})\Bigr| \;\le\; C\,\frac{y_{\max}\,\varphi(W)\,\log R}{W\,D_0}.$$

Lean:

- `PrimeGaps.MaynardOffDiagonal.offDiagonal_bound`

Uses (statement): 7.9, 4.27, 3.39, 3.29, 3.28, 3.37

Uses(notation): k, R, W, D_0, μ, φ, g, $y(r_1,\dots,r_k)$, $y_{\max}$, (a,b), $O(g)$, $\mathbb{N}$, $\mathbf{r}$, ϑ, δ (index)

Proof. If $k \le 1$ the off-diagonal index set is empty and there is nothing to prove, so assume $k \ge 2$.

Step 1: a y-free majorant. On the permissible support every coordinate a_i is positive, squarefree and coprime to W, and $\prod_i a_i \le R$. Since $|y(\mathbf{a})| \le y_{\max}$ and $|\mu| \le 1$, Definition 7.9 gives, for every permissible $\mathbf{a}$,

$$\bigl|T_{m,\mathbf{r}}(\mathbf{a})\bigr| \;\le\; y_{\max}\cdot\Bigl|\prod_{i=1}^{k}\mu(r_i)g(r_i)\Bigr|\cdot\frac{1}{\prod_{i=1}^{k}\varphi(a_i)}\cdot\prod_{i\ne m}\frac{r_i}{\varphi(a_i)}.$$

Summing over the off-diagonal set, the whole off-diagonal sum is at most $y_{\max} K$, where K denotes the sum of the right-hand majorant (without the factor $y_{\max}$) over all permissible off-diagonal $\mathbf{a}$.

Step 2: reduction to one-dimensional sums. Every contributing $\mathbf{a}$ has $r_i \mid a_i$; write $a_i = r_i b_i$ with $b_i \geq 1$. On the permissible support each a_i is squarefree, so $(r_i, b_i) = 1$ and $\varphi(a_i) = \varphi(r_i)\varphi(b_i)$; moreover each b_i is itself squarefree, coprime to W, and at most R, that is, $b_i \in \mathcal{S}$, where $\mathcal{S} := \{n \leq R : n \text{ squarefree}, \ (n, W) = 1\}$. In these variables the majorant factors as

$$\underbrace{\Big|\prod_i \mu(r_i) g(r_i)\Big| \cdot \frac{1}{\prod_i \varphi(r_i)} \cdot \prod_{i \neq m} \frac{r_i}{\varphi(r_i)}}_{=: \Pi(\mathbf{r})} \cdot \frac{1}{\varphi(b_m)} \prod_{i \neq m} \frac{1}{\varphi(b_i)^2}.$$

Now $\Pi(\mathbf{r}) \leq 1$: indeed $|\mu| \leq 1$, the m-th coordinate contributes $g(1)/\varphi(1) = 1$ since $r_m = 1$, and for each $i \neq m$ the squarefreeness of r_i gives $g(r_i)\, r_i \leq \varphi(r_i)^2$, this being the inequality $\prod_{p \mid r_i}(1 - (p-1)^{-2}) \leq 1$. Being off-diagonal means $a_j \neq r_j$, i.e. $b_j \neq 1$, for at least one $j \neq m$; bounding the number of such j by $k - 1$ and summing each coordinate independently,

$$K \ \leq \ (k-1)\, A \cdot T \cdot B^{\,k-2}, \qquad A := \sum_{n \in \mathcal{S}} \frac{1}{\varphi(n)}, \quad B := \sum_{n \in \mathcal{S}} \frac{1}{\varphi(n)^2}, \quad T := \sum_{\substack{n \in \mathcal{S} \\ n \neq 1}} \frac{1}{\varphi(n)^2},$$

where A accounts for the free m-th coordinate, T for the distinguished offending coordinate j, and B for each of the remaining $k - 2$ coordinates.

Step 3: the three sums. First, $B \leq \sum_{n \geq 1} \mu(n)^2/\varphi(n)^2 =: C_B$, which is finite by Lemma 6.11 and absolute. Second, every $n \in \mathcal{S}$ with $n \neq 1$ is squarefree and coprime to W, hence has all of its prime factors exceeding D_0 (Definition 3.29); choosing one such prime factor p and writing $n = pn'$ with $n' \in \mathcal{S}$ coprime to p, we have $\varphi(n) = (p-1)\varphi(n')$, so, over-counting each n at most once per prime factor,

$$T \ \leq \ \Big(\sum_{p > D_0} \frac{1}{(p-1)^2}\Big) \cdot B \ \leq \ \frac{8}{D_0} \cdot C_B,$$

the bound on the prime sum being the one established in the proof of Lemma 7.12. Third, $A \leq C_A\, \varphi(W)(\log R + 1)/W$ exactly as in the proof of Lemma 7.13, via Lemma 6.20.

Step 4: assembly. Combining,

$$K \ \leq \ (k-1)\, C_A\, C_B^{k-1} \cdot \frac{8}{D_0} \cdot \frac{\varphi(W)(\log R + 1)}{W}.$$

Since $R = N^{\vartheta/2 - \delta}$ with $\vartheta/2 - \delta > 0$, we have $\log R \geq 1$ for all N beyond a threshold N_0 depending only on ϑ and δ; for such N, $\log R + 1 \leq 2 \log R$. Multiplying by $y_{\max}$ from Step 1 gives the assertion with $C := 16(k-1)\, C_A\, C_B^{k-1}$. □

Uses (proof): 4.21, 7.12, 7.13, 6.20, 6.11

Lemma 7.15 ($y^{(m)}$ from y)**.** *There is a constant $C = C(k) > 0$ with the following property. Let $\vartheta \in (0,1)$ and $\delta \in (0, \vartheta/2)$. Then there is N_0 such that for every $N \geq N_0$, every λ of permissible support, every $m \in \{1, \dots, k\}$, and every $\mathbf{r} \in \mathbb{N}^k$ with $r_m = 1$ and every r_i squarefree,*

$$\big|y^{(m)}(r_1, \dots, r_k) \ - \ Y^{(m)}(r_1, \dots, r_k)\big| \ \leq \ C\, \frac{y_{\max}\, \varphi(W)\, \log R}{W\, D_0}.$$

Lean:

- `PrimeGaps.maynard_ym_identity`

Uses (statement): 4.25, 7.10, 4.27, 3.39, 3.29, 3.28, 3.37

Uses(notation): k, R, W, D_0, φ, $y(r_1,\dots,r_k)$, $y_{\max}$, $y^{(m)}(r_1,\dots,r_k)$, $\mathbb{N}$, $\mathbf{r}$, ϑ, δ (index)

Proof. Squarefreeness of each r_i gives $r_i \geq 1$, so Lemma 7.11 applies and, writing $c(m,\mathbf{r}) := \prod_{i\neq m} g(r_i) r_i/\varphi(r_i)^2$,

$$y^{(m)}(\mathbf{r}) - Y^{(m)}(\mathbf{r}) \;=\; \big(c(m,\mathbf{r})-1\big)\, Y^{(m)}(\mathbf{r}) \;+\; \sum_{\substack{\mathbf{a}:\ r_i \mid a_i\ \forall i \\ \exists\, j\neq m,\ a_j \neq r_j}} T_{m,\mathbf{r}}(\mathbf{a}).$$

The second term is bounded by Lemma 7.14, applied to the y-variables of λ, which have permissible support by Lemma 4.21; this gives $C_1\, y_{\max}\varphi(W)\log R/(W D_0)$ for all N beyond a threshold. We bound the first term.

If some r_j with $j \neq m$ is *not* coprime to W, then every tuple obtained from $\mathbf{r}$ by varying the m-th coordinate has coordinate product not coprime to W, so y vanishes at all of them by Lemma 4.21; hence $Y^{(m)}(\mathbf{r}) = 0$ and the first term vanishes.

Otherwise every r_i is squarefree and coprime to W, so by Definition 3.29 every prime factor of every r_i exceeds D_0. Applying Lemma 7.12 to the tuple $\mathbf{r}$ — whose m-th factor equals 1, so that the full product and the product over $i \neq m$ agree — gives $|c(m,\mathbf{r}) - 1| \leq C_2/D_0$ once $D_0 \geq 2$, which holds for all large N. Combined with Lemma 7.13,

$$\big|(c(m,\mathbf{r})-1)\, Y^{(m)}(\mathbf{r})\big| \;\leq\; \frac{C_2}{D_0}\cdot C_3\,\frac{y_{\max}\,\varphi(W)(\log R+1)}{W} \;\leq\; 2C_2C_3\,\frac{y_{\max}\,\varphi(W)\log R}{W D_0},$$

using $\log R + 1 \leq 2\log R$, valid for all N beyond a threshold depending only on ϑ and δ. Adding the two contributions, and taking N_0 to be the largest of the finitely many thresholds used, gives the claim with $C := C_1 + 2C_2C_3$. □

Uses (proof): 7.11, 7.14, 7.12, 7.13, 4.21, 7.9, 3.6

7.5 The smooth case

Lemma 7.16 (Substitute smooth y into Lemma 7.15)**.** *There is a constant $C = C(k) \geq 0$ with the following property. Let $\vartheta \in (0,1)$, $\delta \in (0,\vartheta/2)$, and let $F : \mathbb{R}^k \to \mathbb{R}$ be smooth with $\operatorname{supp} F \subseteq \mathcal{R}_k$. Then there is N_0 such that for every $N \geq N_0$, every λ of permissible support whose y-variables satisfy $y_{\max} \leq F_{\max}$, every $m \in \{1,\dots,k\}$, and every $\mathbf{r} \in \mathbb{N}^k$ with $r_m = 1$ and every r_i squarefree,*

$$\big|y^{(m)}(r_1,\dots,r_k) \;-\; Y^{(m)}(r_1,\dots,r_k)\big| \;\leq\; C\,\frac{F_{\max}\,\varphi(W)\,\log R}{W\,D_0}.$$

Lean:
- `MaynardSmoothY.exists_abs_ym_sub_yInverseSum_le`

Uses (statement): 7.15, 7.4, 4.25, 7.10, 4.27, 5.2, 3.39, 3.29, 3.28, 3.37

Uses(notation): k, R, W, D_0, F, $F_{\max}$, $\mathcal{R}_k$, φ, $y_{\max}$, $y^{(m)}(r_1,\dots,r_k)$, $\mathbb{N}$, $\mathbf{r}$, ϑ, δ (index)

Proof. Apply Lemma 7.15, which bounds the left-hand side by $C\,y_{\max}\varphi(W)\log R/(W D_0)$ for all N beyond a threshold. For all such N we have $\log R > 0$, and $\varphi(W)/W > 0$ always, so that bound is a non-decreasing function of $y_{\max}$. Replacing $y_{\max}$ by the larger quantity $F_{\max}$ using the hypothesis $y_{\max} \leq F_{\max}$ yields the stated bound with the same constant. □

Lemma 7.17 (Maximal modulus of the transformed coefficients)**.** *Let $k \geq 1$, let $R \geq 2$, and let $y : \mathbb{N}^k \to \mathbb{R}$ be finitely supported with permissible support at level R and modulus W (Definition 3.39). Then the transformed system λ_y of Definition 4.22 satisfies*

$$\|\lambda_y\|_{\max} \;\leq\; e^{\,1+3k}\,\|y\|_{\max}\,(\log R)^k.$$

Lean:

- `PrimeGaps.max_yToL_le_const_mul_max_y_mul_log_pow`

Uses (statement): 4.22, 4.27, 3.39

Uses(notation): $\lambda_{\max}$, $y_{\max}$, $\lambda_y(d_1,\dots,d_k)$, $y(r_1,\dots,r_k)$, k, R, W (index)

Proof. Fix $\mathbf{d}$. Expanding $\lambda_y(\mathbf{d})$ and bounding each summand by $\|y\|_{\max}$ leaves the divisor sum $\sum_{\mathbf{r}} \prod_i \mu(r_i)^2/\varphi(r_i)$ taken over those $\mathbf{r}$ with $d_i \mid r_i$ and $\prod_i r_i \leq R$. Together with the prefactor $\prod_i \mu(d_i)\,d_i$ this is exactly the quantity bounded in Lemma 6.10, taken at modulus W and at the present $\mathbf{d}$. Each coordinate factor is at most $\sum_{r\leq R} \mu(r)^2/\varphi(r) \leq e\,(1+\log R)$, and the outer product over the k coordinates contributes at most e^{3k} after absorbing the prefactor $\prod_i \mu(d_i)d_i$ against the coprimality constraints. Multiplying the k coordinate bounds and using $1+\log R \leq e \log R$ for $R \geq 2$ gives the stated bound. □

Uses (proof): 6.10

8 Sieve manipulations: the reduction of $S_1(\lambda)$

For the first moment $S_1(\lambda)$, we pass from the coefficients $\lambda(d_1,\dots,d_k)$ to the y-variables. We expand the square defining w_n and interchange the order of summation (Lemma 8.6); we evaluate the resulting inner count by the Chinese remainder theorem and bound the aggregate error (Lemmas 8.7–8.10); we split $1/[d_i,e_i]$ across the two coordinates, detect the residual cross-coprimality by Möbius inversion, and restrict the resulting off-diagonal variables (Lemmas 8.11–8.15); we substitute the inversion formula expressing λ in terms of y, and show that all but the diagonal term $s_{i,j} \equiv 1$ is negligible (Lemmas 8.16–8.21); and finally we insert the F-derived weight, drop the last coprimality constraint, and convert the discrete sum to an integral by partial summation (Lemmas 8.22–8.25). The final statement, Lemma 8.25, is the asymptotic

$$S_1(\lambda_0) \;=\; \frac{\varphi(W)^k\,N\,(\log R)^k}{W^{k+1}}\,I_k(F) \;+\; O_k\Big(\frac{F_{\max}^2\,\varphi(W)^k\,N\,(\log R)^k}{W^{k+1}\,D_0}\Big).$$

Throughout, $k \geq 2$ is fixed, $\mathcal{H} = \{h_1,\dots,h_k\}$ is admissible (Definition 3.15), $\vartheta \in (0,1)$ and $\delta \in (0,\vartheta/2)$ are fixed real parameters, $R = N^{\vartheta/2-\delta}$ (Definition 3.37), $D_0 = \log\log\log N$ (Definition 3.28), $W = \prod_{p\leq D_0} p$ (Definition 3.29), and v_0 is a valid residue class (Lemma 3.34). No level-of-distribution hypothesis is used in this section: ϑ enters only through the truncation R, and the arithmetic input is elementary (Mertens sums and partial summation). Bombieri–Vinogradov is needed only for the second moment.

8.1 Auxiliary objects

Definition 8.1 (The lcm modulus $q(\mathbf{d},\mathbf{e})$)**.** Let $\mathbf{d} = (d_1,\dots,d_k)$ and $\mathbf{e} = (e_1,\dots,e_k)$ be k-tuples of positive integers. Define

$$q(\mathbf{d},\mathbf{e}) \;:=\; W\prod_{i=1}^{k}[d_i,e_i].$$

Lean:
- `PrimeGaps.qMod`

Uses (statement): 3.29, 4.11

Uses(notation): W, $\mathbf{d}$, $\mathbf{e}$, $[a,b]$, $q(\mathbf{d},\mathbf{e})$, k (index)

Definition 8.2 (The CRT remainder $\varrho(\mathbf{d},\mathbf{e})$)**.** For k-tuples $\mathbf{d},\mathbf{e}$ of positive integers, set

$$\varrho(\mathbf{d},\mathbf{e}) \;:=\; \sum_{\substack{N<n\le 2N\\ n\equiv v_0 \pmod W\\ [d_i,e_i]\mid n+h_i\ \forall i}} 1 \;-\; \frac{N}{q(\mathbf{d},\mathbf{e})}\,\mathbf{1}\big[W,[d_1,e_1],\dots,[d_k,e_k]\text{ are pairwise coprime}\big].$$

Uses (statement): 8.1, 3.29, 4.11

Uses(notation): h_i, k, N, v_0, $[a,b]$, W, $\mathbf{d}$, $\mathbf{e}$, $q(\mathbf{d},\mathbf{e})$ (index)

Notation 8.3 (Restricted sum $\sum^*$ over the $s_{i,j}$)**.** Fix positive integers $u_1,\dots,u_k$. For a function f of the variables $(s_{i,j})_{1\le i\ne j\le k}$ we write

$$\sum_{s_{1,2},\dots,s_{k,k-1}}{}^{\!*}\; f\big((s_{i,j})_{i\ne j}\big) \;:=\; \sum_{\substack{s_{i,j}\ge 1,\ 1\le i\ne j\le k\\ (*_1),(*_2),(*_3),(*_4)}} f\big((s_{i,j})_{i\ne j}\big),$$

where the four restrictions are

$$\begin{aligned}
&(*_1)\quad (s_{i,j},\,u_i)=1 && \text{for all } i\ne j,\\
&(*_2)\quad (s_{i,j},\,u_j)=1 && \text{for all } i\ne j,\\
&(*_3)\quad (s_{i,j},\,s_{i,a})=1 && \text{for all } a\notin\{i,j\},\\
&(*_4)\quad (s_{i,j},\,s_{b,j})=1 && \text{for all } b\notin\{i,j\}.
\end{aligned}$$

Lean:
- `PrimeGaps.LemS1RestrictSij.SigmaRestr`

Uses(notation): k, (a,b), u_i, $s_{i,j}$, $\sum^*$ (index)

Definition 8.4 (The exponent tuple $\mathbf{a}$)**.** Given positive integers $u_1,\dots,u_k$ and $(s_{i,j})_{1\le i\ne j\le k}$, set

$$a_j \;:=\; u_j\prod_{i\ne j} s_{j,i} \qquad (1\le j\le k), \qquad \mathbf{a}:=(a_1,\dots,a_k).$$

Lean:
- `PrimeGaps.boldA`

Uses(notation): k, $\mathbf{a}$, u_i, $s_{i,j}$, $\mathbb{N}$ (index)

Definition 8.5 (The exponent tuple $\mathbf{b}$)**.** With the same data, set

$$b_j \;:=\; u_j\prod_{i\ne j} s_{i,j} \qquad (1\le j\le k), \qquad \mathbf{b}:=(b_1,\dots,b_k).$$

Note the transposition of the indices of s relative to Definition 8.4: in a_j the first index of s is fixed to j, in b_j the second.

Lean:
- `PrimeGaps.boldB`

Uses (statement): 8.4

Uses(notation): k, $\mathbf{b}$, u_i, $s_{i,j}$, $\mathbb{N}$ (index)

8.2 Expansion and the Chinese remainder theorem

Lemma 8.6 (Expansion of $S_1(\lambda)$)**.** *Let $\lambda : \mathbb{N}^k \to \mathbb{R}$ have finite support. Then*

$$S_1(\lambda) \;=\; \sum_{\mathbf{d}} \sum_{\mathbf{e}} \lambda(\mathbf{d})\,\lambda(\mathbf{e}) \sum_{\substack{N<n\le 2N \\ n\equiv v_0 \pmod W \\ [d_i,e_i]\mid n+h_i \;\forall i}} 1.$$

Lean:

- `PrimeGaps.lem_S1_expansion_S1_expansion`

Uses (statement): 4.14, 4.12, 3.29, 4.11, 3.35

Uses(notation): h_i, k, w_n, N, $S_1(\lambda)$, $\lambda(d_1,\dots,d_k)$, v_0, $[a,b]$, W, $\mathbf{d}$, $\sum_{\mathbf{d}}$, $\mathbf{e}$ (index)

Proof. We perform three rearrangements.

Step 1: expand the square. By Definition 4.12,

$$w_n \;=\; \Big(\sum_{d_i\mid n+h_i\;\forall i} \lambda(\mathbf{d})\Big)^2 \;=\; \sum_{d_i\mid n+h_i\;\forall i} \sum_{e_i\mid n+h_i\;\forall i} \lambda(\mathbf{d})\,\lambda(\mathbf{e}),$$

the second equality being the distributive law for the product of two finite sums (both inner sums are over divisors of the fixed positive integers $n+h_i$, hence finite).

Step 2: substitute into S_1. By Definition 4.14,

$$S_1(\lambda) \;=\; \sum_{\substack{N<n\le 2N \\ n\equiv v_0 \pmod W}} \sum_{d_i\mid n+h_i\;\forall i} \sum_{e_i\mid n+h_i\;\forall i} \lambda(\mathbf{d})\,\lambda(\mathbf{e}).$$

Step 3: interchange. All three sums are finite: the outer one is over a finite arithmetic progression in $(N,2N]$, and the inner ones are supported in the finite support of λ. Interchanging, and merging the two divisibilities $d_i \mid n+h_i$ and $e_i \mid n+h_i$ into the single condition $[d_i,e_i] \mid n+h_i$, gives the assertion. □

Lemma 8.7 (CRT evaluation of the inner count)**.** *Let λ have R-permissible support (Definition 3.39), let v_0 be a valid residue class, and let $\mathbf{d},\mathbf{e}$ be k-tuples of positive integers with $\lambda(\mathbf{d}) \neq 0$ and $\lambda(\mathbf{e}) \neq 0$. If N is large enough that $|h_i - h_j| \le D_0$ for all $i \neq j$, then*

$$\sum_{\substack{N<n\le 2N \\ n\equiv v_0 \pmod W \\ [d_i,e_i]\mid n+h_i\;\forall i}} 1 \;=\; \begin{cases} \dfrac{N}{q(\mathbf{d},\mathbf{e})} + O(1) & \text{if } W,[d_1,e_1],\dots,[d_k,e_k] \text{ are pairwise coprime,} \\ 0 & \text{otherwise,} \end{cases}$$

with an absolute implied constant.

Lean:

- `PrimeGaps.lem_S1_CRT`

Uses (statement): 3.39, 3.29, 3.28, 8.1, 3.34, 4.11, 3.35

Uses(notation): h_i, k, N, $\lambda(d_1,\dots,d_k)$, v_0, $[a,b]$, $\lfloor x\rfloor$, $O(g)$, D_0, W, $\mathbf{d}$, $\mathbf{e}$, $q(\mathbf{d},\mathbf{e})$, (a,b)(index)

Proof. Write $q := q(\mathbf{d}, \mathbf{e})$. The summation imposes the system of congruences $n \equiv v_0 \pmod{W}$ and $n \equiv -h_i \pmod{[d_i, e_i]}$ for $i = 1, \dots, k$.

Coprime case. If the $k+1$ moduli $W, [d_1, e_1], \dots, [d_k, e_k]$ are pairwise coprime, the Chinese remainder theorem gives a unique solution modulo their product, which is exactly q. The interval $(N, 2N]$ has length N, so it contains $N/q + O(1)$ integers in a fixed residue class modulo q; the error is at most 1 in absolute value.

Non-coprime case. Suppose the moduli are not pairwise coprime, so some prime p divides two of them. If $p \mid W$ and $p \mid [d_i, e_i]$ for some i, then $p \le D_0$ by Definition 3.29, while $[d_i, e_i]$ divides $\prod_\ell d_\ell e_\ell$, which is coprime to W by R-permissibility of the support (Definition 3.39); this is a contradiction, so this case does not occur. Otherwise $p \mid [d_i, e_i]$ and $p \mid [d_j, e_j]$ for some $i \neq j$. Any n counted by the sum satisfies $p \mid n + h_i$ and $p \mid n + h_j$, hence $p \mid h_i - h_j$. As $p \nmid W$ we have $p > D_0$, whereas $0 < |h_i - h_j| \le D_0$ for N large; so $h_i - h_j$ is a nonzero integer of absolute value smaller than p, and $p \mid h_i - h_j$ is impossible. The set of counted n is therefore empty and the sum is 0. □

Lemma 8.8 (Aggregate CRT lookup error)**.** *Let $\delta \in (0, \vartheta/2)$ and let λ have R-permissible support. Then, for all sufficiently large N, with ϱ as in Definition 8.2,*

$$\Big|\sum_{\mathbf{d}}\sum_{\mathbf{e}} \lambda(\mathbf{d})\,\lambda(\mathbf{e})\,\varrho(\mathbf{d},\mathbf{e})\Big| \;\le\; \lambda_{\max}^2 \sum_{\mathbf{d}}\sum_{\mathbf{e}} |\varrho(\mathbf{d},\mathbf{e})| \;\ll_k\; \lambda_{\max}^2\, R^2\,(\log R)^{2k} \;\ll_k\; y_{\max}^2\, R^2\,(\log R)^{4k},$$

the sums being over the (finite) support of λ.

Lean:

- `PrimeGaps.gpy_aggregate_remainder_bound`

Uses (statement): 3.39, 3.37, 8.2, 4.20, 4.27, 8.7, 4.11

Uses(notation): k, N, $\lambda(d_1, \dots, d_k)$, R, $[a, b]$, $f \ll g$, W, δ, $\mathbf{d}$, $\sum_{\mathbf{d}}$, $\mathbf{e}$, $y_{\max}$, $\lambda_{\max}$ (index)

Proof. The first inequality is the triangle inequality together with $|\lambda(\mathbf{d})\lambda(\mathbf{e})| \le \lambda_{\max}^2$.

Step 1: each remainder is bounded. By Lemma 8.7, for every pair $(\mathbf{d}, \mathbf{e})$ in the support of λ we have $|\varrho(\mathbf{d}, \mathbf{e})| \le C_\varrho$ for an absolute constant C_ϱ: in the pairwise coprime case ϱ is the $O(1)$ discrepancy of the count, and in the remaining case both terms defining ϱ vanish.

Step 2: count the tuples in the support. By R-permissibility every $\mathbf{d}$ in the support satisfies $d_i \ge 1$ for all i and $\prod_i d_i \le R$. The number of such tuples is at most

$$\#\Big\{\mathbf{d} \in \mathbb{N}^k : \prod_i d_i \le R\Big\} \;\le\; \big\lfloor R \big\rfloor \cdot \Big(\sum_{1 \le a \le \lfloor R \rfloor} \frac{1}{a}\Big)^k \;\ll_k\; R\,(\log R)^k.$$

The middle bound holds because on the counting set one has $1 \le R \prod_i d_i^{-1}$, so the cardinality is at most $R \sum_{\mathbf{d}} \prod_i d_i^{-1}$ with each d_i running over $[1, \lfloor R \rfloor]$, and this factorises into k harmonic sums; the last step uses $\sum_{a \le x} 1/a \le 1 + \log x$.

Step 3: assemble. Squaring Step 2 bounds the number of pairs by $\ll_k R^2 (\log R)^{2k}$; multiplying by $C_\varrho \lambda_{\max}^2$ gives the second inequality.

Step 4: pass from $\lambda_{\max}$ to $y_{\max}$. By Lemma 7.17, $\lambda_{\max} \ll_k y_{\max} (\log R)^k$, so $\lambda_{\max}^2 \ll_k y_{\max}^2 (\log R)^{2k}$; substituting gives the last bound. □

Uses (proof): 7.17

Lemma 8.9 (Negligibility of the CRT lookup error)**.** *Let $\vartheta \in (0,1)$, $\delta \in (0, \vartheta/2)$ and $Y \geq 0$. Then, for all sufficiently large N,*

$$Y^2\, R^2\, (\log R)^{4k} \;\ll_k\; \frac{Y^2\, \varphi(W)^k\, N\, (\log R)^k}{W^{k+1}\, D_0}.$$

Lean:

- `PrimeGaps.crt_error_negligible`

Uses (statement): 3.37, 3.28, 3.29, 3.4

Uses(notation): k, N, R, φ, $f \ll g$, ϑ, D_0, W, δ (index)

Proof. Both sides carry the factor $Y^2 \geq 0$, so we may divide it out.

Step 1: polynomial saving. By Definition 3.37, $R^2 = N^{\vartheta - 2\delta}$, and $\vartheta - 2\delta < \vartheta < 1$. Put $\eta := 1 - (\vartheta - 2\delta) > 0$, so $R^2 = N^{1-\eta}$ and, using $\log R \leq \log N$,

$$R^2 (\log R)^{4k} \;\leq\; N^{1-\eta} (\log N)^{4k}.$$

Step 2: the W-factor is harmless. By Lemma 3.32, $W \ll (\log\log N)^2 \leq \log N$ for large N, so $W^{k+1} \leq (\log N)^{2k+2}$; combined with $\varphi(W) \geq 1$ this gives $\varphi(W)^k (\log N)^{2k+2} / W^{k+1} \geq 1$.

Step 3: combine. Multiplying the bound of Step 1 by the quantity of Step 2,

$$R^2 (\log R)^{4k} \;\leq\; N^{1-\eta}\, (\log N)^{6k+2}\, \frac{\varphi(W)^k}{W^{k+1}} \;=\; N\, \frac{\varphi(W)^k}{W^{k+1}} \cdot N^{-\eta} (\log N)^{6k+2}.$$

Since $\log R = (\vartheta/2 - \delta) \log N$ and $D_0 = \log\log\log N$, the target factor $(\log R)^k / D_0$ is $\gg_k (\log N)^k / \log\log\log N$, whereas $N^{-\eta}(\log N)^{6k+2} \to 0$ as $N \to \infty$ because $\eta > 0$ dominates any fixed power of $\log N$. Hence $N^{-\eta}(\log N)^{6k+2} \leq (\log R)^k / D_0$ for large N, which is the claim. □

Uses (proof): 3.32

Lemma 8.10 (Main term of $S_1(\lambda)$ after the CRT step)**.** *Let $k \geq 2$, let $\mathcal{H}$ be admissible, let $\vartheta \in (0,1)$ and $\delta \in (0, \vartheta/2)$. Then, for all sufficiently large N, for every λ with R-permissible support and every valid v_0,*

$$S_1(\lambda) \;=\; \frac{N}{W} \sideset{}{'}\sum_{\mathbf{d},\mathbf{e}} \frac{\lambda(\mathbf{d})\, \lambda(\mathbf{e})}{\prod_{i=1}^k [d_i, e_i]} \;+\; O_k\Big(\frac{y_{\max}^2\, \varphi(W)^k\, N\, (\log R)^k}{W^{k+1}\, D_0} \Big),$$

where $\sideset{}{'}\sum$ denotes summation restricted to pairs $(\mathbf{d}, \mathbf{e})$ of tuples of positive integers for which $W, [d_1, e_1], \ldots, [d_k, e_k]$ are pairwise coprime.

Lean:

- `GPYSieveS1.main_S1_asymptotic`

Uses (statement): 4.14, 3.15, 3.39, 3.37, 3.28, 3.29, 4.20, 4.27, 3.34, 4.11, 3.35

Uses(notation): h_i, k, N, $S_1(\lambda)$, $\lambda(d_1, \ldots, d_k)$, R, v_0, $[a,b]$, φ, $f \ll g$, $O(g)$, D_0, W, ϑ, δ, $\mathbf{d}$, $\mathbf{e}$, $y_{\max}$, $\lambda_{\max}$, $q(\mathbf{d}, \mathbf{e})$, $\sideset{}{'}\sum$ (index)

Proof. Step 1. By Lemma 8.6, $S_1(\lambda) = \sum_{\mathbf{d},\mathbf{e}} \lambda(\mathbf{d})\lambda(\mathbf{e})\, T(\mathbf{d},\mathbf{e})$, where $T(\mathbf{d},\mathbf{e})$ is the inner count.
Step 2. Admissibility makes $h_1,\dots,h_k$ distinct, so for N large enough $|h_i - h_j| \le D_0$ for all $i \neq j$, and Lemma 8.7 applies to every pair in the support. By Definition 8.2 this says precisely

$$T(\mathbf{d},\mathbf{e}) \;=\; \frac{N}{q(\mathbf{d},\mathbf{e})}\,\mathbf{1}\big[W,[d_1,e_1],\dots,[d_k,e_k] \text{ pairwise coprime}\big] \;+\; \varrho(\mathbf{d},\mathbf{e}).$$

Substituting into Step 1 and using $N/q(\mathbf{d},\mathbf{e}) = (N/W)\prod_i [d_i,e_i]^{-1}$,

$$S_1(\lambda) \;=\; \frac{N}{W}\sideset{}{'}\sum_{\mathbf{d},\mathbf{e}} \frac{\lambda(\mathbf{d})\lambda(\mathbf{e})}{\prod_i [d_i,e_i]} \;+\; \sum_{\mathbf{d},\mathbf{e}} \lambda(\mathbf{d})\lambda(\mathbf{e})\,\varrho(\mathbf{d},\mathbf{e}).$$

Step 3. By Lemma 8.8 the second sum is $\ll_k y_{\max}^2 R^2 (\log R)^{4k}$.
Step 4. By Lemma 8.9 applied with $Y = y_{\max}$, this is $\ll_k y_{\max}^2 \varphi(W)^k N (\log R)^k/(W^{k+1} D_0)$, which is the stated error term. □

Uses (proof): 8.6, 8.7, 8.8, 8.9, 8.2, 8.1

8.3 Decoupling the lcm and the Möbius inversion

Lemma 8.11 (Removing the automatic constraints)**.** *Let $\lambda : \mathbb{N}^k \to \mathbb{R}$ be such that $\lambda(\mathbf{t}) \neq 0$ implies that every t_i is positive, $\prod_i t_i$ is squarefree, and $(\prod_i t_i, W) = 1$. Then, for every function f of pairs of tuples,*

$$\sideset{}{'}\sum_{\mathbf{d},\mathbf{e}} \lambda(\mathbf{d})\,\lambda(\mathbf{e})\, f(\mathbf{d},\mathbf{e}) \;=\; \sum_{\substack{\mathbf{d},\mathbf{e}\\ (d_i,e_j)=1\ \forall i\neq j}} \lambda(\mathbf{d})\,\lambda(\mathbf{e})\, f(\mathbf{d},\mathbf{e}),$$

where $\sum'$ is the restricted summation of Lemma 8.10.

Lean:

- `PrimeGaps.lem_S1_remove_automatic`

Uses (statement): 8.10, 3.39, 3.29, 4.11

Uses(notation): k, $\lambda(d_1,\dots,d_k)$, (a,b), $[a,b]$, W, $\mathbf{d}$, $\mathbf{e}$, $\sum'$ (index)

Proof. It suffices to show that, on the support of $\lambda(\mathbf{d})\lambda(\mathbf{e})$, the pairwise coprimality of $W,[d_1,e_1],\dots,[d_k,e_k]$ is equivalent to $(d_i,e_j)=1$ for all $i\neq j$; both sides then have the same summands.

Step 1: the intra-coprimalities are automatic. If $\lambda(\mathbf{d})\neq 0$ then $\prod_i d_i$ is squarefree. If a prime p divided both d_i and d_j with $i\neq j$, then $p^2 \mid \prod_\ell d_\ell$, a contradiction. Hence $(d_i,d_j)=1$ for $i\neq j$, and likewise $(e_i,e_j)=1$ for $i\neq j$.

Step 2: coprimality to W is automatic. From $(\prod_i d_i, W)=1$ and $(\prod_i e_i, W)=1$, and since $[d_i,e_i]$ divides $\prod_\ell d_\ell e_\ell$, we get $([d_i,e_i],W)=1$ for every i.

Step 3: the residual condition. Pairwise coprimality of the $k+1$ moduli unwinds into $([d_i,e_i],W)=1$ for all i (Step 2) together with $([d_i,e_i],[d_j,e_j])=1$ for $i\neq j$. A prime p dividing both $[d_i,e_i]$ and $[d_j,e_j]$ divides one of d_i,e_i and one of d_j,e_j; by Step 1 the possibilities $p \mid d_i,d_j$ and $p \mid e_i,e_j$ are excluded, leaving $p \mid (d_i,e_j)$ or $p \mid (d_j,e_i)$. Thus the only non-automatic constraint is $(d_i,e_j)=1$ for all $i\neq j$, and conversely that constraint together with Steps 1–2 gives the full pairwise coprimality. □

Lemma 8.12 (Decoupling the outer lcm product)**.** *Let $\lambda : \mathbb{N}^k \to \mathbb{R}$ be such that $\lambda(\mathbf{t}) \neq 0$ implies that every t_i is positive and squarefree and that $t_1, \dots, t_k$ are pairwise coprime. Then*

$$\sum_{\substack{\mathbf{d},\mathbf{e} \\ (d_i,e_j)=1\ \forall\, i\neq j}} \frac{\lambda(\mathbf{d})\,\lambda(\mathbf{e})}{\prod_{i=1}^{k}[d_i,e_i]} = \sum_{\substack{\mathbf{d},\mathbf{e} \\ (d_i,e_j)=1\ \forall\, i\neq j}} \Big(\sum_{\substack{u_1,\dots,u_k \\ u_i \mid (d_i,e_i)\ \forall i}} \prod_{i=1}^{k} \varphi(u_i) \Big) \frac{\lambda(\mathbf{d})\,\lambda(\mathbf{e})}{\prod_{i=1}^{k} d_i e_i}.$$

Lean:

- `PrimeGaps.lem_S1_decouple_lcm`

Uses (statement): 8.11, 6.3, 3.4, 4.11

Uses(notation): k, $\lambda(d_1,\dots,d_k)$, (a,b), $[a,b]$, φ, $\mathbf{d}$, $\mathbf{e}$, $\mathbf{u}$, u_i (index)

Proof. We may work summand by summand, and we may assume $\lambda(\mathbf{d})\lambda(\mathbf{e}) \neq 0$, since otherwise both summands vanish. Then every d_i and every e_i is positive and squarefree, so Lemma 6.3 gives, for each i,

$$\frac{1}{[d_i,e_i]} = \frac{1}{d_i e_i} \sum_{u_i \mid (d_i,e_i)} \varphi(u_i).$$

Taking the product over $i = 1,\dots,k$ and expanding the product of k finite sums into a single sum over the tuple $(u_1,\dots,u_k)$ with $u_i \mid (d_i,e_i)$,

$$\frac{1}{\prod_{i=1}^{k}[d_i,e_i]} = \frac{1}{\prod_{i=1}^{k} d_i e_i} \sum_{\substack{u_1,\dots,u_k \\ u_i \mid (d_i,e_i)\ \forall i}} \prod_{i=1}^{k} \varphi(u_i).$$

Multiplying by $\lambda(\mathbf{d})\lambda(\mathbf{e})$ and summing over the pairs satisfying the cross-coprimality guard gives the identity; the guard is a condition on $(\mathbf{d},\mathbf{e})$ alone and is untouched by the manipulation, since the $\mathbf{u}$-sum is internal to the summand. □

Lemma 8.13 (Möbius inversion of the constraint $(d_i,e_j)=1$)**.** *For every finite family of coefficients λ and every pair $(\mathbf{d},\mathbf{e})$,*

$$\sum_{\substack{\mathbf{d},\mathbf{e} \\ (d_i,e_j)=1\ \forall\, i\neq j}} \Big(\sum_{\substack{u_1,\dots,u_k \\ u_i \mid (d_i,e_i)}} \prod_{i} \varphi(u_i) \Big) \frac{\lambda(\mathbf{d})\lambda(\mathbf{e})}{\prod_i d_i e_i} = \Sigma_{\mathrm{full}},$$

where Σ_{full} is the finite sum

$$\Sigma_{\mathrm{full}} = \sum_{\mathbf{d},\mathbf{e}} \sum_{\substack{u_1,\dots,u_k \\ u_i \mid (d_i,e_i)}} \sum_{\substack{s_{1,2},\dots,s_{k,k-1} \\ s_{i,j} \mid (d_i,e_j)\ (i\neq j)}} \Big(\prod_{i=1}^{k} \varphi(u_i) \Big) \Big(\prod_{i\neq j} \mu(s_{i,j}) \Big) \frac{\lambda(\mathbf{d})\,\lambda(\mathbf{e})}{\prod_{i=1}^{k} d_i e_i}.$$

Lean:

- `PrimeGaps.LemS1RestrictSij.lem_S1_mobius_coprime`

Uses (statement): 8.12, 3.3, 3.4, 4.11

Uses(notation): k, $\lambda(d_1,\dots,d_k)$, μ, (a,b), φ, $\mathbf{d}$, $\mathbf{e}$, $\mathbf{u}$, u_i, $s_{i,j}$ (index)

Proof. Step 1: factor the indicator. The constraint is a conjunction over the $k(k-1)$ ordered pairs (i,j) with $i \neq j$, so

$$\mathbf{1}\big[(d_i, e_j) = 1 \ \forall\, i \neq j\big] \;=\; \prod_{i \neq j} \mathbf{1}\big[(d_i, e_j) = 1\big].$$

Step 2: detect each factor by Möbius inversion. By Lemma 3.7 applied at $n = (d_i, e_j)$,

$$\mathbf{1}\big[(d_i, e_j) = 1\big] \;=\; \sum_{s_{i,j} \mid (d_i, e_j)} \mu(s_{i,j}),$$

and $s_{i,j} \mid (d_i, e_j)$ is the conjunction $s_{i,j} \mid d_i$ and $s_{i,j} \mid e_j$. Taking the product over the $k(k-1)$ pairs turns Step 1 into a single sum over tuples $(s_{i,j})_{i \neq j}$ of positive integers weighted by $\prod_{i \neq j} \mu(s_{i,j})$.
Step 3: insert and interchange. Insert the expansion of Step 2 into the left-hand side. All sums are finite: $\mathbf{d}, \mathbf{e}$ range over the finite support of λ, the u_i over divisors of (d_i, e_i) and the $s_{i,j}$ over divisors of (d_i, e_j). Interchanging the order of summation and merging the divisibility indicators into the summation ranges yields exactly Σ_{full}. □

Uses (proof): 3.7

Lemma 8.14 (Pointwise coprimalities of the $s_{i,j}$)**.** *Let λ be such that $\lambda(\mathbf{t}) \neq 0$ implies that every t_i is positive and that $t_1, \ldots, t_k$ are pairwise coprime. Let $(\mathbf{d}, \mathbf{e}, \mathbf{u}, (s_{i,j})_{i \neq j})$ be a configuration occurring in Σ_{full}, so that $u_i \mid d_i$, $u_i \mid e_i$ and $s_{i,j} \mid d_i$, $s_{i,j} \mid e_j$ for $i \neq j$, and suppose its summand is nonzero, i.e. $\lambda(\mathbf{d})\lambda(\mathbf{e}) \neq 0$. Then the four restrictions $(*_1)$–$(*_4)$ of Notation 8.3 hold.*

Lean:
- `PrimeGaps.LemS1RestrictSij.lem_S1_sij_coprimality`

Uses (statement): 8.13, 8.3, 3.39, 4.11

Uses(notation): k, $\lambda(d_1, \ldots, d_k)$, (a,b), $\mathbf{d}$, $\mathbf{e}$, u_i, $s_{i,j}$ (index)

Proof. By hypothesis the coordinates of $\mathbf{d}$ are pairwise coprime and so are those of $\mathbf{e}$:

$$(d_i, d_j) = 1, \qquad (e_i, e_j) = 1 \qquad (i \neq j). \tag{$*$}$$

Each of the four cases assumes a prime p divides $s_{i,j}$ and one further listed factor, and contradicts $(*)$.
$(*_1)$. If $p \mid s_{i,j}$ and $p \mid u_i$, then $p \mid s_{i,j} \mid e_j$ and $p \mid u_i \mid e_i$, so $p \mid (e_i, e_j)$ with $i \neq j$: contradiction.
$(*_2)$. If $p \mid s_{i,j}$ and $p \mid u_j$, then $p \mid s_{i,j} \mid d_i$ and $p \mid u_j \mid d_j$, so $p \mid (d_i, d_j)$: contradiction.
$(*_3)$. If $p \mid s_{i,j}$ and $p \mid s_{i,a}$ with $a \notin \{i,j\}$, then $p \mid e_j$ and $p \mid e_a$ with $j \neq a$: contradiction.
$(*_4)$. If $p \mid s_{i,j}$ and $p \mid s_{b,j}$ with $b \notin \{i,j\}$, then $p \mid d_i$ and $p \mid d_b$ with $i \neq b$: contradiction. □

Lemma 8.15 (The restricted and unrestricted s-sums agree)**.** *Under the hypothesis of Lemma 8.14 on λ, let Σ_{restr} denote the sum obtained from Σ_{full} by replacing the unrestricted s-summation by the restricted summation $\sum^{*}$ of Notation 8.3. Then*

$$\Sigma_{\text{restr}} \;=\; \Sigma_{\text{full}}.$$

Lean:
- `PrimeGaps.LemS1RestrictSij.lem_S1_restrict_sij`

Uses (statement): 8.13, 8.14, 8.3, 4.11

Uses(notation): k, $\lambda(d_1,\dots,d_k)$, (a,b), $\mathbf{d}$, $\mathbf{e}$, u_i, $s_{i,j}$, $\sum^*$ (index)

Proof. The restricted sum is obtained from the unrestricted one by deleting the configurations that violate at least one of $(*_1)$–$(*_4)$. By Lemma 8.14, every configuration whose summand is nonzero satisfies all four restrictions; contrapositively, every deleted configuration has summand 0. Deleting terms that vanish does not change the value of a sum, so the two sums are equal. □

8.4 The substitution $\lambda \rightsquigarrow y$

Lemma 8.16 (Substituting y into $S_1(\lambda)$)**.** *Let $k \geq 2$, let $\mathcal{H}$ be admissible, let $\vartheta \in (0,1)$ and $\delta \in (0,\vartheta/2)$. Then, for all sufficiently large N, for every λ with R-permissible support and every valid v_0,*

$$S_1(\lambda) \;=\; \frac{N}{W}\,\Sigma_y(\lambda) \;+\; O_k\Big(\frac{y_{\max}^2\,\varphi(W)^k\,N\,(\log R)^k}{W^{k+1}\,D_0}\Big),$$

where, with $\mathbf{a},\mathbf{b}$ as in Definitions 8.4 and 8.5, $\Sigma_y(\lambda)$ is the finite sum

$$\Sigma_y(\lambda) \;=\; \sum_{\mathbf{u}}\Big(\prod_{i=1}^{k}\frac{\mu(u_i)^2}{\varphi(u_i)}\Big)\sum_{s_{1,2},\dots,s_{k,k-1}}^{*}\Big(\prod_{i\neq j}\frac{\mu(s_{i,j})}{\varphi(s_{i,j})^2}\Big)\,y(a_1,\dots,a_k)\,y(b_1,\dots,b_k).$$

Lean:

- `PrimeGaps.lem_S1_substitute_y`

Uses (statement): 4.14, 3.15, 3.39, 3.37, 3.28, 3.29, 4.20, 4.27, 8.4, 8.5, 3.34, 8.3, 4.11, 3.35

Uses(notation): k, N, $S_1(\lambda)$, $\lambda(d_1,\dots,d_k)$, R, μ, (a,b), φ, $f \ll g$, $O(g)$, D_0, W, ϑ, δ, $\mathbf{d}$, $\mathbf{e}$, $\mathbf{u}$, $\mathbf{a}$, $\mathbf{b}$, $y(r_1,\dots,r_k)$, $y_{\max}$, $\sum^*$, u_i, $s_{i,j}$ (index)

Proof. Step 1: the starting point. By Lemma 8.10, $S_1(\lambda)$ equals (N/W) times the primed sum, up to the stated error. Applying, in turn, Lemma 8.11 (legitimate: R-permissibility of the support gives positivity, squarefreeness of $\prod_i t_i$ and coprimality to W), Lemma 8.12 (legitimate: squarefreeness of $\prod_i t_i$ gives squarefreeness and pairwise coprimality of the coordinates), Lemma 8.13 and Lemma 8.15, we obtain

$$S_1(\lambda) \;=\; \frac{N}{W}\sum_{\mathbf{u}}\Big(\prod_{i=1}^{k}\varphi(u_i)\Big)\sum_{s_{1,2},\dots,s_{k,k-1}}^{*}\Big(\prod_{i\neq j}\mu(s_{i,j})\Big)\sum_{\substack{\mathbf{d},\mathbf{e}\\ u_i\mid d_i,e_i\\ s_{i,j}\mid d_i,e_j}}\frac{\lambda(\mathbf{d})\,\lambda(\mathbf{e})}{\prod_{i=1}^{k}d_ie_i} \;+\; E,$$

with E the error term of Lemma 8.10.

Step 2: collapse the $\mathbf{d}$-sum. Fix $\mathbf{u}$ and $(s_{i,j})$ satisfying the restrictions of Notation 8.3, which by Lemma 8.14 is no loss. Then, for each i, the integers u_i and $\{s_{i,j}\}_{j\neq i}$ are pairwise coprime, so the conjunction of $u_i \mid d_i$ and $s_{i,j} \mid d_i$ for all $j \neq i$ is equivalent to the single divisibility $a_i \mid d_i$ (Definition 8.4). Hence the inner $\mathbf{d}$-sum is over $a_i \mid d_i$ for all i, and by Definition 4.20 and the Möbius inversion of Lemma 4.23,

$$\sum_{\substack{\mathbf{d}\\ a_i\mid d_i\ \forall i}}\frac{\lambda(\mathbf{d})}{\prod_i d_i} \;=\; \frac{y(a_1,\dots,a_k)}{\prod_{i=1}^{k}\mu(a_i)\,\varphi(a_i)}$$

whenever $\prod_i a_i$ is squarefree, the summand vanishing otherwise. The same argument with b_i in place of a_i (using coprimality of u_i and $\{s_{j,i}\}_{j\neq i}$, and Definition 8.5) collapses the $\mathbf{e}$-sum.

Step 3: recombine the arithmetic factors. On the support, $a_i = u_i \prod_{j\neq i} s_{i,j}$ and $b_i = u_i \prod_{j\neq i} s_{j,i}$ are squarefree with pairwise coprime factors, so μ and φ are multiplicative across them. Collecting the three sources of $\mu(s_{i,j})$ — one from $\mu(a_i)$, one from $\mu(b_j)$, and the Möbius weight inserted in Lemma 8.13 — and the factors $\varphi(u_i)^2$ against $\prod_i \varphi(u_i)$, we obtain, for each i, a factor $\mu(u_i)^2/\varphi(u_i)$, and for each ordered pair $i \neq j$ a factor $\mu(s_{i,j})/\varphi(s_{i,j})^2$. This is exactly the summand of $\Sigma_y(\lambda)$, and Step 1's error term is unchanged. □

Uses (proof): 8.10, 8.11, 8.12, 8.13, 8.15, 8.14, 4.23, 4.22

Lemma 8.17 (Contributing $s_{i,j}$ are coprime to W)**.** *Let λ have R-permissible support. If a configuration $(\mathbf{u}, (s_{i,j}))$ contributes to $\Sigma_y(\lambda)$, that is if $y(a_1,\dots,a_k) \neq 0$ and $y(b_1,\dots,b_k) \neq 0$, then $(s_{i,j}, W) = 1$ for every $i \neq j$.*

Lean:

- `PrimeGaps.lem_S1_sij_coprime_to_W`

Uses (statement): 8.16, 3.39, 4.20, 3.29, 8.5

Uses(notation): k, $\lambda(d_1,\dots,d_k)$, (a,b), W, $\mathbf{d}$, $\mathbf{b}$, $y(r_1,\dots,r_k)$, u_i, $s_{i,j}$ (index)

Proof. Fix $i \neq j$. By Definition 4.20, $y(b_1,\dots,b_k) \neq 0$ forces the inner sum defining it to be nonzero, so there is a tuple $\mathbf{d}$ with $\lambda(\mathbf{d}) \neq 0$ and $b_\ell \mid d_\ell$ for every ℓ. By Definition 8.5, $s_{i,j} \mid b_j \mid d_j$. By R-permissibility of the support of λ we have $(\prod_\ell d_\ell, W) = 1$, hence $(d_j, W) = 1$ and therefore $(s_{i,j}, W) = 1$. □

Lemma 8.18 (Dichotomy for the contributing $s_{i,j}$)**.** *Let λ have R-permissible support and let $W = \prod_{p\le D_0} p$. If a configuration contributes to $\Sigma_y(\lambda)$, then for every $i \neq j$ either $s_{i,j} = 1$, or every prime factor of $s_{i,j}$ exceeds D_0; in particular $s_{i,j} = 1$ or $s_{i,j} > D_0$.*

Lean:

- `PrimeGaps.lem_S1_sij_dichotomy`

Uses (statement): 8.17, 3.29, 3.28, 3.39

Uses(notation): (a,b), D_0, W, $\mathbb{P}$, $s_{i,j}$, k (index)

Proof. If $s_{i,j} = 1$ there is nothing to prove, so assume $s_{i,j} > 1$ and let q be a prime factor of $s_{i,j}$. By Lemma 8.17, $(s_{i,j}, W) = 1$, so $q \nmid W$. By Definition 3.29, W is the product of all primes $p \le D_0$, so a prime divides W if and only if it is at most D_0. Hence $q > D_0$. Since $s_{i,j}$ is then divisible by a prime exceeding D_0, we get $s_{i,j} > D_0$. □

Lemma 8.19 (The contribution of $s_{i,j} > D_0$)**.** *Let $k \ge 2$, $\vartheta \in (0,1)$, $\delta \in (0, \vartheta/2)$. For all sufficiently large N and every λ with R-permissible support, the part of $(N/W)\Sigma_y(\lambda)$ supported on configurations with $s_{i,j} > D_0$ for at least one pair $i \neq j$ is*

$$O_k\Big(\frac{y_{\max}^2\, \varphi(W)^k\, N\, (\log R)^k}{W^{k+1}\, D_0}\Big).$$

Lean:

- `PrimeGaps.lem_S1_sij_D0`

Uses (statement): 8.16, 3.39, 3.37, 3.28, 3.29, 4.27, 8.3

Uses(notation): k, N, $\lambda(d_1,\dots,d_k)$, R, μ, (a,b), φ, $f \ll g$, $O(g)$, D_0, W, ϑ, δ, $y(r_1,\dots,r_k)$, $y_{\max}$, u_i, $s_{i,j}$, $\sum^*$ (index)

Proof. There are $k(k-1)$ ordered pairs (i,j) with $i \neq j$, so by the union bound it suffices to bound, for one fixed pair, the contribution of the configurations with $s_{i,j} > D_0$, and then multiply by $k(k-1)$.

Bound $|y(\mathbf{a})\,y(\mathbf{b})| \le y_{\max}^2$ and take absolute values in the summand. Every variable is constrained to be squarefree and coprime to W (Lemma 8.17 and R-permissibility), and each of $\mathbf{a},\mathbf{b}$ has product at most R on the support of y, so each u_ℓ and each $s_{a,b}$ is at most R. Therefore the contribution of the fixed pair is at most

$$y_{\max}^2\Big(\sum_{\substack{u\le R\\(u,W)=1}}\frac{\mu(u)^2}{\varphi(u)}\Big)^k\Big(\sum_{\substack{s\le R\\(s,W)=1}}\frac{\mu(s)^2}{\varphi(s)^2}\Big)^{k(k-1)-1}\sum_{\substack{s>D_0\\(s,W)=1}}\frac{\mu(s)^2}{\varphi(s)^2}.$$

By Lemma 6.30 the first factor is $\ll \big(\varphi(W)(\log R)/W\big)^k$; the middle factor is $O_k(1)$ because $\sum_{s\ge1}\mu(s)^2/\varphi(s)^2$ converges; and the last factor is $\ll 1/D_0$, since $\varphi(s)\gg s/\log\log(3s)$ makes the tail $\sum_{s>D_0}\mu(s)^2/\varphi(s)^2$ comparable to $\sum_{s>D_0}s^{-2+o(1)} \ll 1/D_0$. Multiplying by N/W and by $k(k-1)$ gives the assertion. □

Uses (proof): 8.17, 6.30

Lemma 8.20 (The diagonal term $s_{i,j}\equiv 1$)**.** *For every $\lambda:\mathbb{N}^k\to\mathbb{R}$ of finite support,*

$$\sum_{\mathbf{u}}\Big(\prod_{i=1}^k\frac{\mu(u_i)^2}{\varphi(u_i)}\Big)\,y(u_1,\dots,u_k)^2 \;=\; \sum_{\mathbf{u}}\frac{y(u_1,\dots,u_k)^2}{\prod_{i=1}^k\varphi(u_i)},$$

the sums being over k-tuples of positive integers. (The left-hand side is the summand of $\Sigma_y(\lambda)$ at $s_{i,j}=1$ for all $i\neq j$, since then $\mathbf{a}=\mathbf{b}=\mathbf{u}$, $\mu(1)=\varphi(1)=1$, and the restrictions of Notation 8.3 are vacuous.)

Lean:

- `PrimeGaps.lem_S1_sij_one`

Uses (statement): 8.16, 4.20, 3.3, 3.4, 8.4, 8.5, 8.3

Uses(notation): k, $\lambda(d_1,\dots,d_k)$, μ, φ, $\mathbf{u}$, $\mathbf{a}$, $\mathbf{b}$, $y(r_1,\dots,r_k)$, u_i, $s_{i,j}$, $\mathbb{N}$ (index)

Proof. It suffices to compare the two summands at a fixed tuple $\mathbf{u}$. If $\prod_i u_i$ is squarefree then each $\mu(u_i)^2=1$ and the two summands coincide. If $\prod_i u_i$ is not squarefree, then by Definition 4.20 the value $y(u_1,\dots,u_k)$ vanishes — the y-variables are supported on tuples with squarefree product — so both summands are 0. In either case the summands agree, and hence so do the sums. □

Lemma 8.21 ($S_1(\lambda)$ in terms of y)**.** *Let $k\ge2$, let $\mathcal{H}$ be admissible, let $\vartheta\in(0,1)$ and $\delta\in(0,\vartheta/2)$. Then, for all sufficiently large N, for every λ with R-permissible support and every valid v_0,*

$$S_1(\lambda) \;=\; \frac{N}{W}\sum_{\mathbf{u}}\frac{y(u_1,\dots,u_k)^2}{\prod_{i=1}^k\varphi(u_i)} \;+\; O_k\Big(\frac{y_{\max}^2\,\varphi(W)^k\,N\,(\log R)^k}{W^{k+1}\,D_0}\Big).$$

Lean:

- `PrimeGaps.lem_S1_from_y`

Uses (statement): 4.14, 3.15, 3.39, 3.37, 3.28, 3.29, 4.20, 4.27, 3.34, 4.11, 3.35

Uses(notation): k, N, $S_1(\lambda)$, $\lambda(d_1,\dots,d_k)$, R, φ, $f\ll g$, $O(g)$, D_0, W, ϑ, δ, $\mathbf{u}$, $y(r_1,\dots,r_k)$, $y_{\max}$, u_i (index)

Proof. By Lemma 8.16, $S_1(\lambda) = (N/W)\,\Sigma_y(\lambda) + E_1$ with $E_1 \ll_k y_{\max}^2 \varphi(W)^k N(\log R)^k/(W^{k+1}D_0)$.

Split $\Sigma_y(\lambda)$ according to whether all the $s_{i,j}$ with $i \neq j$ equal 1. By Lemma 8.18, on every contributing configuration each $s_{i,j}$ is either 1 or exceeds D_0; hence the configurations that are not diagonal are exactly those with $s_{i,j} > D_0$ for at least one pair, and their total contribution to $(N/W)\Sigma_y(\lambda)$ is $O_k\big(y_{\max}^2\varphi(W)^k N(\log R)^k/(W^{k+1}D_0)\big)$ by Lemma 8.19.

The diagonal part is (N/W) times the left-hand side of Lemma 8.20, which that lemma identifies with $(N/W)\sum_{\mathbf{u}} y(\mathbf{u})^2/\prod_i \varphi(u_i)$. Adding the two error contributions gives the assertion. □

Uses (proof): 8.16, 8.18, 8.19, 8.20

8.5 The smooth case

Lemma 8.22 (Substituting the smooth y)**.** *Let $k \geq 2$, let $\mathcal{H}$ be admissible, let $\vartheta \in (0,1)$ and $\delta \in (0, \vartheta/2)$, and let $F : \mathbb{R}^k \to \mathbb{R}$ be smooth with support contained in $\mathcal{R}_k$ (Definition 5.2). Let λ_0 be the F-derived sieve weight of Definition 5.11. Then there is a constant $C > 0$, depending only on k, ϑ and δ, such that for all sufficiently large N and every valid v_0,*

$$\Big|S_1(\lambda_0) \;-\; \frac{N}{W} \sum_{\substack{u_1,\dots,u_k \\ \prod_i u_i \text{ squarefree} \\ (\prod_i u_i, W)=1}} \frac{1}{\prod_{i=1}^k \varphi(u_i)}\, F\Big(\frac{\log u_1}{\log R}, \dots, \frac{\log u_k}{\log R}\Big)^2\Big| \;\leq\; C\,\frac{F_{\max}^2\,\varphi(W)^k\,N\,(\log R)^k}{W^{k+1}\,D_0}.$$

Lean:

- `PrimeGaps.lem_S1_substitute_smooth_y`

Uses (statement): 8.21, 7.3, 5.11, 5.2, 7.4, 4.20, 4.27, 3.37, 3.28, 3.29, 3.35

Uses(notation): k, N, $S_1(\lambda)$, R, F, μ, (a,b), φ, D_0, W, $\mathcal{R}_k$, ϑ, δ, $\mathbf{u}$, $y(r_1,\dots,r_k)$, $y_{\max}$, u_i, $F_{\max}$, $\lambda_0(d_1,\dots,d_k)$, $C^\infty(X)$ (index)

Proof. Apply Lemma 8.21 to $\lambda = \lambda_0$, which has R-permissible support by Definition 5.11. By Lemma 7.3, the associated y-variables are given in closed form:

$$y(u_1,\dots,u_k) \;=\; \begin{cases} F\big(\frac{\log u_1}{\log R},\dots,\frac{\log u_k}{\log R}\big) & \text{if } \prod_i u_i \text{ is squarefree and } (\prod_i u_i, W) = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Substituting this into the main term of Lemma 8.21 restricts the $\mathbf{u}$-sum to tuples with $\prod_i u_i$ squarefree and coprime to W and replaces $y(\mathbf{u})^2$ by $F(\log u_1/\log R,\dots)^2$, which is the displayed main term.

For the error term, the same closed form gives $y_{\max} = \sup_{\mathbf{u}} |y(\mathbf{u})| \leq \sup_{[0,1]^k} |F| \leq F_{\max}$: indeed the support condition on F forces F to vanish unless $\log u_i/\log R \in [0,1]$ for every i, and $F_{\max} \geq \sup_{[0,1]^k}|F|$ by Definition 7.4. Replacing $y_{\max}$ by $F_{\max}$ in the error term of Lemma 8.21 gives the stated bound. □

Uses (proof): 8.21, 7.3

Lemma 8.23 (Dropping the constraint $(u_i,u_j)=1$)**.** *Let $\vartheta \in (0,1)$, $\delta \in (0,\vartheta/2)$. There is a constant $C > 0$ depending only on k, ϑ, δ such that for all sufficiently large N and every smooth $F : \mathbb{R}^k \to \mathbb{R}$ supported in $\mathcal{R}_k$,*

$$\Big|\sum_{\substack{\mathbf{u} \\ \prod_i u_i \text{ squarefree} \\ (\prod_i u_i, W)=1}} \frac{F\big(\frac{\log u_1}{\log R},\dots\big)^2}{\prod_i \varphi(u_i)} - \sum_{\substack{\mathbf{u} \\ (u_i,W)=1\ \forall i}} \Big(\prod_{i=1}^k \frac{\mu(u_i)^2}{\varphi(u_i)}\Big) F\Big(\frac{\log u_1}{\log R},\dots\Big)^2\Big| \;\leq\; C\,\frac{F_{\max}^2\,\varphi(W)^k\,(\log R)^k}{W^k\,D_0}.$$

Lean:

- `PrimeGaps.drop_uij_coupling_bound`

Uses (statement): 8.22, 5.2, 7.4, 3.37, 3.28, 3.29, 3.3, 3.4

Uses(notation): k, R, F, μ, (a,b), φ, $f \ll g$, D_0, W, $\mathcal{R}_k$, ϑ, δ, $\mathbf{u}$, u_i, $F_{\max}$ (index)

Proof. Step 1: identify the difference. A tuple $\mathbf{u}$ has $\prod_i u_i$ squarefree and coprime to W if and only if every u_i is squarefree and coprime to W *and* the u_i are pairwise coprime. On such tuples $\mu(u_i)^2 = 1$ for every i, so the two summands coincide. The second sum runs over the larger index set $\{(u_i, W) = 1\ \forall i\}$, on which its summand is nonnegative and vanishes unless each u_i is squarefree. Hence the difference is, in absolute value, exactly

$$\sum_{\substack{\mathbf{u}:\ (u_i,W)=1\ \forall i \\ \exists\, i\neq j:\ (u_i,u_j)>1}} \Big(\prod_{\ell=1}^{k} \frac{\mu(u_\ell)^2}{\varphi(u_\ell)}\Big) F\Big(\tfrac{\log u_1}{\log R}, \dots\Big)^2.$$

Step 2: a shared prime is large. Fix $i \neq j$ and suppose a prime p divides both u_i and u_j. Since $(u_i, W) = 1$ and $W = \prod_{q \le D_0} q$, we have $p > D_0$.

Step 3: bound one pair. Bound $F^2 \le F_{\max}^2$; the support of F inside $\mathcal{R}_k$ forces $u_\ell \le R$ for every ℓ in each nonzero term. For fixed $i \neq j$ and a fixed prime $p > D_0$, summing over u_i, u_j divisible by p and over the remaining coordinates freely,

$$\sum_{\substack{\mathbf{u}\le R,\ (u_\ell,W)=1 \\ p\mid u_i,\ p\mid u_j}} \prod_{\ell} \frac{\mu(u_\ell)^2}{\varphi(u_\ell)} \;\le\; \frac{1}{(p-1)^2} \Big(\sum_{\substack{u\le R \\ (u,W)=1}} \frac{\mu(u)^2}{\varphi(u)} \Big)^k,$$

because for squarefree u_i with $p \mid u_i$ one has $\mu(u_i)^2/\varphi(u_i) = (p-1)^{-1}\mu(u_i/p)^2/\varphi(u_i/p)$.

Step 4: sum over p and over pairs. By Lemma 6.30, $\sum_{u\le R,(u,W)=1} \mu(u)^2/\varphi(u) \ll \varphi(W)(\log R)/W$. Moreover $\sum_{p>D_0}(p-1)^{-2} \ll 1/D_0$. Summing Step 3 over the primes $p > D_0$ and over the $k(k-1)$ ordered pairs (i,j) bounds the quantity of Step 1 by

$$\ll_k \; F_{\max}^2 \cdot \frac{1}{D_0} \cdot \frac{\varphi(W)^k (\log R)^k}{W^k},$$

which is the assertion. □

Uses (proof): 6.30

Lemma 8.24 (Applying partial summation once per coordinate)**.** *Let $\vartheta \in (0,1)$, $\delta \in (0, \vartheta/2)$. There is a constant $C > 0$ depending only on k, ϑ, δ such that for all sufficiently large N and every smooth $F : \mathbb{R}^k \to \mathbb{R}$ supported in $\mathcal{R}_k$,*

$$\Big| \sum_{\substack{u_1,\dots,u_k \\ (u_i,W)=1\ \forall i}} \Big(\prod_{i=1}^{k} \frac{\mu(u_i)^2}{\varphi(u_i)}\Big) F\Big(\frac{\log u_1}{\log R}, \dots, \frac{\log u_k}{\log R}\Big)^2 \;-\; \frac{\varphi(W)^k (\log R)^k}{W^k}\, I_k(F) \Big|$$
$$\le\; C\, \frac{F_{\max}^2\, \varphi(W)^k\, (\log D_0)\, (\log R)^{k-1}}{W^k}.$$

Lean:

- `PrimeGaps.S1_aggregate`

Uses (statement): 8.23, 6.43, 5.2, 5.6, 7.4, 3.37, 3.28, 3.29, 3.41

Uses(notation): k, R, F, μ, $I_k(F)$, (a,b), φ, $f \ll g$, $O(g)$, D_0, W, $\mathcal{R}_k$, $\mathfrak{S}(\gamma)$, γ_W, g_*, G, $G_{\max}$, A_1, A_2, L, A_3, V, τ, u_i, $F_{\max}$ (index)

Proof. We verify the hypotheses of Lemma 6.43 for the density γ_W and then iterate it once per coordinate.

Step 1: the sieve data. Let γ_W be the multiplicative density with $\gamma_W(p) = \mathbf{1}[p \nmid W]$. Then $\gamma_W(p)/p \le 1/2$ for all p, so the growth hypothesis holds with $A_1 = 1/2$. For the logarithmic hypothesis, Mertens' first theorem gives $\sum_{w\le p\le z}(\log p)/p = \log(z/w) + O(1)$, while the primes dividing W contribute at most $\sum_{p\le D_0}(\log p)/p = \log D_0 + O(1)$; subtracting, the hypothesis holds with $A_2 = O(1)$ and $L \ll \log D_0$. Finally the structural hypothesis holds with $V = W$ (squarefree) and $A_3 = 0$, since $\gamma_W(p) = 0$ for $p \mid W$ and $\gamma_W(p) = 1$ otherwise; the associated residual error of Lemma 6.43 is $O(\tau(W)\log(2WR)/\log R)$, negligible because $W \le (\log\log N)^2$ while $\log R$ is a fixed positive multiple of $\log N$.

Step 2: the singular product and the multiplicative weight. By Lemma 6.34, $\mathfrak{S}(\gamma_W) = \varphi(W)/W$. The associated totally multiplicative function is $g_*(p) = \gamma_W(p)/(p - \gamma_W(p))$, i.e. $g_*(p) = 0$ for $p \mid W$ and $g_*(p) = 1/(p-1)$ for $p \nmid W$. Consequently, for squarefree u coprime to W, $g_*(u) = 1/\varphi(u)$, and $\mu(u)^2 g_*(u) = \mu(u)^2/\varphi(u)$ is supported exactly on the squarefree u coprime to W — the vanishing off the integers coprime to the modulus being Lemma 4.7 at $V = W$ — which is precisely the summand appearing in the statement.

Step 3: one coordinate at a time. Take $z := R$ throughout; the support of F inside $\mathcal{R}_k$ makes the summand vanish unless $\log u_i/\log R \le 1$, i.e. $u_i \le R$. Fix values $x_\ell := \log u_\ell/\log R$ for $\ell > 1$ and apply Lemma 6.43 to the smooth profile $G_1(t) := F(t, x_2, \dots, x_k)^2$ on $[0,1]$, whose C^1-norm satisfies $G_{\max} \ll F_{\max}^2$ by Definition 7.4. This yields

$$\sum_{\substack{u_1\le R\\(u_1,W)=1}} \frac{\mu(u_1)^2}{\varphi(u_1)} F\Big(\frac{\log u_1}{\log R}, x_2, \dots, x_k\Big)^2 = \frac{\varphi(W)}{W}(\log R)\int_0^1 F(t_1, x_2, \dots, x_k)^2\, dt_1 + O\Big(\frac{\varphi(W)}{W}(\log D_0)\, F_{\max}^2\Big).$$

Iterate over the coordinates $i = 2, \dots, k$, at stage i applying Lemma 6.43 to the profile $G_i(t) := \int_{[0,1]^{i-1}} F(t_1, \dots, t_{i-1}, t, x_{i+1}, \dots, x_k)^2\, dt_1 \cdots dt_{i-1}$, whose C^1-norm is again $\ll F_{\max}^2$.

Step 4: aggregate. After k iterations the main term is

$$\Big(\frac{\varphi(W)\log R}{W}\Big)^k \int_{[0,1]^k} F(t_1, \dots, t_k)^2\, dt_1 \cdots dt_k = \frac{\varphi(W)^k (\log R)^k}{W^k} I_k(F),$$

using that F is supported in $\mathcal{R}_k \subseteq [0,1]^k$ so that the integral over the cube is $I_k(F)$ (Definition 5.6). The total error is the sum over $i = 1, \dots, k$ of the terms in which the error appears in the i-th coordinate and the main term in the other $k-1$; each such term has size

$$\Big(\frac{\varphi(W)\log R}{W}\Big)^{k-1} \cdot \frac{\varphi(W)}{W}(\log D_0)\, F_{\max}^2 = \frac{F_{\max}^2\, \varphi(W)^k (\log D_0)(\log R)^{k-1}}{W^k},$$

and there are k of them. This is the stated bound. □

Uses (proof): 6.43, 6.34, 4.7, 3.32

Lemma 8.25 (Asymptotic for $S_1(\lambda_0)$ in the smooth case)**.** *Let $k \geq 2$, let $\mathcal{H}$ be admissible, let $\vartheta \in (0,1)$ and $\delta \in (0, \vartheta/2)$, and let $F : \mathbb{R}^k \to \mathbb{R}$ be smooth with support contained in $\mathcal{R}_k$. Let λ_0 be the F-derived sieve weight of Definition 5.11. Then there is a constant $C > 0$, depending only on k, ϑ and δ, such that for all sufficiently large N and every valid v_0,*

$$\Big| S_1(\lambda_0) \;-\; \frac{\varphi(W)^k\, N\,(\log R)^k}{W^{k+1}}\, I_k(F) \Big| \;\leq\; C\, \frac{F_{\max}^2\, \varphi(W)^k\, N\,(\log R)^k}{W^{k+1}\, D_0}.$$

Lean:

- `PrimeGaps.lem_S1_smooth`

Uses (statement): 4.14, 3.15, 5.11, 5.2, 5.6, 7.4, 3.37, 3.28, 3.29, 3.34, 3.35

Uses(notation): k, N, $S_1(\lambda)$, R, F, $I_k(F)$, φ, D_0, W, $\mathcal{R}_k$, ϑ, δ, $F_{\max}$, $\lambda_0(d_1, \dots, d_k)$, $O(g)$ (index)

Proof. Write $A := \sum_{\mathbf{u}} \mathbf{1}[\prod_i u_i \text{ squarefree}, (\prod_i u_i, W) = 1]\, F(\log u_1/\log R, \dots)^2 / \prod_i \varphi(u_i)$ for the sum appearing in Lemma 8.22, B for the sum appearing on the right of Lemma 8.23, and $M := \varphi(W)^k (\log R)^k I_k(F)/W^k$. The three preceding lemmas give, for N large,

$$\Big| S_1(\lambda_0) - \frac{N}{W} A \Big| \leq C_1 \frac{F_{\max}^2 \varphi(W)^k N (\log R)^k}{W^{k+1} D_0}, \qquad |A - B| \leq C_2 \frac{F_{\max}^2 \varphi(W)^k (\log R)^k}{W^k D_0},$$

$$|B - M| \leq C_3 \frac{F_{\max}^2 \varphi(W)^k (\log D_0)(\log R)^{k-1}}{W^k}.$$

Since $N/W \geq 0$, the triangle inequality gives

$$\Big| S_1(\lambda_0) - \frac{N}{W} M \Big| \;\leq\; C_1 \frac{F_{\max}^2 \varphi(W)^k N (\log R)^k}{W^{k+1} D_0} \;+\; \frac{N}{W} \big(|A - B| + |B - M| \big),$$

and $(N/W)M$ is exactly $\varphi(W)^k N (\log R)^k I_k(F)/W^{k+1}$, the asserted main term.

It remains to see that the third error is of the same shape as the first two. Set $K := F_{\max}^2 \varphi(W)^k N (\log R)^{k-1}/W^{k+1} \geq 0$; then the first two errors are $C_1 K (\log R)/D_0$ and $C_2 K (\log R)/D_0$, while the third is $C_3 K \log D_0$. Since $D_0 = \log\log\log N$ and $\log R = (\vartheta/2 - \delta)\log N$, we have $D_0 \log D_0 \leq \log R$ for all large N, that is $\log D_0 \leq (\log R)/D_0$. Hence the third error is at most $C_3 K (\log R)/D_0$, and the total is at most $(C_1 + C_2 + C_3) K (\log R)/D_0$, which is the stated bound with $C := C_1 + C_2 + C_3$. □

Uses (proof): 8.22, 8.23, 8.24

9 Sieve manipulations: the reduction of $S_2^{(m)}$

For each $m \in \{1, \dots, k\}$ we carry out the passage from the sieve coefficients λ to the smooth y-variables in the second moment $S_2^{(m)}(\lambda)$, ending in an asymptotic expressed through the functional $J_k^{(m)}$. The argument parallels the reduction of $S_1(\lambda)$ — expand the square, evaluate the inner sum by the Chinese remainder theorem, decouple the least common multiples, invert the residual coprimality conditions by Möbius, and substitute the transformed variables — but two features are new. First, the inner sum now carries the prime indicator $\chi_{\mathbb{P}}(n + h_m)$, so the Chinese remainder step produces a main term $X_N/\varphi(q)$ and an error $E(N, q)$ rather than an exact count; controlling the aggregate of those errors over the whole support of λ requires the hypothesis that the primes have a level of distribution ϑ, and is the only place where Bombieri–Vinogradov (Theorem 1) is used. Second, the transformed variables are the $y^{(m)}$ of

Definition 4.25, whose normalisation is $\prod_i \varphi(d_i)$ rather than $\prod_i d_i$; this is the source of the multiplicative function g of Definition 3.6 throughout.

Throughout the section $k \geq 2$ is fixed, $\mathcal{H} = \{h_1, \dots, h_k\}$ is admissible, $m \in \{1, \dots, k\}$ is a distinguished index, $D_0 = \log\log\log N$, $W = \prod_{p \leq D_0} p$, v_0 is a residue compatible with $\mathcal{H}$ modulo W, $\vartheta \in (0, 1/2)$ is a level of distribution for the primes, $\delta \in (0, \vartheta/2)$ and $R = N^{\vartheta/2-\delta}$. The sieve coefficients λ are always assumed to have permissible support (Definition 3.39) at the parameters R and W. All implied constants depend only on k and $\mathcal{H}$, never on N, on λ, on the profile F, or on m, unless explicitly stated otherwise.

9.1 Expansion and the Chinese remainder step

Lemma 9.1 (Expansion of $S_2^{(m)}(\lambda)$)**.** *For every finitely supported λ,*

$$S_2^{(m)}(\lambda) \;=\; \sum_{\mathbf{d},\mathbf{e}} \lambda(\mathbf{d})\,\lambda(\mathbf{e}) \sum_{\substack{N < n \leq 2N \\ n \equiv v_0 \pmod W \\ [d_i, e_i] \mid n + h_i \;\forall i}} \chi_{\mathbb{P}}(n + h_m).$$

Lean:

- `PrimeGaps.lem_S2m_expansion`

Uses (statement): 4.16, 4.12, 3.29

Uses(notation): k, h_i, N, v_0, W, w_n, $\lambda(d_1, \dots, d_k)$, $\chi_{\mathbb{P}}$, $\mathbf{d}$, $\mathbf{e}$, $[a, b]$, $S_2^{(m)}(\lambda)$, $\sum_{\mathbf{d}}$ (index)

Proof. This is the same three-step rearrangement that produced the expansion of $S_1(\lambda)$ (Lemma 8.6), with the extra factor $\chi_{\mathbb{P}}(n + h_m)$ carried inertly through every step; only the coordinatewise divisibility identity is used, so the argument applies verbatim to any weight attached to n.

Step 1: expand the square. By Definition 4.12,

$$w_n \;=\; \Big(\sum_{d_i \mid n + h_i \,\forall i} \lambda(\mathbf{d}) \Big)^2 \;=\; \sum_{d_i \mid n + h_i \,\forall i} \; \sum_{e_i \mid n + h_i \,\forall i} \lambda(\mathbf{d})\,\lambda(\mathbf{e}),$$

the two inner sums being finite because λ has finite support.

Step 2: substitute into $S_2^{(m)}$. By Definition 4.16, $S_2^{(m)}(\lambda)$ is the sum of $\chi_{\mathbb{P}}(n + h_m)\, w_n$ over $N < n \leq 2N$ with $n \equiv v_0 \pmod W$; inserting Step 1 gives a triple sum.

Step 3: interchange and merge the divisibilities. All three sums are finite, so the n-summation may be moved inside. For each fixed i the pair of conditions $d_i \mid n + h_i$ and $e_i \mid n + h_i$ holds if and only if $[d_i, e_i] \mid n + h_i$. Merging them coordinatewise yields the displayed identity. □

Uses (proof): 8.6

Lemma 9.2 (Primality forces $d_m = e_m = 1$)**.** *Let x, c be positive integers with $2 \leq c < x$ and $c \mid x$. Then x is not prime. Consequently, for all N large enough that $R < N$, the inner sum of Lemma 9.1 vanishes unless $d_m = e_m = 1$.*

Lean:

- `PrimeGaps.lem_S2m_dm_em_one`

Uses (statement): 9.1, 3.37, 3.39

Uses(notation): h_i, N, R, $\mathbb{P}$, $\chi_{\mathbb{P}}$, $[a, b]$ (index)

Proof. The first assertion is immediate: c is a divisor of x with $c \neq 1$ and $c \neq x$, so x has a non-trivial divisor and is composite.

For the consequence, suppose $[d_m, e_m] > 1$ for a pair $(\mathbf{d}, \mathbf{e})$ in the support of λ, and let n contribute to the inner sum, so that $[d_m, e_m] \mid n + h_m$. By Definition 3.39 the support of λ is contained in the tuples with $\prod_i d_i \leq R$, whence $[d_m, e_m] \leq d_m e_m \leq R^2$; more crudely, $[d_m, e_m] \leq R < N < n + h_m$ once N exceeds a threshold depending only on ϑ and δ. Applying the first assertion with $c = [d_m, e_m]$ and $x = n+h_m$ shows $n+h_m$ is composite, so $\chi_{\mathbb{P}}(n+h_m) = 0$. Hence the entire inner sum vanishes. □

Lemma 9.3 (Chinese remainder evaluation of the inner sum)**.** *There is a constant $C > 0$, depending only on k, $\mathcal{H}$ and m, with the following property. Let $W \geq 1$, let v_0 satisfy $(v_0 + h_i, W) = 1$ for every i, and let $\mathbf{d}, \mathbf{e}$ be tuples of positive integers such that*

(i) $d_m = e_m = 1$;

(ii) *W is coprime to $[d_i, e_i]$ for every i, and the $[d_i, e_i]$ are pairwise coprime;*

(iii) $(h_m - h_i,\, [d_i, e_i]) = 1$ *for every $i \neq m$.*

Put $q = q(\mathbf{d}, \mathbf{e}) = W \prod_{i=1}^{k} [d_i, e_i]$. Then

$$\Bigg| \sum_{\substack{N < n \leq 2N \\ n \equiv v_0 \pmod W \\ [d_i, e_i] \mid n + h_i \ \forall i}} \chi_{\mathbb{P}}(n + h_m) \;-\; \frac{X_N}{\varphi(q)} \Bigg| \;\leq\; C\, E(N, q).$$

Lean:

- `PrimeGaps.lem_S2m_CRT`

Uses (statement): 8.1, 3.24, 3.26, 3.4, 3.29

Uses(notation): k, h_i, N, v_0, W, $\chi_{\mathbb{P}}$, φ, (a, b), $[a, b]$, $\mathbf{d}$, $\mathbf{e}$, $q(\mathbf{d}, \mathbf{e})$, X_N, $E(N, q)$ (index)

Proof. Hypothesis (ii) makes $W, [d_1, e_1], \ldots, [d_k, e_k]$ a pairwise coprime family, so the Chinese remainder theorem identifies the system of congruences

$$n \equiv v_0 \pmod W, \qquad n \equiv -h_i \pmod{[d_i, e_i]} \quad (1 \leq i \leq k)$$

with a single congruence $n \equiv a \pmod q$, for a residue a determined by the data. The inner sum is therefore $\sum_{N < n \leq 2N,\ n \equiv a\ (q)} \chi_{\mathbb{P}}(n + h_m)$, that is, the number of primes in $(N + h_m, 2N + h_m]$ lying in the single residue class $a + h_m \pmod q$, up to at most $2h_k$ boundary terms; the boundary discrepancy is absorbed into the constant C, using $E(N, q) \geq 1$.

It remains to see that the class $a + h_m$ is coprime to q, so that the class is admissible for the definition of $E(N, q)$ (Notation 3.26). Let $p \mid q$. If $p \mid W$ then $a + h_m \equiv v_0 + h_m \pmod p$, which is coprime to W by the choice of v_0. If $p \mid [d_i, e_i]$ for some i, then by (ii) this i is unique; when $i = m$ hypothesis (i) gives $[d_m, e_m] = 1$, so no such p exists; when $i \neq m$ we have $a + h_m \equiv h_m - h_i \pmod p$, which is nonzero modulo p by hypothesis (iii). Hence $(a + h_m, q) = 1$.

By the very definition of $E(N, q)$ as the supremum, over classes coprime to q, of the discrepancy between the prime count in that class and $X_N/\varphi(q)$, the inner sum differs from $X_N/\varphi(q)$ by at most $E(N, q)$, and the claim follows with a constant absorbing the boundary terms. □

9.2 The aggregate error term and Bombieri–Vinogradov

Lemma 9.3 replaces the inner sum by a main term at the cost of one error $E(N, q(\mathbf{d}, \mathbf{e}))$ per pair $(\mathbf{d}, \mathbf{e})$. The aggregate of these errors, weighted by $|\lambda(\mathbf{d})\lambda(\mathbf{e})|$, is negligible. Three ingredients combine: the moduli $q(\mathbf{d}, \mathbf{e})$ are squarefree and lie below R^2W, which is a power $N^{\vartheta-2\delta} \cdot N^{o(1)}$ strictly below $N^{1/2}$; each modulus is attained by at most τ_{3k} pairs; and a level of distribution ϑ controls the sum of $E(N, q)$ over all $q \le N^{\vartheta}$. Only the last of these uses anything beyond elementary arithmetic, and it is exactly Theorem 1.

Sublemma 9.4 (Squarefreeness of the sieve modulus)**.** *Let W be squarefree and let $(\mathbf{d}, \mathbf{e})$ be a pair in the support of a weight λ with permissible support at (R, W). Then $q(\mathbf{d}, \mathbf{e})$ is squarefree.*

Lean:
- `PrimeGaps.qMod_squarefree`

Uses (statement): 3.39, 8.1, 3.29

Uses(notation): k, W, μ, (a, b), $[a, b]$, $\mathbf{d}$, $\mathbf{e}$, $q(\mathbf{d}, \mathbf{e})$, $\lambda(d_1, \dots, d_k)$ (index)

Proof. By Definition 3.39 the products $\prod_i d_i$ and $\prod_i e_i$ are squarefree and coprime to W. Squarefreeness of $\prod_i d_i$ forces each d_i to be squarefree and $(d_i, d_j) = 1$ for $i \ne j$: a prime dividing two distinct coordinates would divide the product twice. The same holds for the e_i.

Fix i. Since d_i and e_i are squarefree, so is $[d_i, e_i]$ — indeed $[d_i, e_i] = (d_i, e_i) \cdot \frac{d_i}{(d_i, e_i)} \cdot \frac{e_i}{(d_i, e_i)}$ is a product of three pairwise coprime squarefree integers. For $i \ne j$, a prime dividing both $[d_i, e_i]$ and $[d_j, e_j]$ would divide one of d_i, e_i and one of d_j, e_j; each of the four cases contradicts one of $(d_i, d_j) = 1$, $(e_i, e_j) = 1$, $(d_i, e_j) = 1$, $(e_i, d_j) = 1$, the last two being part of the permissible-support condition. Hence $\prod_i [d_i, e_i]$ is a product of pairwise coprime squarefree integers, so is squarefree.

Finally W is squarefree (it is a product of distinct primes) and coprime to $\prod_i [d_i, e_i]$, since every prime factor of the latter divides some d_i or e_i. A product of two coprime squarefree integers is squarefree, so $q(\mathbf{d}, \mathbf{e}) = W \prod_i [d_i, e_i]$ is squarefree. □

Sublemma 9.5 (Cardinality of the q-fibre)**.** *Let $W \ge 1$ and let $r \ge 1$ be a multiple of W. Then the set of pairs $(\mathbf{d}, \mathbf{e})$ with $q(\mathbf{d}, \mathbf{e}) = r$ has cardinality at most $\tau_{3k}(r/W)$.*

Lean:
- `PrimeGaps.phiTuple_injOn`
- `PrimeGaps.card_Pr_le`

Uses (statement): 8.1, 3.5, 6.5, 3.39

Uses(notation): k, W, τ_r, (a, b), $[a, b]$, $\mathbf{d}$, $\mathbf{e}$, $q(\mathbf{d}, \mathbf{e})$ (index)

Proof. We construct an injection from the fibre into the set of ordered factorisations of r/W into $3k$ factors, whose cardinality is $\tau_{3k}(r/W)$ by Definition 3.5.

Given a pair in the fibre, apply the per-coordinate lcm decomposition of Sublemma 6.5 in each coordinate: it sends the pair (d_i, e_i) to the triple $\big((d_i, e_i),\ d_i/(d_i, e_i),\ e_i/(d_i, e_i)\big)$, whose entries are pairwise coprime with product $[d_i, e_i]$, and it is a bijection onto such triples, with inverse $(g, a, b) \mapsto (ga, gb)$. Concatenating the k triples gives an ordered $3k$-tuple whose product is $\prod_i [d_i, e_i] = q(\mathbf{d}, \mathbf{e})/W = r/W$.

The map is injective: from the $3k$-tuple one recovers each pair (d_i, e_i) by the inverse of Sublemma 6.5, coordinate by coordinate. Injectivity into a set of cardinality $\tau_{3k}(r/W)$ gives the bound. □

Uses (proof): 6.5

Sublemma 9.6 (Multiplicity of a squarefree modulus)**.** *Let λ have permissible support at (R, W) with $W \geq 1$, and let $r \geq 1$ be squarefree. Then*

$$\big|\{(\mathbf{d}, \mathbf{e}) \text{ in the support of } \lambda : q(\mathbf{d}, \mathbf{e}) = r\}\big| \;\leq\; \tau_{3k}(r).$$

Lean:
- `PrimeGaps.multiplicity_bound`

Uses (statement): 9.5, 8.1, 3.5, 3.39

Uses(notation): k, W, τ_r, μ, $[a,b]$, $\mathbf{d}$, $\mathbf{e}$, $q(\mathbf{d},\mathbf{e})$, $\lambda(d_1,\dots,d_k)$ (index)

Proof. If $W \nmid r$ the fibre is empty, since $W \mid q(\mathbf{d},\mathbf{e})$ always. Otherwise Sublemma 9.5 bounds the fibre by $\tau_{3k}(r/W)$. Since r is squarefree and $W \mid r$, we have $r = W \cdot (r/W)$ with $(W, r/W) = 1$; as τ_{3k} is multiplicative and takes values ≥ 1 on positive integers, $\tau_{3k}(r) = \tau_{3k}(W)\,\tau_{3k}(r/W) \geq \tau_{3k}(r/W)$. □

Uses (proof): 9.5

Sublemma 9.7 (Majorisation of the error sum by a sum over moduli)**.** *Let $k \geq 2$, let ϑ, δ satisfy $0 < \delta < \vartheta/2$, and let λ have permissible support at level $\lfloor R \rfloor$ and modulus W. Then there are a constant $C > 0$ and a threshold N_0, depending only on k, ϑ and δ, such that for all $N \geq N_0$*

$$\sum_{\mathbf{d},\mathbf{e}} |\lambda(\mathbf{d})\,\lambda(\mathbf{e})|\, E\big(N, q(\mathbf{d},\mathbf{e})\big) \;\leq\; C\,\lambda_{\max}^2 \sum_{1 \leq q \leq N^{\vartheta}} \tau_{3k}(q)^2\, E(N,q),$$

the outer sum being over pairs in the support of λ.

Lean:
- `PrimeGaps.sieve_error_majorization`

Uses (statement): 9.4, 9.6, 3.39, 3.37, 3.26, 8.1

Uses(notation): k, N, R, W, μ, τ_r, $[a,b]$, $\mathbf{d}$, $\mathbf{e}$, $\lambda(d_1,\dots,d_k)$, $\lambda_{\max}$, $q(\mathbf{d},\mathbf{e})$, $E(N,q)$ (index)

Proof. Group the pairs according to the value $r = q(\mathbf{d},\mathbf{e})$. By Sublemma 9.4 every such r is squarefree, so the grouped sum is supported on squarefree r, i.e. carries the factor $\mu(r)^2$. By Sublemma 9.6 each value r is attained by at most $\tau_{3k}(r)$ pairs, and $|\lambda(\mathbf{d})\lambda(\mathbf{e})| \leq \lambda_{\max}^2$ throughout. For the range: on the support of λ we have $\prod_i d_i \leq R$ and $\prod_i e_i \leq R$ (Definition 3.39), hence $\prod_i [d_i, e_i] \leq \prod_i d_i e_i \leq R^2$ and $r = W \prod_i [d_i, e_i] \leq R^2 W$. □

Uses (proof): 9.4, 9.6

Sublemma 9.8 (Divisor-weighted sum of local errors)**.** *Let $k \geq 2$, let $\vartheta \in (0,1)$ be a level of distribution for the primes, and let $A > 0$. Then there are $C > 0$ and N_0 such that for every $N \geq N_0$,*

$$\sum_{1 \leq q \leq N^{\vartheta}} \tau_{3k}(q)^2\, E(N,q) \;\leq\; C\,\frac{N}{(\log N)^A}.$$

Uses (statement): 3.21, 3.26, 3.5, 3.27

Uses(notation): k, N, φ, τ_r, ϑ, $O(g)$, $f \ll g$, $E(N,q)$, X_N, $o(g)$ (index)

Proof. Write $E(N,q) = 1 + \mathcal{E}(N,q)$ with $\mathcal{E} \geq 0$ the genuine discrepancy, and treat the two pieces separately.

The constant piece. A crude divisor bound suffices: for every $\alpha > 0$ one has $\tau_r(n) \ll_{r,\alpha} n^{\alpha}$, proved by comparing, at each prime power $p^a \| n$, the polynomial factor $\binom{a+r-1}{r-1}$ with the geometric factor $p^{a\alpha}$, which dominates it for all p beyond a threshold depending on r and α, the finitely many remaining primes contributing a bounded factor. Taking α so small that $\vartheta(1+2\alpha) < 1$ gives $\sum_{q \leq N^{\vartheta}} \tau_{3k}(q)^2 \ll N^{\vartheta(1+2\alpha)} = o\big(N/(\log N)^A\big)$.

The discrepancy piece. By Cauchy–Schwarz,

$$\sum_{q \leq N^{\vartheta}} \tau_{3k}(q)^2\, \mathcal{E}(N,q) \;\leq\; \Big(\sum_{q \leq N^{\vartheta}} \frac{\tau_{3k}(q)^4}{\varphi(q)}\, N \Big)^{1/2} \Big(\sum_{q \leq N^{\vartheta}} \frac{\varphi(q)\, \mathcal{E}(N,q)^2}{N} \Big)^{1/2}.$$

For the first factor, split each q uniquely as a squarefree integer times a squarefull one: on squarefree q the sum $\sum_{q \leq z} \tau_{3k}(q)^4/\varphi(q) \ll_k z\,(\log z)^{(3k)^4}$ by a Mertens-type estimate (Lemma 6.8 applied with the parameter $(3k)^4$, using $\tau_{3k}(q)^4 = \tau_{(3k)^4}(q)$ on squarefree arguments, which is Lemma 3.9 applied twice), while the squarefull part contributes a convergent factor because $\sum_{w \text{ squarefull}} \tau_{3k}(w)^4/\varphi(w) < \infty$. Hence the first factor is $\ll_k N^{(1+\vartheta)/2} (\log N)^B$ for some $B = B(k)$. For the second factor, the trivial bound $\mathcal{E}(N,q) \ll N/\varphi(q)$ — a residue class modulo q meets $(N, 2N]$ in at most $N/q + 1$ integers, and $X_N \leq N$ — gives $\varphi(q)\mathcal{E}(N,q)/N \ll 1$, so the second factor is $\ll \big(\sum_{q \leq N^{\vartheta}} \mathcal{E}(N,q)\big)^{1/2}$. Finally the level of distribution ϑ (Definition 3.21), in the dyadic form of Lemma 3.27, bounds $\sum_{q \leq N^{\vartheta}} \mathcal{E}(N,q)$ by $\ll_{A'} N/(\log N)^{A'}$ for any prescribed $A' > 0$.

Multiplying the two factors gives $\ll_{k,A'} N/(\log N)^{A'/2 - B}$; choosing $A' \geq 2A + 2B$ makes this $\ll N/(\log N)^A$, and adding the constant piece completes the proof. □

Uses (proof): 3.27, 6.8, 3.9

Lemma 9.9 (Aggregate error bound)**.** *Let $k \geq 2$, let $A > 0$, and let $\vartheta < 1/2$ be a level of distribution for the primes, with $0 < \delta < \vartheta/2$. Then there are $C > 0$ and N_0 such that for every $N \geq N_0$ and every λ with permissible support at (R, W),*

$$\sum_{\mathbf{d},\mathbf{e}} |\lambda(\mathbf{d})\, \lambda(\mathbf{e})|\, E\big(N, q(\mathbf{d},\mathbf{e})\big) \;\leq\; C\, \frac{y_{\max}^2\, N}{(\log N)^A}.$$

Lean:

- `MaynardS2Error.exists_totalErrorContribution_le`

Uses (statement): 9.7, 9.8, 7.17, 3.39, 3.37, 3.29, 3.26, 8.1, 1

Uses(notation): k, N, R, W, μ, φ, τ_r, ϑ, δ, $O(g)$, $f \ll g$, $\mathbf{d}$, $\mathbf{e}$, $\lambda(d_1, \dots, d_k)$, $\lambda_{\max}$, $y_{\max}$, $q(\mathbf{d},\mathbf{e})$, $E(N,q)$, D_0 (index)

Proof. By Sublemma 9.7 the left-hand side is at most $\lambda_{\max}^2 \sum_{1 \leq r \leq R^2 W} \mu(r)^2 \tau_{3k}(r) E(N,r)$.

We first check that the range $R^2 W$ lies inside the range N^{ϑ} covered by the level of distribution. Indeed $R^2 = N^{\vartheta - 2\delta}$ and $W \leq e^{2D_0} = N^{o(1)}$ by Lemma 3.32, so $R^2 W \leq N^{\vartheta}$ for all N large enough, $\delta > 0$ being fixed; the same computation, with the sharper bound of Lemma 3.38, places $R^2 W$ below $N^{1/2 - \eta}$ for some $\eta > 0$, which is what makes Bombieri–Vinogradov applicable at all.

Bounding $\mu(r)^2 \tau_{3k}(r) \leq \tau_{3k}(r)^2$ and applying Sublemma 9.8 with the exponent $A + 2k$ gives

$$\sum_{1 \leq r \leq R^2 W} \mu(r)^2\, \tau_{3k}(r)\, E(N,r) \;\ll_{k,A}\; \frac{N}{(\log N)^{A+2k}}.$$

Finally Lemma 7.17 bounds the sieve coefficients in terms of the y-variables, $\lambda_{\max} \ll_k y_{\max}(\log N)^k$, so $\lambda_{\max}^2 \ll_k y_{\max}^2(\log N)^{2k}$. Multiplying the two displays absorbs the $(\log N)^{2k}$ and leaves $C\, y_{\max}^2 N/(\log N)^A$. □

Uses (proof): 9.7, 9.8, 7.17, 3.32, 3.38

9.3 Partial summation and the g-weighted Mertens sum

The $s_{i,j}$-estimates of the next subsection and the evaluation of the smooth case both rest on one arithmetic input: the Mertens-type sum $\sum_{r<R,\,(r,W)=1} \mu(r)^2 G(\log r/\log R)/g(r)$, and its multi-coordinate version. We record these here.

Two refinements of the one-variable partial-summation estimate of Lemma 6.43 are needed. First, iterating that estimate over several coordinates produces, at each intermediate stage, a profile obtained by integrating out the coordinates already processed; such a profile is bounded and Lipschitz, but there is no reason for it to be continuously differentiable, so a version of partial summation valid for Lipschitz profiles is required. Second, the accumulated error over k coordinates is acceptable only if the residual of the one-variable estimate is split into a piece decaying like a power of z and a piece of size $O(1/\log z)$; the crude bound $O(\log(2Vz)/\log z)$ of Lemma 6.43 is not enough to produce a $\log D_0$ saving at each layer.

Sublemma 9.10 (Partial summation for a Lipschitz profile)**.** *There are functions $C_1 = C_1(A_1, A_3) > 0$ and $C_2 = C_2(A_1, A_3) > 0$ such that the following holds for every sieve datum $(\gamma, V, A_1, A_2, A_3, L)$, every real $z \geq 2$, every $M \geq 0$ and every $G : [0,1] \to \mathbb{R}$ that is bounded by M and M-Lipschitz:*

$$\Big|\sum_{0<d<z} \mu(d)^2 g_*(d)\, G\Big(\frac{\log d}{\log z}\Big) \;-\; \mathfrak{S}(\gamma)\,(\log z)\int_0^1 G\Big| \;\leq\; 2M\,\mathcal{L}(\gamma, V, z),$$

where $\mathcal{L}(\gamma, V, z)$ denotes the real number satisfying

$$\mathcal{L}(\gamma, V, z) \;=\; 2C_1\,\mathfrak{S}(\gamma)\,(1+\ell_V) \;+\; C_2\,\tau(V)\Big(z^{-1/8}\log(2Vz) \;+\; \frac{8\log(2V)+64}{\log z}\Big).$$

Uses (statement): 4.3, 4.4, 4.6, 6.18, 6.41, 6.43

Uses(notation): μ, τ, g_*, $h(d)$, $\mathfrak{S}(\gamma)$, γ, A_1, A_2, A_3, L, ℓ_m, V, G, $G_{\max}$, $O(g)$ (index)

Proof. For a C^1 profile the estimate is the Abel-summation argument of Lemma 6.43, run without collapsing the two shapes of residual: writing $H(t) := \sum_{d<t} \mu(d)^2 g_*(d)$ and $H(t) = \mathfrak{S}(\gamma)\log t + r(t)$, Abel summation bounds the left-hand side by $G_{\max}\big(|r(z)| + (\log z)^{-1}\int_1^z |r(t)|\,dt/t\big)$, and Sublemma 6.41 gives $|r(t)| \leq C_1\mathfrak{S}(\gamma)(1+\ell_V) + C_2\tau(V)\,t^{-1/8}\log(2Vt)$. Integrating the first shape against dt/t contributes $\mathfrak{S}(\gamma)(1+\ell_V)$; integrating the second contributes $\tau(V)\log(2V)/\log z$ up to absolute constants, since $\int_1^\infty t^{-9/8}\log(2Vt)\,dt \ll \log(2V) + 1$; and the boundary term $|r(z)|$ retains the shape $\tau(V)z^{-1/8}\log(2Vz)$. This is the displayed bound with $G_{\max}$ in place of $2M$.

For a merely bounded Lipschitz G, approximate: for each $\eta > 0$ there is a C^1 function G_η on $[0,1]$ with $\sup|G - G_\eta| \leq \eta$ and C^1-norm $G_{\eta,\max} \leq 2M$, obtained by convolving G with a smooth bump of width η/M (the Lipschitz bound controls the derivative of the mollification by M, and the sup-norm by M). Both the discrete sum and the integral change by at most η times the total mass, which is bounded uniformly in η; letting $\eta \to 0$ transfers the C^1 estimate to G with the constant $2M$ in place of $G_{\eta,\max}$. □

Uses (proof): 6.43, 6.41

Lemma 9.11 (k-fold partial summation against a smooth profile)**.** *There are positive functions $C = C(A_1, A_3)$, $C_1 = C_1(A_1, A_3)$ and $C_2 = C_2(A_1, A_3)$ such that the following holds for every sieve datum $(\gamma, V, A_1, A_2, A_3, L)$, every $n \geq 1$, every real $z \geq 2$ and every smooth $F : \mathbb{R}^n \to \mathbb{R}$ supported in $\mathcal{R}_n$:*

$$\Big| \sum_{\mathbf{u} \in \mathbb{N}^n} \Big(\prod_{i=1}^{n} \mu(u_i)^2 g_*(u_i) \Big) F\Big(\frac{\log u_1}{\log z}, \dots, \frac{\log u_n}{\log z} \Big)^2 - \big(\mathfrak{S}(\gamma) \log z \big)^n \int_{\mathcal{R}_n} F^2 \Big| \;\leq\; C\, n\, F_{\max}^2\, B^{n-1}\, \mathcal{L}(\gamma, V, z),$$

where $\mathcal{L}(\gamma, V, z)$ and B are the real numbers satisfying

$$\mathcal{L}(\gamma, V, z) = 2C_1 \mathfrak{S}(\gamma)(1 + \ell_V) + C_2 \tau(V) \left(z^{-1/8} \log(2Vz) + \frac{8 \log(2V) + 64}{\log z} \right)$$

and $B = 2\mathfrak{S}(\gamma) \log z + \mathcal{L}(\gamma, V, z)$.

Uses (statement): 9.10, 4.3, 4.4, 7.4, 5.2

Uses(notation): k, μ, g_*, $\mathfrak{S}(\gamma)$, γ, τ, A_1, A_3, L, V, ℓ_m, F, $F_{\max}$, $\mathcal{R}_k$, $O(g)$, $f \ll g$, G, $G_{\max}$ (index)

Proof. For each $0 \leq \ell \leq n$, interpolate between the two sides through the $n+1$ *layered evaluations*

$$E_\ell := \sum_{\mathbf{u} \in \mathbb{N}^\ell} \Big(\prod_{i \leq \ell} \mu(u_i)^2 g_*(u_i) \Big) \big(\mathfrak{S}(\gamma) \log z \big)^{n-\ell} \times \int_{[0,1]^{n-\ell}} \mathbf{1}\Big[\textstyle\sum_{i \leq \ell} \frac{\log u_i}{\log z} + \sum_{j > \ell} t_j \leq 1 \Big] F\Big(\frac{\log u_1}{\log z}, \dots, \frac{\log u_\ell}{\log z}, t_{\ell+1}, \dots, t_n \Big)^2 \, d\mathbf{t}$$

in which the first ℓ coordinates are summed discretely against the sieve weight and the remaining $n - \ell$ are integrated, each integrated coordinate paying the factor $\mathfrak{S}(\gamma) \log z$ that partial summation would have produced. Since F is supported in $\mathcal{R}_n$ the indicator may be dropped in E_0, so $E_0 = (\mathfrak{S}(\gamma) \log z)^n \int_{\mathcal{R}_n} F^2$; and E_n is the discrete sum on the left, the simplex indicator again being automatic from the support of F.

One layer. Fix ℓ and a prefix $\mathbf{u} \in \mathbb{N}^\ell$, and let $G_{\mathbf{u}}(s)$ denote the inner integral of F^2 with the first ℓ coordinates frozen at $\log u_i / \log z$, the $(\ell+1)$-st coordinate frozen at s, and the remaining ones integrated over $[0,1]$ subject to the simplex constraint. Then $|G_{\mathbf{u}}| \leq F_{\max}^2$ pointwise, because the integrand is bounded by $F_{\max}^2$ on a set of volume at most 1; and $G_{\mathbf{u}}$ is $2F_{\max}^2$-Lipschitz in s, because moving s moves both the integrand (with derivative bounded by $2F_{\max}^2$, as $\partial_s F^2 = 2F \partial_s F$) and the simplex constraint (whose symmetric difference has measure at most $|s - s'|$, on which the integrand is bounded by $F_{\max}^2$). Applying Sublemma 9.10 with $M = 2F_{\max}^2$ replaces the discrete $(\ell+1)$-st sum by $\mathfrak{S}(\gamma)(\log z) \int_0^1 G_{\mathbf{u}}$, at a cost $\leq 4F_{\max}^2 \mathcal{L}(\gamma, V, z)$ per prefix. Consequently

$$|E_{\ell+1} - E_\ell| \;\leq\; \big(\mathfrak{S}(\gamma) \log z \big)^{n-\ell-1} \Big(\sum_{\mathbf{u} \in \mathbb{N}^\ell} \prod_{i \leq \ell} \mu(u_i)^2 g_*(u_i) \Big) \cdot 4F_{\max}^2\, \mathcal{L}(\gamma, V, z),$$

the prefix sum being restricted, by the support of F, to $u_i < z$.

The prefix mass. The prefix sum factorises over the ℓ coordinates, and each single-coordinate sum $\sum_{0 < d < z} \mu(d)^2 g_*(d)$ is, by Sublemma 9.10 with $G \equiv 1$, at most $\mathfrak{S}(\gamma) \log z + 2\mathcal{L}(\gamma, V, z) \leq B$. Hence the prefix mass is at most B^ℓ, and since also $\mathfrak{S}(\gamma) \log z \leq B$ the whole bracket is at most B^{n-1}.

Telescoping. Summing $|E_{\ell+1} - E_\ell|$ over $0 \leq \ell < n$ gives $|E_n - E_0| \leq 4nF_{\max}^2 B^{n-1} \mathcal{L}(\gamma, V, z)$, which is the claim. □

Uses (proof): 9.10

Lemma 9.12 (The r_i-coordinate sum)**.** *Let $\vartheta \in (0,1)$ and $\delta \in (0, \vartheta/2)$. There are $C > 0$ and N_0 such that for every $N \geq N_0$ with $D_0 \geq 2$ and every $G \in C^1([0,1])$,*

$$\Big| \sum_{\substack{r<R \\ (r,W)=1}} \frac{\mu(r)^2}{g(r)}\, G\Big(\frac{\log r}{\log R}\Big) \;-\; \frac{\varphi(W)}{W}\,(\log R)\int_0^1 G \Big| \;\leq\; C\,\frac{\varphi(W)}{W}\Big(\frac{\log R}{D_0}\Big|\int_0^1 G\Big| + (\log D_0)\,G_{\max}\Big).$$

Lean:

- `PrimeGaps.lem_S2m_g_coord`

Uses (statement): 9.10, 4.10, 6.31, 6.34, 4.9, 3.6, 4.4, 3.37, 3.29, 3.32

Uses(notation): N, R, W, D_0, μ, φ, g, g_*, γ_g, (a,b), $\mathfrak{S}(\gamma)$, A_1, A_2, A_3, L, V, G, $G_{\max}$, $O(g)$, $f \ll g$, $o(g)$, τ, ℓ_m (index)

Proof. Apply the sharp partial-summation estimate of Sublemma 9.10 at $z = R$ to the Maynard sieve datum built on the W-tricked density γ_g of Definition 4.9, namely the totally multiplicative function with $\gamma_g(p) = 0$ for $p \mid W$ and $\gamma_g(p) = p/(p-1)$ for $p \nmid W$. That this is a sieve datum, with $V = W$, $A_1 = 1/2$, $A_3 = 2$ and $L \ll \log D_0$, is Lemma 4.10; and by Sublemma 6.31 its derived multiplicative function is $g_*(r) = \mu(r)^2/g(r)$ for r coprime to W and 0 otherwise, so the sum in the statement is exactly the sum that estimate governs.

By Sublemma 6.34, $\mathfrak{S}(\gamma_g) = \frac{\varphi(W)}{W}\big(1 + O(1/D_0)\big)$. Substituting, the main term $\mathfrak{S}(\gamma_g)(\log R)\int_0^1 G$ becomes $\frac{\varphi(W)}{W}(\log R)\int_0^1 G$ together with an error $O\big(\frac{\varphi(W)}{W}\frac{\log R}{D_0}\big|\int_0^1 G\big|\big)$ — the first of the two asserted error shapes, and the reason the statement carries $\int_0^1 G$ rather than $G_{\max}$ in that term.

For the residual $\mathcal{L}(\gamma_g, W, R)$ of Sublemma 9.10: the term $\mathfrak{S}(\gamma_g)(1+\ell_W)$ is $O\big(\frac{\varphi(W)}{W}\log D_0\big)$, since $\ell_W = \sum_{p \leq D_0} \frac{\log p}{p-1} = \log D_0 + O(1)$; and the term $\tau(W)\big(R^{-1/8}\log(2WR) + (\log(2W)+1)/\log R\big)$ is negligible in comparison, because $W = N^{o(1)}$ and $\tau(W) = N^{o(1)}$ by Lemma 3.32 while $\log R$ is a fixed positive multiple of $\log N$. Multiplying by $G_{\max}$ gives the second asserted error shape $O\big(\frac{\varphi(W)}{W}(\log D_0)G_{\max}\big)$. □

Uses (proof): 9.10, 4.10, 6.31, 6.34, 3.32

9.4 The main term and the substitution $\lambda \rightsquigarrow y^{(m)}$

Definition 9.13 (The $y^{(m)}$-bilinear sum)**.** For a finitely supported λ and an index m, set

$$\mathcal{Y}_m(\lambda) \;:=\; \sum_{\mathbf{u}} \Big(\prod_{i=1}^{k} \frac{\mu(u_i)^2}{g(u_i)}\Big) \sum_{s_{1,2},\dots,s_{k,k-1}}^{*} \Big(\prod_{i \neq j} \frac{\mu(s_{i,j})}{g(s_{i,j})^2}\Big)\, y^{(m)}(\mathbf{a})\; y^{(m)}(\mathbf{b}),$$

where $\mathbf{a}$ and $\mathbf{b}$ are the exponent tuples $a_j = u_j \prod_{i \neq j} s_{j,i}$ and $b_j = u_j \prod_{i \neq j} s_{i,j}$ of Definitions 8.4 and 8.5, and $\sum^*$ is the restricted sum of Notation 8.3.

Uses (statement): 4.25, 3.6, 3.3, 8.4, 8.5, 8.3

Uses(notation): k, μ, g, $y^{(m)}(r_1,\dots,r_k)$, $\mathbf{u}$, $\mathbf{a}$, $\mathbf{b}$, u_i, $s_{i,j}$, $\sum^*$, $\lambda(d_1,\dots,d_k)$ (index)

Lemma 9.14 (Main term of $S_2^{(m)}(\lambda)$ after the Chinese remainder step)**.** *Let $k \geq 2$, let $\mathcal{H}$ have distinct shifts, and let $\vartheta < 1/2$ be a level of distribution for the primes, with $0 < \delta < \vartheta/2$. For every $A > 0$ there are $C, N_0 > 0$ such that for every $N \geq N_0$, every compatible v_0 and every λ with permissible support at (R, W),*

$$\Big| S_2^{(m)}(\lambda) \;-\; \frac{X_N}{\varphi(W)} \sum_{\substack{\mathbf{d},\mathbf{e} \\ d_m = e_m = 1}}{}' \frac{\lambda(\mathbf{d})\,\lambda(\mathbf{e})}{\prod_{i=1}^{k} \varphi([d_i, e_i])} \Big| \;\leq\; C\, \frac{y_{\max}^2\, N}{(\log N)^A},$$

where $\sum'$ restricts to pairs with $W, [d_1, e_1], \ldots, [d_k, e_k]$ pairwise coprime and, in addition, $(h_m - h_i, [d_i, e_i]) = 1$ for every $i \neq m$.

Lean:

- `PrimeGaps.S2m_CRT_main_term_BV_error`

Uses (statement): 9.1, 9.2, 9.3, 9.9, 4.16, 3.39, 3.24, 3.4, 3.29, 1

Uses(notation): k, h_i, N, v_0, W, φ, (a,b), $[a,b]$, $\lambda(d_1,\ldots,d_k)$, $\mathbf{d}$, $\mathbf{e}$, $S_2^{(m)}(\lambda)$, $y_{\max}$, X_N, $q(\mathbf{d},\mathbf{e})$, $\sum'$, $O(g)$, ϑ (index)

Proof. By Lemma 9.1, $S_2^{(m)}(\lambda) = \sum_{\mathbf{d},\mathbf{e}} \lambda(\mathbf{d})\lambda(\mathbf{e})\, T_m(\mathbf{d},\mathbf{e})$ with $T_m(\mathbf{d},\mathbf{e})$ the inner prime-weighted count.

Step 1: discard the pairs that cannot contribute. By Lemma 9.2, $T_m(\mathbf{d},\mathbf{e}) = 0$ for N large unless $d_m = e_m = 1$. If two of $W, [d_1,e_1], \ldots, [d_k,e_k]$ share a prime p, then the congruence system defining the inner sum is either inconsistent (so $T_m = 0$) or forces $p \mid (h_i - h_j)$ for some $i \neq j$; on the support of λ every prime factor of $[d_i, e_i]$ exceeds D_0, and $D_0 \to \infty$, so for N large this cannot happen for distinct shifts. The same argument shows the extra condition $(h_m - h_i, [d_i, e_i]) = 1$ is automatically satisfied for N large: any common prime would exceed $D_0 > \max_{i \neq j} |h_i - h_j|$ and divide the nonzero integer $h_m - h_i$. Thus, for N large, the outer sum may be restricted to $\sum'$ with $d_m = e_m = 1$, without changing its value.

Step 2: evaluate each retained inner sum. For a retained pair, all three hypotheses of Lemma 9.3 hold, so $T_m(\mathbf{d},\mathbf{e}) = X_N/\varphi(q(\mathbf{d},\mathbf{e})) + O(E(N, q(\mathbf{d},\mathbf{e})))$. Since W and the $[d_i, e_i]$ are pairwise coprime, multiplicativity of φ gives $\varphi(q(\mathbf{d},\mathbf{e})) = \varphi(W) \prod_i \varphi([d_i, e_i])$.

Step 3: split main and error. Substituting Step 2 into Step 1 produces exactly the displayed main term, plus an error at most a constant multiple of $\sum_{\mathbf{d},\mathbf{e}} |\lambda(\mathbf{d})\lambda(\mathbf{e})|\, E(N, q(\mathbf{d},\mathbf{e}))$. Lemma 9.9 bounds this by $C\, y_{\max}^2 N/(\log N)^A$. □

Uses (proof): 9.1, 9.2, 9.3, 9.9

Lemma 9.15 (φ-lcm decoupling)**.** *Let λ be supported on tuples whose coordinates are positive, squarefree and pairwise coprime. Then*

$$\sum_{\substack{\mathbf{d},\mathbf{e} \\ d_m = e_m = 1}}{}' \frac{\lambda(\mathbf{d})\,\lambda(\mathbf{e})}{\prod_{i=1}^{k} \varphi([d_i, e_i])} \;=\; \sum_{\mathbf{u}} \Big(\prod_{i=1}^{k} g(u_i) \Big) \sum_{\substack{\mathbf{d},\mathbf{e} \\ u_i \mid (d_i, e_i)\ \forall i \\ d_m = e_m = 1}}{}' \frac{\lambda(\mathbf{d})\,\lambda(\mathbf{e})}{\prod_{i=1}^{k} \varphi(d_i)\,\varphi(e_i)}.$$

Lean:

- `PrimeGaps.lem_S2m_decouple_phi_lcm`

Uses (statement): 6.2, 3.6, 3.4, 3.39

Uses(notation): k, φ, g, (a,b), $[a,b]$, $\lambda(d_1,\dots,d_k)$, $\mathbf{d}$, $\mathbf{e}$, $\mathbf{u}$, u_i, $\sum'$ (index)

Proof. For squarefree d, e, Lemma 6.2 gives the divisor-sum identity

$$\frac{1}{\varphi([d,e])} = \frac{1}{\varphi(d)\,\varphi(e)} \sum_{u \mid (d,e)} g(u),$$

which is the multiplicative reformulation of $\varphi([d,e])\,\varphi((d,e)) = \varphi(d)\varphi(e)$ together with $\varphi = g * 1$ on squarefree arguments. Applying it in each of the k coordinates — legitimate, since the coordinates of tuples in the support of λ are squarefree — and expanding the resulting product of k finite divisor sums into a single sum over the tuple $\mathbf{u} = (u_1,\dots,u_k)$ with $u_i \mid (d_i, e_i)$ gives

$$\prod_{i=1}^{k} \frac{1}{\varphi([d_i,e_i])} = \frac{1}{\prod_{i=1}^{k} \varphi(d_i)\varphi(e_i)} \sum_{\substack{\mathbf{u} \\ u_i \mid (d_i,e_i)\ \forall i}} \prod_{i=1}^{k} g(u_i).$$

Multiplying by $\lambda(\mathbf{d})\lambda(\mathbf{e})$, summing over the restricted set of pairs and interchanging the two finite summations yields the claim; the restriction $\sum'$ and the pinning $d_m = e_m = 1$ concern only the outer variables and are untouched. □

Uses (proof): 6.2, 3.10

Lemma 9.16 (Möbius inversion of the cross-coprimality conditions)**.** *Let λ be supported on tuples whose coordinates are positive, squarefree and pairwise coprime. Then*

$$\sum_{\substack{\mathbf{d},\mathbf{e} \\ u_i \mid (d_i,e_i)\ \forall i \\ d_m = e_m = 1}}{}' \frac{\lambda(\mathbf{d})\,\lambda(\mathbf{e})}{\prod_{i=1}^{k} \varphi(d_i)\,\varphi(e_i)} = \sum_{s_{1,2},\dots,s_{k,k-1}} \Big(\prod_{i \neq j} \mu(s_{i,j})\Big) \sum_{\substack{\mathbf{d},\mathbf{e} \\ u_i \mid (d_i,e_i)\ \forall i \\ s_{i,j} \mid (d_i,e_j)\ \forall i \neq j \\ d_m = e_m = 1}} \frac{\lambda(\mathbf{d})\,\lambda(\mathbf{e})}{\prod_{i=1}^{k} \varphi(d_i)\,\varphi(e_i)}.$$

Lean:

- `PrimeGaps.lem_S2m_mobius`

Uses (statement): 9.15, 3.7, 3.3, 3.39

Uses(notation): k, μ, φ, (a,b), $\lambda(d_1,\dots,d_k)$, $\mathbf{d}$, $\mathbf{e}$, $\mathbf{u}$, u_i, $s_{i,j}$, $\sum'$ (index)

Proof. By Lemma 8.11, which uses only the permissible-support hypothesis, the pairwise-coprimality restriction recorded by $\sum'$ is, on the support of λ, equivalent to the family of conditions $(d_i, e_j) = 1$ over ordered pairs $i \neq j$: the conditions involving W and the same-index conditions hold automatically.

For each such ordered pair, Lemma 3.7 gives the Möbius detection identity $\mathbf{1}[(d_i,e_j)=1] = \sum_{s_{i,j} \mid (d_i,e_j)} \mu(s_{i,j})$, and $s_{i,j} \mid (d_i,e_j)$ is the conjunction of $s_{i,j} \mid d_i$ and $s_{i,j} \mid e_j$. Taking the product of these $k(k-1)$ identities and interchanging the resulting finite $(s_{i,j})$-summation with the $(\mathbf{d},\mathbf{e})$-summation gives the displayed formula.

Moreover, exactly as in Lemma 8.14, every tuple $(s_{i,j})$ that contributes a nonzero term satisfies the coprimality restrictions recorded by $\sum^{*}$: $(s_{i,j}, u_i) = (s_{i,j}, u_j) = 1$, $(s_{i,j}, s_{i,a}) = 1$ for $a \neq j$ and $(s_{i,j}, s_{b,j}) = 1$ for $b \neq i$. Each is a two-line contradiction with $(d_i, d_j) = 1$ or $(e_i, e_j) = 1$, both of which follow from squarefreeness of $\prod_i d_i$ and of $\prod_i e_i$. By Lemma 8.15 the restricted and unrestricted s-sums therefore agree. □

Uses (proof): 8.11, 8.14, 8.15, 3.7

Lemma 9.17 (Substitution $\lambda(\mathbf{d}) \rightsquigarrow y^{(m)}(\mathbf{r})$)**.** *Let $k \geq 2$, let $\mathcal{H}$ have distinct shifts, and let $\vartheta < 1/2$ be a level of distribution for the primes, with $0 < \delta < \vartheta/2$. For every $A > 0$ there are $C, N_0 > 0$ such that for every $N \geq N_0$, every compatible v_0 and every λ with permissible support at (R, W),*

$$\left| S_2^{(m)}(\lambda) \; - \; \frac{X_N}{\varphi(W)}\, \mathcal{Y}_m(\lambda) \right| \; \leq \; C\, \frac{y_{\max}^2\, N}{(\log N)^A}.$$

Lean:

- `PrimeGaps.lem_S2m_substitute_ym`

Uses (statement): 9.14, 9.15, 9.16, 9.13, 4.25, 3.6, 3.24, 3.39, 1

Uses(notation): k, N, R, W, μ, φ, g, $\lambda(d_1, \dots, d_k)$, $\mathbf{d}$, $\mathbf{e}$, $\mathbf{u}$, $\mathbf{a}$, $\mathbf{b}$, u_i, $s_{i,j}$, $\sum'$, $\sum^*$, $y^{(m)}(r_1, \dots, r_k)$, $y_{\max}$, X_N, $S_2^{(m)}(\lambda)$, ϑ (index)

Proof. By Lemma 9.14 it suffices to identify the restricted double sum, after the transformations of Lemmas 9.15 and 9.16, with $\mathcal{Y}_m(\lambda)$; the error is carried unchanged. The identification is the $S_2^{(m)}$-analogue of the substitution performed for S_1 (Lemma 8.16), the difference being that $\lambda(\mathbf{d})$ is here normalised by $\prod_i \varphi(d_i)$ rather than $\prod_i d_i$, so that g replaces φ throughout.

Step 1: invert the definition of $y^{(m)}$. Definition 4.25 reads $y^{(m)}(\mathbf{r}) = \big(\prod_{i \neq m} \mu(r_i) g(r_i)\big) \sum_{\mathbf{d}:\ r_i \mid d_i,\ d_m = 1} \lambda(\mathbf{d}) / \prod_{i \neq m} \varphi(d_i)$. Möbius inversion, coordinate by coordinate, gives on the support of λ

$$\frac{\lambda(\mathbf{d})}{\prod_i \varphi(d_i)} \; = \; \Big(\prod_i \mu(d_i)\Big) \sum_{\substack{\mathbf{r} \\ d_i \mid r_i \ \forall i \\ r_m = 1}} \frac{y^{(m)}(\mathbf{r})}{\prod_i g(r_i)} \qquad (d_m = 1),$$

the support constraint $r_m = 1$ coming from Lemma 4.26.

Step 2: collapse the $\mathbf{d}$- and $\mathbf{e}$-sums. On the inner sum of Lemma 9.16 the constraints on $\mathbf{d}$ are $u_i \mid d_i$ and $s_{i,j} \mid d_i$ for $j \neq i$; by the coprimality of u_i with the $s_{i,j}$ recorded in Lemma 9.16, these are equivalent to the single divisibility $a_i \mid d_i$ with $a_i = u_i \prod_{j \neq i} s_{i,j}$ (Definition 8.4). Inserting Step 1, swapping the finite $\mathbf{d}$- and $\mathbf{r}$-sums and applying Lemma 3.7 in the form $\sum_{a_i \mid d_i \mid r_i} \mu(d_i) = \mu(a_i) \mathbf{1}[r_i = a_i]$ collapses the $\mathbf{r}$-sum to the single tuple $\mathbf{r} = \mathbf{a}$, leaving $\big(\prod_i \mu(a_i)\big) y^{(m)}(\mathbf{a}) / \prod_i g(a_i)$. The identical computation on the $\mathbf{e}$-side — where the constraints are $u_i \mid e_i$ and $s_{j,i} \mid e_i$, with the two indices of s transposed — leaves $\big(\prod_i \mu(b_i)\big) y^{(m)}(\mathbf{b}) / \prod_i g(b_i)$ with $b_i = u_i \prod_{j \neq i} s_{j,i}$ (Definition 8.5). The pinning $d_m = e_m = 1$ forces $u_m = 1$ and $s_{m,j} = s_{j,m} = 1$, hence $a_m = b_m = 1$.

Step 3: assemble the coefficient. The tuples $\mathbf{a}, \mathbf{b}$ have squarefree entries, and μ and g are multiplicative on them, so $\mu(a_i) = \mu(u_i) \prod_{j \neq i} \mu(s_{i,j})$ and $g(a_i) = g(u_i) \prod_{j \neq i} g(s_{i,j})$, and similarly for b_i. Each $\mu(s_{i,j})$ occurs once through $\mu(a_i)$, once through $\mu(b_j)$, and once through the inversion factor $\prod_{i \neq j} \mu(s_{i,j})$ already present; since $\mu^3 = \mu$, the net exponent is one. The g-factors combine as $\big(\prod_i g(u_i)\big) \prod_i \big(g(a_i) g(b_i)\big)^{-1} = \prod_i g(u_i)^{-1} \prod_{i \neq j} g(s_{i,j})^{-2}$. The resulting coefficient is $\big(\prod_i \mu(u_i)^2 / g(u_i)\big) \prod_{i \neq j} \mu(s_{i,j}) / g(s_{i,j})^2$, which is exactly the summand of Definition 9.13. □

Uses (proof): 9.14, 9.15, 9.16, 8.16, 3.7, 4.26, 8.14, 8.4, 8.5

Lemma 9.18 (Contribution of the terms with some $s_{i,j} > D_0$)**.** *Let $\vartheta \in (0,1)$ and $\delta \in (0, \vartheta/2)$. There are $C > 0$ and N_0 such that for every $N \geq N_0$, every λ with permissible support at (R,W) and every pair of indices $i \neq j$, the absolute contribution to $\frac{X_N}{\varphi(W)}\mathcal{Y}_m(\lambda)$ of the terms with $s_{i,j} > D_0$ is at most*

$$C\,\frac{(y^{(m)}_{\max})^2\,\varphi(W)^{k-2}\,N\,(\log R)^{k-1}}{W^{k-1}\,D_0\,\log N}.$$

Lean:

- `PrimeGaps.lem_S2m_sij_D0`

Uses (statement): 9.13, 4.27, 3.6, 3.28, 3.29, 3.37, 3.24

Uses(notation): k, N, R, W, D_0, μ, φ, g, (a,b), τ, $f \ll g$, $O(g)$, $y^{(m)}(r_1,\dots,r_k)$, $y^{(m)}_{\max}$, X_N, u_i, $s_{i,j}$, ϑ, δ (index)

Proof. The argument is the g-analogue of Lemma 8.19, with three substitutions: the prefactor N/W becomes $X_N/\varphi(W)$, the multiplicative weight φ becomes g, and the convergent series of Lemma 6.11 becomes that of Lemma 6.12.

Fix the offending pair (i,j). Bound $|y^{(m)}(\mathbf{a})\,y^{(m)}(\mathbf{b})|$ by $(y^{(m)}_{\max})^2$, which is finite by Lemma 7.1, and factor the remaining sum into independent coordinate sums. There are three kinds of factor.

(a) The $k-1$ free u-coordinates. Since $u_m = 1$ on the support, the u-sum contributes $\big(\sum_{u \leq R,\ (u,W)=1} \mu(u)^2/g(u)\big)^{k-1}$. The g-analogue of the Mertens estimate of Lemma 6.30 — available here as the $G \equiv 1$ case of Lemma 9.12, or directly by comparing $1/g(p)$ with $1/\varphi(p)$, whose difference $1/((p-1)(p-2)) \ll p^{-2}$ sums to $O(1/D_0)$ over $p > D_0$ — bounds this by $\ll_k \varphi(W)^{k-1}(\log R)^{k-1}/W^{k-1}$.

(b) The distinguished big coordinate. By Lemmas 8.17 and 8.18, every contributing $s_{i,j} \neq 1$ is coprime to W and has all its prime factors $> D_0$. For squarefree s with least prime factor $> D_0 \geq 9$ one has $g(p) = p-2 \geq p/3$, hence $1/g(s)^2 \leq \prod_{p \mid s} 9/p^2$, and summing the resulting multinomial series gives

$$\sum_{\substack{s>1,\ \mu(s)^2=1 \\ p \mid s \Rightarrow p > D_0}} \frac{1}{g(s)^2} \;\leq\; \exp(9/D_0) - 1 \;\ll\; \frac{1}{D_0}.$$

(c) The remaining k^2-k-1 coordinates $s_{a,b}$. Each contributes $\sum_{s \geq 1} \mu(s)^2/g(s)^2 \ll 1$, convergent by Lemma 6.12.

Multiplying (a), (b), (c) by the prefactor and using $X_N \ll N/\log N$ (a consequence of Theorem 3, or of Chebyshev's bound) gives the stated estimate. □

Uses (proof): 8.19, 8.17, 8.18, 6.12, 6.11, 6.30, 9.12, 7.1, 3.14, 3

Lemma 9.19 (Contribution of the diagonal $s_{i,j} = 1$)**.** *For every N, W and every finitely supported λ, the contribution to $\frac{X_N}{\varphi(W)}\,\mathcal{Y}_m(\lambda)$ from the single tuple with $s_{i,j} = 1$ for all $i \neq j$ is*

$$\frac{X_N}{\varphi(W)} \sum_{\mathbf{u}} \frac{\big(y^{(m)}(\mathbf{u})\big)^2}{\prod_{i=1}^{k} g(u_i)}.$$

Lean:

- `PrimeGaps.lem_S2m_sij_one`

Uses (statement): 9.13, 4.25, 3.6, 4.26, 3.24

Uses(notation): k, W, μ, φ, g, $\mathbf{u}$, u_i, $s_{i,j}$, $y^{(m)}(r_1,\dots,r_k)$, X_N, $S_2^{(m)}(\lambda)$ (index)

Proof. Setting $s_{i,j}=1$ for all $i\neq j$ in $a_j=u_j\prod_{i\neq j}s_{j,i}$ and $b_j=u_j\prod_{i\neq j}s_{i,j}$ gives $\mathbf{a}=\mathbf{b}=\mathbf{u}$, so the product $y^{(m)}(\mathbf{a})\,y^{(m)}(\mathbf{b})$ becomes $(y^{(m)}(\mathbf{u}))^2$. The s-coefficient is $\prod_{i\neq j}\mu(1)/g(1)^2=1$ since $g(1)=1$. Finally $y^{(m)}(\mathbf{u})$ vanishes unless every u_i is squarefree (Lemma 4.26), so on the support $\mu(u_i)^2=1$ and the u-coefficient $\prod_i\mu(u_i)^2/g(u_i)$ reduces to $1/\prod_i g(u_i)$. □

Uses (proof): 8.20

Lemma 9.20 (Absorption of the X_N error)**.** *Let $k\geq 2$, $\vartheta\in(0,1)$ and $\delta\in(0,\vartheta/2)$. There are $C>0$ and N_0 such that for every $N\geq N_0$ and every λ with permissible support at (R,W),*

$$\Big|\Big(\frac{X_N}{\varphi(W)}-\frac{N}{\varphi(W)\log N}\Big)\sum_{\mathbf{u}}\frac{(y^{(m)}(\mathbf{u}))^2}{\prod_{i=1}^k g(u_i)}\Big|\;\leq\;C\,\frac{(y^{(m)}_{\max})^2\,\varphi(W)^{k-2}\,N\,(\log N)^{k-2}}{W^{k-1}\,D_0}.$$

Lean:

- `PrimeGaps.lem_S2m_PNT`

Uses (statement): 3, 3.24, 4.25, 4.27, 3.6, 3.37, 3.28, 3.29

Uses(notation): k, N, R, W, D_0, μ, φ, g, (a,b), $O(g)$, $f\ll g$, $\mathbf{u}$, u_i, $y^{(m)}(r_1,\dots,r_k)$, $y^{(m)}_{\max}$, X_N (index)

Proof. By Theorem 3, $X_N=N/\log N+O(N/(\log N)^2)$, so the prefactor difference is $O\big(N/(\varphi(W)(\log N)^2)\big)$.

It remains to bound the diagonal sum. Bounding $(y^{(m)}(\mathbf{u}))^2$ by $(y^{(m)}_{\max})^2$ and using $u_m=1$ together with the coordinate restrictions $u_i\leq R$, $(u_i,W)=1$ and $\mu(u_i)^2=1$ from Lemma 4.26, the sum factorises across the $k-1$ free coordinates:

$$\sum_{\mathbf{u}}\frac{(y^{(m)}(\mathbf{u}))^2}{\prod_i g(u_i)}\;\leq\;(y^{(m)}_{\max})^2\Big(\sum_{\substack{u\leq R\\(u,W)=1}}\frac{\mu(u)^2}{g(u)}\Big)^{k-1}\;\ll_k\;\frac{(y^{(m)}_{\max})^2\,\varphi(W)^{k-1}\,(\log R)^{k-1}}{W^{k-1}},$$

the last step being the $G\equiv 1$ case of Lemma 9.12. Multiplying the two displays and using $\log R\leq\log N$ gives a bound $\ll(y^{(m)}_{\max})^2\varphi(W)^{k-2}N(\log N)^{k-3}/W^{k-1}$, which is smaller than the asserted bound because $(\log N)^{k-3}\,D_0\leq(\log N)^{k-2}$ for N large: D_0 grows only triple-logarithmically. □

Uses (proof): 3, 9.12, 4.26, 7.1

Lemma 9.21 (The second moment in terms of $y^{(m)}$)**.** *Let $k\geq 2$, let $\mathcal{H}$ have distinct shifts, let $\vartheta<1/2$ be a level of distribution for the primes and $0<\delta<\vartheta/2$. For every $A>0$ there are $C>0$ and N_0 such that for every $N\geq N_0$, every compatible v_0 and every λ with permissible support at (R,W),*

$$\Big|S_2^{(m)}(\lambda)\;-\;\frac{N}{\varphi(W)\,\log N}\sum_{\mathbf{r}}\frac{\big(y^{(m)}(\mathbf{r})\big)^2}{\prod_{i=1}^k g(r_i)}\Big|$$
$$\leq\;C\,\frac{(y^{(m)}_{\max})^2\,\varphi(W)^{k-2}\,N\,(\log N)^{k-2}}{W^{k-1}\,D_0}\;+\;C\,\frac{y_{\max}^2\,N}{(\log N)^A}.$$

Lean:

- `PrimeGaps.lem_S2m_from_ym`

Uses (statement): 9.17, 9.18, 9.19, 9.20, 4.16, 4.25, 4.27, 3.6, 3.28, 3.29, 1, 3

Uses(notation): k, N, R, W, D_0, φ, g, $O(g)$, $f \ll g$, $\mathbf{r}$, $S_2^{(m)}(\lambda)$, $y^{(m)}(r_1,\dots,r_k)$, $y^{(m)}_{\max}$, $y_{\max}$, X_N, $s_{i,j}$, ϑ, δ (index)

Proof. By Lemma 9.17, $S_2^{(m)}(\lambda)$ agrees with $\frac{X_N}{\varphi(W)}\mathcal{Y}_m(\lambda)$ up to $C\,y_{\max}^2 N/(\log N)^A$, which is the second error term of the claim.

Split the s-summation in $\mathcal{Y}_m(\lambda)$ into the diagonal tuple, all $s_{i,j}=1$, and its complement. By Lemma 8.18 — whose proof uses only that λ is supported on tuples $\mathbf{d}$ with $\prod_i d_i$ squarefree and coprime to W, and that $s_{i,j} \mid d_i$, both of which hold here — every contributing $s_{i,j}$ is either 1 or has all prime factors $> D_0$; in particular any off-diagonal tuple has some $s_{i,j} > D_0$. Summing the bound of Lemma 9.18 over the $k(k-1)$ ordered pairs (Lemma 3.14) shows the off-diagonal contribution is

$$\ll_k \frac{(y^{(m)}_{\max})^2\,\varphi(W)^{k-2}\,N\,(\log R)^{k-1}}{W^{k-1}\,D_0\,\log N} \;\le\; \frac{(y^{(m)}_{\max})^2\,\varphi(W)^{k-2}\,N\,(\log N)^{k-2}}{W^{k-1}\,D_0},$$

using $\log R \le \log N$.

The diagonal contribution is evaluated by Lemma 9.19 as $\frac{X_N}{\varphi(W)}\sum_{\mathbf{r}}(y^{(m)}(\mathbf{r}))^2/\prod_i g(r_i)$, and Lemma 9.20 replaces X_N by $N/\log N$ at a cost within the same first error term. Adding the three contributions gives the claim. □

Uses (proof): 9.17, 8.18, 9.18, 9.19, 9.20, 3.14

9.5 The smooth case

Lemma 9.21 reduces $S_2^{(m)}(\lambda)$ to the diagonal quadratic form $\sum_{\mathbf{r}}(y^{(m)}(\mathbf{r}))^2/\prod_{i\neq m} g(r_i)$ — the $i=m$ factor being $g(1)=1$ because $y^{(m)}$ is supported on $r_m=1$. We now evaluate that form for the F-derived weight λ_0 of Definition 5.11.

The evaluation does *not* go through a per-tuple asymptotic for $y^{(m)}(\mathbf{r})$. Up to a negligible error, $y^{(m)}(\mathbf{r})$ equals the inverse sum $Y^{(m)}(\mathbf{r})$ of Definition 7.10, whose defining coprimality condition carries the large modulus $W\prod_{i\neq m} r_i$; a per-tuple evaluation would need partial summation at that modulus, which is not available. Instead we square first and sum second: the large modulus then appears as a *coupling* between the $k+1$ summation coordinates $u, u', (r_i)_{i\neq m}$, and every such coupling can be dropped at a relative cost $O(1/D_0)$, exactly as in the S_1 analysis, because two coordinates coprime to W can only share a prime $> D_0$. Once the couplings are dropped, each coordinate is coprime to W alone, and the $(k+1)$-fold sum is evaluated coordinate by coordinate at modulus W.

Definition 9.22 (The decoupled $(k+1)$-fold sum)**.** Let $F:[0,1]^k\to\mathbb{R}$. For an index m set

$$\mathcal{D}_m(F) \;:=\; \sum_{\substack{u,u',(\rho_i)_{i\neq m}\\ \text{all squarefree and coprime to } W}} \frac{\mu(u)^2\,\mu(u')^2}{\varphi(u)\,\varphi(u')}\Big(\prod_{i\neq m}\frac{\mu(\rho_i)^2}{g(\rho_i)}\Big)\,\Phi\Big(\big(\tfrac{\log\rho_i}{\log R}\big)_{i\neq m};\ \tfrac{\log u}{\log R},\ \tfrac{\log u'}{\log R}\Big),$$

where $\Phi\big((t_i)_{i\neq m};s,s'\big) := F(\dots,s,\dots)\,F(\dots,s',\dots)$ places s, respectively s', in the m-th slot of two copies of F and t_i in the others. No coprimality is imposed between the summation variables.

Uses (statement): 3.29, 3.37, 3.3, 3.4, 3.6

Uses(notation): k, R, W, μ, φ, g, (a,b), F, Φ (index)

Lemma 9.23 (Reduction of the diagonal form to a second moment of Y). *Let $k \geq 2$ and let $F : [0,1]^k \to \mathbb{R}$ be smooth and supported in $\mathcal{R}_k$. There is a constant $C \geq 0$, depending only on k, and a threshold N_0 depending on F, ϑ, δ, such that for every $N \geq N_0$, every λ with permissible support at (R,W) and $y_{\max} \leq F_{\max}$, and every m,*

$$\Big|\sum_{\mathbf{r}} \frac{(y^{(m)}(\mathbf{r}))^2}{\prod_{i\neq m} g(r_i)} - \sum_{\mathbf{r}} \frac{Y^{(m)}(\mathbf{r})^2}{\prod_{i\neq m} g(r_i)}\Big| \leq C\,\frac{F_{\max}^2\,\varphi(W)^{k+1}\,(\log R)^{k+1}}{W^{k+1}\,D_0},$$

both sums being over tuples $\mathbf{r}$ with $r_m = 1$ and $r_i \geq 1$ coprime to W for $i \neq m$.

Lean:

- `PrimeGaps.lem_S2m_second_moment`

Uses (statement): 7.10, 7.16, 9.12, 4.25, 3.6, 7.4, 4.27, 3.39, 3.37, 3.29, 5.2

Uses(notation): k, R, W, D_0, μ, φ, g, (a,b), $f \ll g$, $\mathbf{r}$, $y^{(m)}(r_1,\dots,r_k)$, $y_{\max}$, F, $F_{\max}$, $\mathcal{R}_k$, $Y_{\mathbf{r}}$ (index)

Proof. By Lemma 7.16, $y^{(m)}(\mathbf{r}) = Y^{(m)}(\mathbf{r}) + \delta_{\mathbf{r}}$ with a uniform bound $|\delta_{\mathbf{r}}| \leq \eta := c\,F_{\max}\,\varphi(W)(\log R)/(WD_0)$: the passage from $y^{(m)}$ to the inverse sum costs only the terms in which the auxiliary tuple differs from $\mathbf{r}$ in some coordinate $j \neq m$, and those are controlled by a $1/D_0$ saving of the same shape as in the S_1 analysis.

Next, $|Y^{(m)}(\mathbf{r})| \leq F_{\max}\sum_{(a,W)=1}\mu(a)^2/\varphi(a) \ll F_{\max}\,\varphi(W)(\log R)/W$ by the Mertens estimate of Lemma 6.30, and hence also $|y^{(m)}(\mathbf{r})| \ll F_{\max}\varphi(W)(\log R)/W$. Therefore

$$(y^{(m)}(\mathbf{r}))^2 - Y^{(m)}(\mathbf{r})^2 = 2Y^{(m)}(\mathbf{r})\delta_{\mathbf{r}} + \delta_{\mathbf{r}}^2 = O\big(\eta\,F_{\max}\varphi(W)(\log R)/W\big).$$

Summing this against $1/\prod_{i\neq m} g(r_i)$ and using

$$\sum_{\mathbf{r}}\prod_{i\neq m}\frac{\mu(r_i)^2}{g(r_i)} \leq \prod_{i\neq m}\sum_{\substack{r<R\\(r,W)=1}}\frac{\mu(r)^2}{g(r)} \ll_k \Big(\frac{\varphi(W)\log R}{W}\Big)^{k-1}$$

— the case $G \equiv 1$ of Lemma 9.12 — gives the total $\ll \eta\,F_{\max}\,\varphi(W)^k(\log R)^k/W^k = F_{\max}^2\varphi(W)^{k+1}(\log R)^{k+1}/(W^{k+1}D_0)$, as claimed. □

Uses (proof): 7.16, 6.30, 9.12

Lemma 9.24 (Expanding the square and dropping the couplings). *Let $k \geq 2$ and let $F : [0,1]^k \to \mathbb{R}$ be smooth and supported in $\mathcal{R}_k$, and let λ_0 be the associated F-derived weight. There is a constant $C \geq 0$ depending only on k, and a threshold N_0 depending on F, ϑ, δ, such that for every $N \geq N_0$ and every m,*

$$\Big|\sum_{\mathbf{r}} \frac{Y^{(m)}(\mathbf{r})^2}{\prod_{i\neq m} g(r_i)} - \mathcal{D}_m(F)\Big| \leq C\,\frac{F_{\max}^2\,\varphi(W)^{k+1}\,(\log R)^{k+1}}{W^{k+1}\,D_0}.$$

Lean:

- `PrimeGaps.lem_S2m_expand_drop`

Uses (statement): 7.10, 9.22, 5.11, 7.3, 9.12, 3.6, 7.4, 3.37, 3.29, 5.2

Uses(notation): k, R, W, D_0, μ, φ, g, (a,b), $f \ll g$, $O(g)$, $\mathbf{r}$, F, $F_{\max}$, $\mathcal{R}_k$, $\lambda_0(d_1,\dots,d_k)$, $Y_{\mathbf{r}}$, Φ (index)

Proof. For the F-derived weight, Lemma 7.3 evaluates the y-variables in closed form, and $Y^{(m)}(\mathbf{r})$ becomes

$$Y^{(m)}(\mathbf{r}) \;=\; \sum_{\substack{u\ge 1\\ (u,\ W\prod_{i\neq m} r_i)=1}} \frac{\mu(u)^2}{\varphi(u)}\, F\Big(\dots,\tfrac{\log u}{\log R},\dots\Big).$$

Step 1: square. The condition $(u, W\prod_{i\neq m} r_i) = 1$ is the conjunction of $(u,W)=1$ and $(u,r_i)=1$ for $i\neq m$. Since each r_i is squarefree, $\mathbf{1}[(u,r_i)=1]\,\mathbf{1}[(u',r_i)=1] = \mathbf{1}[(uu',r_i)=1]$, so

$$Y^{(m)}(\mathbf{r})^2 = \sum_{\substack{u,u'\\ (uu',W)=1}} \frac{\mu(u)^2\mu(u')^2}{\varphi(u)\varphi(u')}\Big[\prod_{i\neq m}\mathbf{1}[(uu',r_i)=1]\Big]\,\Phi.$$

All variables are $\le R$ on the support of F, so summing over $\mathbf{r}$ and interchanging the finite sums is legitimate. The result is $\mathcal{D}_m(F)$ with the couplings $(uu',r_i)=1$ $(i\neq m)$ and $(r_i,r_j)=1$ $(i\neq j)$ still imposed.

Step 2: drop the couplings. All of u,u',r_i are coprime to W, so a prime p shared by two of them satisfies $p > D_0$. On squarefree arguments the local weight at such a prime is $1/\varphi(p) = 1/(p-1)$ for the u-coordinates and $1/g(p) = 1/(p-2)$ for the r-coordinates, so restoring a forbidden pair costs a factor at most $1/((p-1)(p-2))$ or $1/(p-2)^2$, and $\sum_{p>D_0}(p-2)^{-2} \ll 1/D_0$. Exactly as in Lemma 8.23, summing over the $O(k^2)$ pairs bounds the total cost of removing all couplings by $O(1/D_0)$ times the fully unrestricted sum with $|\Phi|$ in place of Φ.

Step 3: size of the unrestricted sum. Using $|\Phi| \le F_{\max}^2$, the Mertens estimate of Lemma 6.30 for the two u-coordinates and Lemma 9.12 with $G\equiv 1$ for the $k-1$ coordinates r_i,

$$\sum_{\substack{u,u',(r_i)_{i\neq m}\\ \text{coprime to } W}} \frac{\mu(u)^2\mu(u')^2}{\varphi(u)\varphi(u')}\Big(\prod_{i\neq m}\frac{\mu(r_i)^2}{g(r_i)}\Big)|\Phi| \;\ll_k\; F_{\max}^2\Big(\frac{\varphi(W)\log R}{W}\Big)^{k+1}.$$

Combining with Step 2 gives the stated error. □

Uses (proof): 7.3, 8.23, 6.30, 9.12

Lemma 9.25 (Evaluation of the decoupled $(k+1)$-fold sum)**.** *Let $k\ge 2$. There is a constant $C\ge 0$, depending only on k, such that for every smooth $F:[0,1]^k\to\mathbb{R}$ supported in $\mathcal{R}_k$ there is a threshold N_0 with*

$$\Big|\mathcal{D}_m(F) \;-\; \frac{\varphi(W)^{k+1}\,(\log R)^{k+1}}{W^{k+1}}\,J_k^{(m)}(F)\Big| \;\le\; C\,\frac{F_{\max}^2\,\varphi(W)^{k+1}\,(\log R)^{k+1}}{W^{k+1}\,D_0}$$

for every $N\ge N_0$ and every m.

Lean:

- `PrimeGaps.lem_S2m_eval`

Uses (statement): 9.22, 5.7, 9.11, 9.10, 9.12, 6.31, 6.34, 7.4, 3.37, 3.29, 5.2

Uses(notation): k, R, W, D_0, μ, φ, g, γ_W, γ_g, $\mathfrak{S}(\gamma)$, (a,b), $O(g)$, F, $F_{\max}$, $F^{(m)}_{r_1,\dots,r_k}$, $J_k^{(m)}(F)$, $\mathcal{R}_k$, $\mathcal{R}_k^{(m)}$, Φ, $G_{\max}$ (index)

Proof. Separate the two m-slot coordinates u, u', whose weight is μ^2/φ, from the $k-1$ outer coordinates $(\rho_i)_{i\neq m}$, whose weight is μ^2/g.

Step 1: the two m-slot coordinates. With the outer coordinates frozen, the u-sum is $\sum_{u<R,\ (u,W)=1} \frac{\mu(u)^2}{\varphi(u)} F\big(\dots, \frac{\log u}{\log R}, \dots\big)$, a partial sum for the datum γ_W with $\mathfrak{S}(\gamma_W) = \varphi(W)/W$; the profile is F with the m-th slot free, whose C^1-norm is at most $F_{\max}$. One application of Sublemma 9.10 replaces it by $\frac{\varphi(W)}{W}(\log R)\, F^{(m)}\big((t_i)_{i\neq m}\big)$, where $F^{(m)}\big((t_i)_{i\neq m}\big) = \int_0^1 F(\dots, t_m, \dots)\, dt_m$ is the marginal of F in the m-th coordinate, with error $O\big(\frac{\varphi(W)}{W}(\log D_0)F_{\max}\big)$; the residual of Sublemma 9.10 is of that size because $V = W = N^{o(1)}$ while $\log R$ is a fixed positive multiple of $\log N$. Doing the same for u' and multiplying, the two m-slot coordinates contribute the factor $\big(\frac{\varphi(W)}{W}\log R\big)^2$ and replace the coupling Φ by $\big(F^{(m)}\big)^2$, with a relative error $O(\log D_0/\log R) = O(1/D_0)$.

Step 2: the $k-1$ outer coordinates. What remains is the $(k-1)$-fold sum

$$\sum_{(\rho_i)_{i\neq m}} \Big(\prod_{i\neq m} \mu(\rho_i)^2\, g_*(\rho_i)\Big)\, F^{(m)}\Big(\big(\tfrac{\log \rho_i}{\log R}\big)_{i\neq m}\Big)^2,$$

where g_* is the derived function of the datum γ_g, equal to μ^2/g on integers coprime to W and 0 otherwise by Sublemma 6.31. The marginal $F^{(m)}$ is smooth and supported in $\mathcal{R}_k^{(m)}$, the $(k-1)$-dimensional simplex, so Lemma 9.11 applies with $n = k-1$, $z = R$ and profile $F^{(m)}$, giving

$$\big(\mathfrak{S}(\gamma_g)\log R\big)^{k-1} \int_{\mathcal{R}_k^{(m)}} \big(F^{(m)}\big)^2 \;+\; O\Big(F_{\max}^2\Big(\frac{\varphi(W)\log R}{W}\Big)^{k-2}\frac{\varphi(W)}{W}\log D_0\Big),$$

again using $V = W = N^{o(1)}$ to reduce the residual $\mathcal{L}(\gamma_g, W, R)$ to $O\big(\frac{\varphi(W)}{W}\log D_0\big)$ and B to $O\big(\frac{\varphi(W)}{W}\log R\big)$. By Definition 5.7, $\int_{\mathcal{R}_k^{(m)}}(F^{(m)})^2 = J_k^{(m)}(F)$.

Step 3: assembly. By Sublemma 6.34, $\mathfrak{S}(\gamma_g) = \frac{\varphi(W)}{W}\big(1 + O(1/D_0)\big)$, and for fixed k, $(1 + O(1/D_0))^{k-1} = 1 + O(1/D_0)$. Multiplying Steps 1 and 2 therefore gives the main term $\frac{\varphi(W)^{k+1}(\log R)^{k+1}}{W^{k+1}} J_k^{(m)}(F)$. Since $J_k^{(m)}(F) \le F_{\max}^2$, the deviation of the singular-product factor from 1 costs $O\big(F_{\max}^2\varphi(W)^{k+1}(\log R)^{k+1}/(W^{k+1}D_0)\big)$. Each remaining error is a partial-summation residual at one coordinate multiplied by the leading factors at the other k, hence of size $O\big(F_{\max}^2\varphi(W)^{k+1}(\log D_0)(\log R)^k/W^{k+1}\big)$, whose ratio to the main term is $O(\log D_0/\log R) = O(1/D_0)$. Collecting the errors gives the claim. □

Uses (proof): 9.11, 9.10, 6.31, 6.34, 3.32

Lemma 9.26 (Asymptotic for $S_2^{(m)}$ at the smooth weight)**.** *Let $k \ge 2$, let $\mathcal{H}$ have distinct shifts, let $F : [0,1]^k \to \mathbb{R}$ be smooth and supported in $\mathcal{R}_k$, and let λ_0 be the associated F-derived weight. Let $\vartheta < 1/2$ be a level of distribution for the primes and $0 < \delta < \vartheta/2$. Then there are $C > 0$ and N_0 such that for every $N \ge N_0$ and every compatible v_0,*

$$\Big|S_2^{(m)}(\lambda_0) \;-\; \frac{\varphi(W)^k\, N\,(\log R)^{k+1}}{W^{k+1}\log N}\, J_k^{(m)}(F)\Big| \;\le\; C\,\frac{F_{\max}^2\,\varphi(W)^k\, N\,(\log R)^k}{W^{k+1}\,D_0}.$$

Lean:

- `PrimeGaps.lem_S2m_smooth`

Uses (statement): 9.21, 9.23, 9.24, 9.25, 5.11, 5.7, 4.16, 7.4, 3.37, 3.29, 3.28, 5.2, 1, 3

Uses(notation): k, N, R, W, D_0, φ, g, $O(g)$, $f \ll g$, $S_2^{(m)}(\lambda)$, $y^{(m)}(r_1,\dots,r_k)$, $y_{\max}^{(m)}$, $y_{\max}$, F, $F_{\max}$, $\lambda_0(d_1,\dots,d_k)$, $J_k^{(m)}(F)$, $\mathcal{R}_k$, ϑ, δ (index)

Proof. Apply Lemma 9.21 to $\lambda = \lambda_0$ with the exponent $A = 2k+3$, and then chain Lemmas 9.23, 9.24 and 9.25 to evaluate the diagonal form $\sum_{\mathbf{r}} (y^{(m)}(\mathbf{r}))^2 / \prod_{i \neq m} g(r_i)$. The three chained lemmas each cost $O\big(F_{\max}^2 \varphi(W)^{k+1} (\log R)^{k+1} / (W^{k+1} D_0)\big)$, so their sum does too, and the main term is $\frac{\varphi(W)^{k+1}(\log R)^{k+1}}{W^{k+1}} J_k^{(m)}(F)$. Multiplying by the prefactor $N/(\varphi(W)\log N)$ from Lemma 9.21 gives the asserted main term and an error

$$O\Big(\frac{F_{\max}^2\, \varphi(W)^k\, N\, (\log R)^{k+1}}{W^{k+1}\, D_0\, \log N}\Big) \;=\; O\Big(\frac{F_{\max}^2\, \varphi(W)^k\, N\, (\log R)^k}{W^{k+1}\, D_0}\Big),$$

using $\log R \le \log N$.

Two further errors must be absorbed into the same shape. The first error term of Lemma 9.21 is $(y_{\max}^{(m)})^2 \varphi(W)^{k-2} N (\log N)^{k-2} / (W^{k-1} D_0)$; for the F-derived weight $y_{\max}^{(m)} \ll_k F_{\max}\, \varphi(W)(\log R)/W$ by Lemma 9.12 applied to the defining sum of $y^{(m)}$, and $\log N \le (\vartheta/2 - \delta)^{-1} \log R$, so this term is again $O\big(F_{\max}^2 \varphi(W)^k N (\log R)^k / (W^{k+1} D_0)\big)$. The second error term is $y_{\max}^2 N/(\log N)^A$ with $y_{\max} \le F_{\max}$; since $\varphi(W)^k (\log R)^k / (W^{k+1} D_0) \gg (\log N)^{-(2k+3)}$ for large N — the left side decays only like a power of $\log N$ times $1/(WD_0)$, and $W = N^{o(1)}$ — the choice $A = 2k+3$ absorbs it as well. □

Uses (proof): 9.21, 9.23, 9.24, 9.25, 9.12, 3.32

9.6 The two-weight form and the outer support split

The preceding subsections treat $S_2^{(m)}(\lambda)$ as a quadratic form in a single weight. Two further pieces of the reduction are bilinear, and are recorded here in the two-weight form in which they are proved and used.

The first is that $\lambda \mapsto S_2^{(m)}(\lambda)$ is a quadratic form, so its polarisation is a bilinear prime-weighted sum to which the Chinese remainder step of Lemma 9.3 applies pair by pair, exactly as in Lemma 9.14 but without requiring the two weights to coincide.

The second is a device for controlling the size of the sieve moduli. The aggregate error bound of Lemma 9.9 is available only because every modulus $q(\mathbf{d}, \mathbf{e})$ occurring in the double sum is below $R^2 W$, hence below $N^{1/2-\eta}$: this is what makes Bombieri–Vinogradov applicable across the whole range. The modulus $q(\mathbf{d}, \mathbf{e})$ is controlled by the *product* of the two outer truncations, one from each weight, not by the individual truncations; so if one weight is truncated more severely in its outer coordinates, the other may be truncated correspondingly less without moving the moduli at all. To exploit this one needs to split a weight into a part supported below a prescribed outer cutoff and a complementary part. The split is performed on the y-side — cut off y, then transform back — rather than on the λ-side, because it is exactly the y-side cutoff that is transparent to the transforms and to the sup-norm $y_{\max}$.

Definition 9.27 (Retained weight)**.** Let λ be a finitely supported weight with y-transform y, let m be an index and let $B > 0$. The *retained weight* $\lambda_{B,m}^{\mathrm{ret}}$ is the Möbius transform (Definition 4.22) of the cut-off family

$$y_{B,m}^{\mathrm{ret}}(\mathbf{r}) \;:=\; \begin{cases} y(\mathbf{r}), & \prod_{i \neq m} r_i \le B, \\ 0, & \text{otherwise.} \end{cases}$$

Uses (statement): 4.20, 4.22

Uses(notation): k, $\mathbf{r}$, $y(r_1, \dots, r_k)$, $\lambda(d_1, \dots, d_k)$, $\lambda_y(d_1, \dots, d_k)$, $\mathbb{N}$ (index)

Definition 9.28 (Discarded weight)**.** With λ, m and B as in Definition 9.27, the *discarded weight* is $\lambda_{B,m}^{\mathrm{disc}} := \lambda - \lambda_{B,m}^{\mathrm{ret}}$.

Uses (statement): 9.27

Uses(notation): k, $\lambda(d_1,\dots,d_k)$, $\mathbf{d}$ (index)

Lemma 9.29 (The split is transparent to the y-transform)**.** *Let λ have permissible support at (R,W), let m be an index and let $B>0$. Then for every $\mathbf{r}$,*

$$y_{\lambda^{\mathrm{ret}}_{B,m}}(\mathbf{r}) \;=\; \begin{cases} y(\mathbf{r}), & \prod_{i\neq m} r_i \le B,\\ 0, & \text{otherwise,}\end{cases} \qquad \text{and} \qquad y_{\lambda^{\mathrm{disc}}_{B,m}}(\mathbf{r}) \;=\; y(\mathbf{r}) - y_{\lambda^{\mathrm{ret}}_{B,m}}(\mathbf{r}).$$

In particular both pieces have permissible support at (R,W) and their y-transforms have sup-norm at most $y_{\max}$.

Uses (statement): 9.27, 9.28, 4.24, 4.21, 3.39, 4.27

Uses(notation): k, R, W, $\mathbf{r}$, $y(r_1,\dots,r_k)$, $y_{\max}$, $\lambda(d_1,\dots,d_k)$, $\lambda_y(d_1,\dots,d_k)$, $\mathbb{N}$ (index)

Proof. The y- and λ-transforms are mutually inverse on weights supported on tuples with squarefree coordinates (Lemma 4.24); applying the y-transform to $\lambda^{\mathrm{ret}}_{B,m}$, which by Definition 9.27 is the λ-transform of $y^{\mathrm{ret}}_{B,m}$, returns $y^{\mathrm{ret}}_{B,m}$ on the relevant support. Linearity of the y-transform then gives the second identity. Since the cut-off family is pointwise dominated by y in absolute value, its sup-norm is at most $y_{\max}$, and so is that of the difference after the triangle inequality is applied coordinatewise; and both transforms inherit permissible support from y by Lemma 4.21. □

Uses (proof): 4.24, 4.23, 4.21

Lemma 9.30 (Additivity of the diagonal form along the split)**.** *With λ, m, B as above,*

$$\sum_{\mathbf{r}} \frac{Y^{(m)}(\mathbf{r})^2}{\prod_{i\neq m} g(r_i)} \;=\; \sum_{\mathbf{r}} \frac{\big(Y^{(m)}_{\mathrm{ret}}(\mathbf{r})\big)^2}{\prod_{i\neq m} g(r_i)} \;+\; \sum_{\mathbf{r}} \frac{\big(Y^{(m)}_{\mathrm{disc}}(\mathbf{r})\big)^2}{\prod_{i\neq m} g(r_i)},$$

where $Y^{(m)}_{\mathrm{ret}}$ and $Y^{(m)}_{\mathrm{disc}}$ are the free-coordinate marginals (Definition 7.10) of $\lambda^{\mathrm{ret}}_{B,m}$ and $\lambda^{\mathrm{disc}}_{B,m}$.

Uses (statement): 9.27, 9.28, 7.10, 9.29, 3.6

Uses(notation): k, W, g, $\mathbf{r}$, $y(r_1,\dots,r_k)$, $Y_{\mathbf{r}}$, $\lambda(d_1,\dots,d_k)$ (index)

Proof. The marginal $Y^{(m)}(\mathbf{r})$ sums $y(\mathbf{r}\,[m\mapsto a])/\varphi(a)$ over the single free coordinate a, and the cutoff $\prod_{i\neq m} r_i \le B$ does not involve that coordinate. Hence by Lemma 9.29, for each fixed $\mathbf{r}$ either $Y^{(m)}_{\mathrm{ret}}(\mathbf{r}) = Y^{(m)}(\mathbf{r})$ and $Y^{(m)}_{\mathrm{disc}}(\mathbf{r}) = 0$ (when the cutoff holds), or the reverse (when it fails). In either case $Y^{(m)}(\mathbf{r})^2 = Y^{(m)}_{\mathrm{ret}}(\mathbf{r})^2 + Y^{(m)}_{\mathrm{disc}}(\mathbf{r})^2$ pointwise, and all three families are summable against $1/\prod_{i\neq m} g(r_i)$, so the identity survives summation. □

Uses (proof): 9.29

Lemma 9.31 (The modulus depends only on the product of the outer truncations)**.** *Let λ_1 be supported on tuples with $\prod_{i\neq m} d_i \le B_1$ and λ_2 on tuples with $\prod_{i\neq m} e_i \le B_2$. Then every pair $(\mathbf{d},\mathbf{e})$ with $\lambda_1(\mathbf{d})\lambda_2(\mathbf{e}) \neq 0$ and $d_m = e_m = 1$ satisfies*

$$q(\mathbf{d},\mathbf{e}) \;\le\; B_1\, B_2\, W.$$

Uses (statement): 8.1, 3.29, 9.27

Uses(notation): k, W, R, $[a,b]$, $\mathbf{d}$, $\mathbf{e}$, $q(\mathbf{d},\mathbf{e})$, $\lambda(d_1,\dots,d_k)$ (index)

Proof. Since $d_m = e_m = 1$ we have $[d_m, e_m] = 1$, so $q(\mathbf{d},\mathbf{e}) = W\prod_{i\neq m}[d_i,e_i] \le W\prod_{i\neq m} d_i e_i \le B_1 B_2 W$. □

The bound depends on B_1 and B_2 only through their product. For weights of permissible support one may take $B_1 = B_2 = R$, recovering the range R^2W of Sublemma 9.7; but the same range R^2W — and hence the same appeal to Bombieri–Vinogradov — is available for any pair of outer truncations with $B_1B_2 \le R^2$, even when one of them exceeds R. This is precisely what the split of Definitions 9.27 and 9.28 makes possible: the modulus range is governed by $\vartheta < 1/2$ together with the split, and no further hypothesis relating the truncation exponents is needed.

Lemma 9.32 (Two-weight Chinese remainder bound)**.** *There is a constant $C > 0$, depending only on k, $\mathcal{H}$ and m, such that for all sufficiently large N, every compatible v_0 and all finitely supported weights λ_1, λ_2 dominated by a common weight of permissible support at (R, W),*

$$\Big|\tfrac{1}{2}\big(S_2^{(m)}(\lambda_1+\lambda_2) - S_2^{(m)}(\lambda_1) - S_2^{(m)}(\lambda_2)\big) \;-\; \frac{X_N}{\varphi(W)} \sum_{\substack{\mathbf{d},\mathbf{e}\\ d_m=e_m=1}}{}' \frac{\lambda_1(\mathbf{d})\,\lambda_2(\mathbf{e})}{\prod_{i=1}^k \varphi([d_i,e_i])}\Big|$$
$$\le\; C \sum_{\substack{\mathbf{d},\mathbf{e}\\ d_m=e_m=1}} |\lambda_1(\mathbf{d})|\,|\lambda_2(\mathbf{e})|\,E\big(N, q(\mathbf{d},\mathbf{e})\big),$$

where $\sum'$ restricts to pairs for which $W, [d_1,e_1],\dots,[d_k,e_k]$ are pairwise coprime and $(h_m - h_i, [d_i,e_i]) = 1$ for every $i \neq m$.

Uses (statement): 9.3, 9.2, 9.1, 9.14, 4.16, 8.1, 3.24, 3.26, 3.39, 3.29

Uses(notation): k, h_i, N, v_0, W, φ, $[a,b]$, $\lambda(d_1,\dots,d_k)$, $\mathbf{d}$, $\mathbf{e}$, $S_2^{(m)}(\lambda)$, X_N, $q(\mathbf{d},\mathbf{e})$, $\sum'$, $E(N,q)$ (index)

Proof. The map $\lambda \mapsto S_2^{(m)}(\lambda)$ is a quadratic form: by Lemma 9.1 it is the double sum of $\lambda(\mathbf{d})\lambda(\mathbf{e})\,T_m(\mathbf{d},\mathbf{e})$ with a kernel T_m not depending on λ. Its polarisation, the left-hand quantity, is therefore $\sum_{\mathbf{d},\mathbf{e}} \lambda_1(\mathbf{d})\lambda_2(\mathbf{e})\,T_m(\mathbf{d},\mathbf{e})$.

From here the proof of Lemma 9.14 applies verbatim, pair by pair, since it never used that the two weights coincide: Lemma 9.2 discards the pairs with $d_m \neq 1$ or $e_m \neq 1$; the pairs violating the pairwise-coprimality or shift-coprimality conditions contribute 0 for N large, because every prime factor of a $[d_i,e_i]$ occurring on the common dominating support exceeds D_0; and Lemma 9.3 evaluates each retained $T_m(\mathbf{d},\mathbf{e})$ as $X_N/\varphi(q(\mathbf{d},\mathbf{e}))$ with an error at most $C\,E(N,q(\mathbf{d},\mathbf{e}))$. Multiplicativity of φ across the pairwise coprime factorisation $q(\mathbf{d},\mathbf{e}) = W\prod_i [d_i,e_i]$ turns the main terms into the displayed restricted sum, and the triangle inequality collects the errors. □

Uses (proof): 9.1, 9.2, 9.3, 9.14

10 From positivity to primes, and L^2-continuity

We pass from positivity of the sieve difference to a bound on prime gaps, and establish the mean-square continuity used to compare the variational supremum with the smooth-profile asymptotics.

The sieve difference $S(\lambda) = S_2(\lambda) - \rho\, S_1(\lambda)$ counts, over the sieve window, the excess over ρ of the number of primes among $n + h_1, \dots, n + h_k$. If it is positive then some n in the window has more than ρ of the shifts prime, and, this number being an integer, at least $\lfloor \rho + 1 \rfloor$ of them. Carrying this out for each large N yields infinitely many such n; since the shifts $n + h_i$ lie in an interval of length $\operatorname{diam}(\mathcal{H})$, infinitely many intervals of that length contain $\lfloor \rho + 1 \rfloor$ primes, bounding the gap between consecutive primes. Proposition 10.5 records this for a general tuple and a general prime count; the two bounds follow from it by choosing a tuple.

M_k is a supremum over square-integrable profiles, whereas the sieve asymptotics for S_1 and $S_2^{(m)}$ apply only to *smooth* profiles supported on the simplex. The comparison rests on two facts: the functionals I_k and $J_k^{(m)}$ are continuous for the L^2 topology, and smooth simplex-supported profiles are L^2-dense. A profile that nearly attains M_k may therefore be replaced by a smooth one that nearly attains it. The asymptotic for S follows, and from it the eventual positivity of S.

10.1 From $S > 0$ to primes

Lemma 10.1 ($S > 0$ for all large N gives infinitely many prime-rich translates)**.** *Let $\mathcal{H} = \{h_1, \dots, h_k\}$ and $\rho > 0$. Suppose that for every sufficiently large N there exists a finitely supported $\lambda : \mathbb{N}^k \to \mathbb{R}$ (depending on N) with $S_2(\lambda) - \rho\, S_1(\lambda) > 0$. Then the set*

$$\big\{\, n \in \mathbb{N} \;:\; \#\{\, i : n + h_i \in \mathbb{P} \,\} \;\geq\; \lfloor \rho + 1 \rfloor \,\big\}$$

is infinite.

Lean:

- `PrimeGaps.maynardTao_endgame`

Uses (statement): 4.14, 4.15

Uses(notation): $\mathcal{H}$, h_i, k, $\lambda(d_1, \dots, d_k)$, ρ, N, $S_1(\lambda)$, $S_2(\lambda)$, $S(\lambda)$, $\lfloor x \rfloor$, $\mathbb{P}$, $\mathbb{N}$ (index)

Proof. Write $\mathcal{A}$ for the displayed set. It suffices to show that $\mathcal{A}$ is unbounded above, since an unbounded subset of $\mathbb{N}$ is infinite.

Let $M \in \mathbb{N}$ be arbitrary. Choose $N \geq M$ large enough that the hypothesis supplies a finitely supported λ with $S_2(\lambda) - \rho\, S_1(\lambda) > 0$ for that N. Lemma 4.19 then produces an integer n with $N < n \leq 2N$ (and $n \equiv v_0 \pmod W$, which we discard) such that at least $\lfloor \rho + 1 \rfloor$ of the $n + h_i$ are prime. Thus $n \in \mathcal{A}$ and $n > N \geq M$. As M was arbitrary, $\mathcal{A}$ is unbounded. □

Uses (proof): 4.19

Definition 10.2 (Prime-rich interval)**.** For $n, L, r \in \mathbb{N}$, say that the closed interval $[n, n + L]$ is *r-prime-rich* if it contains at least r distinct primes, that is, if

$$\#\big(\, [n, n + L] \cap \mathbb{P} \,\big) \;\geq\; r.$$

Uses(notation): $\mathbb{N}$, $\mathbb{P}$ (index)

Lemma 10.3 (Prime-rich translates give prime-rich intervals)**.** *Let $\mathcal{H} = \{h_1, \dots, h_k\}$ with $k \geq 1$ and $h_1 < \cdots < h_k$, and let $r \in \mathbb{N}$. If the set $\{\, n : \#\{i : n + h_i \in \mathbb{P}\} \geq r \,\}$ is infinite, then for arbitrarily large n' the interval $[n', n' + \mathrm{diam}(\mathcal{H})]$ is r-prime-rich in the sense of Definition 10.2.*

Uses (statement): 10.2, 3.18

Uses(notation): $\mathcal{H}$, h_i, k, $\mathrm{diam}(\mathcal{H})$, $\mathbb{P}$, $\mathbb{N}$ (index)

Proof. Write $\mathcal{A}$ for the hypothesised infinite set and $D := \mathrm{diam}(\mathcal{H}) = h_k - h_1$. Let M be arbitrary and pick $n \in \mathcal{A}$ with $n > M$; put $n' := n + h_1 \geq n > M$.

Consider the map $i \mapsto n + h_i$ from $\{i : n + h_i \in \mathbb{P}\}$ to $[n', n' + D] \cap \mathbb{P}$. It is well defined: if $n + h_i$ is prime then it is a prime, and $h_1 \leq h_i \leq h_k$ gives $n' = n + h_1 \leq n + h_i \leq n + h_k = n' + D$. It is injective because the h_i are distinct. Hence

$$\#\big([n', n' + D] \cap \mathbb{P}\big) \;\geq\; \#\{i : n + h_i \in \mathbb{P}\} \;\geq\; r,$$

so $[n', n' + D]$ is r-prime-rich. Since M was arbitrary and $n' > M$, such n' occur arbitrarily late. □

Lemma 10.4 (Prime-rich intervals imply small prime gaps)**.** *Let $r \geq 1$ and $L \geq 0$ be integers. If the interval $[n, n + L]$ is r-prime-rich for arbitrarily large n, then*

$$p_{m+r-1} - p_m \;\leq\; L$$

for infinitely many m.

Lean:

- `PrimeGaps.frequently_prime_gap_le_of_frequently_interval`

Uses (statement): 10.2

Uses(notation): p_n, $\pi(x)$, $\mathbb{P}$, $\mathbb{N}$ (index)

Proof. For $n \geq 1$ let $\mu(n) := \min\{\, j \geq 1 : p_j \geq n \,\}$ be the index of the least prime that is at least n; this is well defined because there are infinitely many primes. Since exactly $\pi(n-1)$ primes are smaller than n, we have $\mu(n) = \pi(n-1) + 1$, so μ is non-decreasing and $\mu(n) \to \infty$ as $n \to \infty$.

Let M be arbitrary; we produce $m \geq M$ with $p_{m+r-1} - p_m \leq L$. Choose n_0 with $\mu(n_0) \geq M$, and then, by hypothesis, some $n \geq n_0$ for which $[n, n + L]$ is r-prime-rich. Put $m := \mu(n)$, so that $m \geq \mu(n_0) \geq M$.

By construction $p_m \geq n$, and every prime in $[n, n + L]$ has index at least m. Since the primes in $[n, n + L]$ are consecutive in the enumeration, they are $p_m, p_{m+1}, \dots, p_{\pi(n+L)}$, and there are at least r of them, so $\pi(n + L) \geq m + r - 1$ and therefore $p_{m+r-1} \leq n + L$. Subtracting,

$$p_{m+r-1} - p_m \;\leq\; (n + L) - n \;=\; L.$$

As M was arbitrary, this holds for infinitely many m. □

Proposition 10.5 (Bounded gaps from prime-rich translates)**.** *Let $\mathcal{H} = \{h_1, \dots, h_k\}$ with $k \geq 1$ and $h_1 < \cdots < h_k$, and let $r \geq 1$ be an integer. If the set*

$$\big\{\, n \in \mathbb{N} \;:\; \#\{\, i : n + h_i \in \mathbb{P} \,\} \;\geq\; r \,\big\}$$

is infinite, then $p_{m+r-1} - p_m \leq \mathrm{diam}(\mathcal{H})$ for infinitely many m.

Uses (statement): 3.18

Uses(notation): $\mathcal{H}$, h_i, k, $\operatorname{diam}(\mathcal{H})$, p_n, $\mathbb{P}$, $\mathbb{N}$ (index)

Proof. Apply Lemma 10.3 to obtain, for arbitrarily large n', an r-prime-rich interval $[n', n' + \operatorname{diam}(\mathcal{H})]$; then apply Lemma 10.4 with $L = \operatorname{diam}(\mathcal{H})$. □

Uses (proof): 10.3, 10.4

10.2 L^2-continuity of the functionals

Throughout this subsection $k \geq 1$ and, for $s > 0$, $\mathcal{R}_k(s) := \{t \in [0,\infty)^k : t_1 + \cdots + t_k \leq s\}$ denotes the simplex of scale s, so that $\mathcal{R}_k = \mathcal{R}_k(1)$ and the enlarged simplex $\mathcal{T}_\varepsilon$ is $\mathcal{R}_k(1+\varepsilon)$. Every $\mathcal{R}_k(s)$ is compact, being closed and contained in the ball of radius s about the origin. Functions in $L^2(\mathcal{R}_k(s))$ are always regarded as defined on all of $\mathbb{R}^k$ by extension by zero.

Definition 10.6 (Partial-integration operator T_m)**.** Let $s > 0$ and $m \in \{1, \dots, k\}$. The *m-th partial-integration operator* is the linear map

$$T_m : L^2(\mathcal{R}_k(s)) \longrightarrow L^2(\mathbb{R}^{k-1}), \qquad (T_m f)(t_1, \dots, \widehat{t_m}, \dots, t_k) := \int_{\mathbb{R}} f(t_1, \dots, t_k)\, dt_m,$$

the hat denoting omission of the m-th coordinate.

Uses (statement): 5.2

Uses(notation): T_m, k, $\mathcal{R}_k$, $\|f\|_{L^2(X)}$ (index)

10.3 Smooth approximation on the simplex

Definition 10.7 (The main scale)**.** For $k \in \mathbb{N}$ and $N \in \mathbb{N}$ set

$$\mathfrak{M}_k(N) := \frac{\varphi(W)^k\, N\, (\log R)^k}{W^{k+1}},$$

the common scale of the first and second sieve moments.

Uses (statement): 3.29, 3.37

Uses(notation): k, N, W, R, φ (index)

Proposition 10.8 (Simultaneous asymptotics for the sieve moments)**.** *Let $k \geq 2$, let $\mathcal{H} = \{h_1, \dots, h_k\}$ be admissible, and let $F \in C^\infty(\mathbb{R}^k)$ have $\operatorname{supp} F \subset \mathcal{R}_k$. Let $0 < \vartheta < 1/2$ and $0 < \delta < \vartheta/2$, and suppose the primes have level of distribution ϑ. Write λ_F for the sieve weight attached to F at truncation $R = N^{\vartheta/2-\delta}$. Then there are functions $e_1, e_2 : \mathbb{N} \to \mathbb{R}$ with $e_1(N) \to 0$ and $e_2(N) \to 0$, and a threshold N_0, such that for every $N \geq N_0$ and every residue v_0 modulo W with $(v_0 + h_i, W) = 1$ for all i, the two estimates*

$$\bigl|S_1(\lambda_F) - \mathfrak{M}_k(N)\, I_k(F)\bigr| \leq e_1(N)\, \mathfrak{M}_k(N)$$

and

$$\Bigl|S_2(\lambda_F) - \frac{\varphi(W)^k\, N\, (\log R)^{k+1}}{W^{k+1} \log N} \sum_{m=1}^{k} J_k^{(m)}(F)\Bigr| \leq e_2(N)\, \mathfrak{M}_k(N)$$

hold simultaneously.

Lean:

- `PrimeGaps.MainProp.prop_main_prop`

Uses (statement): 3.15, 5.2, 3.21, 5.11, 3.37, 4.14, 4.15, 5.6, 5.7, 10.7, 3.29

Uses(notation): k, $\mathcal{H}$, h_i, F, $\mathcal{R}_k$, ϑ, δ, R, N, W, v_0, φ, $S_1(\lambda)$, $S_2(\lambda)$, $S_2^{(m)}(\lambda)$, $I_k(F)$, $J_k^{(m)}(F)$, $\lambda_0(d_1,\dots,d_k)$, (a,b), $o(g)$, $C^\infty(X)$, $F_{\max}$, D_0 (index)

Proof. The two estimates are the smooth-profile asymptotics for the two moments, made uniform in N and in v_0.

For the first, Lemma 8.25 supplies a constant $C_1 > 0$ and a threshold N_1 such that for all $N \ge N_1$ and all admissible residues v_0,

$$\Big|S_1(\lambda_F) - \frac{\varphi(W)^k N(\log R)^k}{W^{k+1}} \int_{\mathcal{R}_k} F^2\Big| \;\le\; \frac{C_1 F_{\max}^2\, \varphi(W)^k N(\log R)^k}{W^{k+1} D_0}.$$

The main term is $\mathfrak{M}_k(N)\, I_k(F)$ by Definitions 10.7 and 5.6, and the error is $e_1(N)\,\mathfrak{M}_k(N)$ with $e_1(N) := C_1 F_{\max}^2/D_0$. Since $D_0 = \log\log\log N \to \infty$, we have $e_1(N) \to 0$.

For the second, Lemma 9.26 supplies, for each $m \in \{1,\dots,k\}$, a constant $C^{(m)} > 0$ and a threshold $N^{(m)}$ such that for all $N \ge N^{(m)}$ and all admissible v_0,

$$\Big|S_2^{(m)}(\lambda_F) - \frac{\varphi(W)^k N(\log R)^{k+1}}{W^{k+1}\log N}\, J_k^{(m)}(F)\Big| \;\le\; \frac{C^{(m)} F_{\max}^2\, \varphi(W)^k N(\log R)^k}{W^{k+1} D_0}.$$

Summing over the finitely many m and using the decomposition $S_2(\lambda) = \sum_{m=1}^k S_2^{(m)}(\lambda)$ of Lemma 4.17, the triangle inequality gives the second estimate with $e_2(N) := \big(\sum_m C^{(m)}\big) F_{\max}^2/D_0$, which again tends to 0.

Finally take N_0 to be the largest of N_1 and the finitely many $N^{(m)}$ (and large enough that $D_0 \ge 2$, so that the error scales are meaningful). Both estimates then hold for all $N \ge N_0$, simultaneously and uniformly in v_0. □

Uses (proof): 8.25, 9.26, 4.17, 5.25, 3.28, 4.16

Lemma 10.9 (Asymptotic for the sieve difference)**.** *Under the hypotheses of Proposition 10.8 and for any $\rho > 0$, there is a function $e : \mathbb{N} \to \mathbb{R}$ with $e(N) \to 0$ and a threshold N_0 such that for every $N \ge N_0$ and every residue v_0 modulo W with $(v_0 + h_i, W) = 1$ for all i,*

$$\Big|S(\lambda_F) - \mathfrak{M}_k(N)\Big(\big(\tfrac{\vartheta}{2} - \delta\big)\sum_{m=1}^k J_k^{(m)}(F) - \rho\, I_k(F)\Big)\Big| \;\le\; e(N)\,\mathfrak{M}_k(N),$$

where $S(\lambda_F) = S_2(\lambda_F) - \rho\, S_1(\lambda_F)$.

Lean:

- `PrimeGaps.MainProp.lem_S_asymptotic`

Uses (statement): 10.8, 10.7, 5.6, 5.7, 4.14, 4.15, 3.37

Uses(notation): k, F, ϑ, δ, β, ρ, N, W, v_0, R, $S_1(\lambda)$, $S_2(\lambda)$, $S(\lambda)$, $I_k(F)$, $J_k^{(m)}(F)$, (a,b), $o(g)$ (index)

Proof. The point is that the two main terms of Proposition 10.8 share the scale $\mathfrak{M}_k(N)$. Indeed $\log R = (\vartheta/2-\delta)\log N$ by the definition $R = N^{\vartheta/2-\delta}$, so

$$\frac{\varphi(W)^k N(\log R)^{k+1}}{W^{k+1}\log N} \;=\; \frac{\varphi(W)^k N(\log R)^k}{W^{k+1}} \cdot \frac{\log R}{\log N} \;=\; \mathfrak{M}_k(N)\Big(\frac{\vartheta}{2}-\delta\Big).$$

Let e_1, e_2, N_0 be as in Proposition 10.8 and put $e := e_2 + \rho\, e_1$, which tends to 0. Writing $P := \mathfrak{M}_k(N)(\vartheta/2-\delta)$, $\mathcal{J} := \sum_m J_k^{(m)}(F)$ and $\mathcal{I} := I_k(F)$, we have the algebraic identity

$$\big(S_2(\lambda_F) - \rho\, S_1(\lambda_F)\big) - \mathfrak{M}_k(N)\big((\vartheta/2-\delta)\mathcal{J} - \rho\,\mathcal{I}\big) \;=\; \big(S_2(\lambda_F) - P\,\mathcal{J}\big) - \rho\,\big(S_1(\lambda_F) - \mathfrak{M}_k(N)\,\mathcal{I}\big).$$

The triangle inequality and the two estimates of Proposition 10.8 bound the absolute value of the right-hand side by

$$e_2(N)\,\mathfrak{M}_k(N) + \rho\, e_1(N)\,\mathfrak{M}_k(N) \;=\; e(N)\,\mathfrak{M}_k(N),$$

as required. □

Lemma 10.10 (Eventual positivity of the sieve difference)**.** *Assume the hypotheses of Lemma 10.9, and assume in addition the strict inequality*

$$\rho\, I_k(F) \;<\; \Big(\frac{\vartheta}{2}-\delta\Big)\sum_{m=1}^{k} J_k^{(m)}(F).$$

Then there is a threshold N_0 such that for every $N \ge N_0$ and every residue v_0 modulo W with $(v_0+h_i, W) = 1$ for all i,

$$S_2(\lambda_F) - \rho\, S_1(\lambda_F) \;>\; 0.$$

Lean:

- `PrimeGaps.MainProp.lem_S_positive`

Uses (statement): 10.9, 5.6, 5.7, 4.14, 4.15, 10.7

Uses(notation): k, F, ϑ, δ, ρ, N, W, v_0, $S_1(\lambda)$, $S_2(\lambda)$, $S(\lambda)$, $I_k(F)$, $J_k^{(m)}(F)$, (a,b) (index)

Proof. Set $c := (\vartheta/2-\delta)\sum_m J_k^{(m)}(F) - \rho\, I_k(F)$, which is strictly positive by hypothesis. Let e and N_1 be as in Lemma 10.9. Since $e(N) \to 0$, there is N_2 with $e(N) < c$ for all $N \ge N_2$.

The scale $\mathfrak{M}_k(N) = \varphi(W)^k N(\log R)^k / W^{k+1}$ is strictly positive once N is large enough that $\log R > 0$, which holds as soon as $N > 1$, because $\vartheta/2-\delta > 0$. Let N_0 be the largest of N_1, N_2 and this threshold.

For $N \ge N_0$ and any admissible v_0, the lower half of the estimate of Lemma 10.9 gives

$$S_2(\lambda_F) - \rho\, S_1(\lambda_F) \;\ge\; \mathfrak{M}_k(N)\,c - e(N)\,\mathfrak{M}_k(N) \;=\; \mathfrak{M}_k(N)\,\big(c - e(N)\big) \;>\; 0,$$

since $\mathfrak{M}_k(N) > 0$ and $c - e(N) > 0$. □

Proposition 10.11 (The criterion on the standard simplex)**.** *Let $k \ge 2$, let $\mathcal{H} = \{h_1, \dots, h_k\}$ be admissible, let $\rho > 0$, and let $\vartheta \in (0,1)$ be a level of distribution for the primes. Suppose that*

$$M_k \;>\; \frac{2\rho}{\vartheta}.$$

Then the set $\{\, n \in \mathbb{N} : \#\{i : n + h_i \in \mathbb{P}\} \ge \lfloor \rho \rfloor + 1 \,\}$ is infinite, and consequently $p_{m+\lfloor\rho\rfloor} - p_m \le \operatorname{diam}(\mathcal{H})$ for infinitely many m.

Uses (statement): 5.8, 3.15, 3.21, 3.18

Uses(notation): M_k, $\mathcal{H}$, h_i, k, ρ, ϑ, $\operatorname{diam}(\mathcal{H})$, $\chi_{\mathbb{P}}$, p_n (index)

Proof. Apply Lemma 5.22 at $\varepsilon = 0$. Since $M_k > 2\rho/\vartheta$, it produces a smooth F supported in the closed simplex $\mathcal{R}_k$, together with a $\delta \in (0, \vartheta/2)$, such that

$$\rho\, I_k(F) \;<\; \Big(\frac{\vartheta}{2} - \delta\Big) \sum_{m=1}^{k} J_k^{(m)}(F).$$

Because F is supported in the standard simplex, no enlargement of the sieve support is required, and the tensor coefficients derived from F as in Definition 5.11 are admissible at truncation $R = N^{\vartheta/2-\delta}$.

The strict inequality above is precisely the additional hypothesis of Lemma 10.10; its remaining hypotheses are those of Lemma 10.9, which hold for this F and this δ at the level ϑ. Lemma 10.10 therefore yields an N_0 such that $S_2(\lambda) - \rho\, S_1(\lambda) > 0$ for every $N \geq N_0$, where λ is the coefficient system attached to F at scale N.

Each such λ is finitely supported, so Lemma 10.1 applies and shows that $\{\, n : \#\{i : n + h_i \in \mathbb{P}\} \geq \lfloor \rho + 1 \rfloor \,\}$ is infinite. Since $\lfloor \rho + 1 \rfloor = \lfloor \rho \rfloor + 1$ for every real ρ, this gives the first assertion. The second follows by Proposition 10.5 applied with $r = \lfloor \rho \rfloor + 1$.

The hypothesis $M_k > 2\rho/\vartheta$ permits every $\rho < \vartheta M_k/2$, and the count $\lfloor \rho \rfloor + 1$ is largest when ρ is taken just below that threshold: for $\rho = \vartheta M_k/2 - \eta$ with $\eta > 0$ small enough one has $\lfloor \rho \rfloor + 1 = \lceil \vartheta M_k/2 \rceil = r_k(\vartheta)$, the prime count of Definition 5.9. Thus $r_k(\vartheta)$ is exactly the number of primes this proposition delivers at level ϑ. □

Uses (proof): 5.22, 10.10, 10.9, 10.1, 10.5, 5.11, 5.6, 5.7, 5.9

11 Instantiation I: the standard simplex and the bound $H_1 \leq 600$

We fix the data for the first case. The domain is the standard simplex $\mathcal{R}_k$; the dimension is $k = 105$; the profile is an explicit symmetric polynomial supported on $\mathcal{R}_{105}$; and the admissible tuple is an explicit 105-tuple of diameter 600. The variational criterion requires $M_{105} > 4$, and the profile certifies it. The only analytic input is the Bombieri–Vinogradov theorem (Theorem 1); no hypothesis on the level of distribution beyond $\vartheta < 1/2$ is assumed.

11.1 The tuple $\mathcal{H}_{105}$

Definition 11.1 (The 105-tuple)**.** Let

$$\begin{aligned}\mathcal{H}_{105} \;:=\; \{\, & 0,\ 10,\ 12,\ 24,\ 28,\ 30,\ 34,\ 42,\ 48,\ 52,\ 54,\ 64,\ 70,\\ & 72,\ 78,\ 82,\ 90,\ 94,\ 100,\ 112,\ 114,\ 118,\ 120,\ 124,\ 132,\ 138,\\ & 148,\ 154,\ 168,\ 174,\ 178,\ 180,\ 184,\ 190,\ 192,\ 202,\ 204,\ 208,\ 220,\\ & 222,\ 232,\ 234,\ 250,\ 252,\ 258,\ 262,\ 264,\ 268,\ 280,\ 288,\ 294,\ 300,\\ & 310,\ 322,\ 324,\ 328,\ 330,\ 334,\ 342,\ 352,\ 358,\ 360,\ 364,\ 372,\ 378,\\ & 384,\ 390,\ 394,\ 400,\ 402,\ 408,\ 412,\ 418,\ 420,\ 430,\ 432,\ 442,\ 444,\\ & 450,\ 454,\ 462,\ 468,\ 472,\ 478,\ 484,\ 490,\ 492,\ 498,\ 504,\ 510,\ 528,\\ & 532,\ 534,\ 538,\ 544,\ 558,\ 562,\ 570,\ 574,\ 580,\ 582,\ 588,\ 594,\ 598,\\ & 600\,\},\end{aligned}$$

a set of 105 non-negative integers listed in increasing order, with least element 0 and greatest element 600. It is the narrowest admissible 105-tuple recorded in the tables of Engelsma and Sutherland.

Lean:
- `PrimeGaps.H105`

Uses(notation): $\mathcal{H}_{105}$, h_i, k, $\mathbb{N}$ (index)

Lemma 11.2 (Cardinality of $\mathcal{H}_{105}$)**.** $\#\mathcal{H}_{105} = 105$.

Lean:
- `PrimeGaps.card_H105`

Uses (statement): 11.1

Uses(notation): $\mathcal{H}_{105}$, k (index)

Proof. The list in Definition 11.1 is strictly increasing, hence has pairwise distinct entries; counting them gives 105. □

Lemma 11.3 (Admissibility of $\mathcal{H}_{105}$)**.** $\mathcal{H}_{105}$ *is admissible in the sense of Definition 3.15.*

Lean:
- `PrimeGaps.admissible_H105`

Uses (statement): 11.1, 3.15

Uses(notation): $\mathcal{H}_{105}$, $\mathbb{P}$ (index)

Proof. By Lemma 3.17 it suffices to exhibit, for each prime $p \le \#\mathcal{H}_{105} = 105$, a residue class modulo p omitted by $\mathcal{H}_{105}$: for a prime $p > 105$ the reduction map $\mathcal{H}_{105} \to \mathbb{Z}/p\mathbb{Z}$ has image of cardinality at most $\#\mathcal{H}_{105} = 105 < p$ by Lemma 11.2, so by Lemma 3.16 the omission is automatic. There are exactly twenty-seven primes $p \le 105$, namely

2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97, 101, 103,

and for each the set $\{h \bmod p : h \in \mathcal{H}_{105}\}$ is an explicit finite subset of $\mathbb{Z}/p\mathbb{Z}$ which may be computed by direct enumeration of the 105 listed integers; in every case the enumeration omits a class. (For instance every element of $\mathcal{H}_{105}$ is even, so the class 1 mod 2 is omitted; the classes 2 mod 3, 1 mod 5 and 4 mod 7 are likewise omitted.) Since the check is a finite computation on an explicit finite set, it is decidable. □

Uses (proof): 3.17, 3.16, 11.2

Lemma 11.4 (Diameter of $\mathcal{H}_{105}$)**.** $\operatorname{diam}(\mathcal{H}_{105}) = 600$.

Lean:
- `PrimeGaps.diameter_H105`

Uses (statement): 11.1, 3.18

Uses(notation): $\mathcal{H}_{105}$, $\operatorname{diam}(\mathcal{H})$, h_i (index)

Proof. By Lemma 3.19, $\operatorname{diam}(\mathcal{H}_{105}) = \max \mathcal{H}_{105} - \min \mathcal{H}_{105}$. The list of Definition 11.1 is increasing, so its minimum is its first entry 0 and its maximum is its last entry 600; the difference is 600. □

Uses (proof): 3.19

11.2 The variational datum at $k = 105$

The profile lies in the space spanned by the symmetric monomials $(1-P_1)^b P_2^c$ of the two power sums $P_1 = t_1 + \cdots + t_{105}$ and $P_2 = t_1^2 + \cdots + t_{105}^2$, restricted to total weight $b + 2c \le 11$; this is the basis $\mathcal{B}_{11}$ of Definition 5.5. A profile in this span is determined by its coefficient vector, and both I_{105} and $J_{105}^{(m)}$ restrict to explicit rational quadratic forms in that vector, so a single strict inequality between two rational numbers certifies the required bound.

Definition 11.5 (The certificate profile at $k = 105$)**.** Let $(b_i^*, c_i^*;\, a_i^*)_{1 \le i \le 32}$ be the following list of triples, in which $b_i^*, c_i^* \in \mathbb{Z}_{\ge 0}$ satisfy $b_i^* + 2c_i^* \le 11$ and $a_i^* \in \mathbb{Z}$:

$$
\begin{aligned}
&(1,0;\,-330),\ (1,1;\,78029),\ (4,0;\,5743874),\\
&(1,2;\,-7297879),\ (0,3;\,16574),\ (2,2;\,-12057779),\\
&(4,1;\,-1391698982),\ (6,0;\,-1110129088),\ (1,3;\,340699862),\\
&(3,2;\,896499841),\ (5,1;\,6664560083),\ (7,0;\,3420088505),\\
&(0,4;\,-1685613),\ (2,3;\,1231124500),\ (4,2;\,98419064701),\\
&(6,1;\,126065419782),\ (1,4;\,-8050576592),\ (3,3;\,-90260561416),\\
&(5,2;\,-636913875947),\ (9,0;\,305538813668),\ (0,5;\,42387884),\\
&(2,4;\,-33232242480),\ (4,3;\,-2440220187876),\ (6,2;\,-7132620815184),\\
&(8,1;\,-6762976569892),\ (10,0;\,-1975749959806),\ (1,5;\,78797053343),\\
&(3,4;\,2504524184641),\ (5,3;\,24468760733568),\ (7,2;\,35184372088832),\\
&(9,1;\,18268089505780),\ (11,0;\,3188339316861).
\end{aligned}
$$

The *certificate polynomial* at $k = 105$ is the symmetric polynomial

$$
P^*(t_1, \ldots, t_{105}) \;:=\; \sum_{i=1}^{32} a_i^* \,\big(1 - P_1\big)^{b_i^*}\, P_2^{\,c_i^*}, \qquad P_j \;=\; \sum_{\ell=1}^{105} t_\ell^{\,j}.
$$

Uses (statement): 5.5

Uses(notation): P^*, $\mathbf{a}^*$, $\mathbf{b}^*$, $\mathbf{c}^*$, $\mathcal{B}_d$, P_j, k (index)

Definition 11.6 (The I-value of the certificate)**.** With the data of Definition 11.5, set

$$
A^* \;:=\; \sum_{i=1}^{32}\sum_{j=1}^{32} a_i^*\, a_j^*\, \frac{(b_i^* + b_j^*)!\; G_{c_i^*+c_j^*,\,2}(105)}{\big(105 + b_i^* + b_j^* + 2c_i^* + 2c_j^*\big)!} \;\in\; \mathbb{Q},
$$

where $G_{b,2}$ is the polynomial of Definition 5.3.

Uses (statement): 11.5, 5.3

Uses(notation): $\mathbf{a}^*$, $\mathbf{b}^*$, $\mathbf{c}^*$, $G_{b,j}(x)$ (index)

Definition 11.7 (The J-value of the certificate)**.** With the data of Definition 11.5, set

$$
B^* \;:=\; \sum_{i=1}^{32}\sum_{j=1}^{32} a_i^*\, a_j^*\, \frac{\displaystyle\sum_{c'=0}^{c_i^*}\sum_{c''=0}^{c_j^*} \binom{c_i^*}{c'}\binom{c_j^*}{c''}\, \gamma\big(b_i^*, b_j^*, c_i^*, c_j^*, c', c''\big)\, G_{c'+c'',\,2}(104)}{\big(105 + b_i^* + b_j^* + 2c_i^* + 2c_j^* + 1\big)!} \;\in\; \mathbb{Q},
$$

where $G_{b,2}$ is the polynomial of Definition 5.3 and, for non-negative integers b_1, b_2, c_1, c_2 and $c' \le c_1$, $c'' \le c_2$,

$$
\gamma(b_1, b_2, c_1, c_2, c', c'') \;:=\; \frac{b_1!\; b_2!\; (2c_1 - 2c')!\; (2c_2 - 2c'')!\; \big(b_1 + b_2 + 2c_1 + 2c_2 - 2c' - 2c'' + 2\big)!}{\big(b_1 + 2c_1 - 2c' + 1\big)!\; \big(b_2 + 2c_2 - 2c'' + 1\big)!}.
$$

Uses (statement): 11.5, 5.3

Uses(notation): $\mathbf{a}^*$, $\mathbf{b}^*$, $\mathbf{c}^*$, $G_{b,j}(x)$ (index)

Lemma 11.8 (The certificate profile realises A^*)**.** *Let F_{P^*} be the extension of P^* by zero outside $\mathcal{R}_{105}$ (Definition 5.4). Then*

$$I_{105}\big(F_{P^*}\big) \;=\; A^*.$$

Uses (statement): 11.5, 11.6, 5.4, 5.6

Uses(notation): $I_k(F)$, F_P, P^*, $\mathcal{R}_k$, k (index)

Proof. By Definition 5.6, $I_{105}(F_{P^*}) = \int_{\mathcal{R}_{105}} (P^*)^2$. Expanding the square of the finite sum of Definition 11.5 and using $(1-P_1)^{b_i^*}P_2^{c_i^*}\cdot(1-P_1)^{b_j^*}P_2^{c_j^*} = (1-P_1)^{b_i^*+b_j^*}P_2^{c_i^*+c_j^*}$, we get

$$I_{105}\big(F_{P^*}\big) \;=\; \sum_{i,j} a_i^* a_j^* \int_{\mathcal{R}_{105}} (1-P_1)^{b_i^*+b_j^*} P_2^{\,c_i^*+c_j^*}.$$

Lemma 5.25 evaluates each of these simplex integrals in closed form,

$$\int_{\mathcal{R}_k} (1-P_1)^b P_2^c \;=\; \frac{b!\; G_{c,2}(k)}{(k+b+2c)!},$$

and substituting $k = 105$, $b = b_i^* + b_j^*$, $c = c_i^* + c_j^*$ reproduces exactly the double sum of Definition 11.6. □

Uses (proof): 5.25

Lemma 11.9 (The certificate profile realises B^* in every coordinate)**.** *Let F_{P^*} be the extension of P^* by zero outside $\mathcal{R}_{105}$ (Definition 5.4). Then for every $m \in \{1,\dots,105\}$,*

$$J_{105}^{(m)}\big(F_{P^*}\big) \;=\; B^*.$$

Uses (statement): 11.5, 11.7, 5.4, 5.7

Uses(notation): $J_k^{(m)}(F)$, F_P, P^*, $\mathcal{R}_k^{(m)}$, k (index)

Proof. Fix m. Since F_{P^*} vanishes off $\mathcal{R}_{105}$, for $(t_i)_{i\neq m} \in \mathcal{R}_{105}^{(m)}$ the inner integral $\int_0^\infty F_{P^*}\,dt_m$ is the integral of P^* over the segment $0 \le t_m \le 1 - \sum_{i\neq m} t_i$. The integrand is linear in the coefficient vector, so squaring and integrating over $\mathcal{R}_{105}^{(m)}$ exhibits $J_{105}^{(m)}(F_{P^*})$ as the quadratic form $\sum_{i,j} a_i^* a_j^* \Lambda_m(b_i^*, b_j^*, c_i^*, c_j^*)$, where Λ_m is the pairing of the m-th marginals of two basis monomials. Lemma 5.26 evaluates Λ_m in closed form: the inner one-dimensional integral of $(1-P_1)^b P_2^c$ in the variable t_m is computed by expanding $P_2 = t_m^2 + \sum_{i\neq m} t_i^2$ binomially and integrating each resulting Beta-type integral, and the remaining $(105-1)$-dimensional integral over $\mathcal{R}_{105}^{(m)}$ is again a simplex integral of the shape appearing in Lemma 5.25, now in dimension 104. The value obtained is

$$\Lambda_m(b_1,b_2,c_1,c_2) \;=\; \frac{\sum_{c'\le c_1}\sum_{c''\le c_2}\binom{c_1}{c'}\binom{c_2}{c''}\,\gamma(b_1,b_2,c_1,c_2,c',c'')\,G_{c'+c'',2}(104)}{(105+b_1+b_2+2c_1+2c_2+1)!},$$

with γ as in Definition 11.7. This expression does not involve m: the symmetry of P^* in $t_1,\dots,t_{105}$ means the marginal in any one coordinate is obtained from the marginal in any other by a permutation of the remaining variables, which preserves $\mathcal{R}_{105}^{(m)}$ and the Lebesgue measure. Summing over i,j gives $J_{105}^{(m)}(F_{P^*}) = B^*$ for every m. □

Uses (proof): 5.26, 5.25

Lemma 11.10 (The certificate inequality)**.** $4A^* < 105\,B^*$.

Uses (statement): 11.6, 11.7

Uses(notation): $\mathbf{a}^*$, $\mathbf{b}^*$, $\mathbf{c}^*$, $G_{b,j}(x)$ (index)

Proof. Both A^* and B^* are explicit rational numbers: each is a finite double sum of 32^2 terms whose entries are the integers a_i^* and quotients of factorials and values $G_{c,2}(105)$, $G_{c,2}(104)$, all of which are determined by Definitions 11.6 and 11.7 and computable by exact integer arithmetic. Every denominator occurring in A^* divides $(105+L+2M)!$, and every denominator occurring in B^* divides the same factorial, where $L := 2\max_i b_i^* + 1 = 23$ and $M := 2\max_i c_i^* + 1 = 11$; multiplying through by this single positive common denominator turns the asserted inequality into a comparison of two integers,

$$4 \cdot \big[(105+L+2M)!\,A^*\big] \;<\; 105 \cdot \big[(105+L+2M)!\,B^*\big],$$

both sides of which are integers built from the a_i^* by additions and multiplications. That integer comparison holds, and dividing back by the positive common denominator gives the claim. □

Proposition 11.11 (The spectral bound at $k = 105$)**.** $M_{105} > 4$.

Lean:

- `PrimeGaps.four_lt_M_verified`

Uses (statement): 5.8

Uses(notation): M_{105}, M_k, k, $\mathcal{R}_k$, $I_k(F)$, $J_k^{(m)}(F)$, $\|f\|_{L^2(X)}$ (index)

Proof. Write $F := F_{P^*}$ for the profile of Lemma 11.8. It is a polynomial multiplied by the indicator of the compact set $\mathcal{R}_{105}$, hence bounded and measurable, hence square-integrable on $\mathcal{R}_{105}$; it is symmetric in $t_1, \dots, t_{105}$ and supported in $\mathcal{R}_{105}$. Thus F is a competitor in the supremum defining M_{105} (Definition 5.8), and consequently

$$M_{105} \;\geq\; \frac{\sum_{m=1}^{105} J_{105}^{(m)}(F)}{I_{105}(F)}$$

as soon as $I_{105}(F) \neq 0$.

We first check $I_{105}(F) > 0$. By Lemma 11.8, $I_{105}(F) = A^*$, and $A^* = \int_{\mathcal{R}_{105}} (P^*)^2 \geq 0$. Suppose $A^* = 0$. Then P^* vanishes almost everywhere on $\mathcal{R}_{105}$, so $F = 0$ almost everywhere, so every marginal $\int_0^\infty F\,dt_m$ vanishes almost everywhere and $J_{105}^{(m)}(F) = 0$ for each m; by Lemma 11.9 this forces $B^* = 0$. But then Lemma 11.10 reads $0 < 0$, a contradiction. Hence $A^* > 0$.

By Lemma 11.9 the 105 marginal functionals all take the value B^*, so $\sum_{m=1}^{105} J_{105}^{(m)}(F) = 105\,B^*$. Combining with Lemma 11.10 and $A^* > 0$,

$$M_{105} \;\geq\; \frac{105\,B^*}{A^*} \;>\; \frac{4A^*}{A^*} \;=\; 4. \qquad \square$$

Uses (proof): 11.8, 11.9, 11.10, 11.5, 11.6, 11.7

11.3 The bound $H_1 \le 600$

The criterion of Proposition 10.11 states that for an admissible k-tuple, a level of distribution ϑ and a parameter $\rho > 0$ with $M_k > 2\rho/\vartheta$, there are infinitely many n for which at least $\lfloor \rho \rfloor + 1$ of the shifts $n + h_i$ are prime. To extract two primes it suffices to take $\rho = 1$, which requires

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

Bombieri–Vinogradov gives every $\vartheta < 1/2$ and nothing more, so the threshold is $\inf_{\vartheta<1/2} 2/\vartheta = 4$: the criterion yields two primes precisely when $M_k > 4$, and yields nothing when $M_k \le 4$. This is the exact threshold certified by Proposition 11.11. The next lemma converts the strict inequality $M > 4$ into an admissible choice of ϑ, with no limiting argument in ϑ.

Lemma 11.12 (A level of distribution matched to a spectral bound)**.** *Let $M \in \mathbb{R}$ with $M > 4$. Then there exists ϑ with $0 < \vartheta < 1/2$ and $\lceil \vartheta M/2 \rceil \ge 2$.*

Uses(notation): ϑ, $\lceil x \rceil$, M_k (index)

Proof. Take $\vartheta := \frac{1}{4} + \frac{1}{M}$. Since $M > 4 > 0$ we have $\vartheta > 0$, and $1/M < 1/4$ gives $\vartheta < \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$. Moreover

$$\frac{\vartheta M}{2} \;=\; \frac{1}{2}\Big(\frac{M}{4} + 1\Big) \;=\; \frac{M}{8} + \frac{1}{2} \;>\; \frac{4}{8} + \frac{1}{2} \;=\; 1,$$

using $M > 4$. A real number strictly greater than 1 has ceiling at least 2, so $\lceil \vartheta M/2 \rceil \ge 2$. □

Theorem 4 (Bounded gaps: the bound 600)**.**

$$\liminf_{n\to\infty} (p_{n+1} - p_n) \;\le\; 600.$$

Lean:

- `PrimeGaps.thm_main`

Uses (statement): 10.11, 11.3, 11.4, 11.2, 11.11, 1

Uses(notation): p_n, $\mathcal{H}_{105}$, $\operatorname{diam}(\mathcal{H})$, h_i, k, M_{105}, r_k, ϑ (index)

Proof. By Proposition 11.11 we have $M_{105} > 4$, so Lemma 11.12 applied with $M = M_{105}$ produces $\vartheta \in (0, 1/2)$ with $r_{105} = \lceil \vartheta M_{105}/2 \rceil \ge 2$. For this ϑ the Bombieri–Vinogradov theorem (Theorem 1) supplies the level of distribution ϑ required by the criterion.

Write $\mathcal{H}_{105} = \{h_1 < h_2 < \cdots < h_{105}\}$; by Lemma 11.2 this is a tuple of length $k = 105 \ge 2$, and by Lemma 11.3 it is admissible. Since $M_{105} > 4$ and ϑ was chosen with $2/\vartheta < M_{105}$, the hypothesis $M_{105} > 2\rho/\vartheta$ of Proposition 10.11 holds at $\rho = 1$. Applying that proposition at $k = 105$ with this tuple, this ϑ and $\rho = 1$ yields $p_{m+1} - p_m \le \operatorname{diam}(\mathcal{H}_{105})$ for infinitely many m.

By Lemma 11.4 that diameter is 600, so

$$\liminf_{m\to\infty} (p_{m+1} - p_m) \;\le\; 600.$$ □

Uses (proof): 11.11, 11.12, 11.2, 11.3, 11.4, 10.11, 1

12 The enlarged variational data, sieve weight, and sieve estimate

This section deforms the variational data and the sieve estimates of the standard simplex by a parameter ε with $0 \le \varepsilon < 1$. Every object below is defined for all such ε and reduces to its standard counterpart at $\varepsilon = 0$. At $\varepsilon = 0$ the estimates are the standard ones; at $\varepsilon > 0$ they give a strictly larger variational problem at the cost of one extra hypothesis on the level of distribution.

The deformation applies two independent scalings to two different pieces of the data. The region on which the profile F may live is *enlarged* from $\{\sum_i t_i \le 1\}$ to $\{\sum_i t_i \le 1+\varepsilon\}$, which increases the numerator of the Rayleigh quotient by admitting more competitors. Simultaneously, the region over which the m-th marginal is integrated is *shrunk* from $\{\sum_{i\neq m} t_i \le 1\}$ to $\{\sum_{i\neq m} t_i \le 1-\varepsilon\}$, which decreases it. The reason the two scalings are tied to the same ε is arithmetic, not geometric, and is the content of Lemma 12.20 below: the enlargement pushes the sieve divisors out to $R^{1+\varepsilon}$, which by itself would carry the moduli of the second-moment bilinear form past the Bombieri–Vinogradov threshold, and the shrinking is exactly the compensating restriction that brings them back.

12.1 The enlarged simplex and the shrunken slice

Both regions are corner regions of the same shape as $\mathcal{R}_k$ (Definition 5.2), with the total-mass bound 1 replaced by $1+\varepsilon$ and $1-\varepsilon$ respectively; at $\varepsilon = 0$ they are $\mathcal{R}_k$ and $\mathcal{R}_{k-1}$ exactly. Both are compact, hence Lebesgue measurable and of finite measure, so every integral written below over $\mathcal{T}_\varepsilon$ or $\mathcal{S}_\varepsilon$ of a continuous integrand converges absolutely.

Lemma 12.1 (Enlarged simplex is a homothety of the standard one)**.** *For every $k \ge 2$ and every $\varepsilon \ge 0$, $\mathcal{T}_\varepsilon = (1+\varepsilon)\cdot\mathcal{R}_k = \{(1+\varepsilon)\,u : u \in \mathcal{R}_k\}$.*

Uses (statement): 5.2

Uses(notation): $\mathcal{T}_\varepsilon$, ε, $\mathcal{R}_k$, k (index)

Proof. Write $c := 1+\varepsilon > 0$. If $u \in \mathcal{R}_k$ then cu has non-negative coordinates and $\sum_i (cu)_i = c\sum_i u_i \le c\cdot 1 = 1+\varepsilon$, so $cu \in \mathcal{T}_\varepsilon$. Conversely, if $t \in \mathcal{T}_\varepsilon$ put $u := c^{-1}t$; then $u_i = c^{-1}t_i \ge 0$ for each i and $\sum_i u_i = c^{-1}\sum_i t_i \le c^{-1}(1+\varepsilon) = 1$, so $u \in \mathcal{R}_k$ and $t = cu$. The two inclusions give the claimed equality. □

The same argument with $c := 1-\varepsilon > 0$ gives $\mathcal{S}_\varepsilon = (1-\varepsilon)\cdot\mathcal{R}_{k-1}$, which is used without further comment below.

Definition 12.2 (Enlarged profile class)**.** For $k \ge 2$ and $0 \le \varepsilon < 1$, the *enlarged profile class* $\mathcal{F}_\varepsilon$ is the set of $F \in C^\infty(\mathbb{R}^k;\mathbb{R})$ with $\operatorname{supp} F \subseteq \mathcal{T}_\varepsilon$.

Uses (statement): 5.2

Uses(notation): $\mathcal{T}_\varepsilon$, ε, F, k, $C_c^\infty(X)$ (index)

Members of $\mathcal{F}_\varepsilon$ are automatically compactly supported, since $\mathcal{T}_\varepsilon$ is compact; and $\mathcal{F}_0$ is exactly the class of smooth profiles supported on $\mathcal{R}_k$ used for the standard sieve. No symmetry or sign condition is imposed on F.

Definition 12.3 (Rescaling operator)**.** For $\varepsilon \ge 0$ and $F : \mathbb{R}^k \to \mathbb{R}$, define $G_\varepsilon F : \mathbb{R}^k \to \mathbb{R}$ by $(G_\varepsilon F)(u) := F\big((1+\varepsilon)\,u\big)$.

Uses(notation): ε, F, k (index)

The operator G_ε reduces the enlarged development to the standard one. By Lemma 12.1, G_ε maps $\mathcal{F}_\varepsilon$ bijectively onto $\mathcal{F}_0$: if $\operatorname{supp} F \subseteq \mathcal{T}_\varepsilon$ and $(G_\varepsilon F)(u) \neq 0$ then $(1+\varepsilon)u \in \mathcal{T}_\varepsilon = (1+\varepsilon)\cdot\mathcal{R}_k$, whence $u \in \mathcal{R}_k$; smoothness is preserved because G_ε is precomposition with a linear map. On the arithmetic side G_ε absorbs the change of the normalising truncation from $\log R$ to $(1+\varepsilon)\log R$ (Lemma 12.13); on the analytic side it turns integrals over $\mathcal{T}_\varepsilon$ into integrals over $\mathcal{R}_k$ at the price of an explicit Jacobian, as follows.

Lemma 12.4 (Change of variables on the enlarged simplex)**.** *For every $\varepsilon \geq 0$ and every continuous $F : \mathbb{R}^k \to \mathbb{R}$,*

$$\int_{\mathcal{T}_\varepsilon} F(t)^2\, dt \;=\; (1+\varepsilon)^k \int_{\mathcal{R}_k} \big(G_\varepsilon F\big)(u)^2\, du.$$

Uses (statement): 5.2, 12.3

Uses(notation): $\mathcal{T}_\varepsilon$, ε, $\mathcal{R}_k$, F, k, $\mathrm{vol}(A)$ (index)

Proof. By Lemma 12.1 the substitution $t = (1+\varepsilon)u$ maps $\mathcal{R}_k$ bijectively onto $\mathcal{T}_\varepsilon$. It is the restriction of the linear map $u \mapsto (1+\varepsilon)u$ of $\mathbb{R}^k$, whose determinant is $(1+\varepsilon)^k$; the scaling law for Lebesgue measure on $\mathbb{R}^k$ therefore gives, for any non-negative measurable H, $\int_{(1+\varepsilon)\cdot\mathcal{R}_k} H(t)\, dt = (1+\varepsilon)^k \int_{\mathcal{R}_k} H\big((1+\varepsilon)u\big)\, du$. Apply this with $H = F^2$ and use $F\big((1+\varepsilon)u\big)^2 = (G_\varepsilon F)(u)^2$. □

Uses (proof): 12.1

12.2 The enlarged functionals

The denominator of the variational problem is *not* deformed. For $F \in \mathcal{F}_\varepsilon$ the quantity $\int_{\mathcal{T}_\varepsilon} F^2$ equals $\int_{[0,\infty)^k} F^2 = I_k(F)$, because F vanishes off $\mathcal{T}_\varepsilon$; so I_k of Definition 5.6 serves unchanged for every ε, and the deformation enters only through the widened class of competitors. The numerator, by contrast, is deformed: the outer integration is over the shrunken slice.

Lemma 12.5 (The enlarged marginal in rescaled coordinates)**.** *Let $0 \leq \varepsilon < 1$, let $F \in \mathcal{F}_\varepsilon$, and let $m \in \{1, \dots, k\}$. Then*

$$J_\varepsilon^{(m)}(F) \;=\; (1+\varepsilon)^{k+1} \int_{\mathcal{R}_{k-1}} \mathbf{1}\Big[\textstyle\sum_{i\neq m} u_i \leq \frac{1-\varepsilon}{1+\varepsilon}\Big] \Big(\int_0^\infty (G_\varepsilon F)(u)\, du_m\Big)^2\, du_1 \cdots du_{m-1}\, du_{m+1} \cdots du_k.$$

Uses (statement): 5.7, 12.3, 12.2, 5.2

Uses(notation): $J_\varepsilon^{(i)}(F)$, ε, $\mathcal{S}_\varepsilon$, $\mathcal{R}_k$, F, k, T_m (index)

Proof. Substitute $t = (1+\varepsilon)u$ in both the inner and the outer integral. In the inner integral over the single variable t_m this contributes one factor $1+\varepsilon$, so $\int_0^\infty F(t)\, dt_m = (1+\varepsilon)\int_0^\infty (G_\varepsilon F)(u)\, du_m$; squaring gives $(1+\varepsilon)^2$. In the outer integral over the $k-1$ variables $(t_i)_{i\neq m}$ it contributes $(1+\varepsilon)^{k-1}$. Together these give the factor $(1+\varepsilon)^{k+1}$.

It remains to identify the domain. Under $t = (1+\varepsilon)u$ the constraint $(t_i)_{i\neq m} \in \mathcal{S}_\varepsilon$, that is $t_i \geq 0$ for $i \neq m$ and $\sum_{i\neq m} t_i \leq 1-\varepsilon$, becomes $u_i \geq 0$ for $i \neq m$ and $\sum_{i\neq m} u_i \leq (1-\varepsilon)/(1+\varepsilon)$. Since $(1-\varepsilon)/(1+\varepsilon) \leq 1$, this domain is the subset of $\mathcal{R}_{k-1}$ cut out by the displayed indicator, which is the stated right-hand side. □

Lemma 12.5 exhibits the enlarged marginal as a *restricted* standard marginal: after rescaling by G_ε the profile lives on $\mathcal{R}_k$, and the only remaining trace of the enlargement is the sharp cutoff $\sum_{i\neq m} u_i \le (1-\varepsilon)/(1+\varepsilon)$ on the outer variables. That cutoff is the image under the $\lambda \mapsto y$ transform of the divisor restriction $\prod_{i\neq m} d_i \le R^{1-\varepsilon}$ imposed in Definition 12.16.

Lemma 12.6 (Recovery of the standard data at $\varepsilon = 0$)**.** *Let F be compactly supported and Riemann-integrable with* $\operatorname{supp} F \subseteq \mathcal{R}_k$*. Then $J_0^{(m)}(F) = J_k^{(m)}(F)$ for every $m \in \{1, \dots, k\}$.*

Uses (statement): 5.7, 5.2

Uses(notation): $J_\varepsilon^{(i)}(F)$, $J_k^{(m)}(F)$, $\mathcal{S}_\varepsilon$, $\mathcal{R}_k$, F, k (index)

Proof. By Definition 5.2 with $\varepsilon = 0$ we have $\mathcal{S}_0 = \mathcal{R}_{k-1} = \mathcal{R}_k^{(m)}$, so the two outer domains agree. The two inner integrands also agree: the standard $J_k^{(m)}$ integrates F in the variable t_m over $[0, \infty)$, and so does $J_0^{(m)}$. (The hypothesis $\operatorname{supp} F \subseteq \mathcal{R}_k$ is what makes this identification legitimate: it forces F to vanish for t_m outside a bounded range, so both inner integrals converge and are computed by the same one-dimensional integral. Without a support hypothesis the value of $J_0^{(m)}$ genuinely depends on the values of F off $\mathcal{R}_k$, which the standard marginal does not see.) The outer integrals therefore have equal integrands on equal domains. □

Uses (proof): 5.2

Consequently, for such F with $I_k(F) \neq 0$ one has $\mathcal{Q}_0(F) = \big(\sum_m J_k^{(m)}(F)\big)/I_k(F)$: at $\varepsilon = 0$ the enlarged witness ratio is the standard Rayleigh quotient, and the variational problem introduced next is the standard one.

The supremum is finite: an application of the Cauchy–Schwarz inequality in the variable t_m, whose range is contained in an interval of length $1+\varepsilon$, gives $J_\varepsilon^{(m)}(F) \le 2(1+\varepsilon)\|F\|^2_{L^2(\mathcal{T}_\varepsilon)}$ for each m, whence $0 \le M_{k,\varepsilon} \le 2(1+\varepsilon)k$. Taking $\varepsilon = 0$, Lemma 12.6 identifies $M_{k,0}$ with M_k of Definition 5.8, and the bound with $M_k \le 2k$. In this notation the criterion to be met is $M_{k,\varepsilon} > 4$ for a single pair (k, ε); the first case meets it at $(105, 0)$ and the second at $(50, \frac{1}{25})$.

Throughout what follows, $M_{k,\varepsilon}$ plays a purely nominal role. Every estimate is stated and used in terms of the ratio $\mathcal{Q}_\varepsilon(F)$ of a concrete competitor F; the supremum is never evaluated, and appears only so that the criterion can be phrased as a property of (k, ε) alone.

Lemma 12.7 (A witness certifies the enlarged constant)**.** *Let $0 \le \varepsilon < 1$ and suppose $F \in L^2(\mathcal{T}_\varepsilon)$ satisfies*

$$4\,\|F\|^2_{L^2(\mathcal{T}_\varepsilon)} \;<\; \sum_{m=1}^{k} J_\varepsilon^{(m)}(F).$$

Then $M_{k,\varepsilon} > 4$.

Uses (statement): 5.8, 5.7

Uses(notation): $J_\varepsilon^{(i)}(F)$, $\mathcal{T}_\varepsilon$, ε, k, $\|f\|_{L^2(X)}$, M_k (index)

Proof. Each $J_\varepsilon^{(m)}$ is non-negative, being an integral of a square; so the right-hand side is non-negative, and the strict inequality forces $\|F\|_{L^2(\mathcal{T}_\varepsilon)} \neq 0$ (otherwise the left side would be 0 and the right side would be 0 as well, since $J_\varepsilon^{(m)}$ vanishes on the zero element of L^2). Dividing the hypothesis by $\|F\|^2_{L^2(\mathcal{T}_\varepsilon)} > 0$ gives $4 < \sum_m J_\varepsilon^{(m)}(F)/\|F\|^2$, and the right-hand side is at most $M_{k,\varepsilon}$ because F is one of the competitors in Definition 5.8. □

Lemma 12.8 (L^2-continuity of the enlarged witness ratio)**.** *Let $0 \le \varepsilon < 1$, let $F_1 \in L^2(\mathcal{T}_\varepsilon)$ with $I_k(F_1) \neq 0$, and let $\eta > 0$. Then there is $\zeta > 0$ such that every $F_2 \in L^2(\mathcal{T}_\varepsilon)$ with $\|F_2 - F_1\|_{L^2(\mathcal{T}_\varepsilon)} < \zeta$ satisfies $I_k(F_2) \neq 0$ and $\mathcal{Q}_\varepsilon(F_2) > \mathcal{Q}_\varepsilon(F_1) - \eta$.*

Uses (statement): 5.8, 5.2

Uses(notation): $\mathcal{Q}_\varepsilon(F)$, $\mathcal{T}_\varepsilon$, ε, $I_k(F)$, $\|f\|_{L^2(X)}$, k (index)

Proof. On $L^2(\mathcal{T}_\varepsilon)$ the denominator of $\mathcal{Q}_\varepsilon$ is $F \mapsto \|F\|^2_{L^2(\mathcal{T}_\varepsilon)}$, which is continuous (Lemma 5.14), and the numerator is $F \mapsto \sum_m J_\varepsilon^{(m)}(F)$, a finite sum of continuous functionals (Lemma 5.15 applied on $\mathcal{T}_\varepsilon$; each $J_\varepsilon^{(m)}$ is the squared norm of the composite of the partial-integration map $L^2(\mathcal{T}_\varepsilon) \to L^2(\mathbb{R}^{k-1})$ with the restriction map onto $\mathcal{S}_\varepsilon$, both bounded linear). A quotient of continuous real functions is continuous at every point where the denominator is non-zero; since $\|F_1\|^2 = I_k(F_1) \neq 0$, the map $\mathcal{Q}_\varepsilon$ is continuous at F_1. Continuity at F_1 furnishes $\zeta_1 > 0$ with $|\mathcal{Q}_\varepsilon(F_2) - \mathcal{Q}_\varepsilon(F_1)| < \eta$ whenever $\|F_2 - F_1\| < \zeta_1$; shrinking ζ_1 further so that $\|F_2 - F_1\| < \zeta$ forces $\|F_2\| \ge \|F_1\| - \zeta > 0$, hence $I_k(F_2) \neq 0$, yields the claim. □

Uses (proof): 5.14, 5.15

12.3 The certificate-admissible level

Two of the estimates below — the aggregate Chinese remainder error of Lemma 12.14 and the Bombieri–Vinogradov modulus bound of Lemma 12.20 — require the enlarged divisor range $R^{1+\varepsilon}$ to stay within the ranges those estimates tolerate. Both requirements, together with the requirement that the resulting sieve criterion be satisfiable at all, are collected in the following single condition on the level ϑ.

Definition 12.9 (Certificate-admissible level)**.** Let $\varepsilon \ge 0$ and $Q \in \mathbb{R}$. A real number ϑ is (ε, Q)*-certificate-admissible* if

(i) $0 < \vartheta < \frac{1}{2}$;

(ii) $2/\vartheta < Q$;

(iii) $1 + \varepsilon < 1/\vartheta$;

and the primes have level of distribution ϑ.

Uses (statement): 3.21

Uses(notation): ϑ, ε (index)

Both ε and the threshold Q are parameters of the definition, not constants: nothing in the theory below depends on a particular numerical value of either.

The three clauses are consumed in three different places.

Clause (iii), equivalently $(1 + \varepsilon)\vartheta < 1$, is the *enlargement room*. It makes the enlarged truncation harmless in the first moment: with $R = N^{\vartheta/2-\delta}$ one has $R^{2(1+\varepsilon)} = N^{(1+\varepsilon)(\vartheta-2\delta)}$, and (iii) gives $(1 + \varepsilon)(\vartheta - 2\delta) < (1 + \varepsilon)\vartheta < 1$, so the aggregate error of Lemma 12.14 is $o(N)$ and Proposition 12.15 needs no equidistribution input at all. Clause (iii) also guarantees $(1+\varepsilon)(\vartheta/2-\delta) < \frac{1}{2}$, i.e. $R^{1+\varepsilon} < N^{1/2}$, which lets the enlarged weight be rewritten as a standard weight at a legitimate truncation level (Lemma 12.13).

Clause (i) is the *Bombieri–Vinogradov room*. The moduli occurring in the second-moment bilinear form are bounded by R^{a+b} where a, b are the exponents of the outer divisor products

of the two weights; the support split of Definition 12.16 arranges $a+b\le 2$, so the moduli are at most $R^2 = N^{\vartheta-2\delta}$, and (i) puts this below $N^{1/2}$ (Lemma 12.20). Note that without the split the bound would only be $R^{2(1+\varepsilon)}$, which (iii) keeps below N but *not* below $N^{1/2}$; the split is therefore the mechanism that makes the scaling compatible with Bombieri–Vinogradov.

Clause (ii) is the *sieve criterion*. It makes the positivity window of Lemma 12.25 non-empty, and hence gives the value a numerical $\mathcal{Q}_\varepsilon$ must exceed. Since ϑ may be taken as close to $\frac12$ as one likes — but not equal to it — clause (ii) is satisfiable for some ϑ exactly when $Q>4$, which is why the target of the enlarged variational problem is $M_{k,\varepsilon}>4$ and not $M_{k,\varepsilon}\ge 4$.

Clause (iii) is implied by clause (i) whenever $\varepsilon\le 1$: then $1+\varepsilon\le 2<1/\vartheta$. It is nevertheless stated separately because it is the clause the scaling consumes, and because it is the one that would bind first if ε were pushed towards or beyond 1.

12.4 The enlarged sieve weight and its support

Definition 12.10 (ε-permissible support)**.** Let $\varepsilon\ge 0$. A function $\lambda:\mathbb{N}^k\to\mathbb{R}$ has *ε-permissible support* if $\lambda(d_1,\dots,d_k)=0$ whenever one of the following fails:

(i) $d_i\ge 1$ for every i, and $\prod_{i=1}^k d_i\le R^{1+\varepsilon}$;

(ii) $\gcd\bigl(\prod_{i=1}^k d_i,\,W\bigr)=1$;

(iii) $\prod_{i=1}^k d_i$ is squarefree.

Uses (statement): 3.37, 3.29, 3.39

Uses(notation): $\lambda(d_1,\dots,d_k)$, R, W, ε, k, $\mathbf{d}$, (a,b) (index)

This is Definition 3.39 with the single bound R relaxed to $R^{1+\varepsilon}$; at $\varepsilon=0$ the two conditions coincide.

Definition 12.11 (Enlarged sieve weight)**.** Let $\varepsilon\ge 0$ and $F\in\mathcal{F}_\varepsilon$. Define $\lambda_\varepsilon:\mathbb{N}^k\to\mathbb{R}$ by

$$\lambda_\varepsilon(d_1,\dots,d_k)\;:=\;\Bigl(\prod_{i=1}^k\mu(d_i)\,d_i\Bigr)\sum_{\substack{r_1,\dots,r_k\\ d_i\mid r_i\ \forall i\\ \prod_i r_i\le R^{1+\varepsilon}\\ \gcd(\prod_i r_i,W)=1,\ \mu(\prod_i r_i)^2=1}}\frac{1}{\prod_{i=1}^k\varphi(r_i)}\,F\Bigl(\frac{\log r_1}{\log R},\dots,\frac{\log r_k}{\log R}\Bigr).$$

Uses (statement): 12.2, 3.37, 3.29, 5.11

Uses(notation): λ_ε, $\mathcal{T}_\varepsilon$, ε, R, W, μ, φ, k, $\mathbf{d}$, (a,b) (index)

Two features of this definition are used below, and neither is forced by the shape of the formula. First, the $\mathbf{r}$-sum ranges over the enlarged region $\prod_i r_i\le R^{1+\varepsilon}$ rather than $\prod_i r_i\le R$: this is the only change relative to the standard F-derived weight of Definition 5.11. Second, the argument of F is normalised by $\log R$ with the *original* $R=N^{\vartheta/2-\delta}$, not by $\log R^{1+\varepsilon}$. Consequently the point $\bigl(\frac{\log r_1}{\log R},\dots,\frac{\log r_k}{\log R}\bigr)$ lies in $\mathcal{T}_\varepsilon$ exactly when $\prod_i r_i\le R^{1+\varepsilon}$: the sieve range and the support of F match, so that $\mathcal{T}_\varepsilon$ is the correct region.

Lemma 12.12 (The enlarged weight is ε-permissible)**.** *For every $\varepsilon\ge 0$ and every $F\in\mathcal{F}_\varepsilon$, the weight λ_ε has ε-permissible support.*

Uses (statement): 12.11, 12.10

Uses(notation): λ_ε, ε, R, W, k (index)

Proof. Fix $\mathbf{d}$ with $\lambda_\varepsilon(\mathbf{d}) \neq 0$. If some $d_i = 0$ then the factor $\prod_i \mu(d_i)d_i$ vanishes, contradiction; so $d_i \geq 1$ for all i. Since $\lambda_\varepsilon(\mathbf{d}) \neq 0$, the defining sum is non-empty, so there is at least one $\mathbf{r}$ with $d_i \mid r_i$ for every i satisfying the three summation constraints. From $\prod_i r_i \leq R^{1+\varepsilon}$ and $d_i \mid r_i$ (hence $d_i \leq r_i$) we get $\prod_i d_i \leq \prod_i r_i \leq R^{1+\varepsilon}$, which is clause (i). From $\gcd(\prod_i r_i, W) = 1$ and $\prod_i d_i \mid \prod_i r_i$ we get $\gcd(\prod_i d_i, W) = 1$, which is clause (ii). From $\prod_i r_i$ squarefree and $\prod_i d_i \mid \prod_i r_i$ we get $\prod_i d_i$ squarefree, which is clause (iii). □

Lemma 12.13 (The enlarged weight is a standard weight at a shifted level)**.** *Let $\varepsilon \geq 0$, let $F \in \mathcal{F}_\varepsilon$, and let ϑ, δ satisfy $0 < \vartheta < 1$, $0 < \delta < \vartheta/2$ and $1 + \varepsilon < 1/\vartheta$. Let Θ and Δ be the real numbers satisfying*

$$\Theta = \tfrac{1}{2} + (1+\varepsilon)\big(\tfrac{\vartheta}{2} - \delta\big), \qquad \Delta = \tfrac{1}{4} - \tfrac{1}{2}(1+\varepsilon)\big(\tfrac{\vartheta}{2} - \delta\big).$$

Then $0 < \Theta < 1$, $0 < \Delta < \Theta/2$, the truncation attached to (Θ, Δ) is $N^{\Theta/2-\Delta} = R^{1+\varepsilon}$, and λ_ε is the standard F-derived weight of Definition 5.11 at truncation $R^{1+\varepsilon}$ applied to the rescaled profile $G_\varepsilon F$.

Uses (statement): 12.11, 12.3, 5.11, 3.37, 12.2

Uses(notation): λ_ε, ε, R, ϑ, δ, N, k (index)

Proof. Write $\alpha := (1+\varepsilon)(\vartheta/2 - \delta)$, so $\Theta = \frac{1}{2} + \alpha$ and $\Delta = \frac{1}{4} - \frac{\alpha}{2}$, and hence $\Theta/2 - \Delta = \alpha$ identically. We have $\alpha > 0$ because $\delta < \vartheta/2$; and $\alpha < (1+\varepsilon)\vartheta/2 < \frac{1}{2}$ by $1 + \varepsilon < 1/\vartheta$. From $0 < \alpha < \frac{1}{2}$ the two parameter constraints $0 < \Theta < 1$ and $0 < \Delta < \Theta/2$ follow by inspection. Also $N^{\Theta/2-\Delta} = N^\alpha = \big(N^{\vartheta/2-\delta}\big)^{1+\varepsilon} = R^{1+\varepsilon}$.

For the last assertion, compare Definition 12.11 with Definition 5.11 at truncation $R' := R^{1+\varepsilon}$. The two Möbius prefactors and the two summation ranges agree verbatim, the latter because the standard weight at truncation R' sums over $\prod_i r_i \leq R' = R^{1+\varepsilon}$ subject to the same coprimality and squarefreeness conditions. Only the argument of the profile differs: Definition 5.11 at truncation R' evaluates the profile at $\big(\frac{\log r_i}{\log R'}\big)_i$, whereas Definition 12.11 evaluates F at $\big(\frac{\log r_i}{\log R}\big)_i$. Since $\log R' = (1+\varepsilon)\log R$, we have $\big(\frac{\log r_i}{\log R}\big)_i = (1+\varepsilon)\big(\frac{\log r_i}{\log R'}\big)_i$, so by Definition 12.3

$$F\Big(\tfrac{\log r_1}{\log R}, \dots, \tfrac{\log r_k}{\log R}\Big) \;=\; (G_\varepsilon F)\Big(\tfrac{\log r_1}{\log R'}, \dots, \tfrac{\log r_k}{\log R'}\Big),$$

which is exactly the profile value the standard weight at truncation R' would use for $G_\varepsilon F$. □

Lemma 12.13 shows the enlarged sieve weight is not a new object but the standard weight, evaluated at a different truncation level and applied to a rescaled profile. Every estimate of the standard development that is uniform in the level parameters therefore transfers to λ_ε without new work, and the only new arguments are those where the enlargement meets a fixed threshold — the Bombieri–Vinogradov threshold in the second moment.

12.5 The enlarged S_1 asymptotic

Lemma 12.14 (Aggregate CRT error for ε-permissible weights)**.** *Let $\varepsilon \geq 0$ and let λ, λ' have ε-permissible support with $|\lambda|, |\lambda'| \leq 1$ pointwise, and suppose $(1+\varepsilon)\vartheta < 1$. Then*

$$\Bigg|\sum_{\mathbf{d},\mathbf{e}} \lambda(\mathbf{d})\,\lambda'(\mathbf{e}) \Bigg(\sum_{\substack{N < n \leq 2N \\ n \equiv v_0 \pmod W \\ [d_i,e_i] \mid n + h_i \ \forall i}} 1 \;-\; \frac{N}{W \prod_i [d_i, e_i]}\, \mathbf{1}\big[W, [d_1,e_1], \dots, [d_k,e_k] \text{ pairwise coprime}\big] \Bigg)\Bigg|$$

$$\ll_k \; R^{2(1+\varepsilon)} (\log R)^{2(k-1)} \;=\; o(N).$$

Uses (statement): 12.10, 3.37

Uses(notation): ε, R, W, ϑ, N, v_0, $\mathbf{d}$, $\mathbf{e}$, $[a,b]$, h_i, $o(g)$, $f \ll g$, τ_r (index)

Proof. Write $\rho(\mathbf{d},\mathbf{e})$ for the parenthesised difference. By Lemma 8.7 the inner counting sum equals $N/(W\prod_i[d_i,e_i]) + O(1)$ when $W,[d_1,e_1],\dots,[d_k,e_k]$ are pairwise coprime and equals 0 otherwise, so in either branch $|\rho(\mathbf{d},\mathbf{e})| \le C_k$ for a constant depending only on k. With $|\lambda|,|\lambda'| \le 1$ this gives

$$\Big|\sum_{\mathbf{d},\mathbf{e}} \lambda(\mathbf{d})\lambda'(\mathbf{e})\rho(\mathbf{d},\mathbf{e})\Big| \;\le\; C_k \cdot \#\big\{(\mathbf{d},\mathbf{e}) : \lambda(\mathbf{d})\lambda'(\mathbf{e}) \neq 0\big\}.$$

By ε-permissibility each factor is supported on tuples with $\prod_i d_i \le R^{1+\varepsilon}$ and $\prod_i d_i$ squarefree. Writing $D = \prod_i d_i$, the number of ordered k-tuples with $\prod_i d_i = D$ is $\tau_k(D)$, so the number of admissible $\mathbf{d}$ is at most $\sum_{D\le R^{1+\varepsilon}} \tau_k(D) \ll_k R^{1+\varepsilon}(\log R)^{k-1}$ by the standard divisor-sum bound; squaring gives the pair count.

Finally $R = N^{\vartheta/2-\delta}$ gives $R^{2(1+\varepsilon)} = N^{(1+\varepsilon)(\vartheta-2\delta)}$, and $(1+\varepsilon)(\vartheta-2\delta) < (1+\varepsilon)\vartheta < 1$; the logarithmic factor is $N^{o(1)}$, so the whole bound is $o(N)$. □

Uses (proof): 8.7

Proposition 12.15 (Enlarged S_1 asymptotic)**.** *Let $k \ge 2$, let $\mathcal{H}$ be admissible, let $\varepsilon \ge 0$, and let $F \in \mathcal{F}_\varepsilon$. Let $0 < \vartheta < 1$, $0 < \delta < \vartheta/2$ and $1+\varepsilon < 1/\vartheta$. Then, with $R = N^{\vartheta/2-\delta}$ the* unenlarged *truncation,*

$$S_1(\lambda_\varepsilon) \;=\; \big(1 + O(D_0^{-1})\big)\,\frac{\varphi(W)^k\, N\,(\log R)^k}{W^{k+1}} \int_{\mathcal{T}_\varepsilon} F(t)^2\,dt$$

as $N \to \infty$, the implied constant depending only on k, $\mathcal{H}$ and $F_{\max}(G_\varepsilon F)$.

Uses (statement): 12.11, 4.14, 12.2, 5.2, 3.37

Uses(notation): λ_ε, $S_1(\lambda)$, $\mathcal{T}_\varepsilon$, ε, R, W, φ, N, ϑ, δ, D_0, $F_{\max}$, k (index)

Proof. By Lemma 12.13 the parameters (Θ,Δ) are legitimate level parameters and λ_ε is the standard F-derived weight at truncation $R' = R^{1+\varepsilon} = N^{\Theta/2-\Delta}$ applied to $G_\varepsilon F$. The rescaled profile $G_\varepsilon F$ is smooth and, by Lemma 12.1, supported on $\mathcal{R}_k$; it is therefore an admissible input to the standard smooth first-moment asymptotic Lemma 8.25, whose hypotheses involve only k, $\mathcal{H}$ and the level parameters. That lemma gives

$$S_1(\lambda_\varepsilon) \;=\; \big(1 + O(D_0^{-1})\big)\,\frac{\varphi(W)^k\, N\,(\log R')^k}{W^{k+1}} \int_{\mathcal{R}_k} \big(G_\varepsilon F\big)(u)^2\,du.$$

(The equidistribution input needed here is only the aggregate CRT error of Lemma 12.14, which is $o(N)$ by $(1+\varepsilon)\vartheta < 1$; no Bombieri–Vinogradov estimate is used in the first moment.)

It remains to restore the original normalisation. Since $\log R' = (1+\varepsilon)\log R$ we have $(\log R')^k = (1+\varepsilon)^k(\log R)^k$, and by Lemma 12.4 $\int_{\mathcal{R}_k}(G_\varepsilon F)^2 = (1+\varepsilon)^{-k}\int_{\mathcal{T}_\varepsilon} F^2$. The two factors $(1+\varepsilon)^{\pm k}$ cancel, and the same cancellation applies to the error term, giving the stated form with the unenlarged $\log R$ and the integral over $\mathcal{T}_\varepsilon$. □

Uses (proof): 12.12, 12.13, 12.14, 12.1, 12.4, 8.25

The cancellation of $(1+\varepsilon)^{\pm k}$ makes the enlarged first moment identical in form to the standard one, with $I_k(F)$ now computed over the larger region. Since I_k is the denominator of the variational problem, enlarging the simplex costs nothing in the denominator; the entire effect of the enlargement is in the numerator, treated next.

12.6 The support split and the enlarged S_2 lower bound

Definition 12.16 (Support-split weight)**.** Fix $m \in \{1, \dots, k\}$ and $\varepsilon \geq 0$. Set

$$\lambda_{1,m}(\mathbf{d}) \;:=\; \lambda_\varepsilon(\mathbf{d})\,\mathbf{1}\Big[\textstyle\prod_{i\neq m} d_i \leq R^{1-\varepsilon}\Big].$$

Uses (statement): 12.11, 3.37

Uses(notation): $\lambda_{1,m}$, λ_ε, ε, R, k, $\mathbf{d}$ (index)

Definition 12.17 (Complementary weight)**.** With the notation of Definition 12.16, set $\lambda_{2,m} := \lambda_\varepsilon - \lambda_{1,m}$.

Uses (statement): 12.11, 12.16

Uses(notation): λ_ε, $\lambda_{1,m}$, k (index)

Explicitly, $\lambda_{2,m}(\mathbf{d}) = \lambda_\varepsilon(\mathbf{d})$ when $\prod_{i\neq m} d_i > R^{1-\varepsilon}$ and 0 otherwise; in particular $\lambda_{2,m}$ inherits from λ_ε the outer bound $\prod_{i\neq m} d_i \leq \prod_i d_i \leq R^{1+\varepsilon}$. The two outer exponents are therefore $1-\varepsilon$ for $\lambda_{1,m}$ and $1+\varepsilon$ for $\lambda_{2,m}$, and the point of the split is the identity $(1-\varepsilon)+(1+\varepsilon) = 2$.

Notation 12.18 (Split divisor sums)**.** For $\ell \in \{1, 2\}$ and n in the sieve range, set

$$A_{\ell,m}(n) \;:=\; \sum_{\substack{\mathbf{d} \\ d_i \mid n+h_i \ \forall i}} \lambda_{\ell,m}(\mathbf{d}),$$

so that the inner divisor sum defining w_n is $A_{1,m}(n) + A_{2,m}(n)$.

Uses (statement): 12.16, 12.17, 4.12

Uses(notation): $A_{\ell,m}(n)$, $\lambda_{1,m}$, w_n, h_i, $\mathbf{d}$, k (index)

Lemma 12.19 (Pointwise lower bound)**.** *For every n in the sieve range, $w_n \geq A_{1,m}(n)^2 + 2\,A_{1,m}(n)\,A_{2,m}(n)$.*

Uses (statement): 12.18, 4.12

Uses(notation): $A_{\ell,m}(n)$, w_n (index)

Proof. Put $x_1 := A_{1,m}(n)$ and $x_2 := A_{2,m}(n)$. By Notation 12.18, $w_n = (x_1 + x_2)^2 = x_1^2 + 2x_1x_2 + x_2^2 \geq x_1^2 + 2x_1x_2$, since $x_2^2 \geq 0$. □

Lemma 12.19 discards the one term, $A_{2,m}(n)^2$, that the scaling makes unavailable: it is a bilinear form in $(\lambda_{2,m}, \lambda_{2,m})$, whose outer exponents sum to $2(1+\varepsilon) > 2$, and whose moduli therefore exceed the Bombieri–Vinogradov threshold. The two terms that are kept have outer exponent sums $2(1-\varepsilon) \leq 2$ and $(1-\varepsilon)+(1+\varepsilon) = 2$ respectively, and are handled by the following bound.

Lemma 12.20 (Bombieri–Vinogradov modulus bound)**.** *Fix $m \in \{1, \dots, k\}$ and $A > 0$. Let ϑ satisfy Definition 12.9(i) and (iii) and have level of distribution ϑ, and let $a, b \geq 0$ with $a+b \leq 2$. Let λ, λ' have ε-permissible support and be supported additionally on $\{\prod_{i\neq m} d_i \leq R^a\}$ and $\{\prod_{i\neq m} d_i \leq R^b\}$ respectively. Then*

$$\sum_{N<n\leq 2N} \chi_{\mathbb{P}}(n+h_m) \sum_{\mathbf{d},\mathbf{e}} \lambda(\mathbf{d})\,\lambda'(\mathbf{e}) \prod_{i=1}^{k} \mathbf{1}\big[[d_i, e_i] \mid n+h_i\big]$$

differs from its expected main term by $O_A\big(N/(\log N)^A\big)$.

Uses (statement): 12.10, 12.9, 3.37

Uses(notation): ε, R, W, ϑ, $\chi_{\mathbb{P}}$, h_i, N, $\mathbf{d}$, $\mathbf{e}$, $[a,b]$, $O(g)$, k (index)

Proof. The factor $\chi_{\mathbb{P}}(n+h_m)$ restricts n to integers with $n+h_m$ prime, so any divisor d_m of $n+h_m$ lies in $\{1, n+h_m\}$. By ε-permissibility, $d_m \le \prod_i d_i \le R^{1+\varepsilon} = N^{(1+\varepsilon)(\vartheta/2-\delta)}$, and clause (iii) of Definition 12.9 gives $(1+\varepsilon)(\vartheta/2-\delta) < (1+\varepsilon)\vartheta/2 < \frac{1}{2}$, so $d_m < N^{1/2} < N < n+h_m$ for large N; hence $d_m = 1$, and likewise $e_m = 1$. The m-th coordinate therefore contributes only the trivial divisor, and the effective modulus of the bilinear form is $q = \prod_{i\neq m}[d_i, e_i]$.

For $(\mathbf{d}, \mathbf{e})$ in the joint support, $[d_i, e_i] \le d_i e_i$ and the extra outer support hypotheses give $q \le \prod_{i\neq m} d_i \cdot \prod_{i\neq m} e_i \le R^a R^b = R^{a+b} \le R^2 = N^{\vartheta-2\delta}$; this is Lemma 9.31 with $B_1 = R^a$ and $B_2 = R^b$, whose point is precisely that the modulus depends on the two outer truncations only through their product, so that $a+b \le 2$ suffices even when one of a, b exceeds 1. By clause (i) of Definition 12.9, $\vartheta < \frac{1}{2}$, so $\vartheta - 2\delta < \frac{1}{2}$, and Lemma 3.38 supplies $\eta_R > 0$ with $qW \le N^{1/2-\eta_R}$ for all large N. Every modulus occurring in the bilinear form thus lies below the Bombieri–Vinogradov threshold, and the Cauchy–Schwarz-plus-Bombieri–Vinogradov bilinear estimate of Lemma 9.9 bounds the aggregate error by $O_A(N/(\log N)^A)$ for every $A > 0$. (The pointwise divisor-weight cost of size $(\log N)^{O(k)}$ incurred in applying that estimate is absorbed by requesting the exponent $A+2k$ in place of A from the same level-of-distribution hypothesis.) □

Uses (proof): 3.38, 9.31, 9.9

Lemma 12.21 (Retained bilinear forms have negligible error)**.** *Fix $m \in \{1,\dots,k\}$ and $A > 0$, let ϑ satisfy Definition 12.9(i) and (iii) and have level of distribution ϑ, and let $F \in \mathcal{F}_\varepsilon$. Then for each $\lambda' \in \{\lambda_{1,m}, \lambda_{2,m}\}$ the sum*

$$\sum_{N<n\le 2N} \chi_{\mathbb{P}}(n+h_m)\, A_{1,m}(n)\, A'_m(n), \qquad A'_m(n) = \sum_{d_i \mid n+h_i \forall i} \lambda'(\mathbf{d}),$$

differs from its expected main term by $O_A\big(N/(\log N)^A\big)$.

Uses (statement): 12.16, 12.17, 12.18, 12.20, 12.9

Uses(notation): $A_{\ell,m}(n)$, $\lambda_{1,m}$, λ_ε, ε, R, ϑ, $\chi_{\mathbb{P}}$, h_i, N, $O(g)$, k (index)

Proof. Expanding the product of the two divisor sums exhibits the displayed quantity as the bilinear form of Lemma 12.20 with the pair $(\lambda_{1,m}, \lambda')$. Both weights are ε-permissible: $\lambda_{1,m}$ and $\lambda_{2,m}$ are obtained from λ_ε by multiplying by an indicator, respectively by subtracting such a product, and ε-permissibility (Definition 12.10) is preserved by both operations; λ_ε itself is ε-permissible by Lemma 12.12.

It remains to check the outer exponents. By Definition 12.16, $\lambda_{1,m}$ is supported on $\prod_{i\neq m} d_i \le R^{1-\varepsilon}$, so $a = 1-\varepsilon$. For $\lambda' = \lambda_{1,m}$ we may take $b = 1-\varepsilon$, and $a+b = 2(1-\varepsilon) \le 2$. For $\lambda' = \lambda_{2,m}$ we may take $b = 1+\varepsilon$: indeed $\lambda_{2,m}(\mathbf{d}) \neq 0$ implies $\lambda_\varepsilon(\mathbf{d}) \neq 0$, whence $\prod_{i\neq m} d_i \le \prod_i d_i \le R^{1+\varepsilon}$; and then $a+b = (1-\varepsilon)+(1+\varepsilon) = 2$. In both cases $a+b \le 2$ and Lemma 12.20 applies. □

Uses (proof): 12.20, 12.12, 12.10

Lemma 12.22 (Main term of the retained square)**.** *Fix $m \in \{1,\dots,k\}$, let $\varepsilon \ge 0$, let $F \in \mathcal{F}_\varepsilon$, and let ϑ satisfy Definition 12.9(i) and (iii) and have level of distribution ϑ. Then, for every $A > 0$,*

$$\sum_{N<n\le 2N} \chi_{\mathbb{P}}(n+h_m)\, A_{1,m}(n)^2 \;=\; \big(1+o(1)\big)\, \frac{\varphi(W)^k\, N\, (\log R)^{k+1}}{W^{k+1} \log N}\, J_\varepsilon^{(m)}(F) \;+\; O_A\Big(\frac{N}{(\log N)^A}\Big).$$

Uses (statement): 12.16, 12.18, 5.7, 12.21, 12.9

Uses(notation): $A_{\ell,m}(n)$, $\lambda_{1,m}$, $J_\varepsilon^{(i)}(F)$, ε, R, W, ϑ, $\chi_{\mathbb{P}}$, h_i, N, φ, $o(g)$, $O(g)$, k (index)

Proof. By Lemma 12.21 with $\lambda' = \lambda_{1,m}$, replacing the sum by its arithmetic main term commits an error $O_A(N/(\log N)^A)$; here the level of distribution enters, through Lemma 12.20. The identification of that arithmetic main term, for a bilinear expression in two weights rather than a square, is the two-weight Chinese remainder bound of Lemma 9.32, applied to the pair $(\lambda_{1,m}, \lambda_{1,m})$.

The arithmetic main term is evaluated by the standard $S_2^{(m)}$ chain, applied to $\lambda_{1,m}$ regarded — via Lemma 12.13 — as a standard weight at truncation $R' = R^{1+\varepsilon}$ built from $G_\varepsilon F$, and carrying the additional outer restriction $\prod_{i\neq m} r_i \le R^{1-\varepsilon}$. The chain consists of the extraction of the diagonal form (Lemma 9.21), the reduction to a second moment (Lemma 9.23), the expansion and decoupling step (Lemma 9.24), and the evaluation of the resulting decoupled $(k+1)$-fold sum (Lemma 9.25). Each of these uses the divisor range only through the coordinate-wise bound $\prod_i r_i \le R'$ and its logarithm, and so applies verbatim at the enlarged truncation.

The sharp cutoff $\prod_{i\neq m} r_i \le R^{1-\varepsilon}$ is not directly admissible in the chain, which requires a smooth profile. It is therefore replaced by a smooth ramp of width $\eta > 0$ in the logarithmic variables, i.e. $\lambda_{1,m}$ is replaced by the weight built from the profile $u \mapsto \chi_\eta\big(\sum_{i\neq m} u_i\big)\,(G_\varepsilon F)(u)$ where χ_η is a smooth non-increasing transition equal to 1 below $\frac{1-\varepsilon}{1+\varepsilon} - \eta$ and to 0 above $\frac{1-\varepsilon}{1+\varepsilon}$. This is a legitimate smooth profile supported on $\mathcal{R}_k$, so the chain applies to it for each fixed η, and produces the prefactor $\varphi(W)^k N(\log R')^{k+1}/(W^{k+1}\log N)$ times the standard marginal $J_k^{(m)}$ of the ramped rescaled profile. Letting $\eta \to 0$, dominated convergence in the outer $(k-1)$-fold integral replaces the ramp by the sharp indicator, so the marginal converges to

$$\int_{\mathcal{R}_{k-1}} \mathbf{1}\Big[\textstyle\sum_{i\neq m} u_i \le \frac{1-\varepsilon}{1+\varepsilon}\Big] \Big(\int_0^\infty (G_\varepsilon F)(u)\,du_m\Big)^2 du_{\hat m} \;=\; \frac{J_\varepsilon^{(m)}(F)}{(1+\varepsilon)^{k+1}},$$

the last equality being Lemma 12.5. Since $(\log R')^{k+1} = (1+\varepsilon)^{k+1}(\log R)^{k+1}$, the factors $(1+\varepsilon)^{\pm(k+1)}$ cancel and the main term takes the displayed form. □

Uses (proof): 12.21, 12.20, 9.32, 12.13, 12.5, 9.21, 9.23, 9.24, 9.25, 7.3

Lemma 12.23 (Smallness of the retained–complementary cross term)**.** *In the situation of Lemma 12.22, for every $\kappa > 0$ there is a choice of the smoothing width for which*

$$\Big|\sum_{N<n\le 2N} \chi_{\mathbb{P}}(n+h_m)\,A_{1,m}(n)\,A_{2,m}(n)\Big| \;\le\; \kappa\cdot\frac{\varphi(W)^k\,N\,(\log R)^{k+1}}{W^{k+1}\log N} \;+\; O_A\Big(\frac{N}{(\log N)^A}\Big)$$

for every $A > 0$ and all sufficiently large N.

Uses (statement): 12.16, 12.17, 12.18, 12.21, 12.22

Uses(notation): $A_{\ell,m}(n)$, $\lambda_{1,m}$, ε, R, W, $\chi_{\mathbb{P}}$, h_i, N, φ, $O(g)$, k (index)

Proof. By Lemma 12.21 with $\lambda' = \lambda_{2,m}$, replacing the sum by its arithmetic main term commits an error $O_A(N/(\log N)^A)$, and the main term is produced by the same $S_2^{(m)}$ chain as in Lemma 12.22, now applied to the pair $(\lambda_{1,m}, \lambda_{2,m})$; in particular it carries the same prefactor $\varphi(W)^k N(\log R)^{k+1}/(W^{k+1}\log N)$.

What remains is to bound the resulting integral. In the rescaled logarithmic coordinates $u_i = \log r_i/\log R'$, the profile attached to $\lambda_{1,m}$ is supported where $\sum_{i\neq m} u_i \le \frac{1-\varepsilon}{1+\varepsilon}$ and the

profile attached to $\lambda_{2,m} = \lambda_\varepsilon - \lambda_{1,m}$ is supported where $\sum_{i\neq m} u_i > \frac{1-\varepsilon}{1+\varepsilon}$. The two supports are disjoint — this is the same disjointness that makes the diagonal form additive along the split in Lemma 9.30, where it leaves no cross term at all — so with the sharp cutoff the product would vanish identically; with the smooth ramp of width η used in Lemma 12.22, the product is supported on the boundary strip

$$\mathcal{B}_\eta \;:=\; \Big\{u \in \mathcal{R}_k : \tfrac{1-\varepsilon}{1+\varepsilon} - \eta \le \textstyle\sum_{i\neq m} u_i \le \tfrac{1-\varepsilon}{1+\varepsilon}\Big\},$$

whose $(k-1)$-dimensional outer slice has Lebesgue measure at most $\eta/(k-2)!$. Since $G_\varepsilon F$ is bounded on the compact set $\mathcal{R}_k$ and the ramp takes values in $[0,1]$, the integrand is $O(1)$ on $\mathcal{B}_\eta$ and the integral is $O(\eta)$, with an implied constant depending only on k and $\sup|G_\varepsilon F|$. Choosing η so small that this is at most κ gives the claim. □

Uses (proof): 12.21, 12.22, 9.30, 9.24, 7.3

Proposition 12.24 (Enlarged $S_2^{(m)}$ lower bound)**.** *Let $k \ge 2$, let $\mathcal{H}$ be admissible, let $\varepsilon \ge 0$, let $F \in \mathcal{F}_\varepsilon$, and let ϑ satisfy Definition 12.9(i) and (iii) and have level of distribution ϑ. Then for each $m \in \{1,\dots,k\}$,*

$$S_2^{(m)}(\lambda_\varepsilon) \;\ge\; \big(1+o(1)\big)\,\frac{\varphi(W)^k\, N\,(\log R)^{k+1}}{W^{k+1}\log N}\, J_\varepsilon^{(m)}(F)$$

as $N \to \infty$.

Uses (statement): 12.11, 4.16, 5.7, 12.9, 12.2

Uses(notation): λ_ε, $S_2^{(m)}(\lambda)$, $J_\varepsilon^{(i)}(F)$, ε, R, W, ϑ, N, φ, $o(g)$, k (index)

Proof. Multiplying Lemma 12.19 by $\chi_{\mathbb{P}}(n+h_m) \ge 0$ and summing over $N < n \le 2N$ with $n \equiv v_0 \pmod W$ gives

$$S_2^{(m)}(\lambda_\varepsilon) \;\ge\; \sum_n \chi_{\mathbb{P}}(n+h_m) A_{1,m}(n)^2 \;+\; 2\sum_n \chi_{\mathbb{P}}(n+h_m) A_{1,m}(n) A_{2,m}(n).$$

Write $P := \varphi(W)^k N (\log R)^{k+1} W^{-k-1} (\log N)^{-1}$ for the common prefactor. By Lemma 12.22 the first sum is $(1+o(1))\,P\,J_\varepsilon^{(m)}(F) + O_A(N/(\log N)^A)$. By Lemma 12.23, for any $\kappa > 0$ the second sum is at most $2\kappa P + O_A(N/(\log N)^A)$ in absolute value. Given $\eta > 0$, if $J_\varepsilon^{(m)}(F) > 0$ choose $\kappa < \eta J_\varepsilon^{(m)}(F)/2$; then

$$S_2^{(m)}(\lambda_\varepsilon) \;\ge\; \big(1-\eta+o(1)\big) P\, J_\varepsilon^{(m)}(F) \;+\; O_A\big(N/(\log N)^A\big).$$

As $\eta > 0$ was arbitrary the claim follows. If $J_\varepsilon^{(m)}(F) = 0$ the right-hand side of the claim is 0 while the left-hand side is a sum of squares against a non-negative weight, hence non-negative, and the claim is trivial. □

Uses (proof): 12.19, 12.22, 12.23, 4.17

12.7 The criterion in enlarged form

Lemma 12.25 (The positivity window is non-empty)**.** *Let $0 < \vartheta < \frac{1}{2}$ and $Q \in \mathbb{R}$ with $2/\vartheta < Q$. Then there exist $\delta \in (0, \vartheta/2)$ and $\rho \in (1,2)$ with $\big(\frac{\vartheta}{2} - \delta\big)Q > \rho$.*

Uses(notation): ϑ, δ, ρ (index)

Proof. From $2/\vartheta < Q$ and $\vartheta > 0$ we get $2 < Q\vartheta$, i.e. $M := \frac{\vartheta}{2}Q > 1$. Put $u := \min(2, M) \in (1,2]$ and $\rho := \frac{1+u}{2}$; then $1 < \rho < u \le 2$ and $\rho < M$, so $\rho \in (1,2)$ and $\rho/Q < \vartheta/2$ (note $Q > 2/\vartheta > 0$). Finally set $\delta := \frac{1}{2}\big(\frac{\vartheta}{2} - \frac{\rho}{Q}\big)$, which lies in $(0, \vartheta/2)$ because $0 < \rho/Q < \vartheta/2$. Then $\frac{\vartheta}{2} - \delta = \frac{1}{2}\big(\frac{\vartheta}{2} + \frac{\rho}{Q}\big)$ and therefore $\big(\frac{\vartheta}{2} - \delta\big)Q = \frac{1}{2}(M + \rho) > \frac{1}{2}(\rho + \rho) = \rho$, using $M > \rho$. □

The window is non-empty for *every* ϑ with $2/\vartheta < Q$, with no further constraint tying δ to ε. Had one insisted instead on the reparametrisation condition $(1+\varepsilon)(\vartheta/2 - \delta) < \vartheta/2$, the window would have been empty in the decisive regime $\delta \to 0$, $\vartheta \to \frac{1}{2}$, and the scaling would have yielded no improvement. It is the weak room $(1+\varepsilon)\vartheta < 1$ of Definition 12.9(iii), together with the support split, that keeps that regime available.

Proposition 12.26 (Witness form of the enlarged criterion)**.** *Let $k \ge 2$, let $\mathcal{H} = \{h_1, \dots, h_k\}$ be admissible, let $0 \le \varepsilon < 1$, let $F \in \mathcal{F}_\varepsilon$ with $I_k(F) \neq 0$, and let ϑ be (ε, Q)-certificate-admissible for some Q with $\mathcal{Q}_\varepsilon(F) \ge Q$. Then* DHL$[k, 2]$ *holds: there are infinitely many n for which at least two of $n + h_1, \dots, n + h_k$ are prime.*

Uses (statement): 12.15, 12.24, 5.8, 12.9, 12.2, 3.15

Uses(notation): $\mathcal{Q}_\varepsilon(F)$, $\mathcal{T}_\varepsilon$, ε, ϑ, ρ, $I_k(F)$, h_i, k, $S_1(\lambda)$, $S_2(\lambda)$ (index)

Proof. By clauses (i) and (ii) of Definition 12.9 we have $0 < \vartheta < \frac{1}{2}$ and $2/\vartheta < Q \le \mathcal{Q}_\varepsilon(F)$, so Lemma 12.25 supplies $\delta \in (0, \vartheta/2)$ and $\rho \in (1,2)$ with $\big(\frac{\vartheta}{2} - \delta\big)\mathcal{Q}_\varepsilon(F) > \rho$. Fix these and set $R := N^{\vartheta/2 - \delta}$, so $\log R / \log N = \vartheta/2 - \delta$.

Clause (iii) gives $1 + \varepsilon < 1/\vartheta$, so Propositions 12.15 and 12.24 both apply. Summing Proposition 12.24 over $m \in \{1, \dots, k\}$ and using $\sum_m S_2^{(m)}(\lambda) = S_2(\lambda)$ (Lemma 4.17) together with Definition 5.8,

$$S_2(\lambda_\varepsilon) \;\ge\; \big(1 + o(1)\big)\,\frac{\varphi(W)^k N (\log R)^k}{W^{k+1}}\,\big(\tfrac{\vartheta}{2} - \delta\big)\,I_k(F)\,\mathcal{Q}_\varepsilon(F),$$

where the exchange of one power of $\log R$ for the factor $\vartheta/2 - \delta$ uses $(\log R)^{k+1}/(\log N \cdot (\log R)^k) = \vartheta/2 - \delta$. Proposition 12.15 gives $S_1(\lambda_\varepsilon) = (1 + o(1))\,\varphi(W)^k N (\log R)^k W^{-k-1} I_k(F)$, using $\int_{\mathcal{T}_\varepsilon} F^2 = I_k(F)$ for $F \in \mathcal{F}_\varepsilon$. Subtracting ρ times the latter from the former,

$$S_2(\lambda_\varepsilon) - \rho\, S_1(\lambda_\varepsilon) \;\ge\; \big(1 + o(1)\big)\,\frac{\varphi(W)^k N (\log R)^k}{W^{k+1}}\,I_k(F)\Big[\big(\tfrac{\vartheta}{2} - \delta\big)\mathcal{Q}_\varepsilon(F) - \rho\Big].$$

The bracket is a fixed positive number by the choice of ρ, and the prefactor is positive since $I_k(F) \neq 0$ forces $I_k(F) > 0$. Hence $S(\lambda_\varepsilon) = S_2(\lambda_\varepsilon) - \rho\, S_1(\lambda_\varepsilon) > 0$ for every sufficiently large N.

Finally, admissibility of $\mathcal{H}$ supplies for each N a residue v_0 modulo W with $\gcd(v_0 + h_i, W) = 1$ for all i, so the sieve sums above are the ones attached to a legitimate residue class, and Lemma 10.1 converts eventual positivity of S into infinitely many n for which at least $\lfloor \rho + 1 \rfloor$ of the $n + h_i$ are prime. Since $\rho \in (1,2)$ we have $\lfloor \rho + 1 \rfloor = 2$, which is DHL$[k, 2]$. □

Uses (proof): 12.25, 12.15, 12.24, 4.17, 10.1

Proposition 12.26 demands a *smooth* witness, whereas the available certificates are polynomials cut off at the boundary of $\mathcal{T}_\varepsilon$, hence not smooth. The gap is closed by mollification, using the L^2-continuity of $\mathcal{Q}_\varepsilon$.

Lemma 12.27 (Mollification of an L^2 witness)**.** *Let $0 \le \varepsilon < 1$, let $F_0 \in L^2(\mathcal{T}_\varepsilon)$ with $I_k(F_0) \neq 0$, and let $c < \mathcal{Q}_\varepsilon(F_0)$. Then there exists $\widetilde{F} \in \mathcal{F}_\varepsilon$ with $I_k(\widetilde{F}) \neq 0$ and $\mathcal{Q}_\varepsilon(\widetilde{F}) > c$.*

Uses (statement): 5.8, 12.2, 12.8

Uses(notation): $\mathcal{Q}_\varepsilon(F)$, $\mathcal{T}_\varepsilon$, ε, $I_k(F)$, $\|f\|_{L^2(X)}$, k (index)

Proof. Set $\eta := \mathcal{Q}_\varepsilon(F_0) - c > 0$ and let $\zeta > 0$ be the radius supplied by Lemma 12.8 for $F_1 := F_0$ and this η: every $F_2 \in L^2(\mathcal{T}_\varepsilon)$ with $\|F_2 - F_0\|_{L^2(\mathcal{T}_\varepsilon)} < \zeta$ has $I_k(F_2) \neq 0$ and $\mathcal{Q}_\varepsilon(F_2) > \mathcal{Q}_\varepsilon(F_0) - \eta = c$.

It therefore suffices to produce $\widetilde{F} \in \mathcal{F}_\varepsilon$ within ζ of F_0 in $L^2(\mathcal{T}_\varepsilon)$. Transport the problem to the standard simplex by the homothety of Lemma 12.1: the map $F \mapsto (1+\varepsilon)^{k/2}\, G_\varepsilon F$ is an isometry of $L^2(\mathcal{T}_\varepsilon)$ onto $L^2(\mathcal{R}_k)$, by Lemma 12.4 applied to differences, and it carries $\mathcal{F}_\varepsilon$ onto the class of smooth functions supported on $\mathcal{R}_k$. On $L^2(\mathcal{R}_k)$, Lemma 5.19 provides smooth approximants to any given element within any prescribed L^2 distance, and Lemma 5.18 lets one multiply such an approximant by a smooth cutoff supported on $\mathcal{R}_k$ without increasing the distance by more than an arbitrarily small amount; the resulting function is smooth and supported on $\mathcal{R}_k$. Pulling back through the isometry gives the required $\widetilde{F} \in \mathcal{F}_\varepsilon$. □

Uses (proof): 12.8, 12.1, 12.4, 5.19, 5.18

Lemma 12.28 (Bombieri–Vinogradov supplies a certificate-admissible level)**.** *Let $0 \le \varepsilon \le 1$ and let $Q > 4$. Then there is an (ε, Q)-certificate-admissible ϑ; one may take $\vartheta = \frac{1}{2}\big(\frac{2}{Q} + \frac{1}{2}\big)$.*

Uses (statement): 12.9, 1

Uses(notation): ϑ, ε (index)

Proof. Since $Q > 4$ we have $2/Q < \frac{1}{2}$, so the interval $(2/Q, \frac{1}{2})$ is non-empty; let ϑ be its midpoint. Then $0 < 2/Q < \vartheta < \frac{1}{2}$, which is clause (i). From $\vartheta > 2/Q$ and $\vartheta, Q > 0$ we get $2/\vartheta < Q$, clause (ii). For clause (iii), $\vartheta < \frac{1}{2}$ gives $1/\vartheta > 2 \ge 1+\varepsilon$ whenever $\varepsilon \le 1$; the inequality is strict because $\vartheta < \frac{1}{2}$ strictly. Finally $\vartheta < \frac{1}{2}$, so the primes have level of distribution ϑ by Bombieri–Vinogradov (Theorem 1). □

Uses (proof): 1

The proof of Lemma 12.28 chooses ϑ as a function of Q alone, and the closer Q is to 4 the closer ϑ must be pushed to $\frac{1}{2}$; this is the only quantitative demand the enlarged theory makes on a witness. A witness attaining $\mathcal{Q}_\varepsilon(F_0) > 4.0043$ permits $\vartheta = 0.4995$, for which $2/\vartheta = 4.00400\ldots$ and $(1+\varepsilon)\vartheta < 1$ for every $\varepsilon \le 1$.

Proposition 12.29 (The enlarged criterion)**.** *Let $k \ge 2$, let $0 \le \varepsilon < 1$, and let $\mathcal{H} = \{h_1, \ldots, h_k\}$ be admissible. Suppose there exists $F_0 \in L^2(\mathcal{T}_\varepsilon)$ with*

$$4\,\|F_0\|^2_{L^2(\mathcal{T}_\varepsilon)} \;<\; \sum_{m=1}^{k} J_\varepsilon^{(m)}(F_0),$$

equivalently $\mathcal{Q}_\varepsilon(F_0) > 4$, which by Lemma 12.7 certifies $M_{k,\varepsilon} > 4$. Then $\mathrm{DHL}[k, 2]$ holds.

Uses (statement): 5.7, 5.8, 12.7, 12.26, 5.2, 3.15

Uses(notation): $\mathcal{Q}_\varepsilon(F)$, $J_\varepsilon^{(i)}(F)$, M_k, $\mathcal{T}_\varepsilon$, ε, k, $\|f\|_{L^2(X)}$, h_i (index)

Proof. As in the proof of Lemma 12.7, the hypothesis forces $\|F_0\|_{L^2(\mathcal{T}_\varepsilon)} \neq 0$, i.e. $I_k(F_0) \neq 0$, and dividing gives $\mathcal{Q}_\varepsilon(F_0) > 4$. Choose Q with $4 < Q < \mathcal{Q}_\varepsilon(F_0)$. By Lemma 12.28 there is an (ε, Q)-certificate-admissible ϑ. By Lemma 12.27 applied with $c := Q$ there is $\widetilde{F} \in \mathcal{F}_\varepsilon$ with $I_k(\widetilde{F}) \neq 0$ and $\mathcal{Q}_\varepsilon(\widetilde{F}) > Q$. Then ϑ is (ε, Q)-certificate-admissible and $\mathcal{Q}_\varepsilon(\widetilde{F}) \geq Q$, so Proposition 12.26 applied to $\widetilde{F}$ and $\mathcal{H}$ gives DHL$[k, 2]$.

At $\varepsilon = 0$ every step above is the corresponding step of the standard development: by Lemma 12.6 the hypothesis reads $4\, I_k(F_0) < \sum_m J_k^{(m)}(F_0)$, which is the standard criterion $M_k > 4$ witnessed by F_0. □

Uses (proof): 12.7, 12.28, 12.27, 12.26, 12.6

Proposition 12.29 is the criterion of the development, stated once in (k, ε). At $\varepsilon = 0$ it is verbatim the standard criterion — the demand $M_k > 4$, witnessed by an explicit competitor rather than by the supremum — and the proof specialises to the standard proof, with the support split of Definition 12.16 becoming trivial ($\lambda_{2,m} = 0$) and clause (iii) of Definition 12.9 becoming vacuous. For $\varepsilon > 0$ the competitor is allowed to live on a strictly larger region, at the cost of a strictly smaller marginal slice and of clause (iii); whether that trade is profitable at a given k is a question about the competitor, settled by exhibiting one. The two cases settle it at $(k, \varepsilon) = (105, 0)$ and at $(k, \varepsilon) = (50, \frac{1}{25})$ respectively.

13 The numerical certificate

The variational input to the ε-enlarged sieve is the single inequality $M_{k,\varepsilon} > 4$. At $k = 50$ this inequality is not established by an asymptotic estimate; one exhibits an explicit test function and computes its Rayleigh quotient exactly. The objects below reduce the inequality to an exact, finite computation with rational numbers.

There are two forms of certificate. The *abstract* form (Definition 13.5) is the form the argument uses: a single element of $L^2(\mathcal{T}_\varepsilon)$ the sum of whose enlarged marginal functionals exceeds four times its squared norm. The *explicit* form (Definition 13.10) is what is checked: a finite list of exponent data together with a finite list of rational coefficients, subject to one inequality between two explicitly computable rational numbers. Lemmas 13.13 and 13.14 evaluate the two sides as exact rational quadratic forms. The analytic content lies in using the resulting test function; the certificate itself is arithmetic.

13.1 The certificate basis

Notation 13.1 (Signature)**.** A *signature* is a finite multiset α of positive integers, written in non-increasing order as $\alpha = (\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_j \geq 1)$; the empty signature $j = 0$ is admitted. Its *size* is $|\alpha| := \sum_{i=1}^{j} \alpha_i$, its *parts* are the values $\alpha_1, \ldots, \alpha_j$, and for $r \geq 1$ the *deletion* $\alpha \setminus r$ is the signature obtained by removing one copy of the part r from α. By convention $\alpha \setminus r := \alpha$ when r is not a part of α; in particular $\alpha \setminus 0 = \alpha$.

Uses(notation): α (index)

Definition 13.2 (Monomial-symmetric polynomial)**.** Let $k \geq 1$ and let α be a signature (Notation 13.1). Write

$$E_k(\alpha) \;:=\; \big\{\, A = (A_1, \ldots, A_k) \in \mathbb{Z}_{\geq 0}^k \;:\; \text{the multiset of non-zero entries of } A \text{ equals } \alpha \,\big\},$$

a finite set, empty unless α has at most k parts. The *monomial-symmetric polynomial* of signature α in k variables is

$$m_\alpha(t) \;:=\; \sum_{A \in E_k(\alpha)} \prod_{i=1}^{k} t_i^{A_i} \;\in\; \mathbb{Z}[t_1, \ldots, t_k].$$

Each distinct way of distributing the parts of α among the k variables contributes exactly one monomial, so m_α is symmetric, homogeneous of degree $|\alpha|$, and $m_\emptyset = 1$.

Uses (statement): 13.1

Uses(notation): α, k (index)

Definition 13.3 (Certificate basis element)**.** Let $k \geq 1$, let $\varepsilon \in \mathbb{R}$, let $a \in \mathbb{Z}_{\geq 0}$ and let α be a signature. The associated *certificate basis element* is the symmetric polynomial

$$b_{a,\alpha}(t) \;:=\; \bigl(1 + \varepsilon - P_{(1)}(t)\bigr)^a \, m_\alpha(t), \qquad P_{(1)}(t) = t_1 + \cdots + t_k,$$

with m_α as in Definition 13.2. Its total degree is $a + |\alpha|$.

Uses (statement): 13.2, 13.1

Uses(notation): $P_{(1)}$, ε, α, k (index)

Definition 13.4 (Parametric certificate polynomial)**.** Let $k \geq 1$, $\varepsilon \in \mathbb{R}$, and let $n \in \mathbb{N}$. Fix basis indices $(a_1, \alpha_1), \ldots, (a_n, \alpha_n)$, each consisting of an integer $a_i \geq 0$ and a signature α_i, and let $c = (c_1, \ldots, c_n) \in \mathbb{Q}^n$. The associated *certificate polynomial* is

$$F_c \;:=\; \sum_{i=1}^{n} c_i \, b_{a_i,\alpha_i} \;\in\; \mathbb{Q}[t_1, \ldots, t_k],$$

with $b_{a,\alpha}$ as in Definition 13.3. It is a symmetric polynomial, hence continuous, hence bounded on the compact set $\mathcal{T}_\varepsilon$; we regard it also as an element of $L^2(\mathcal{T}_\varepsilon)$, and where a value off $\mathcal{T}_\varepsilon$ is needed we use the extension of F_c by zero.

Uses (statement): 13.3, 5.2

Uses(notation): $\mathcal{T}_\varepsilon$, ε, k, α, $\|f\|_{L^2(X)}$ (index)

13.2 The two forms of certificate

Definition 13.5 (Abstract certificate)**.** Let $k \geq 1$ and $\varepsilon \in \mathbb{R}$. An *abstract certificate* for (k, ε) is a function $F \in L^2(\mathcal{T}_\varepsilon)$ such that

$$4 \, \|F\|^2_{L^2(\mathcal{T}_\varepsilon)} \;<\; \sum_{m=1}^{k} J_\varepsilon^{(m)}(F),$$

where $J_\varepsilon^{(m)}$ is the enlarged marginal functional of Definition 5.7. Equivalently, since $\|F\|^2_{L^2(\mathcal{T}_\varepsilon)} = I_k(F)$ for F extended by zero off $\mathcal{T}_\varepsilon$, an abstract certificate is an F with $\mathcal{Q}_\varepsilon(F) > 4$.

Uses (statement): 5.2, 5.7, 5.6, 5.8

Uses(notation): $\mathcal{T}_\varepsilon$, $\mathcal{S}_\varepsilon$, $I_k(F)$, $J_\varepsilon^{(i)}(F)$, $\mathcal{Q}_\varepsilon(F)$, $\|f\|_{L^2(X)}$, ε, k (index)

Definition 13.6 (Factorial moment of two signatures)**.** Let $k \geq 1$ and let α, β be signatures. Their *factorial moment* in k variables is

$$\Phi_k(\alpha, \beta) \;:=\; \sum_{A \in E_k(\alpha)} \; \sum_{B \in E_k(\beta)} \; \prod_{i=1}^{k} (A_i + B_i)! \;\in\; \mathbb{Z}_{\geq 0},$$

the exponent sets $E_k(\cdot)$ being those of Definition 13.2. It is a finite sum of positive integers, and $\Phi_k(\emptyset, \emptyset) = 1$.

Uses (statement): 13.1, 13.2

Uses(notation): α, k (index)

Definition 13.7 (Explicit enlarged-simplex pairing)**.** Let $k \geq 1$, $\varepsilon \in \mathbb{Q}$, let $a_1, a_2 \in \mathbb{Z}_{\geq 0}$ and let α_1, α_2 be signatures. Put $e := a_1 + a_2$ and $S := |\alpha_1| + |\alpha_2|$. The *explicit enlarged-simplex pairing* is the rational number

$$\mathcal{I}_k^{\varepsilon}(a_1, \alpha_1; a_2, \alpha_2) \;:=\; (1+\varepsilon)^{\,k+e+S}\, \frac{e!}{(k+e+S)!}\, \Phi_k(\alpha_1, \alpha_2),$$

with Φ_k as in Definition 13.6.

Uses (statement): 13.6, 13.1

Uses(notation): ε, k, α (index)

Definition 13.8 (Radial factor)**.** For $q, e \in \mathbb{Z}_{\geq 0}$ and $\varepsilon \in \mathbb{Q}$, set

$$\varrho_q(\varepsilon, e) \;:=\; (1-\varepsilon)^q \sum_{r=0}^{e} \binom{e}{r} (2\varepsilon)^{e-r} (1-\varepsilon)^r\, \frac{r!}{(r+q)!} \;\in\; \mathbb{Q}.$$

Uses(notation): ε (index)

Definition 13.9 (Explicit shrunken-slice pairing)**.** Let $k \geq 1$, $\varepsilon \in \mathbb{Q}$, let $a_1, a_2 \in \mathbb{Z}_{\geq 0}$ and let α_1, α_2 be signatures. For parts $r_1, r_2 \geq 0$ write

$$q(r_1, r_2) \;:=\; (k-1) + \big(|\alpha_1| - r_1\big) + \big(|\alpha_2| - r_2\big), \qquad e(r_1, r_2) \;:=\; (a_1 + r_1 + 1) + (a_2 + r_2 + 1).$$

The *explicit shrunken-slice pairing* is the rational number

$$\mathcal{J}_k^{\varepsilon}(a_1, \alpha_1; a_2, \alpha_2) \;:=\; \sum_{r_1} \sum_{r_2} \frac{r_1!\, a_1!}{(a_1 + r_1 + 1)!}\, \frac{r_2!\, a_2!}{(a_2 + r_2 + 1)!}\, \varrho_{q(r_1, r_2)}\big(\varepsilon, e(r_1, r_2)\big)\, \Phi_{k-1}\big(\alpha_1 \backslash r_1,\; \alpha_2 \backslash r_2\big),$$

where r_1 runs over 0 together with the distinct parts of α_1, and r_2 over 0 together with the distinct parts of α_2; the deletions $\alpha \setminus r$ and the convention $\alpha \setminus 0 = \alpha$ are those of Notation 13.1, ϱ is Definition 13.8 and Φ is Definition 13.6.

Uses (statement): 13.1, 13.8, 13.6

Uses(notation): ε, k, α (index)

Definition 13.10 (Explicit certificate)**.** Let $k \geq 1$ and $\varepsilon \in \mathbb{Q}$. An *explicit certificate* for (k, ε) consists of

- an integer $n \geq 0$;
- exponents $a_1, \ldots, a_n \in \mathbb{Z}_{\geq 0}$ and signatures $\alpha_1, \ldots, \alpha_n$ (Notation 13.1);
- coefficients $c_1, \ldots, c_n \in \mathbb{Q}$;

subject to the single inequality

$$4 \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j\, \mathcal{I}_k^{\varepsilon}(a_i, \alpha_i; a_j, \alpha_j) \;<\; k \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j\, \mathcal{J}_k^{\varepsilon}(a_i, \alpha_i; a_j, \alpha_j),$$

with $\mathcal{I}_k^{\varepsilon}$ as in Definition 13.7 and $\mathcal{J}_k^{\varepsilon}$ as in Definition 13.9. Both sides are rational numbers determined by the data by finitely many additions and multiplications.

Uses (statement): 13.1, 13.7, 13.9

Uses(notation): ε, k, α (index)

13.3 From a certificate to the bound $M_{k,\varepsilon} > 4$

Lemma 13.11 (Closed form of the enlarged-simplex pairing)**.** *Let $k \geq 1$, let $\varepsilon \in \mathbb{Q}$ with $1+\varepsilon \geq 0$, let $a_1, a_2 \in \mathbb{Z}_{\geq 0}$ and let α_1, α_2 be signatures (Notation 13.1). Then*

$$\int_{\mathcal{T}_\varepsilon} b_{a_1,\alpha_1}(t)\, b_{a_2,\alpha_2}(t)\, dt \;=\; \mathcal{I}_k^{\varepsilon}(a_1,\alpha_1;a_2,\alpha_2),$$

with $b_{a,\alpha}$ as in Definition 13.3 and $\mathcal{I}_k^{\varepsilon}$ as in Definition 13.7. In particular the pairing is rational.

Uses (statement): 13.3, 13.7, 5.2, 13.1

Uses(notation): $\mathcal{T}_\varepsilon$, ε, k, $P_{(1)}$, α (index)

Proof. Expanding both monomial-symmetric factors (Definition 13.2) and collecting the two slack powers,

$$b_{a_1,\alpha_1}(t)\, b_{a_2,\alpha_2}(t) \;=\; \sum_{A\in E_k(\alpha_1)} \sum_{B\in E_k(\alpha_2)} \big(1+\varepsilon-\textstyle\sum_i t_i\big)^{a_1+a_2} \prod_{i=1}^{k} t_i^{A_i+B_i},$$

a finite sum of continuous functions, so the integral may be taken term by term. Here A and B denote exponent vectors, as in Definition 13.2. Fix A, B, put $C_i := A_i + B_i$ and write $e := a_1 + a_2$. Every $A \in E_k(\alpha_1)$ satisfies $\sum_i A_i = |\alpha_1|$ and likewise for B, so $\sum_i C_i = |\alpha_1| + |\alpha_2| =: S$. Substituting $t = (1+\varepsilon)u$, which maps $\mathcal{R}_k$ onto $\mathcal{T}_\varepsilon$ and has Jacobian $(1+\varepsilon)^k$, and using $1+\varepsilon-\sum_i t_i = (1+\varepsilon)(1-\sum_i u_i)$,

$$\int_{\mathcal{T}_\varepsilon} \big(1+\varepsilon-\textstyle\sum_i t_i\big)^{e} \prod_i t_i^{C_i}\, dt \;=\; (1+\varepsilon)^{\,k+e+S} \int_{\mathcal{R}_k} \big(1-\textstyle\sum_i u_i\big)^{e} \prod_i u_i^{C_i}\, du.$$

Lemma 5.23 evaluates the remaining integral as $e!\prod_i C_i!/(k+e+S)!$. Summing over $A \in E_k(\alpha_1)$ and $B \in E_k(\alpha_2)$ turns $\prod_i (A_i+B_i)!$ into $\Phi_k(\alpha_1,\alpha_2)$ (Definition 13.6) and produces exactly $\mathcal{I}_k^{\varepsilon}(a_1,\alpha_1;a_2,\alpha_2)$. □

Uses (proof): 13.2, 13.6, 5.2, 5.23

Lemma 13.12 (Closed form of the shrunken-slice pairing)**.** *Let $k \geq 1$, let $\varepsilon \in \mathbb{Q}$ with $0 \leq \varepsilon \leq 1$, let $m \in \{1,\dots,k\}$, let $a_1, a_2 \in \mathbb{Z}_{\geq 0}$ and let α_1, α_2 be signatures (Notation 13.1). Extending the basis elements by zero off $\mathcal{T}_\varepsilon$,*

$$\int_{\mathcal{S}_\varepsilon} \Big(\int_0^\infty b_{a_1,\alpha_1}(t)\, dt_m\Big)\Big(\int_0^\infty b_{a_2,\alpha_2}(t)\, dt_m\Big)\, dt_{\hat{m}} \;=\; \mathcal{J}_k^{\varepsilon}(a_1,\alpha_1;a_2,\alpha_2),$$

with $b_{a,\alpha}$ as in Definition 13.3 and $\mathcal{J}_k^{\varepsilon}$ as in Definition 13.9. In particular the value is rational and independent of m.

Uses (statement): 13.3, 13.9, 5.2, 13.1

Uses(notation): $\mathcal{T}_\varepsilon$, $\mathcal{S}_\varepsilon$, ε, k, $P_{(1)}$, α (index)

Proof. Write $\hat{t} = (t_i)_{i\neq m}$ for a point of $\mathcal{S}_\varepsilon$ and set $c(\hat{t}) := 1+\varepsilon-\sum_{i\neq m} t_i$, the available length in the isolated coordinate. Since $\sum_{i\neq m} t_i \leq 1-\varepsilon$ on $\mathcal{S}_\varepsilon$, we have $c(\hat{t}) \geq 2\varepsilon \geq 0$.

Step 1: the marginal of one basis element. A point with m-th coordinate s and remaining coordinates $\hat{t}$ lies in $\mathcal{T}_\varepsilon$ if and only if $0 \le s \le c(\hat{t})$, so the zero-extension makes $\int_0^\infty b_{a,\alpha}\,dt_m = \int_0^{c(\hat{t})} b_{a,\alpha}\,dt_m$. Expanding m_α (Definition 13.2) and separating the isolated coordinate,

$$b_{a,\alpha}(s;\hat{t}) \;=\; \sum_{A\in E_k(\alpha)} \big(c(\hat{t})-s\big)^{a}\, s^{A_m} \prod_{i\neq m} t_i^{A_i}.$$

By Lemma 5.30, after the substitution $s = c(\hat{t})\,v$,

$$\int_0^{c} (c-s)^a s^r\,ds \;=\; \frac{a!\,r!}{(a+r+1)!}\,c^{\,a+r+1} \qquad (c\ge 0,\ a,r\in\mathbb{Z}_{\ge 0}),$$

so that

$$\Big(\int_0^\infty b_{a,\alpha}\,dt_m\Big)(\hat{t}) \;=\; \sum_{A\in E_k(\alpha)} \frac{a!\,A_m!}{(a+A_m+1)!}\,c(\hat{t})^{\,a+A_m+1} \prod_{i\neq m} t_i^{A_i}.$$

Step 2: a shifted monomial integral over the shrunken slice. For $e\in\mathbb{Z}_{\ge 0}$ and exponents $(C_i)_{i\neq m}$ we claim

$$\int_{\mathcal{S}_\varepsilon} c(\hat{t})^{\,e} \prod_{i\neq m} t_i^{C_i}\,dt_{\hat{m}} \;=\; \varrho_q(\varepsilon,e)\prod_{i\neq m} C_i!, \qquad q := (k-1)+\sum_{i\neq m} C_i,$$

with ϱ as in Definition 13.8. Indeed $c(\hat{t}) = 2\varepsilon + \big((1-\varepsilon)-\sum_{i\neq m} t_i\big)$, so the binomial theorem gives $c(\hat{t})^e = \sum_{r=0}^{e}\binom{e}{r}(2\varepsilon)^{e-r}\big((1-\varepsilon)-\sum_{i\neq m}t_i\big)^r$. Each summand is integrated by the substitution $\hat{t} = (1-\varepsilon)u$, which maps $\mathcal{R}_{k-1}$ onto $\mathcal{S}_\varepsilon$ (here $0\le\varepsilon\le 1$ is used), followed by Lemma 5.23; this contributes $(1-\varepsilon)^{\,q+r}\,r!\prod_i C_i!/(r+q)!$, and summing over r gives the claim.

Step 3: assembly. Multiply the two marginals furnished by Step 1 and integrate over $\mathcal{S}_\varepsilon$. The result is the double sum over $A\in E_k(\alpha_1)$, $B\in E_k(\alpha_2)$ of

$$\frac{a_1!\,A_m!}{(a_1+A_m+1)!}\,\frac{a_2!\,B_m!}{(a_2+B_m+1)!}\int_{\mathcal{S}_\varepsilon} c(\hat{t})^{\,(a_1+A_m+1)+(a_2+B_m+1)} \prod_{i\neq m} t_i^{A_i+B_i}\,dt_{\hat{m}},$$

which Step 2 evaluates, since $\sum_{i\neq m}(A_i+B_i) = (|\alpha_1|-A_m)+(|\alpha_2|-B_m)$. Finally group the double sum by the two isolated exponents $r_1 := A_m$ and $r_2 := B_m$: deleting the m-th coordinate is a bijection

$$E_k(\alpha) \;\longleftrightarrow\; \coprod_{r} \{r\}\times E_{k-1}(\alpha\setminus r),$$

where r runs over 0 together with the distinct parts of α (Notation 13.1). Under this bijection the residual factorial products $\prod_{i\neq m}(A_i+B_i)!$ sum to $\Phi_{k-1}(\alpha_1\setminus r_1,\alpha_2\setminus r_2)$ (Definition 13.6), and the remaining factors are exactly those of Definition 13.9. No step depends on which coordinate was isolated, so the value is independent of m. □

Uses (proof): 13.2, 13.6, 13.8, 5.2, 5.30, 5.23

Lemma 13.13 ($I_k(F_c)$ as a rational quadratic form)**.** *Let $k\ge 1$, let $\varepsilon\in\mathbb{Q}$ with $1+\varepsilon\ge 0$, and let F_c be the certificate polynomial attached to basis indices $(a_i,\alpha_i)_{i\le n}$ and coefficients $c\in\mathbb{Q}^n$ (Definition 13.4). Then*

$$I_k(F_c) \;=\; \|F_c\|^2_{L^2(\mathcal{T}_\varepsilon)} \;=\; \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j\, \mathcal{I}_k^{\varepsilon}(a_i,\alpha_i;a_j,\alpha_j) \;\in\; \mathbb{Q},$$

where F_c is extended by zero off $\mathcal{T}_\varepsilon$ in the definition of I_k.

Uses (statement): 13.4, 13.7, 5.2, 5.6

Uses(notation): $\mathcal{T}_\varepsilon$, ε, k, $I_k(F)$, $\|f\|_{L^2(X)}$ (index)

Proof. F_c is continuous, hence square-integrable on the compact set $\mathcal{T}_\varepsilon$, and $\|F_c\|^2_{L^2(\mathcal{T}_\varepsilon)} = \int_{\mathcal{T}_\varepsilon} F_c^2$. Expanding the square of the finite sum,

$$F_c^2 \;=\; \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \, b_{a_i,\alpha_i}\, b_{a_j,\alpha_j},$$

and integrating term by term (a finite sum of continuous functions on a compact set), Lemma 13.11 evaluates each $\int_{\mathcal{T}_\varepsilon} b_{a_i,\alpha_i} b_{a_j,\alpha_j}$ as $\mathcal{I}_k^\varepsilon(a_i,\alpha_i;a_j,\alpha_j)$. Rationality follows since each pairing and each c_i is rational. □

Uses (proof): 13.3, 13.11

Lemma 13.14 ($\sum_m J_\varepsilon^{(m)}(F_c)$ as a rational quadratic form)**.** *Let $k \geq 1$, let $\varepsilon \in \mathbb{Q}$ with $0 \leq \varepsilon \leq 1$, and let F_c be the certificate polynomial attached to basis indices $(a_i,\alpha_i)_{i\leq n}$ and coefficients $c \in \mathbb{Q}^n$ (Definition 13.4), extended by zero off $\mathcal{T}_\varepsilon$. Then*

$$\sum_{m=1}^{k} J_\varepsilon^{(m)}(F_c) \;=\; k \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \, \mathcal{J}_k^\varepsilon(a_i,\alpha_i;a_j,\alpha_j) \;\in\; \mathbb{Q}.$$

Uses (statement): 13.4, 13.9, 5.7, 5.2

Uses(notation): $\mathcal{S}_\varepsilon$, $\mathcal{T}_\varepsilon$, ε, k, $J_\varepsilon^{(i)}(F)$ (index)

Proof. Fix m. The marginal $\hat{t} \mapsto \int_0^\infty F_c \, dt_m$, taken for the extension of F_c by zero off $\mathcal{T}_\varepsilon$, is linear in the coefficient vector:

$$\int_0^\infty F_c \, dt_m \;=\; \sum_{i=1}^{n} c_i \int_0^\infty b_{a_i,\alpha_i}\, dt_m.$$

Squaring and integrating over $\mathcal{S}_\varepsilon$, and interchanging the finite sums with the integral,

$$J_\varepsilon^{(m)}(F_c) \;=\; \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \int_{\mathcal{S}_\varepsilon} \Big(\int_0^\infty b_{a_i,\alpha_i} dt_m\Big)\Big(\int_0^\infty b_{a_j,\alpha_j} dt_m\Big) d\hat{t} \;=\; \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \, \mathcal{J}_k^\varepsilon(a_i,\alpha_i;a_j,\alpha_j)$$

by Lemma 13.12. The right-hand side does not depend on m, so summing over the k coordinates multiplies it by k. □

Uses (proof): 13.3, 13.12

13.4 Existence of a certificate at $k = 50$

Proposition 13.15 (Existence of an explicit certificate for $k = 50$)**.** *There exist a rational number ε with $0 \le \varepsilon \le 1$ and an explicit certificate for $(50, \varepsilon)$ in the sense of Definition 13.10.*

Uses (statement): 13.10

Uses(notation): ε, k, α, P_α, $\alpha \cup \alpha'$, $P_{(1)}$, P_j, $\mathcal{T}_\varepsilon$ (index)

Proof. Take $\varepsilon = 1/25$, so that $0 \le \varepsilon \le 1$. Let the index set be

$$\mathcal{B} \;:=\; \big\{\,(a,\alpha)\;:\; a \in \mathbb{Z}_{\ge 0},\ \alpha \text{ a signature all of whose parts are even and } \ge 2,\ a + |\alpha| \le 27\,\big\},$$

a finite set; enumerate it in any fixed order as $(a_1,\alpha_1),\dots,(a_n,\alpha_n)$ with $n = |\mathcal{B}| = 2526$. Definitions 13.7 and 13.9 attach to this enumeration two explicit rational symmetric matrices,

$$(A_I)_{ij} := \mathcal{I}_{50}^{\varepsilon}(a_i,\alpha_i;a_j,\alpha_j), \qquad (A_J)_{ij} := 50\,\mathcal{J}_{50}^{\varepsilon}(a_i,\alpha_i;a_j,\alpha_j),$$

each entry being a finite sum of products of factorials, binomial coefficients and powers of ε. What has to be produced is a vector $c \in \mathbb{Q}^n$ with $4\,c^{\top}A_I\,c < c^{\top}A_J\,c$; once c is written down, this is a comparison of two explicitly given rational numbers and is decided by exact integer arithmetic after clearing denominators.

Such a vector is found as follows. By Lemma 13.11, $c^{\top}A_I\,c = \int_{\mathcal{T}_\varepsilon} F_c^2 \ge 0$, and it vanishes only for $c = 0$ because the polynomials $b_{a,\alpha}$, $(a,\alpha) \in \mathcal{B}$, are linearly independent; so A_I is positive definite and the generalised eigenvalue problem $A_J v = \lambda A_I v$ has real eigenvalues, the largest being $\lambda_{\max} = \sup_{v \ne 0}(v^{\top}A_J v)/(v^{\top}A_I v)$. Computing a top eigenvector to sufficient precision and rounding it coordinatewise to a rational vector over a common denominator yields c. None of this is part of the verification: the precision analysis, the eigenvector and the rounding only serve to produce a candidate c, and once c is fixed the sole assertion to be checked is the displayed rational inequality. Carrying this out at $(k,\varepsilon,d) = (50, 1/25, 27)$ gives a c with

$$\frac{c^{\top}A_J\,c}{c^{\top}A_I\,c} \;>\; 4.0043 \;>\; 4,$$

so the inequality of Definition 13.10 holds. The degree bound cannot be relaxed by much: the supremum of the quotient over the span of the degree-d basis increases with d, and at $(k,\varepsilon) = (50, 1/25)$ it is still below 4 for $d \le 24$, its value there being $3.9995\ldots$; the exhibited c lives at $d = 27$, where the exact quotient is $4.00430\ldots$. The restriction to signatures with even parts is likewise not cosmetic: it is what the test functions used here look like, and dropping the multi-part signatures (allowing only α with at most two parts, say) lowers the attainable quotient below 4.

It is equivalent, and closer to the classical formulation, to search in the power-sum monomials $P_\alpha = \prod_j P_{\alpha_j}$ with $P_r = \sum_i t_i^r$ instead of the monomial-symmetric polynomials m_α: expanding a product of power sums groups the coordinates into blocks, so P_α is a non-negative integer combination of the m_μ with μ obtained by merging parts of α; such μ again have even parts ≥ 2 and $|\mu| = |\alpha|$. Hence $(1 + \varepsilon - P_{(1)})^a P_\alpha$ lies in the span of $\{b_{a',\alpha'} : (a',\alpha') \in \mathcal{B}\}$ for every $(a,\alpha) \in \mathcal{B}$, and a rational coefficient vector in the power-sum family transforms, by a finite integer matrix, into one in the family $\mathcal{B}$. □

Uses (proof): 13.7, 13.9, 13.3, 13.4, 13.1, 13.11

Proposition 13.15 gives an $\varepsilon \in \mathbb{Q} \cap [0,1]$ and an explicit polynomial $F \in L^2(\mathcal{T}_\varepsilon)$ whose enlarged Rayleigh quotient exceeds 4; the polynomial is then smoothed and used in the enlarged sieve estimates.

14 The admissible 50-tuple

Definition 14.1 (The tuple $\mathcal{H}_{50}$)**.** Let

$$\begin{aligned}\mathcal{H}_{50} := \{\,&0,\,4,\,6,\,16,\,30,\,34,\,36,\,46,\,48,\,58,\,60,\,64,\,70,\,78,\,84,\,88,\,90,\\ &94,\,100,\,106,\,108,\,114,\,118,\,126,\,130,\,136,\,144,\,148,\,150,\,156,\,160,\\ &168,\,174,\,178,\,184,\,190,\,196,\,198,\,204,\,210,\,214,\,216,\,220,\,226,\,228,\\ &234,\,238,\,240,\,244,\,246\,\} \subset \mathbb{Z}_{\geq 0},\end{aligned}$$

a set of 50 distinct non-negative integers with $\min \mathcal{H}_{50} = 0$ and $\max \mathcal{H}_{50} = 246$. It is the narrowest known admissible 50-tuple, normalised so that its least element is 0; see the Engelsma–Sutherland tables of admissible tuples, `https://math.mit.edu/~primegaps/tuples/admissible_50_246.txt`.

Uses(notation): $\mathcal{H}_{50}$, h_i, $\mathbb{N}$ (index)

Lemma 14.2 (Diameter of $\mathcal{H}_{50}$)**.** $\operatorname{diam}(\mathcal{H}_{50}) = 246$.

Uses (statement): 14.1, 3.18

Uses(notation): $\mathcal{H}_{50}$, $\operatorname{diam}(\mathcal{H})$ (index)

Proof. The elements of $\mathcal{H}_{50}$ are listed in increasing order in Definition 14.1, so $\min \mathcal{H}_{50} = 0$ and $\max \mathcal{H}_{50} = 246$. By Definition 3.18, $\operatorname{diam}(\mathcal{H}_{50}) = \max \mathcal{H}_{50} - \min \mathcal{H}_{50} = 246 - 0 = 246$. □

Lemma 14.3 (Admissibility of $\mathcal{H}_{50}$)**.** $\mathcal{H}_{50}$ *is admissible.*

Uses (statement): 14.1, 3.15

Uses(notation): $\mathcal{H}_{50}$, $\mathbb{P}$ (index)

Proof. By Lemma 3.16 it suffices to show that $\#\{h \bmod p : h \in \mathcal{H}_{50}\} < p$ for every prime p.

For $p > 50$ this is automatic: the reduction map $\mathcal{H}_{50} \to \mathbb{Z}/p\mathbb{Z}$ has image of cardinality at most $\#\mathcal{H}_{50} = 50 < p$. Thus only the fifteen primes

$$p \in \{2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47\}$$

require checking, and for each of them the reduction of the explicit list of Definition 14.1 is a finite computation. It omits a residue class in every case; for instance every element of $\mathcal{H}_{50}$ is even, so the image modulo 2 is $\{0\}$, and modulo 3 the image is $\{0, 1\}$, missing 2. □

Uses (proof): 14.1, 3.16

15 Assembly

The two bounds below are two cases of one argument. Its variational input is a lower bound for the variational supremum; its arithmetic input is an admissible tuple of the corresponding size. Taking the standard simplex, $k = 105$, the bound $M_{105} > 4$, and the tuple $\mathcal{H}_{105}$ of diameter 600 gives Theorem 4. Taking the ε-enlarged simplex, $k = 50$, a numerical certificate of $\mathcal{Q}_\varepsilon > 4$, and the tuple $\mathcal{H}_{50}$ of diameter 246 gives Theorem 8. In the second case the variational input is a finite, exact verification (Definition 15.1, Theorem 7).

Definition 15.1 (A certificate at $k = 50$)**.** A 50-*certificate* is a datum consisting of a rational number ε with $0 \le \varepsilon \le 1$, finitely many elements $b_1, \dots, b_n$ of the certificate basis of Definition 13.3 in the 50 variables $t_1, \dots, t_{50}$, and a rational vector $a = (a_1, \dots, a_n) \in \mathbb{Q}^n$, such that the polynomial $F_a = \sum_{i=1}^{n} a_i b_i$ of Definition 13.4 satisfies the strict inequality

$$4\, I_{50}(F_a) \;<\; \sum_{\ell=1}^{50} J_\varepsilon^{(\ell)}(F_a).$$

Uses (statement): 13.3, 13.4, 5.6, 5.7

Uses(notation): ε, $I_k(F)$, $J_\varepsilon^{(i)}(F)$, k (index)

Both sides of the displayed inequality are rational numbers determined by the finite datum $(\varepsilon, (b_i)_{i \le n}, a)$ through the closed-form pairings of Lemmas 13.13 and 13.14, so the condition defining a 50-certificate is a single exact comparison of two rationals. The analytic argument uses only that a 50-certificate exists, not the particular ε, basis, or coefficient vector.

Theorem 5 (DHL$[50, 2]$ from a certificate)**.** *Suppose the primes have level of distribution ϑ for every $\vartheta < \frac{1}{2}$ (Theorem 1), and suppose a 50-certificate exists (Definition 15.1). Then* DHL$[50, 2]$ *holds: for every admissible 50-tuple $\mathcal{H} = \{h_1 < \cdots < h_{50}\}$ there are infinitely many $n \in \mathbb{N}$ such that at least two of $n + h_1, \dots, n + h_{50}$ are prime.*

Lean:

- `Gaps246.thm_dhl`

Uses (statement): 15.1, 3.15, 1

Uses(notation): $\mathcal{H}$, h_i, k, $\mathbb{N}$, $\mathbb{P}$, ϑ, ε, $\mathcal{T}_\varepsilon$, $\mathcal{Q}_\varepsilon(F)$, F, ρ (index)

Proof. Fix a 50-certificate $(\varepsilon, (b_i)_{i \le n}, a)$ and set $F_0 := F_a$, so that

$$4\, I_{50}(F_0) \;<\; \sum_{\ell=1}^{50} J_\varepsilon^{(\ell)}(F_0). \tag{15.2}$$

The functional I_{50} is an integral of a square, so $I_{50}(F_0) \ge 0$; and $I_{50}(F_0) = 0$ would force $F_0 = 0$ almost everywhere, hence $J_\varepsilon^{(\ell)}(F_0) = 0$ for every ℓ, contradicting (15.2). Therefore $I_{50}(F_0) > 0$ and the witness ratio $Q := \mathcal{Q}_\varepsilon(F_0)$ of Definition 5.8 is defined and satisfies $Q > 4$ by (15.2).

Apply Lemma 12.28 with this $\varepsilon \in [0, 1]$ and this $Q > 4$: it produces a level $\vartheta < \frac{1}{2}$ with $2/\vartheta < Q$ and $1 + \varepsilon < 1/\vartheta$, at which the primes have level of distribution ϑ. Note $\vartheta > 0$, since $1/\vartheta > 1 + \varepsilon \ge 1 > 0$.

Apply Lemma 12.27 at this ϑ: since $\mathcal{Q}_\varepsilon(F_0) > 2/\vartheta$ and $1 + \varepsilon < 1/\vartheta$, it produces a smooth F supported on the enlarged simplex $\mathcal{T}_\varepsilon$ and a margin $\delta \in (0, \vartheta/2)$ with

$$\int_{\mathcal{T}_\varepsilon} F^2 \;<\; \Big(\frac{\vartheta}{2} - \delta\Big) \sum_{m=1}^{50} J_\varepsilon^{(m)}(F).$$

Now let $\mathcal{H} = \{h_1 < \cdots < h_{50}\}$ be an arbitrary admissible 50-tuple. The data ε, F, ϑ, δ satisfy every hypothesis of Proposition 12.26 at $k = 50$ with $\rho = 1$: $\varepsilon \in [0, 1]$; F is smooth and supported on $\mathcal{T}_\varepsilon$; $\vartheta \in (0, 1)$, $\vartheta < \frac{1}{2}$, and the primes have level of distribution ϑ; $\delta \in (0, \vartheta/2)$; $1 + \varepsilon < 1/\vartheta$; and the displayed inequality is exactly the witness inequality with $\rho = 1$. Proposition 12.26 therefore gives infinitely many n with at least two of $n + h_1, \dots, n + h_{50}$ prime. As $\mathcal{H}$ was arbitrary, this is DHL$[50, 2]$. □

Uses (proof): 5.8, 12.28, 12.27, 12.26

Theorem 6 ($H_1 \le 246$, conditional on a certificate)**.** *Suppose the primes have level of distribution ϑ for every $\vartheta < \frac{1}{2}$ (Theorem 1), and suppose a 50-certificate exists (Definition 15.1). Then*

$$\liminf_{m\to\infty}(p_{m+1}-p_m) \;\le\; 246.$$

Uses (statement): 1, 15.1

Uses(notation): p_n, ϑ (index)

Proof. The tuple $\mathcal{H}_{50}$ is admissible (Lemma 14.3) and has $\operatorname{diam}(\mathcal{H}_{50}) = 246$ (Lemma 14.2). Both hypotheses of Theorem 5 hold, so DHL[50, 2] holds, and applying it to $\mathcal{H}_{50}$ shows that

$$\mathcal{N} \;:=\; \{n \in \mathbb{N} :\ \text{at least two of the shifts } n+h,\ h \in \mathcal{H}_{50},\ \text{are prime}\}$$

is infinite.

For $n \in \mathcal{N}$ every shift $n+h$ with $h \in \mathcal{H}_{50}$ lies in $[\,n+\min\mathcal{H}_{50},\ n+\max\mathcal{H}_{50}\,]$, an interval of length $\operatorname{diam}(\mathcal{H}_{50}) = 246$. Hence $\mathcal{N}' := \{n+\min\mathcal{H}_{50} : n \in \mathcal{N}\}$ is an infinite set of natural numbers such that $[n', n'+246]$ contains at least two distinct primes for every $n' \in \mathcal{N}'$.

Lemma 10.4 applied to $\mathcal{N}'$ with $r = 2$ and $L = 246$ gives the conclusion. □

Uses (proof): 5, 14.3, 14.2, 10.4

Theorem 7 (Existence of a 50-certificate)**.** *A 50-certificate exists (Definition 15.1).*

Uses (statement): 15.1

Uses(notation): ε, $I_k(F)$, $J_\varepsilon^{(i)}(F)$, $\mathcal{Q}_\varepsilon(F)$ (index)

Proof. Take $\varepsilon = \frac{1}{25}$, which is rational and lies in $[0,1]$, and take for $b_1, \dots, b_n$ the certificate basis of Definition 13.3 at $k = 50$, $\varepsilon = \frac{1}{25}$ and total-degree bound 27; here $n = 2526$. Proposition 13.15 supplies a vector $a^* \in \mathbb{Q}^n$ with $I_{50}(F_{a^*}) > 0$ and $\mathcal{Q}_{1/25}(F_{a^*}) > 4.0043$. By Definition 5.8 this reads

$$\sum_{\ell=1}^{50} J^{(\ell)}_{1/25}(F_{a^*}) \;>\; 4.0043\, I_{50}(F_{a^*}) \;>\; 4\, I_{50}(F_{a^*}),$$

the last step because $I_{50}(F_{a^*}) > 0$. Thus $(\frac{1}{25}, (b_i)_{i\le n}, a^*)$ is a 50-certificate.

The verification is finite and exact. By Lemmas 13.13 and 13.14 the two sides are the rational quadratic forms $(a^*)^{\top} A_I\, a^*$ and $(a^*)^{\top} A_J\, a^*$ in the pairing matrices of Definitions 13.7 and 13.9, whose entries are given in closed form by Lemmas 13.11 and 13.12. With a^* taken over a common denominator, clearing denominators turns the required inequality into a single comparison of two integers, namely $10^4 \cdot (a^*)^{\top} A_J\, a^* > 4\,0043 \cdot (a^*)^{\top} A_I\, a^*$; its exact value is $\mathcal{Q}_{1/25}(F_{a^*}) = 4.00430625\ldots$, which clears 4.0043, and a fortiori 4, strictly. No analytic input enters at this point: the statement is a finite assertion about rational numbers. □

Uses (proof): 13.15, 5.8, 13.13, 13.14, 13.7, 13.9, 13.11, 13.12, 13.3

Theorem 8 ($H_1 \le 246$)**.** *Suppose the primes have level of distribution ϑ for every $\vartheta < \frac{1}{2}$ (Theorem 1). Then*

$$\liminf_{m\to\infty}(p_{m+1}-p_m) \;\le\; 246.$$

Lean:

- `Gaps246.thm_main`

Uses (statement): 1

Uses(notation): p_n, ϑ (index)

Proof. Theorem 7 provides a 50-certificate, which is the second hypothesis of Theorem 6; the first is the present hypothesis. Theorem 6 then gives the conclusion. □

Uses (proof): 6, 7

References

[1] J. Maynard, *Small gaps between primes*, Ann. of Math. (2) **181** (2015), no. 1, 383–413.

[2] D. H. J. Polymath, *Variants of the Selberg sieve, and bounded intervals containing many primes*, Res. Math. Sci. **1** (2014), Art. 12, arXiv:1407.4897.

Notation index

Notation	Definition
p_n	$p_n \in \mathbb{P},\ p_1 < p_2 < p_3 < \cdots,\ \{p_n : n \in \mathbb{Z}_{\geq 1}\} = \mathbb{P}$
$\mathcal{H}$	$\mathcal{H} \subset \mathbb{Z},\ \lvert\mathcal{H}\rvert < \infty,\ \forall q \text{ prime} :\ \#\{h \bmod q : h \in \mathcal{H}\} < q$
h_i	$h_i \in \mathbb{Z},\ \mathcal{H} = \{h_1, \ldots, h_{\lvert\mathcal{H}\rvert}\},\ h_1 < h_2 < \cdots < h_{\lvert\mathcal{H}\rvert}$ **Uses:** $\mathcal{H}$
k	$k := \lvert\mathcal{H}\rvert \in \mathbb{Z}_{\geq 2}$ **Uses:** $\mathcal{H}$
w_n	$w_n := \Big(\sum_{\substack{d_i \mid n+h_i \\ 1 \leq i \leq k}} \lambda(d_1, \ldots, d_k)\Big)^2, \quad n \in \mathbb{Z}_{>0}$ **Uses:** h_i, k, $\lambda(d_1, \ldots, d_k)$
N	$N \in \mathbb{N},\ N \to \infty$ **Uses:** $\mathbb{N}$
$S(N, \rho)$	$S(N,\rho) := \sum_{N < n \leq 2N} \Big(\sum_{i=1}^{k} \chi_{\mathbb{P}}(n + h_i) - \rho\Big)\, w_n, \quad \rho \in \mathbb{R}_{\geq 0}$ **Uses:** h_i, k, w_n, N, $\chi_{\mathbb{P}}$
$\chi_{\mathbb{P}}$	$\chi_{\mathbb{P}} : \mathbb{Z}_{\geq 1} \to \{0, 1\}, \quad \chi_{\mathbb{P}}(n) = 1 \iff n \geq 2$ and $(d \in \mathbb{Z}_{\geq 1},\ d \mid n \implies d \in \{1, n\})$
ρ	$\rho \in \mathbb{R}_{>0},\ \liminf_{N \to \infty} \big(S_2(\lambda) - \rho\, S_1(\lambda)\big) > 0$ **Uses:** N, $S_1(\lambda)$, $S_2(\lambda)$, $\lambda(d_1, \ldots, d_k)$
$S_1(\lambda)$	$S_1(\lambda) := \sum_{\substack{N < n \leq 2N \\ n \equiv v_0 \pmod W}} w_n, \quad \lambda : \mathbb{N}^k \to \mathbb{R}$ **Uses:** w_n, N, v_0, W, $\lambda(d_1, \ldots, d_k)$, k
$S_2(\lambda)$	$S_2(\lambda) := \sum_{\substack{N < n \leq 2N \\ n \equiv v_0 \pmod W}} \Big(\sum_{i=1}^{k} \chi_{\mathbb{P}}(n + h_i)\Big)\, w_n, \quad \lambda : \mathbb{N}^k \to \mathbb{R}$ **Uses:** N, v_0, W, k, $\chi_{\mathbb{P}}$, h_i, w_n, $\lambda(d_1, \ldots, d_k)$
$\lambda(d_1, \ldots, d_k)$	$\lambda(d_1, \ldots, d_k) : \mathbb{N}^k \to \mathbb{R},\ d_i \in \mathbb{Z}_{\geq 1},\ \lambda(d_1, \ldots, d_k) = 0$ unless $\prod_{i=1}^{k} d_i \leq R,\ \gcd\Big(\prod_{i=1}^{k} d_i,\ W\Big) = 1$, and $\prod_{i=1}^{k} d_i$ is squarefree **Uses:** k, R, W
$\mathcal{R}_k$	$\mathcal{R}_k := \{(x_1, \ldots, x_k) \in [0,1]^k : x_1 + \cdots + x_k \leq 1\}$ **Uses:** k

Notation	Definition
$\lambda_0(d_1,\dots,d_k)$	$\lambda_0(d_1,\dots,d_k):\mathbb{N}^k\to\mathbb{R},\ \lambda_0(d_1,\dots,d_k):=$ $\Big(\prod_{i=1}^k \mu(d_i)\,d_i\Big)\sum_{\substack{r_1,\dots,r_k\\ d_i\mid r_i\ \forall i\\ (r_i,W)=1\ \forall i}} \frac{\mu(\prod_{i=1}^k r_i)^2}{\prod_{i=1}^k \varphi(r_i)}\,F\Big(\frac{\log r_1}{\log R},\dots,\frac{\log r_k}{\log R}\Big)$ when $\gcd(\prod d_i, W)=$ 1, $\lambda_0(d_1,\dots,d_k):=0$ otherwise **Uses:** k, F, R, W, μ, φ, $\mathcal{R}_k$
$\lambda_y(d_1,\dots,d_k)$	$\lambda_y:\mathbb{N}^k\to\mathbb{R},\ \lambda_y(d_1,\dots,d_k):=\Big(\prod_{i=1}^k \mu(d_i)\,d_i\Big)\sum_{\substack{r_1,\dots,r_k\\ d_i\mid r_i\ \forall i}} \frac{y(r_1,\dots,r_k)}{\prod_{i=1}^k \varphi(r_i)},\ y:$ $\mathbb{N}^k\to\mathbb{R}$ of finite support **Uses:** k, $y(r_1,\dots,r_k)$, μ, φ
R	$R:=N^{\vartheta/2-\delta}$ **Uses:** N, ϑ, δ
F	$F:[0,1]^k\to\mathbb{R}$, Riemann-integrable, $F(t_1,\dots,t_k)=$ 0 unless $t_1+\cdots+t_k\le 1$ **Uses:** k
μ	$\mu:\mathbb{Z}_{\ge 1}\to\{-1,0,1\},\quad \mu(1):=1,\quad \mu(n):=(-1)^r$ if $n=$ $q_1\cdots q_r$ with q_i distinct primes, $\quad \mu(n):=0$ if $\exists q$ prime $: q^2\mid n$
$I_k(F)$	$I_k(F):=\int_{\mathcal{R}_k} F(t_1,\dots,t_k)^2\,dt_1\cdots dt_k$ **Uses:** k, F, $\mathcal{R}_k$
$J_k^{(m)}(F)$	$J_k^{(m)}(F):=\int_{\mathcal{R}_k^{(m)}}\Big(\int_0^\infty F(t_1,\dots,t_k)\,dt_m\Big)^2\prod_{i\ne m}dt_i,\quad m\in\{1,\dots,k\}$ **Uses:** k, F, $\mathcal{R}_k^{(m)}$
$\mathcal{R}_k^{(m)}$	$\mathcal{R}_k^{(m)}:=\{(t_1,\dots,t_{m-1},t_{m+1},\dots,t_k)\in[0,1]^{k-1}:\sum_{\substack{i=1\\ i\ne m}}^k t_i\le 1\},\ m\in$ $\{1,\dots,k\}$ **Uses:** k
M_k	$M_k:=$ $\sup\Big\{\sum_{m=1}^k J_k^{(m)}(F)\Big/I_k(F)\ \Big|\ F\in C_c^\infty([0,\infty)^k;\mathbb{R}),\ F$ symmetric, $\operatorname{supp}F\subset\mathcal{R}_k,\ I_k(F)\ne 0\Big\}$ $[0,\infty]$ **Uses:** k, F, $\mathcal{R}_k$, $I_k(F)$, $J_k^{(m)}(F)$
M_{105}	$M_{105}:=M_k\big\vert_{k=105}$ **Uses:** k, M_k
r_k	$r_k:=\lceil\vartheta\,M_k/2\rceil\in\mathbb{Z}_{>0}$ **Uses:** ϑ, M_k
v_0	$v_0\in\{0,1,\dots,W-1\}$, $\gcd(v_0+h_i,W)=1$ for all $i\in\{1,\dots,k\}$ **Uses:** W, h_i, k
$\mathbf{a}^*$	$\mathbf{a}^*=(a_1^*,\dots,a_{42}^*)\in\mathbb{Q}^{42}$
$\mathbb{N}$	$\mathbb{N}:=\{1,2,3,\dots\}$
$\mathbb{P}$	$\mathbb{P}:=\{n\in\mathbb{Z}_{\ge 2}:\forall d\in\mathbb{Z}_{\ge 1},\ d\mid n\implies d\in\{1,n\}\}$
(a,b)	$(a,b):=\max\{d\in\mathbb{Z}_{\ge 1}:d\mid a\text{ and }d\mid b\},\quad a,b\in\mathbb{Z},\ (a,b)\ne(0,0)$
$[a,b]$	$[a_1,\dots,a_n]:=\operatorname{lcm}(a_1,\dots,a_n)=\min\{m\in\mathbb{Z}_{\ge 1}:a_i\mid m\ \forall i\},\quad a_i\in\mathbb{Z}_{\ge 1}$
$\lfloor x\rfloor$	$\lfloor x\rfloor:=\max\{n\in\mathbb{Z}:n\le x\},\ x\in\mathbb{R}$
$\lceil x\rceil$	$\lceil x\rceil:=\min\{n\in\mathbb{Z}:n\ge x\},\ x\in\mathbb{R}$
φ	$\varphi:\mathbb{N}\to\mathbb{N},\ \varphi(n):=\#\{a\in\{1,\dots,n\}:(a,n)=1\}$ **Uses:** $\mathbb{N}$, (a,b)
τ_r	$\tau_r(n):=\#\{(d_1,\dots,d_r)\in\mathbb{N}^r:d_1d_2\cdots d_r=n\},\ n\in\mathbb{N},\ r\in\mathbb{N}$ **Uses:** $\mathbb{N}$
$f\ll g$	$f\ll g\iff\exists C=C(k,\mathcal{H})>0:\lvert f\rvert\le C\lvert g\rvert,\ C$ independent of N **Uses:** k, $\mathcal{H}$, N
$O(g)$	$f(x)=O(g(x))\iff\exists C>0,\ x_0\in\mathbb{R}:\lvert f(x)\rvert\le C\,\lvert g(x)\rvert\ \forall x\ge x_0$
$o(g)$	$f=o(g)\iff\lim_{N\to\infty}f(N)/g(N)=0$ **Uses:** N

Notation	Definition
$\operatorname{diam}(\mathcal{H})$	$\operatorname{diam}(\mathcal{H}) := \max_{i,j}(h_i - h_j) = h_k - h_1$ **Uses:** $\mathcal{H}$, h_i, k
$\pi(x;q,a)$	$\pi(x;q,a) := \#\{p \le x : p \in \mathbb{P},\ p \equiv a \pmod q\}, \quad x \in \mathbb{R}_{\ge 2},\ q \in \mathbb{Z}_{\ge 1},\ a \in \mathbb{Z}$ **Uses:** $\mathbb{P}$
$\pi(x)$	$\pi(x) := \#\{p \le x : p \in \mathbb{P}\},\ x \in \mathbb{R}$ **Uses:** $\mathbb{P}$
ϑ	$\vartheta \in (0,1),\ \forall A > 0 : \sum_{q \le x^{\vartheta}} \max_{(a,q)=1} \lvert\pi(x;q,a) - \pi(x)/\varphi(q)\rvert \ll_A x(\log x)^{-A}$ **Uses:** $\pi(x;q,a)$, $\pi(x)$, φ
D_0	$D_0 := \log\log\log N$ **Uses:** N
W	$W := \prod_{\substack{p \le D_0 \\ p \text{ prime}}} p$ **Uses:** D_0
$\vartheta(x)$	$\vartheta(x) := \sum_{\substack{p \le x \\ p \in \mathbb{P}}} \log p$ **Uses:** $\mathbb{P}$
δ	$\delta \in \mathbb{R},\ 0 < \delta < \vartheta/2$ **Uses:** ϑ
$\mathbf{d}$	$\mathbf{d} := (d_1, \dots, d_k) \in \mathbb{Z}_{>0}^k$ **Uses:** k
$\sum_{\mathbf{d}}$	$\sum_{\mathbf{d}} := \sum_{d_1 \ge 1} \sum_{d_2 \ge 1} \cdots \sum_{d_k \ge 1}$ **Uses:** k, $\mathbf{d}$
$\mathbf{e}$	$\mathbf{e} := (e_1, \dots, e_k) \in \mathbb{Z}_{\ge 1}^k$ **Uses:** k
$\mathbf{r}$	$\mathbf{r} := (r_1, \dots, r_k) \in \mathbb{Z}_{\ge 0}^k$ **Uses:** k
$\mathbf{u}$	$\mathbf{u} := (u_1, \dots, u_k) \in \mathbb{Z}_{\ge 1}^k,\ u_i \mid (d_i, e_i),\ \mathbf{d} = (d_1, \dots, d_k),\ \mathbf{e} = (e_1, \dots, e_k)$ **Uses:** k, $\mathbf{d}$, $\mathbf{e}$, (a,b)
$\mathbf{a}$	$\mathbf{a} := (a_1, \dots, a_k) \in \mathbb{Z}_{>0}^k,$ where each $a_j = u_j \prod_{i \ne j} s_{j,i}$ **Uses:** k, u_i, $s_{i,j}$
$\mathbf{b}$	$\mathbf{b} := (b_1, \dots, b_k) \in \mathbb{Z}_{>0}^k,$ where each $b_j = u_j \prod_{i \ne j} s_{i,j}$ **Uses:** k, u_i, $s_{i,j}$
$S_2^{(m)}(\lambda)$	$S_2^{(m)}(\lambda) := \sum_{\substack{N < n \le 2N \\ n \equiv v_0 \pmod W}} \chi_{\mathbb{P}}(n + h_m)\, w_n, \quad \lambda : \mathbb{N}^k \to \mathbb{R},\ 1 \le m \le k$ **Uses:** N, v_0, W, $\chi_{\mathbb{P}}$, h_i, w_n, k, $\lambda(d_1, \dots, d_k)$
$y(r_1, \dots, r_k)$	$y : \mathbb{N}^k \to \mathbb{R},\ y := y_\lambda$ (the y-transform of λ, Notation `y_lambda`), $\lambda : \mathbb{N}^k \to \mathbb{R}$ of finite support **Uses:** k, $\lambda(d_1, \dots, d_k)$, $y_\lambda(r_1, \dots, r_k)$, $\mathbb{N}$
$y_\lambda(r_1, \dots, r_k)$	$y_\lambda : \mathbb{N}^k \to \mathbb{R},\ y_\lambda(r_1, \dots, r_k) := \left(\prod_{i=1}^k \mu(r_i)\,\varphi(r_i)\right) \sum_{\substack{\mathbf{d} \in \mathbb{Z}_{>0}^k \\ r_i \mid d_i\ \forall i}} \frac{\lambda(d_1, \dots, d_k)}{\prod_{i=1}^k d_i}, \quad \lambda : \mathbb{N}^k \to \mathbb{R}$ of finite support **Uses:** k, $\lambda(d_1, \dots, d_k)$, μ, φ, $\mathbf{d}$, $\mathbf{r}$, $\mathbb{N}$
$y_{\max}$	$y_{\max} := \sup_{\mathbf{r} \in \mathbb{N}^k} \lvert y(r_1, \dots, r_k)\rvert$ **Uses:** $\mathbf{r}$, $\mathbb{N}$, k, $y(r_1, \dots, r_k)$
g	$g := \varphi * \mu \colon \mathbb{N} \to \mathbb{N}$ (Dirichlet convolution), equivalently the multiplicative function with $g(p) = p - 2$ and $g(p^a) = p^{a-2}(p-1)^2$ for $a \ge 2,\ p \in \mathbb{P}$ **Uses:** $\mathbb{N}$, $\mathbb{P}$, φ, μ
$y^{(m)}(r_1, \dots, r_k)$	$y^{(m)} : \mathbb{N}^k \to \mathbb{R},\ y^{(m)}(r_1, \dots, r_k) := \left(\prod_{i \ne m} \mu(r_i)\, g(r_i)\right) \sum_{\substack{\mathbf{d}:\ r_i \mid d_i\ \forall i \ne m \\ d_m = 1}} \frac{\lambda(d_1, \dots, d_k)}{\prod_{i \ne m} \varphi(d_i)}, \quad r_m = 1,\ m \in \{1, \dots, k\}$ **Uses:** k, $\lambda(d_1, \dots, d_k)$, μ, φ, $\mathbf{d}$, g, $\mathbb{N}$
$y_{\max}^{(m)}$	$y_{\max}^{(m)} := \sup_{\mathbf{r} \in \mathbb{N}^k} \lvert y^{(m)}(r_1, \dots, r_k)\rvert$ **Uses:** k, $\mathbb{N}$, $\mathbf{r}$, $y^{(m)}(r_1, \dots, r_k)$
$\mathfrak{S}(\gamma)$	$\mathfrak{S}(\gamma) := \prod_{p \in \mathbb{P}} \left(1 - \frac{\gamma(p)}{p}\right)^{-1} \left(1 - \frac{1}{p}\right)$ **Uses:** $\mathbb{P}$, γ

Notation	Definition
γ	$\gamma : \mathbb{N} \to \mathbb{R}_{\geq 0},\ \gamma(1) = 1,\ \gamma(mn) = \gamma(m)\gamma(n)$ if $(m,n) = 1,\ \gamma(p) \in [0, p)\ \forall p \in \mathbb{P}$ **Uses:** $\mathbb{N}$, $\mathbb{P}$, (a,b)
$E(N,q)$	$E(N,q) := 1 + \sup_{(a,q)=1} \left\lvert \sum_{\substack{N<n\leq 2N \\ n\equiv a \pmod q}} \chi_{\mathbb{P}}(n) - \varphi(q)^{-1} \sum_{N<n\leq 2N} \chi_{\mathbb{P}}(n) \right\rvert,\ q \in \mathbb{N}$ **Uses:** $\chi_{\mathbb{P}}$, φ, $\mathbb{N}$, N, (a,b)
X_N	$X_N := \sum_{N<n\leq 2N} \chi_{\mathbb{P}}(n)$ **Uses:** N, $\chi_{\mathbb{P}}$
$\omega(n)$	$\omega(n) := \#\{p \in \mathbb{P} : p \mid n\},\ n \in \mathbb{N}$ **Uses:** $\mathbb{P}$, $\mathbb{N}$
τ	$\tau(n) := \#\{d \in \mathbb{Z}_{\geq 1} : d \mid n\},\ n \in \mathbb{Z}_{\geq 1}$
$\zeta(2)$	$\zeta(2) := \sum_{n\geq 1} \frac{1}{n^2}$
$A(M)$	$A(M) := \sum_{\substack{e\geq 1 \\ (e,M)=1}} \mu(e)/e^2, \quad M \in \mathbb{Z}_{\geq 1}$ **Uses:** μ, $\mathbb{P}$, (a,b)
T_d	$T_d := \sum_{\substack{v\leq R/d \\ (v,dW)=1}} \frac{\mu(v)^2}{v}, \quad d \in \mathbb{Z}_{\geq 1}$ **Uses:** R, W, μ, (a,b)
$B(M)$	$B(M) := \sum_{\substack{e\geq 1 \\ (e,M)=1}} \frac{\mu(e)\log e}{e^2}, \quad M \in \mathbb{Z}_{\geq 1}$ **Uses:** μ, (a,b)
P_j	$P_j := \sum_{i=1}^{k} t_i^j, \quad j \in \mathbb{Z}_{\geq 1},\ (t_1,\dots,t_k) \in \mathbb{R}_{\geq 0}^k,\ \sum_{i=1}^{k} t_i \leq 1$ **Uses:** k
$\Gamma(s)$	$\Gamma(s) := \int_0^\infty t^{s-1}e^{-t}\,dt, \quad s \in \mathbb{R}_{>0}$
$B(a,b)$	$B(a,b) := \int_0^1 t^{a-1}(1-t)^{b-1}\,dt, \quad a,b \in \mathbb{Z}_{\geq 1}$ **Uses:** $\Gamma(s)$
$\lambda_{\max}$	$\lambda_{\max} := \sup_{\mathbf{d}} \lvert\lambda(d_1,\dots,d_k)\rvert$ **Uses:** $\lambda(d_1,\dots,d_k)$, $\mathbf{d}$
$q(\mathbf{d},\mathbf{e})$	$q(\mathbf{d},\mathbf{e}) := W \prod_{i=1}^{k} [d_i, e_i]$ **Uses:** k, W, $\mathbf{d}$, $\mathbf{e}$, $[a,b]$
$\sum'$	$\sum'_{\mathbf{d},\mathbf{e}} f(\mathbf{d},\mathbf{e}) := \sum_{\substack{\mathbf{d},\mathbf{e}\in\mathbb{N}^k \\ (W,[d_1,e_1]\cdots[d_k,e_k])=1 \\ [d_i,e_i] \text{ pairwise coprime}}} f(\mathbf{d},\mathbf{e})$ **Uses:** k, $\mathbb{N}$, W, $\mathbf{d}$, $\mathbf{e}$, (a,b), $[a,b]$
u_i	$u_i \in \mathbb{Z}_{\geq 1},\ u_i \mid (d_i,e_i),\ [d_i,e_i] = d_i e_i/u_i,\ \mathbf{d} = (d_i),\ \mathbf{e} = (e_i)$ **Uses:** (a,b), $[a,b]$, $\mathbf{d}$, $\mathbf{e}$
$s_{i,j}$	$s_{i,j} \in \mathbb{Z}_{\geq 1},\ 1 \leq i,j \leq k,\ i \neq j; \quad \mathbf{1}_{(d_i,e_j)=1} = \sum_{s_{i,j}\mid(d_i,e_j)} \mu(s_{i,j})$ **Uses:** k, $\mathbf{d}$, $\mathbf{e}$, μ, (a,b)
$\sum^*$	$\sum^*_{s_{1,2},\dots,s_{k,k-1}} f := \sum_{\substack{s_{i,j}\geq 1,\ 1\leq i\neq j\leq k \\ (s_{i,j},u_i)=1,\ (s_{i,j},u_j)=1 \\ (s_{i,j},s_{i,a})=1\ \forall a\notin\{i,j\} \\ (s_{i,j},s_{b,j})=1\ \forall b\notin\{i,j\}}} f\big((s_{i,j})_{i\neq j}\big)$ **Uses:** k, $s_{i,j}$, u_i, (a,b)
$H(z)$	$H(z) := \sum_{d<z} \mu(d)^2 g_*(d),\ z \in \mathbb{R}_{>0}$ **Uses:** μ, g_*
$H_p(y)$	$H_p(y) := \sum_{\substack{d<y \\ \mu(d)^2=1 \\ (d,p)=1}} \mu(d)^2 g_*(d), \quad p \in \mathbb{P},\ y \in \mathbb{R}_{>0}$ **Uses:** μ, $\mathbb{P}$, (a,b), g_*
A_1	$A_1 \in (0,1),\ \gamma(p)/p \leq 1 - A_1$ for all $p \in \mathbb{P}$ **Uses:** γ, $\mathbb{P}$
A_2	$A_2 \in \mathbb{R}_{>0},\ \sum_{\substack{w\leq p\leq z \\ p\in\mathbb{P}}} \gamma(p)\frac{\log p}{p} - \log(z/w) \leq A_2,\ \forall\, 2 \leq w \leq z$ **Uses:** γ, $\mathbb{P}$

Notation	Definition
L	$L \in \mathbb{R}_{\geq 0},\ \sum_{\substack{p \leq z \\ p \in \mathbb{P}}} \frac{\log p}{p} - \log z \geq -L,\ \forall z \geq 2$ **Uses:** $\mathbb{P}$
g_*	$g_* \colon \mathbb{N} \to \mathbb{R}$ totally multiplicative, $\quad g_*(p) := \gamma(p)/(p - \gamma(p))$ for $p \in \mathbb{P}$ **Uses:** γ, $\mathbb{N}$, $\mathbb{P}$
$T(z)$	$T(z) := \sum_{\substack{p \leq z \\ p \in \mathbb{P}}} \frac{\gamma(p) \log p}{p},\ z \in \mathbb{R}_{\geq 2}$ **Uses:** $\mathbb{P}$, γ
G	$G : [0,1] \to \mathbb{R},\ G \in C^\infty([0,1])$
$G_{\max}$	$G_{\max} := \sup_{t \in [0,1]} \big(\lvert G(t)\rvert + \lvert G'(t)\rvert\big)$ **Uses:** G
$F_{\max}$	$F_{\max} := \sup_{(t_1,\dots,t_k) \in [0,1]^k} \Big(\lvert F(t_1,\dots,t_k)\rvert + \sum_{i=1}^k \big\lvert \frac{\partial F}{\partial t_i}(t_1,\dots,t_k)\big\rvert\Big)$ **Uses:** k, F
γ_W	$\gamma_W : \mathbb{N} \to \{0,1\}$ multiplicative, $\gamma_W(p) := \mathbf{1}[p \nmid W];\ \gamma_W(n) = \mathbf{1}[(n, W) = 1]$ **Uses:** $\mathbb{N}$, W, (a, b)
$F^{(m)}_{r_1,\dots,r_k}$	$F^{(m)}_{r_1,\dots,r_k} := \int_0^1 F(x_1,\dots,x_{m-1},t_m,x_{m+1},\dots,x_k)\,dt_m, \quad x_i := \log r_i / \log R$ for $i \neq m,\ m \in \{1,\dots,k\}$ **Uses:** F, R, k
F_P	$F_P(t_1,\dots,t_k) := P(t_1,\dots,t_k) \cdot \mathbf{1}[(t_1,\dots,t_k) \in \mathcal{R}_k], \quad P \in \mathbb{R}[t_1,\dots,t_k]$ **Uses:** k, $\mathcal{R}_k$
$\mathcal{B}_d$	$\mathcal{B}_d := \{(1-P_j)_1^b\,(P_j)_2^c : b,c \in \mathbb{Z}_{\geq 0},\ b + 2c \leq d\}, \quad d \in \mathbb{Z}_{\geq 0}$ **Uses:** P_j
$G_{b,j}(x)$	$G_{b,j}(x) := b! \sum_{r=1}^b \binom{x}{r} \sum_{\substack{b_1+\cdots+b_r=b \\ b_i \geq 1}} \prod_{i=1}^r \frac{(jb_i)!}{b_i!}, \quad b,j \in \mathbb{Z}_{\geq 0},\ x \in \mathbb{R}$
$\mathbf{b}^*$	$\mathbf{b}^* = (b_1^*,\dots,b_{42}^*) \in \mathbb{Z}^{42}_{\geq 0}, \quad \#\mathcal{B}_d\big\rvert_{d=11} = 42$ **Uses:** $\mathcal{B}_d$
$\mathbf{c}^*$	$\mathbf{c}^* = (c_1^*,\dots,c_{42}^*) \in \mathbb{Z}^{42}_{\geq 0}, \quad \{(\mathbf{b}_i^*, c_i^*) : 1 \leq i \leq 42\}$ enumerates $\mathcal{B}_{11}$ **Uses:** $\mathbf{b}^*$, $\mathcal{B}_d$
P^*	$P^* := \sum_{i=1}^{42} a_i^*\,(1-P_1)^{b_i^*}\,P_2^{c_i^*} \in \mathbb{R}[t_1,\dots,t_{105}]$ **Uses:** P_j, $\mathbf{a}^*$, $\mathbf{b}^*$, $\mathbf{c}^*$
N^*	$N^* \in \mathbb{Z}_{>0}, \quad \frac{N^*}{D^*} = \frac{\sum_{m=1}^{105} J_k^{(m)}(F_P)}{I_k(F_P)}\Big\rvert_{k=105,\ P=P^*}, \quad (N^*, D^*) = 1$ **Uses:** k, $I_k(F)$, $J_k^{(m)}(F)$, F_P, P^*, D^*, (a,b)
D^*	$D^* \in \mathbb{Z}_{>0}, \quad \sum_{m=1}^k J_k^{(m)}(F_P)/I_k(F_P)\big\rvert_{k=105,\ P=P^*} = N^*/D^*, \quad (N^*, D^*) = 1$ **Uses:** k, $I_k(F)$, $J_k^{(m)}(F)$, F_P, P^*, N^*, (a,b)
χ_τ	$\tau \in (0, 1/(2k)),\ \chi_\tau \in C^\infty(\mathbb{R}^k, [0,1]),\ \chi_\tau \equiv 1$ on $\mathcal{T}_\varepsilon^{(\tau)}$, $\operatorname{supp}(\chi_\tau) \subset \mathcal{T}_\varepsilon$, the smooth cut-off of $\mathcal{T}_\varepsilon$ to its τ-inset **Uses:** k, $\mathcal{T}_\varepsilon$, $\mathcal{T}_\varepsilon^{(\tau)}$
$\mathcal{R}_k^{(\varepsilon)}$	$\mathcal{R}_k^{(\varepsilon)} := \{t \in \mathcal{R}_k : t_i \geq \varepsilon\ \forall\, 1 \leq i \leq k,\ \sum_{i=1}^k t_i \leq 1 - \varepsilon\},\ \varepsilon > 0$ **Uses:** $\mathcal{R}_k$, k
T_m	$(T_m F)(t_1,\dots,\widehat{t_m},\dots,t_k) := \int_0^1 F(t_1,\dots,t_k)\,dt_m, \quad m \in \{1,\dots,k\}$ **Uses:** k, F
β	$\beta := \vartheta/2 - \delta, \quad \log R = \beta \log N$ **Uses:** ϑ, δ, R, N
$S(\lambda)$	$S(\lambda) := S_2(\lambda) - \rho\, S_1(\lambda), \quad \lambda : \mathbb{N}^k \to \mathbb{R}$ **Uses:** $S_1(\lambda)$, $S_2(\lambda)$, ρ, $\lambda(d_1,\dots,d_k)$, k
$m(n)$	$m(n) := \min\{j \geq 1 : p_{n+j} \geq n + h_1\},\ n \in \mathbb{Z}_{\geq 1}$ **Uses:** p_n, h_i
$J(n)$	$J(n) := \max\{j \geq 1 : p_{n+j} \leq n + \operatorname{diam}(\mathcal{H})\}, \quad n \in \mathbb{N}$ **Uses:** p_n, $\operatorname{diam}(\mathcal{H})$, $\mathbb{N}$
$\mathcal{H}_{105}$	$\mathcal{H}_{105} = \{h_1,\dots,h_{105}\} \subset \mathbb{Z},\ \operatorname{diam}(\mathcal{H}_{105}) = 600,\ \forall p \in \mathbb{P} : \#\{h_i \bmod p : 1 \leq i \leq 105\} < p$ **Uses:** h_i, $\mathbb{P}$, $\operatorname{diam}(\mathcal{H})$

Notation	Definition
A_3	$A_3 \in \mathbb{R}_{\geq 0},\ \exists\, V \in \mathbb{Z}_{\geq 1}$ squarefree : $\gamma(p) = 0$ for $p \mid V,\ \lvert\gamma(p) - 1\rvert \leq A_3/p$ for $p \nmid V$ **Uses:** γ, V
V	$V \in \mathbb{Z}_{\geq 1}$ squarefree, $\gamma(p) = 0$ for all primes $p \mid V,\ \lvert\gamma(p) - 1\rvert \leq A_3/p$ for all primes $p \nmid V$ **Uses:** γ
$h(d)$	$h(d) := \mu(d)^2\, g_*(d), \quad d \in \mathbb{Z}_{\geq 1}$ **Uses:** μ, g_*
$a(d)$	$a(d) := \mu(d)^2/\varphi(d), \quad d \in \mathbb{Z}_{\geq 1}$ **Uses:** μ, φ
$b(p)$	$b : \mathbb{N} \to \mathbb{R}$ multiplicative, supported on squarefree integers coprime to V, $b(p) := g_*(p) - \frac{1}{p-1}$ for $p \nmid V$ **Uses:** g_*, V
ℓ_m	$\ell_m := \sum_{p \mid m} \frac{\log p}{p-1}, \quad m \in \mathbb{Z}_{\geq 1}$ squarefree
γ_0	$\gamma_0 := \lim_{N\to\infty}\left(\sum_{n=1}^{N} \frac{1}{n} - \log N\right) \in \mathbb{R}$ (Euler–Mascheroni constant; also written γ_E in some proofs)
$A_m(x)$	$A_m(x) := \sum_{\substack{f<x \\ (f,m)=1}} \frac{\mu(f)^2}{\varphi(f)},\ A_m(x) = 0$ for $x \leq 1, \quad m \in \mathbb{Z}_{\geq 1}$ squarefree, $x \in \mathbb{R}$ **Uses:** μ, φ, (a, b)
γ_g	$\gamma_g : \mathbb{N} \to \mathbb{R}_{\geq 0}$ totally multiplicative, $\gamma_g(p) := 0$ for $p \mid W,\ \gamma_g(p) := \frac{p}{p-1}$ for $p \nmid W$ **Uses:** $\mathbb{N}$, W
$Y_{\mathbf{r}}$	$Y_{\mathbf{r}} := \sum_{(u, W\prod_{i\neq m} r_i)=1}^{u} \frac{\mu(u)^2}{\varphi(u)}\, F\left(\dots, \frac{\log u}{\log R}, \dots\right), \quad \mathbf{r} = (r_1, \dots, r_k),\ r_m = 1$ **Uses:** μ, φ, F, R, W, $\mathbf{r}$, (a, b)
Φ	$\Phi\big((t_i)_{i\neq m};\, s, s'\big) := F(\dots, s, \dots)\, F(\dots, s', \dots), \quad s, s'$ in the m-th slot **Uses:** F
P_j'	$P_j' := \sum_{i=2}^{k} t_i^j, \quad j \in \mathbb{Z}_{\geq 1},\ (t_2, \dots, t_k) \in \mathbb{R}_{\geq 0}^{k-1}$ **Uses:** k
$\lVert f\rVert_{L^2(X)}$	$\lVert f\rVert_{L^2(X)} := \left(\int_X \lvert f\rvert^2\right)^{1/2},\ L^2(X) := \{f : X \to \mathbb{R}$ measurable : $\lVert f\rVert_{L^2(X)} < \infty\},\ X \subset \mathbb{R}^n$ measurable
$\lVert f\rVert_{L^\infty(X)}$	$\lVert f\rVert_{L^\infty(X)} := \operatorname{ess\,sup}_{x\in X} \lvert f(x)\rvert,\ L^\infty(X) := \{f : X \to \mathbb{R}$ measurable : $\lVert f\rVert_{L^\infty(X)} < \infty\},\ X \subset \mathbb{R}^n$ measurable
$\operatorname{vol}(A)$	$\operatorname{vol}(A) := \int_A 1\, dx_1 \cdots dx_n,\ A \subset \mathbb{R}^n$ Lebesgue measurable
$C^\infty(X)$	$C^\infty(X) := \{f : X \to \mathbb{R} :$ f has continuous partial derivatives of every order on $X\},\ X \subset \mathbb{R}^n$ open
$C_c^\infty(X)$	$C_c^\infty(X) := \{f \in C^\infty(X) : \operatorname{supp}(f)$ is compact$\},\ X \subset \mathbb{R}^n$ open **Uses:** $C^\infty(X)$
ε	$\varepsilon = 1/25$, the enlargement parameter
k	$k = 50$, the tuple dimension used in Section 14
R	$R := N^{\vartheta/2-\delta}$, the truncation level **Uses:** ϑ
W	$W := \prod_{p \leq D_0} p,\ D_0 = \log\log\log N$
ρ	sieve target; here $\rho \in (1, 2)$, so $\lfloor \rho + 1 \rfloor = 2$
$\mathcal{R}_k$	$\{t \in [0, \infty)^k : \sum_i t_i \leq 1\}$ **Uses:** k
$\mathcal{T}_\varepsilon$	$\{t \in [0, \infty)^k : \sum_i t_i \leq 1 + \varepsilon\}$, the enlarged simplex **Uses:** ε, k

Notation	Definition
$\mathcal{T}_\varepsilon^{(\tau)}$	$\{t \in \mathcal{T}_\varepsilon : t_i \geq \tau\ \forall i,\ \sum_i t_i \leq 1+\varepsilon-\tau\}$, the τ-inset of $\mathcal{T}_\varepsilon$ (mollifier width) **Uses:** $\mathcal{T}_\varepsilon$, ε, k
$\mathcal{S}_\varepsilon$	$\{s \in [0,\infty)^{k-1} : \sum_j s_j \leq 1-\varepsilon\}$, the shrunken slice **Uses:** ε, k
$I_k(F)$	$I_k(F) = \int_{[0,\infty)^k} F^2\,dt$ **Uses:** k
$J_\varepsilon^{(i)}(F)$	$J_\varepsilon^{(i)}(F) = \int_{\mathcal{S}_\varepsilon} \left(\int_0^\infty F\,dt_i\right)^2 dt_{\hat{i}}$ **Uses:** $\mathcal{S}_\varepsilon$
$\mathcal{Q}_\varepsilon(F)$	$\mathcal{Q}_\varepsilon(F) = \sum_{i=1}^k J_\varepsilon^{(i)}(F)/I_k(F)$ **Uses:** $J_\varepsilon^{(i)}(F)$, $I_k(F)$
λ_ε	the F-derived sieve weight on $\mathcal{T}_\varepsilon$ **Uses:** $\mathcal{T}_\varepsilon$
$\lambda_{1,m}$	$\lambda_\varepsilon \cdot \mathbf{1}[\prod_{i\neq m} d_i \leq R^{1-\varepsilon}]$ **Uses:** λ_ε, ε, R
$A_{\ell,m}(n)$	$A_{\ell,m}(n) := \sum_{d_i \mid n+h_i} \lambda_{\ell,m}(\mathbf{d})$, $\ell \in \{1,2\}$ **Uses:** $\lambda_{1,m}$
$P_{(1)}$	power sum $P_{(1)}(t) = t_1 + \cdots + t_k$ **Uses:** k
$\mathcal{H}_{50}$	an admissible 50-tuple with $\mathrm{diam}(\mathcal{H}_{50}) = 246$
α	a signature: a finite non-increasing tuple $(\alpha_1,\ldots,\alpha_j)$ of positive integers; $\lvert\alpha\rvert := \sum_i \alpha_i$ is its size; $\alpha \setminus i$ removes the i-th part
P_α	$P_\alpha := \prod_{i=1}^j P_{\alpha_i}$, the power-sum monomial for a signature $\alpha = (\alpha_1,\ldots,\alpha_j)$; $P_\emptyset = 1$ **Uses:** α, $P_{(1)}$
$\alpha \cup \alpha'$	signature merger: $\lvert\alpha\cup\alpha'\rvert = \lvert\alpha\rvert + \lvert\alpha'\rvert$ and $P_{\alpha\cup\alpha'} = P_\alpha P_{\alpha'}$ **Uses:** α, P_α
α_S	signature restriction: $(\alpha_i)_{i\in S}$ re-listed non-increasing; partial size $\lvert\alpha\rvert_S := \sum_{i\in S} \alpha_i$ **Uses:** α
$H_\alpha(s)$	set-partition weight: $\sum_{\pi\in\mathrm{Part}(r,s)} \prod_{B\in\pi} (\sum_{i\in B} \alpha_i)!$ (a non-negative integer), for a signature α of length r and $1 \leq s \leq r$ **Uses:** α
$G_\alpha(x)$	signature polynomial $G_\alpha \in \mathbb{Q}[x]$: $\sum_{s=1}^r x(x-1)\cdots(x-s+1)H_\alpha(s)$ for signature α of length $r \geq 1$; $G_\emptyset(x) := 1$ **Uses:** α, $H_\alpha(s)$