AN OUTLINE OF THE SHORTENED PROOF OF THE RIEMANN HYPOTHESIS

MICHAEL O. RUBINSTEIN

Abstract. We give an outline of the shortened formal proof of the Riemann hypothesis. Steps 1–4 reduce the Riemann hypothesis to a one-sided bound for a sum over the primes inert in an imaginary quadratic field $\mathbb{Q}(\sqrt{-d})$: Landau's criterion, the prime sum, a positivity argument, and a zero-density estimate for quadratic twists. Steps 5–8 explain the mechanism: Chebyshev's principle, the forced divisibility by inert primes of derivatives of functions with complex multiplication, a Gaussian norm and the Chowla–Selberg period that let size be compared with divisibility, and a Gram determinant in place of a single integer. Steps 9–10 show how three theorems, an upper bound (A, B) and a lower bound (C) relative to one enormous main term that is never evaluated, give the main inequality and with it the bound $y^{-1/2+\varepsilon}$, the exponent $1/2$ coming from $\sqrt{x}$ iterated theta lifts. Steps 11–14 outline the proofs of the three theorems: a seesaw identity that makes the lifts computable, the accounting of size against divisibility, a second-moment argument, and a lower bound from the product formula using twists, Hecke L-values and Euler systems. Step 15 extends the argument to Dirichlet L-functions. Numerical examples are provided in the appendix.

Version 1.5.6, 29 November 2026

0. The outline

1. **One-sided bound (Landau, 1905).** The error in counting primes only needs to be bounded from above, because a zero off the critical line would force it to be large in both directions. (§1)
2. **A sum over inert primes.** The proof counts primes through a weighted sum over the odd powers $p^h \equiv 1 \pmod{Q}$ of the primes p that stay prime in a well-chosen imaginary quadratic field $\mathbb{Q}(\sqrt{-d})$, and it shows that the error in this sum is at most $y^{-1/2+\varepsilon}$. (§2)
3. **No circularity.** Zeros of the twisted L-functions enter the sum with the opposite sign to zeros of ζ, so to exclude a given zero of ζ it is enough to choose one d for which the $\varphi(Q)$ twists $L(s, \psi\chi_d)$, $\psi \bmod Q$, have no zeros in one fixed box to the right of the critical line around it. (§3)
4. **Zero-density for quadratic twists.** Twist zeros near the critical line multiply as d grows, but zeros a fixed distance to its right are rare, so most primes d have none in the box. (§4)
5. **Chebyshev's principle.** A positive integer that is divisible by many primes but is not too large bounds a sum over those primes from above. (§5)
6. **Complex multiplication forces divisibility by the inert primes.** Derivatives of functions with complex multiplication are divisible by prescribed powers of the inert primes, with rounding once per prime instead of the rounding at every prime power that limits Chebyshev's factorials. (§6)
7. **A Gaussian norm and the period.** Measuring vectors of derivatives, written as polynomials in the variables of the theta functions, in a Gaussian norm in which the

monomial z^j has squared length $j!$ lets size and divisibility be compared at leading order, and in the linear term the period of the lattice supplies, by the Chowla–Selberg formula, the constant $\frac{L'}{L}(1,\chi_d)-\gamma$ of the prime sum. (§7)

8. **A Gram determinant.** The squared volume of many such vectors is bounded above by Hadamard's inequality and below through the product formula, the higher-dimensional version of "a nonzero integer is at least 1". (§8)
9. **One main term, bounded from both sides.** Theorems A and B give an upper bound and Theorem C a lower bound relative to the same enormous main term h, so h cancels without being evaluated and leaves an inequality controlling the prime-sum error. (§9)
10. **Iteration and the square root.** Each of m theta lifts adds the inert-prime divisibility once more, a gain linear in m, while the bounds for the accumulated costs are quadratic in m; taking $m \asymp \sqrt{x}$ keeps these within the error budget, and dividing by m gives the exponent $\frac{1}{2}$ and with it the Riemann hypothesis. (§10)
11. **The seesaw identity.** Lifts along a chain of unitary groups that starts with an indefinite one, which are defined by integrals rather than finite class sums, equal a fixed constant times lifts along a chain of definite lattices, which are finite matrices of CM theta derivatives, so their size and divisibility are computable. (§11)
12. **Short forms (Theorem A).** Adding up, over all residues, the size of every derivative against its guaranteed inert-prime divisibility, and choosing an auxiliary prime N to absorb the leftover linear term, shows that the lifts are shorter than the main term h by $m \cdot \mathrm{rem}$, up to a small error. (§12)
13. **Enough forms (Theorem B).** A second-moment computation shows that the Poincaré series is nearly an isometry whose translates spread evenly, so a controlled number of translates of the short forms fill almost all of the target space. (§13)
14. **Integral forms are long (Theorem C).** The product formula, read at each prime ℓ in a family twisted by a character chosen for ℓ, turns the lower bound into ratios of Hecke L-values, and the denominators that appear are bounded with Euler systems. (§14)
15. **Dirichlet L-functions.** The same argument, with Q replaced by Q times the modulus, proves the generalized Riemann hypothesis for every Dirichlet L-function. (§15)

Worked examples. Appendix A carries out the construction of Steps 6–10 on three small models with actual numbers: a single vector, an 8×8 Gram determinant, and a chain of lifts over three classes of lattices.

Steps 1–4 make up the *analytic reduction*: they reduce the Riemann hypothesis to an upper bound for one prime sum, for every prime d in a fixed residue class. Steps 5–8 describe the object that proves that bound and why it works. Steps 9–10 derive the bound, and with it the Riemann hypothesis, from three theorems, A, B and C, and Steps 11–14 outline their proofs. The analytic reduction and the construction interact only through the bound (2.9) below, and the construction reaches it only through the main inequality (M) of Section 9.

The argument in brief. Before the details, here is how the argument fits together. The notation is collected later in this section, and each point is taken up in the section indicated.

The prime sum (Step 2). Let $Q = 18480$, let $d \equiv Q/2 - 1 \pmod{Q}$ be a prime, and put $K = \mathbb{Q}(\sqrt{-d})$. Let χ_d be the Legendre symbol modulo d, Λ the von Mangoldt function and φ

Euler's function. On the analytic side the shortened proof is classical. It considers the sum

$$S_d(y) = \sum_{\substack{n\le y\\ n\equiv 1 \pmod Q}} \Lambda(n)\bigl(1-\chi_d(n)\bigr)\Bigl(\frac{1}{n}-\frac{1}{y}\Bigr)$$

(see (2.3)), with a single parameter, the prime d. (The earlier two-prime formulation uses the factors $1+\chi_{q_{\mathrm{old}}}(n)$ and $1-\chi_d(n)$, with two prime parameters q_{old} and d.) Every prime dividing Q splits in K. The construction requires Q to be a multiple of $1680 = 2^4\cdot 3\cdot 5\cdot 7$, for reasons that come from the arithmetic side: the factors 16 and 3 give $d\equiv 7 \pmod 8$ and $d\equiv -1 \pmod 3$, so that 2 and 3 split, as the construction requires; and the factors 5 and 7, together with the powers of 2 and 3 in an auxiliary multiple q of Q, ensure that the cyclotomic field used in the construction contains the roots of unity of order $5040 = 7!$. The Lean proof takes $Q = 18480 = 1680\cdot 11$, and so do we, although the factor 11 is unnecessary (Section 6).

Apart from powers of the ramified prime d, which are treated separately (Section 12), the factor $1-\chi_d(n)$ restricts the sum to odd powers of primes that are inert in K. For these, theta functions with complex multiplication by K force divisibility by specified powers. The particular smoothing weight $\frac{1}{n}-\frac{1}{y}$ comes from the count of prime powers that arises on the arithmetic side.

The condition $n\equiv 1 \pmod Q$ also comes from the arithmetic side. Fix a large integer x, the value of y at which the sum is to be bounded. The proof pairs differentiation coordinates indexed by the set Φ_a of residues $\tau\equiv 1 \pmod Q$, among the units modulo q, with coordinates indexed by a sparse subset Φ_b of the residues $\equiv -1 \pmod Q$. For odd powers p^h of inert primes, the partner-residue congruence (Lemma 6.1), which comes from the Frobenius automorphism at p, pairs τ with $-p^h\tau$. For $\tau\in\Phi_a$ the partner can lie in Φ_b only when $p^h\equiv 1 \pmod Q$. This produces the desired contribution of the $n\equiv 1 \pmod Q$ at the cutoff x. The remaining coordinates contribute at a much larger cutoff x_1, for all classes modulo Q, and that contribution is handled by the classical prime number theorem (Section 9).

The zeros (Steps 1, 3 and 4). Because of the congruence condition, the sum involves all the characters ψ modulo Q, including the principal character, which gives rise to ζ, and, through $\chi_d(n)$, all the twisted characters $\psi\chi_d$. The class $n\equiv 1$ gives every ψ the same weight 1. So if several $L(s,\psi)$ vanish at the same point ρ, their contributions add, and they cannot cancel the contribution of a zero of ζ. Zeros of the $L(s,\psi\chi_d)$ enter with the opposite sign and could cancel it. But for a fixed zero ρ of ζ to the right of the critical line, one can select d such that no $L(s,\psi\chi_d)$ has a zero in a box $\operatorname{Re} s\ge\beta$, $|\operatorname{Im} s|\le H$ that contains ρ, where $\frac12<\beta<\operatorname{Re}\rho$ and $H\ge|\operatorname{Im}\rho|$. The box also contains the real segment $(\beta,1)$, which is needed for Landau's lemma (Lemma 1.1): excluding real zeros of the twists there ensures that the transform has no real pole whose sign could obstruct the positivity argument. (In terms of oscillations, a real zero $u>\operatorname{Re}\rho$ of a twist would contribute a negative term, of order y^{u-1}, that could hide the positive part of the oscillation coming from ρ.) Although the zeros of the twists become denser as d increases, a zero-density estimate (Theorem 4.1) shows that, for fixed $\beta>\frac12$ and H, almost all primes d in the class have no zeros of any $L(s,\psi\chi_d)$ in the box (Sections 3 and 4).

It is therefore enough to prove a one-sided bound, consistent with the Riemann hypothesis, for the remainder

$$R(y) = S_d(y) - \frac{\log y}{\varphi(Q)} - c_d,$$

where c_d is the constant for which $R(y)\to 0$: for every fixed prime $d\equiv Q/2-1 \pmod Q$ and every $\varepsilon>0$, one needs $R(y)\le y^{-1/2+\varepsilon}$ for all sufficiently large y. The one-sided criterion (Theorem 3.2) then rules out the zero ρ.

Divisibility and size (Steps 5–7). The principle is Chebyshev's (Section 5): a quantity that is forced to be divisible by many prime powers, but is not too large, bounds a sum over those prime powers from above. At an inert prime p, Frobenius acts on K as complex conjugation, and elliptic curves with complex multiplication by the integers of K have supersingular reduction. This forces extra divisibility by p in suitable combinations of high-order derivatives of the associated theta functions. The phenomenon is classical for $\mathbb{Q}(i)$: the derivatives of the lemniscate sine at 0, of order at least p, are divisible by p for every prime $p \equiv 3 \pmod 4$. The gains come from odd powers of p, which is what the factor $1 - \chi_d(n)$ selects.

The basic function is the translated and normalized sigma function

$$\theta(z) = \exp\Big(-\frac{\pi z\,\bar{t}_0}{A_\Lambda}\Big)\frac{\sigma_\Lambda(z+t_0)}{\sigma_\Lambda(t_0)}$$

of a lattice Λ with complex multiplication by the integers of K, where A_Λ is the area of a period parallelogram. The half-period t_0, the normalization of σ_Λ and the scale of Λ are specified after (6.18). The proof differentiates products of twisted translates of θ by torsion points, composed with linear forms. Three properties are used.

- *Algebraicity.* The derivatives of θ at 0 are algebraic numbers, so one can speak of divisibility and use the product formula.
- *Galois conjugates.* The conjugate theta derivatives are described by theta functions of conjugate lattices, with controlled normalization factors. This allows the same size estimates at every conjugate.
- *Size against divisibility.* The construction compares logarithmic squared norms with twice the logarithmic divisibility, as encoded by the weight $1 - \chi_d$ at odd powers of inert primes. The Gaussian norm supplies the factorial size charge, and the Chowla–Selberg period supplies the matching $\frac{L'}{L}(1, \chi_d) - \gamma$ term. The assigned weights and the auxiliary prime N complete the comparison (Sections 7 and 12).

The Gram determinant, the lifts and the three theorems (Steps 8–14). In place of a single integer, the proof uses the Gram determinant

$$\det\big(\langle v_i, v_j\rangle\big) > 0$$

of many vectors (Section 8). Choose independent *target forms* y_i among the finite translates of normalized Poincaré images of *source forms* on a definite unitary group. The y_i lie on an indefinite unitary group over the cyclotomic extension $E = K(e^{2\pi i/q})$ of K, and $v_i = \Theta_1 y_i$ is the image of y_i under a chain of $m \asymp \sqrt{x}$ theta lifts, ending at a definite group. This chain is injective on the target space, so the v_i are independent and their Gram determinant is positive. The lifts along the chain that starts with the indefinite group are defined by automorphic integrals rather than finite class sums. The seesaw identity (Section 11) expresses the lifts of the Poincaré images, up to a fixed constant, as finite combinations of lifts of the source forms along chains of definite hermitian lattices of large rank, whose matrices are finite sums of theta derivatives. Each lift contributes the prime sum once more, while the accumulated costs of the m lifts, whose bounds are quadratic in m, stay within the total error allowance for m of this size. Three theorems, stated relative to a common and very large main term h, give the bound.

- *Theorem A* supplies a spanning family of integral source forms whose target images, and all their translates by a finite group, have m-step lifts with logarithm of the squared length, at every conjugate, at most $h - m \cdot \mathrm{rem}$ plus an error, where rem is a combination of the remainders of the prime sum. It also bounds the index of finite families of translated images.

- *Theorem B* shows that a controlled number of translates of these target images span almost all of the target space, of dimension D.
- *Theorem C* gives $c + 2\delta \geq D(h - \text{error})$ for independent target forms that nearly fill the target space. Here c, the *covolume*, is an upper bound, uniform over all automorphisms σ of $\mathbb{C}$, for the logarithm of the Gram determinant of the σ-conjugates of their m-step lifts, and δ, the *index*, measures the cost of clearing denominators. The proof uses the product formula and twists, together with Hecke L-values, Galois representations and Euler systems.

Theorems A and B supply enough target forms with short lifts, and Hadamard's inequality then gives an admissible c, with δ also controlled by A and B. Comparing with C, the main term $h \geq 0$ drops out, leaving the main inequality (M), an upper bound for $m \cdot \text{rem}$ and so for rem. Now rem is expressed through the remainders at x and x_1 by the two-scale identity (9.6). Dividing by its normalizing factor gives

$$R(x) \leq \frac{\varphi(Q)(\log b)^6}{m} + |R(x_1)| + (\log b)\,|R_1(x_1)|,$$

where b is an auxiliary weight and R_1 is the remainder without the congruence condition. Given $\varepsilon > 0$, choose η with $0 < 6\eta < \varepsilon$ and $\eta < \frac{1}{100}$, before letting x tend to infinity. The parameters allow $\log b \leq 2x^\eta$, while $\log x_1$ grows like a positive power of x, so the prime number theorem makes the last two terms negligible. Since $m \geq \sqrt{x}$, this gives $R(x) \leq x^{-1/2+\varepsilon}$ for all large integers x, and the bound for real y follows (Section 10). Finally, replacing Q by Q times the modulus of a Dirichlet character χ gives the generalized Riemann hypothesis for $L(s, \chi)$ in the same way (Step 15, Section 15).

Notation. The main letters, with the section where each is defined. Some letters are used twice, as noted; the context always decides.

Q, d, χ_d, K	the modulus 18480, a prime $d \equiv Q/2 - 1 \pmod{Q}$, the Legendre symbol modulo d, the field $\mathbb{Q}(\sqrt{-d})$ (§2)
$S_d(y)$, c_d, $R(y)$	the prime sum, its constant, and its remainder (§2); R_1, c_1 are the same for $Q = 1$ (§9)
ψ, ξ	Dirichlet characters modulo Q; in §§8–14 ξ is a character of the norm-one ideles of E (a "twist"); $\psi(x)$ is also Chebyshev's function (§1)
q, E, f, τ	a multiple of Q, the field $K(e^{2\pi i/q})$, $f = \varphi(q)$, a residue in $(\mathbb{Z}/q)^\times$ indexing an embedding of E (§6)
x, x_1, b, a, m	the scale, the second cutoff, the weight, the order $a = xb$, the number of lifts (§9)
Φ_a, Φ_b, A_0, B_0, κ	the special residues, their numbers, the weights (§9)
s_{fin}, s_{main}, rem, s_∞	guaranteed divisibility, its average, the remainder (§9); the size charged to a derivative (§12)
$h = h_{\text{main}}$, $\mathcal{E}$	the main term (§12) and the error $fa(\log b)^4$; h is also an exponent in p^h
r, k, t, M, N	ranks and levels (§11); N is a prime there
X_j, Y_j, Θ_ξ, $\mathcal{P}$, G	the two chains of spaces, the chain of lifts, the Poincaré map, the finite group (§8)
S, D, n, H, H_σ	the source space, the dimension of the target space, the number of vectors, the Gram determinant and its conjugated versions (§8); H is also the height of the box (§3)
ω, Ω	a complex period of a CM lattice, and the positive real scale (7.11) (§7)

Background and status. On 20 November 2026 L. Alpöge announced a proof of the Riemann hypothesis and of the generalized Riemann hypothesis for Dirichlet L-functions, produced with

Anthropic's Claude and formalized in the Lean proof assistant [1]. Five days later L. Sernau and collaborators released a much shorter formal proof [39], which we follow: about 300,000 lines of Lean code, or 76,000 in a compact edition (not counting comments and blank lines), against about 17 million lines in all for the original. Both have been checked repeatedly by independent proof checkers [41].

The shortened proof retains Alpöge's central idea: use derivatives of theta functions to construct algebraic quantities with large, prescribed prime-power divisibility, then compare their divisibility with their size. Several arguments are replaced. An explicit formula in a local parameter supplies the divisibility estimates; Poisson summation supplies the sieve estimate; and the product formula for a Gram determinant replaces the earlier comparison of lattice slopes. The final analytic argument is also shorter: zero-density estimates let us choose an auxiliary quadratic character for each proposed zero, after which Landau's positivity argument gives a contradiction. The earlier Vandiver-type cohomology bound is avoided, although Galois representations, p-adic Hodge theory and Euler systems remain in the control of denominators.

We prove or sketch the analytic reduction (Steps 1–4), the two-scale identity, the deduction of the main inequality (M) from Theorems A, B and C, and the deduction of the final bound from (M). Theorems A, B and C we state precisely, and we outline their proofs in Sections 11–14, naming each step and its inputs, without proving them; they are where the length of the Lean development lies.

The outline was extracted from the formal proof with the help of the AI models Claude Fable 5.2 and Claude Opus 5.5, Codex Astra 6.2 and Codex Sol 6.1, and Gemini Carbon, and they did most of the writing for this document, though I reshaped it and coaxed them to flesh out or simplify certain parts, with the goal of creating a study guide to understand the proof in the language of classical analytic number theory wherever possible, and to make the various steps as concrete as possible. This is ongoing, and this document will continue to evolve.

1. A one-sided bound is enough

Let Λ be the von Mangoldt function and $\psi(x) = \sum_{n \le x} \Lambda(n)$. Landau observed that if

$$\psi(x) - x \;\le\; C_\varepsilon\, x^{1/2+\varepsilon} \qquad \text{for every } \varepsilon > 0, \tag{1.1}$$

then the Riemann hypothesis holds; no lower bound is needed. The reason is positivity.

Lemma 1.1 (Landau [22]; [38, Lemma 15.1])**.** *Let $g\colon [0,\infty) \to [0,\infty)$ be measurable and bounded on bounded sets, and suppose that*

$$F(s) = \int_0^\infty g(v) e^{-sv}\, dv \tag{1.2}$$

converges for some real s. Let σ_c be the infimum of the real s for which it converges. Then F is analytic in $\operatorname{Re} s > \sigma_c$, and if $\sigma_c > -\infty$ then F cannot be continued analytically to a neighbourhood of the real point $s = \sigma_c$.

How it gives RH from (1.1). Fix $\frac{1}{2} < \theta < 1$. Then $Cx^\theta - \psi(x) + x \ge 0$ for large x, and since

$$\int_1^\infty \psi(x) x^{-s-1}\, dx = -\frac{1}{s}\frac{\zeta'}{\zeta}(s) \qquad (\operatorname{Re} s > 1), \tag{1.3}$$

we have

$$\int_1^\infty (Cx^\theta - \psi(x) + x) x^{-s-1}\, dx = \frac{C}{s-\theta} + \frac{1}{s}\frac{\zeta'}{\zeta}(s) + \frac{1}{s-1} \qquad (\operatorname{Re} s > 1). \tag{1.4}$$

With $x = e^v$ this is a Laplace transform of a function that is nonnegative for large v; the part from small x is entire. The right side is analytic on the real half-line (θ, ∞): ζ has no zeros in $(0,1)$ (because $(1-2^{1-s})\zeta(s) = \sum_{n\ge1}(-1)^{n-1}n^{-s} > 0$ there), and at $s=1$ the poles cancel, since

$$\frac{1}{s}\frac{\zeta'}{\zeta}(s) + \frac{1}{s-1} = 1 + \gamma + O(|s-1|) \qquad (s \to 1). \tag{1.5}$$

By Lemma 1.1 the abscissa of convergence is at most θ, so the transform is analytic in $\operatorname{Re} s > \theta$. A zero ρ of ζ with $\operatorname{Re}\rho > \theta$ would give a pole there. Since $\theta > \frac12$ was arbitrary, ζ has no zeros with real part greater than $\frac12$, and by the functional equation, which makes the zeros symmetric about the line $\operatorname{Re} s = \frac12$, RH follows. □

Put differently: a zero $\rho = \beta + i\gamma$ with $\beta > \frac12$ makes $\psi(x) - x$ oscillate with amplitude about x^β, positively and negatively. Excluding the large positive excursions already excludes the zero. The proof applies this not to ψ but to the sum of Step 2, in the form of Theorem 3.2.

2. A sum over inert primes

Fix

$$Q = 18480 = 2^4 \cdot 3 \cdot 5 \cdot 7 \cdot 11, \qquad \varphi(Q) = 3840, \tag{2.1}$$

and a prime $d \equiv 9239 \pmod Q$. Then $d \equiv 7 \pmod 8$ and $d \equiv -1 \pmod p$ for $p = 3, 5, 7, 11$. Let $\chi_d(n) = (\frac{n}{d})$ be the Legendre symbol. Since $d \equiv 3 \pmod 4$, quadratic reciprocity shows that χ_d is the character of the imaginary quadratic field

$$K = \mathbb{Q}(\sqrt{-d}), \tag{2.2}$$

so that a prime $p \neq d$ is a non-square modulo d exactly when it is *inert* in K, and a square exactly when it *splits*. Equivalently, by quadratic reciprocity, an odd prime $p \neq d$ is inert exactly when $-d$ is a non-square modulo p; the prime 2 splits, since $d \equiv 7 \pmod 8$. The congruence on d makes $2, 3, 5, 7, 11$ split.

The construction of Steps 6–14 works in a cyclotomic extension $E = K(e^{2\pi i/q})$ of K, for a modulus q built from Q. That field, and the reason for the congruence on d, are described in Section 6, where they are first used; the analytic reduction of Sections 2–4 needs only Q, d and χ_d.

Definition 2.1. For $y > 0$ put

$$S_d(y) = \sum_{\substack{n \le y \\ n \equiv 1 \pmod Q}} \Lambda(n)\bigl(1 - \chi_d(n)\bigr)\Bigl(\frac{1}{n} - \frac{1}{y}\Bigr), \qquad R(y) = S_d(y) - \frac{\log y}{\varphi(Q)} - c_d, \tag{2.3}$$

where γ is Euler's constant, ψ_0 the principal character modulo Q, and

$$c_d = \frac{1}{\varphi(Q)}\Bigl[\frac{L'}{L}(1,\chi_d) - \gamma - 1 + \sum_{\psi \neq \psi_0}\Bigl(\frac{L'}{L}(1,\psi\chi_d) - \frac{L'}{L}(1,\psi)\Bigr)\Bigr]. \tag{2.4}$$

The weight $\Lambda(n)(1-\chi_d(n))$ at a prime power $n = p^h$ is

p split	every h	0
p inert	h odd	$2\log p$
p inert	h even	0
$p = d$	every h	$\log d$

So S_d counts, twice, the odd powers $p^h \equiv 1 \pmod Q$ of inert primes, with the smoothing weight $\frac{1}{n} - \frac{1}{y}$, a ramp that fades to 0 at $n = y$. The congruence is on the prime power, not on the prime: for $d = 9239$ the prime $p = 21121$ is inert and $p \equiv 2641 \pmod Q$, but $p^3 \equiv 1 \pmod Q$. (The powers of d with $d^h \equiv 1 \pmod Q$ also occur; they are few.) The expected value of S_d is computed by the prime number theorem in progressions.

Proposition 2.2 (Expected value; [*proved*])**.** *As $y \to \infty$, $S_d(y) = \dfrac{\log y}{\varphi(Q)} + c_d + o(1)$; that is, $R(y) \to 0$.*

Proof. By orthogonality of the characters modulo Q,

$$S_d(y) = \frac{1}{\varphi(Q)} \sum_{\psi \bmod Q} \Big(\Sigma_\psi(y) - \Sigma_{\psi\chi_d}(y)\Big), \qquad \Sigma_\xi(y) = \sum_{n \le y} \xi(n)\Lambda(n)\Big(\frac{1}{n} - \frac{1}{y}\Big). \tag{2.5}$$

For a non-principal character ξ, $L(s,\xi)$ has no zeros on $\operatorname{Re} s = 1$, and the prime number theorem for ξ gives $\sum_{n\le y} \xi(n)\Lambda(n)/n \to -\frac{L'}{L}(1,\xi)$ and $\frac{1}{y}\sum_{n\le y} \xi(n)\Lambda(n) \to 0$, so $\Sigma_\xi(y) \to -\frac{L'}{L}(1,\xi)$. This applies to $\xi = \psi \ne \psi_0$ and to every $\xi = \psi\chi_d$. For ψ_0, the prime number theorem in the forms $\sum_{n\le y} \Lambda(n)/n = \log y - \gamma + o(1)$ and $\frac{1}{y}\sum_{n\le y} \Lambda(n) \to 1$, after removing the powers of the primes dividing Q, gives

$$\Sigma_{\psi_0}(y) = \log y - \gamma - 1 - \sum_{p \mid Q} \frac{\log p}{p-1} + o(1). \tag{2.6}$$

Finally $L(s, \psi_0\chi_d) = L(s,\chi_d)\prod_{p\mid Q}(1 - \chi_d(p)p^{-s})$ and $\chi_d(p) = 1$ for $p \mid Q$, so

$$-\frac{L'}{L}(1,\psi_0\chi_d) = -\frac{L'}{L}(1,\chi_d) - \sum_{p\mid Q} \frac{\log p}{p-1}. \tag{2.7}$$

Subtracting, the sums over $p \mid Q$ cancel, and the result is (2.4). The constant c_d is real, since the terms for ψ and $\bar\psi$ are complex conjugates. □

What R measures. Let $\Psi_d(t) = \sum_{n\le t,\ n\equiv 1\,(Q)} \Lambda(n)(1 - \chi_d(n))$ be the weighted count of prime powers. Since $\frac{1}{n} - \frac{1}{y} = \int_n^y t^{-2}\,dt$,

$$S_d(y) = \int_1^y \frac{\Psi_d(t)}{t^2}\,dt, \qquad R(y) = -\int_y^\infty \frac{\Psi_d(t) - t/\varphi(Q)}{t^2}\,dt, \tag{2.8}$$

the second because $R(y) \to 0$. So R is a smoothed tail of the error in the count Ψ_d. A bound $\Psi_d(t) - t/\varphi(Q) = O(t^{1/2+\varepsilon})$ would give $R(y) = O(y^{-1/2+\varepsilon})$; the proof obtains the one-sided bound for the smoothed error directly.

The constant c_d has two kinds of terms. The principal character contributes $\frac{L'}{L}(1,\chi_d) - \gamma$ (and -1, from the smoothing); this is the term that the period of Section 7 matches. The characters $\psi \ne \psi_0$ contribute the remaining sum. In the Lean development the same constant appears as $\operatorname{Re} G_Q(1)$ for a function G_Q built from the logarithmic derivatives of the $L(s,\psi)$ and $L(s,\psi\chi_d)$.

The goal. *For every prime $d \equiv 9239 \pmod Q$ and every $\varepsilon > 0$,*

$$R(y) \le y^{-1/2+\varepsilon} \qquad \text{for all sufficiently large } y. \tag{2.9}$$

Only this upper bound is needed (Step 1), the threshold may depend on d, and the bound is proved for *every* d in the class (Steps 5–14).

3. Why it is not circular

The first sum in S_d sees the zeros of the $L(s,\psi)$, $\psi \bmod Q$, among them those of ζ. The second sees the zeros of the twisted functions $L(s,\psi\chi_d)$.

One d, but $\varphi(Q)=3840$ twists. For a given zero of ζ a single prime d is chosen. But the sum S_d runs over $n \equiv 1 \pmod Q$, and detecting that congruence uses all the characters modulo Q:

$$\mathbf{1}_{n\equiv 1\,(Q)} = \frac{1}{\varphi(Q)} \sum_{\psi \bmod Q} \psi(n). \tag{3.1}$$

So the twisted half of S_d involves $L(s,\psi\chi_d)$ for every $\psi \bmod Q$, and all of them must be zero-free in the box. (The congruence is imposed because the construction of Steps 6–14 produces the odd prime powers $p^h \equiv 1 \pmod Q$ of inert primes, through the choice of residues τ in Section 9.) This is why we say "the twists": one quadratic character χ_d, multiplied by each of the 3840 characters ψ.

So one might fear that bounding R needs the Riemann hypothesis for the twists. It does not, because the two kinds of zeros enter with opposite signs. Let

$$\Phi_d(s) = \int_1^\infty R(y)\, y^{-s}\, dy \qquad (\operatorname{Re} s > 1). \tag{3.2}$$

Proposition 3.1 ([*proved*])**.** *Φ_d is meromorphic in $\operatorname{Re} s > 0$ and analytic at $s=1$. Its possible poles in $\operatorname{Re} s>0$ are simple and lie at the zeros $z \neq 1$ of the $L(s,\psi)$ and the $L(s,\psi\chi_d)$, with residue*

$$\operatorname*{Res}_{s=z} \Phi_d(s) = \frac{N(z) - N_d(z)}{\varphi(Q)\, z(1-z)}, \tag{3.3}$$

where $N(z)$ and $N_d(z)$ are the numbers of zeros at z, with multiplicity, of the $L(s,\psi)$ and of the $L(s,\psi\chi_d)$, $\psi \bmod Q$. (Where $N(z)=N_d(z)$ there is no pole.)

Proof. Put

$$A_d(s) = \frac{1}{\varphi(Q)} \sum_{\psi \bmod Q} \Big(\frac{L'}{L}(s,\psi\chi_d) - \frac{L'}{L}(s,\psi) \Big) = \sum_{n\equiv 1\,(Q)} \frac{\Lambda(n)\big(1-\chi_d(n)\big)}{n^s} \qquad (\operatorname{Re} s > 1), \tag{3.4}$$

by orthogonality of the characters. Since $\int_n^\infty (\frac{1}{n} - \frac{1}{y}) y^{-s}\, dy = n^{-s}/(s(s-1))$, integrating (2.3) term by term gives

$$\Phi_d(s) = \frac{A_d(s)}{s(s-1)} - \frac{1}{\varphi(Q)(s-1)^2} - \frac{c_d}{s-1}. \tag{3.5}$$

The right side is meromorphic in $\operatorname{Re} s > 0$. By the definition (2.4) of c_d, and the cancellation of the Euler factors at $p \mid Q$ as in the proof of Proposition 2.2, $A_d(s) = \frac{1}{\varphi(Q)(s-1)} + c_d + \frac{1}{\varphi(Q)} + O(|s-1|)$, so the singularity at $s=1$ is removable. At a point $z \neq 1$ the logarithmic derivatives give A_d a simple pole of residue $(N_d(z) - N(z))/\varphi(Q)$, or none, and dividing by $z(z-1)$ gives the stated residue. □

Formally, moving the line of integration in the inverse transform turns (3.5) into an explicit formula

$$R(y) = \frac{1}{\varphi(Q)} \sum_z \big(N(z) - N_d(z)\big) \frac{y^{z-1}}{z(1-z)} + (\text{terms from } \operatorname{Re} s \le 0), \tag{3.6}$$

the sum over the zeros z with $0 < \operatorname{Re} z < 1$ [*heuristic: no convergence is claimed, and the formula is not used*]. It shows the mechanism at a glance: a zero of an $L(s,\psi)$ at $z = \beta' + i\gamma'$

contributes an oscillation of amplitude about $y^{\beta'-1}$, a zero of a twist contributes one of the opposite sign, and for a real zero $z \in (0,1)$ the coefficient $1/(z(1-z))$ is positive.

On the real segment $0 < z < 1$ we have $z(1-z) > 0$. So zeros of ζ give residues of one sign, and zeros of twists give residues of the other.

No cancellation within a family. Since

$$\mathbf{1}_{n \equiv a\,(Q)} = \frac{1}{\varphi(Q)} \sum_{\psi \bmod Q} \bar{\psi}(a)\psi(n), \tag{3.7}$$

for a prime to Q, a zero of $L(s,\psi)$ enters with weight $\bar{\psi}(a)$. For a general class a these are roots of unity, and zeros shared by several $L(s,\psi)$ could cancel. For $a = 1$ every weight is $\bar{\psi}(1) = 1$, so shared zeros add: this is why $N(z)$ is simply the total multiplicity, and a zero of ζ gives $N(\rho) \geq 1$. The same holds for the twists. Cancellation is possible only between the two families, and the choice of d removes it.

Theorem 3.2 (One-sided criterion; [*proved*])**.** *Let $\frac{1}{2} < \beta < 1$ and $H \geq 0$. Suppose that*

(i) $R(y) \leq y^{\beta-1}$ for all sufficiently large y, and
(ii) no $L(s, \psi\chi_d)$, $\psi \bmod Q$, vanishes in the box $\operatorname{Re} s \geq \beta$, $|\operatorname{Im} s| \leq H$.

Then no $L(s,\psi)$, $\psi \bmod Q$, vanishes at a point $\rho \neq 1$ with $\operatorname{Re}\rho > \beta$ and $|\operatorname{Im}\rho| \leq H$.

Proof. For large Y let

$$F(s) = \int_Y^\infty (y^{\beta-1} - R(y)) y^{-s}\, dy = \frac{1}{s-\beta} - \Phi_d(s) + (\text{entire}), \tag{3.8}$$

the transform of a nonnegative function. If the abscissa of convergence σ_c exceeded β, Landau's lemma would force a singularity at the real point σ_c, necessarily a simple pole of Φ_d with negative residue: $F(s) \to +\infty$ as s decreases to σ_c along the real axis, because the integrand is nonnegative, so the pole of F has positive residue. But on the real segment $(\beta, 1)$ the twists do not vanish, by (ii), so every residue of Φ_d there is $N(u)/(\varphi(Q)u(1-u)) \geq 0$, and Φ_d is analytic for $u \geq 1$. Hence $\sigma_c \leq \beta$, and Φ_d is analytic in $\operatorname{Re} s > \beta$. At a point ρ of the box, $N_d(\rho) = 0$ by (ii), so the residue is $N(\rho)/(\varphi(Q)\rho(1-\rho))$, which must vanish; so $N(\rho) = 0$. □

A real zero of some $L(s,\psi)$ in $(\beta,1)$ would give a residue of the favourable sign. So no information about Siegel zeros is needed. The only zeros to be avoided are those of the twists, and only in a bounded box. The box has two jobs: it contains the real segment $(\beta,1)$, where Landau's lemma is applied, and the point ρ itself. A small neighbourhood of a nonreal ρ would not do, since it would not supply the argument on the real axis.

Corollary 3.3 ([*proved, given* (2.9) *and Theorem 4.1*])**.** *The Riemann hypothesis holds.*

Proof. Let $\zeta(\rho) = 0$ with $\frac{1}{2} < \operatorname{Re}\rho < 1$. Choose $\frac{1}{2} < \beta < \operatorname{Re}\rho$ and $H = |\operatorname{Im}\rho|$; this fixes the box. Take d from Theorem 4.1, so that (ii) holds. By (2.9), (i) holds. Theorem 3.2 shows that $\zeta(s)\prod_{p|Q}(1-p^{-s})$, the L-function of the principal character, does not vanish at ρ, a contradiction. □

The order of the choices matters. The bound (2.9) is proved for every d in the class by a construction that uses no information about zeros. Only afterwards, given the zero ρ, are the box and then d chosen, and d is then held fixed as $y \to \infty$. The same argument gives the generalized Riemann hypothesis for every $L(s,\psi)$, $\psi \bmod Q$; for a general modulus Q is enlarged (Section 15).

4. Choosing d by a zero-density estimate

Theorem 4.1 (Choice of the prime; [*from the Lean; outlined below*])**.** *Let $\beta > \frac{1}{2}$ and H be real. There is a prime $d \equiv 9239 \pmod{Q}$ such that for every character $\xi \bmod Q$,*

$$L(s, \xi\chi_d) \neq 0 \qquad \text{for } \operatorname{Re} s \geq \beta,\ |\operatorname{Im} s| \leq H. \tag{4.1}$$

More precisely, with $\beta = \frac{1}{2} + 9\delta$, at most $KX^{1-\delta/2}$ of the primes $d \leq X$ fail this, where K depends on Q, δ and H.

Why this is plausible although zeros get denser. For a primitive character χ of conductor q, the number of zeros of $L(s,\chi)$ with $0 < \operatorname{Im} s \leq H$ is [6, Ch. 16]

$$\frac{H}{2\pi}\log\frac{qH}{2\pi e} + O(\log qH). \tag{4.2}$$

For a fixed height this is $O_H(\log q)$, and it is expected to grow on this scale. So as d grows, the twists $L(s, \xi\chi_d)$ have more and more zeros in any fixed height range. But the box of Theorem 4.1 is not a neighbourhood of the critical line. It is the region $\operatorname{Re} s \geq \beta$ with $\beta > \frac{1}{2}$ *fixed*, a fixed distance $\beta - \frac{1}{2}$ to the right of the line. The relevant fact is of a different kind: for each fixed $\beta > \frac{1}{2}$, a zero-density estimate bounds the number of characters that have any zero in the box by a quantity smaller than the size of the family. Under GRH there are no zeros in the box at all; the theorem asks only that this hold for *one* d, and proves it for almost all.

Zero-density estimates make this quantitative. Let $N(\sigma, T, \chi)$ be the number of zeros of $L(s,\chi)$ in the rectangle $\sigma \leq \operatorname{Re} s \leq 1$, $|\operatorname{Im} s| \leq T$, and let $S(X)$ be the set of real primitive characters of conductor at most X. Heath-Brown's large sieve for real characters gives [11, Theorem 3]

$$\sum_{\chi \in S(X)} N(\sigma, T, \chi) \ \ll_\varepsilon\ (XT)^{\varepsilon}\, X^{3(1-\sigma)/(2-\sigma)}\, T^{(3-2\sigma)/(2-\sigma)}, \tag{4.3}$$

improving in the X-aspect Jutila's earlier bound [12]

$$\sum_{\chi \in S(X)} N(\sigma, T, \chi) \ \ll_\varepsilon\ (XT)^{(7-6\sigma)/(6-4\sigma)+\varepsilon}. \tag{4.4}$$

For fixed $T = \max(1, H)$ the exponent of X is less than 1 exactly when $\sigma > \frac{1}{2}$ (and the same holds for Jutila's exponent). Since $S(X)$ has about $\frac{6}{\pi^2}X$ elements, the average number of zeros per character in the box tends to 0, although the total number of zeros up to height H grows like $\log X$. The twists $\psi\chi_d$ by a fixed character $\psi \bmod Q$ are handled in the same way. The proof uses a much weaker estimate, proved from the Pólya–Vinogradov inequality alone, which is all that is needed: the *number* of bad $d \leq X$ is $O(X^{1-\delta/2})$, while the number of admissible primes $d \leq X$ is of order $X/\log X$. The point is not that the zeros of $L(s,\chi_d)$ are few, but that zeros to the right of $\frac{1}{2} + 9\delta$ are.

The proof in outline. The proof is in `Lib/Analytic/QuadraticTwist.lean` of [39]; we describe it [*from the Lean*]. For a sequence c with $|c(n)| \leq 1$ and a parity $j \in \{0, 1\}$, let

$$\theta_c(u) = u^{1/2}\sum_{n\geq 1} c(n)\,(nu)^j e^{-\pi n^2 u^2}, \qquad J(c,T) = \int_{1/T}^{\infty} (|\theta_c(u)| + |\theta_{\mu c}(u)|)\,\frac{du}{u}. \tag{4.5}$$

(1) *A zero is detected.* Let $\chi \neq 1$ be a primitive character with conductor at most T^2, let j be its parity ($\chi(-1) = (-1)^j$), and let μ be the Möbius function. If $L(s,\chi)$ has a zero with $\operatorname{Re} s \geq \frac{1}{2} + 9\delta$, $|\operatorname{Im} s| \leq H$, then

$$J(\chi, T) > T^{2\delta} \tag{4.6}$$

for large T. One applies the three-lines theorem on $\frac{1}{2} \le \operatorname{Re} z \le 2$ to

$$\Psi(z) = \Gamma_j(z) L(z,\chi) M(z) - \Gamma_j(z)^2, \tag{4.7}$$

where

$$\Gamma_j(z) = \tfrac{1}{2}\,\pi^{-(z+j)/2}\,\Gamma\Big(\frac{z+j}{2}\Big), \qquad M(z) = \int_{1/T}^{\infty} \theta_{\mu\chi}(u)\, u^{z-1/2}\,\frac{du}{u}. \tag{4.8}$$

The Mellin transform of θ_χ at $z - \frac{1}{2}$ is $\Gamma_j(z)L(z,\chi)$, and M is the corresponding transform for $\mu\chi$, cut off at $1/T$: a mollifier, approximating $\Gamma_j(z)/L(z,\chi)$ on $\operatorname{Re} z = 2$. So Ψ is $O(1/T)$ on $\operatorname{Re} z = 2$. On $\operatorname{Re} z = \frac{1}{2}$ it is at most about J^2: the functional equation of θ_χ folds the range $u < q^{-1/2}$ onto $u > q^{-1/2} \ge 1/T$, which is why J integrates only over $u \ge 1/T$ and why the conductor must be at most T^2. At the zero, Ψ equals $-\Gamma_j^2$, which is bounded away from 0.

(2) *J is small on average.* Step 1 needs a primitive character, so fix a character $\xi \bmod Q$ and let ξ^* be the primitive character inducing it, of conductor $q_\xi \mid Q$. Then $\xi^*\chi_d$ is primitive of conductor $q_\xi d$, and $L(s, \xi\chi_d)$ differs from $L(s, \xi^*\chi_d)$ only by Euler factors at the primes dividing Q, whose zeros lie on $\operatorname{Re} s = 0$. For an integer $v \ge 1$ let $\alpha_v(n) = \xi^*(n)\big(\frac{-v}{n^\circ}\big)$, a Jacobi symbol, where n° is the odd part of n. Since $d \equiv 7 \pmod 8$, quadratic reciprocity gives $\big(\frac{n}{d}\big) = \big(\frac{-d}{n^\circ}\big)$, so $\alpha_d = \xi^*\chi_d$: the twists are among the α_v. (This is where the congruence $d \equiv 7 \pmod 8$ is used in this section.) A mean-square bound for the theta series of the α_v, obtained by expanding the square, bounding the pairs with $(nn')^\circ$ a square trivially and the others by the Pólya–Vinogradov inequality, gives

$$\sum_{v \le T^2} J(\alpha_v, T) \ll T^{2+\delta}. \tag{4.9}$$

(3) *Counting.* Take $T^2 = QX$, so that the conductor of $\xi^*\chi_d$ is at most T^2 for $d \le X$. Each bad d contributes more than $T^{2\delta}$, so for each of the finitely many ξ there are $\ll X^{1-\delta/2}$ of them with $d \le X$. By partial summation the bad d then have $\sum (\log d)/d < \infty$, while over all primes $d \equiv 9239 \pmod Q$ this sum diverges (Mertens's theorem in arithmetic progressions [6]). So not all of them are bad.

5. Chebyshev's principle

Chebyshev's argument, in Erdős's version [23], rests on a simple observation about the binomial coefficient $\binom{2n}{n}$: every prime $n < p \le 2n$ divides it, and it is at most 4^n, so

$$\sum_{n < p \le 2n} \log p \le \log \binom{2n}{n} \le n \log 4. \tag{5.1}$$

The divisibility of one nonzero integer, compared with its size, bounds a weighted sum over primes from above. This is exactly the direction Step 1 asks for. The same reasoning applies to a nonzero algebraic integer Δ, for instance a determinant, in a number field of degree n. Suppose $p^{e_p} \mid \Delta$ for each p and $|\sigma(\Delta)| \le e^U$ for every embedding σ into $\mathbb{C}$ ("at every conjugate"). Then:

- *Norm.* $N(\Delta) = \prod_\sigma \sigma(\Delta)$ is a nonzero rational integer, so $|N(\Delta)| \ge 1$.
- *Divisibility.* $p^{ne_p} \mid N(\Delta)$, since the norm of p is p^n.
- *Combining.*

$$n \sum_p e_p \log p \le \log |N(\Delta)| = \sum_\sigma \log |\sigma(\Delta)| \le nU, \tag{5.2}$$

so $\sum_p e_p \log p \le U$.

A bound at one embedding is not enough: $1-\sqrt{2}$ is a nonzero algebraic integer of absolute value less than 1. The proof uses exactly this scheme, with two changes (Sections 8 and 9). Its determinant is a Gram determinant H of a positive definite hermitian form; its covolume bound c gives $\log|\sigma(H)| \le c$ at every conjugate (how the conjugates of a Gram determinant are bounded is explained in Section 8). And since H may have denominators, the norm is replaced by the product formula (Section 8), the denominators being controlled by an index.

Chebyshev's method gives only the order of magnitude because of rounding. The exponent of p in $a!$ is

$$v_p(a!) = \Big\lfloor \frac{a}{p} \Big\rfloor + \Big\lfloor \frac{a}{p^2} \Big\rfloor + \cdots , \tag{5.3}$$

and the floors cost up to 1 per prime power, as much as the main term. The construction here first proves a smooth divisibility bound, a sum of *ramps*

$$\Big(\frac{a}{p^h} - w\Big)_+ = a\Big(\frac{1}{p^h} - \frac{1}{x}\Big)_+, \qquad x = \frac{a}{w}, \quad u_+ = \max(u,0), \tag{5.4}$$

and only then rounds, once per prime, when the forced factors are divided out. Because the order $a = xb$ is much larger than the cutoff x (Section 9), this final rounding is negligible after division by a: it costs $O(b/a) = O(1/x)$ [*from the Lean*]. Here a is the order of differentiation, w the weight spent at the partner residue $-p^h\tau$ (Lemma 6.1), and $x = a/w$ the length of the prime sum (Sections 6 and 9). Summing $\log p$ times these over the odd powers $n = p^h$ of inert primes gives

$$a \sum_{\substack{n=p^h \\ p \text{ inert},\ h \text{ odd}}} \log p \,\Big(\frac{1}{n} - \frac{1}{x}\Big)_+, \tag{5.5}$$

which is $\frac{a}{2}$ times a sum of the shape $S_d(x)$ of (2.3): the ramp is exactly the weight $\frac{1}{n} - \frac{1}{x}$. (The congruence $n \equiv 1 \pmod{Q}$ is supplied by the choice of residues in Section 9. Residues and their partners are defined in Section 6: a residue τ modulo q indexes a complex embedding of the field E, and so one of the variables of the theta functions.)

No single integer that we know does the job (Section 9), and the rest of the proof supplies the three ingredients in a stronger form: forced divisibility by the inert primes, given by ramps (Step 6), a size that matches it (Step 7), and a determinant whose upper bound (Hadamard, Theorems A and B) and lower bound (the product formula, Theorem C) agree up to a small error (Steps 8 and 9).

6. Complex multiplication forces divisibility by the inert primes

The fields K and E, and the partner congruence. *Why this background is needed.* The construction of Steps 6–14 uses a cyclotomic extension E of K to organize its coordinates, produces algebraic numbers (the particular entries used lie in a finite extension F of $\mathbb{Q}$, to which the product formula is applied), and reads off their divisibility by an inert prime p one complex embedding at a time. The local fact about E that defines the partner residues is a congruence modulo p linking the embedding indexed by a residue τ to the one indexed by $-p^h\tau$ (Lemma 6.1). That congruence is the source of three features of the proof: the ramps (5.4) of Section 5, the coupling of variables in Section 6, and the restriction to $n \equiv 1 \pmod{Q}$ in the prime sum S_d of Section 2. It also explains the congruence condition on d. We describe E concretely.

The fields. Fix a modulus q that is a multiple of Q with the same prime factors as Q; in the proof q is Q times powers of 2 and 3. Let $\zeta = e^{2\pi i/q}$ and

$$E = K(\zeta) = \mathbb{Q}(\sqrt{-d}, \zeta), \tag{6.1}$$

a field of degree $2\varphi(q)$ over $\mathbb{Q}$ (since d does not divide q, whose prime factors are those of Q, $\sqrt{-d} \notin \mathbb{Q}(\zeta)$). Every element of E can be written as

$$\alpha = \sum_{k \bmod q} c_k \zeta^k, \qquad c_k = u_k + v_k\sqrt{-d} \in K, \quad u_k, v_k \in \mathbb{Q}. \tag{6.2}$$

Embeddings indexed by residues. The embeddings of E into $\mathbb{C}$ that send $\sqrt{-d}$ to $i\sqrt{d}$ are indexed by the residues $\tau \in (\mathbb{Z}/q)^\times$: the embedding τ sends ζ to $e^{2\pi i\tau/q}$. We write $\alpha(\tau)$ for the image of α, and call it the *coordinate* of α at τ:

$$\alpha(\tau) = \sum_k c_k\, e^{2\pi i k\tau/q} \qquad (\sqrt{-d} \mapsto i\sqrt{d}). \tag{6.3}$$

The remaining $\varphi(q)$ embeddings are the complex conjugates of these. So α is determined by its $f = \varphi(q)$ coordinates. (In the literature a residue τ indexes an infinite *place* of E.) The theta functions of the proof have one complex variable z_τ for each τ, and "differentiating at τ" means differentiating in z_τ.

Lemma 6.1 (The partner residue; [*proved*])**.** *Let p be a prime with $p \nmid q$ (so p is odd, since $2 \mid q$) and $-d$ a non-square modulo p. Let*

$$R_p = \mathbb{Z}_{(p)}[\zeta, \sqrt{-d}] \subset E \tag{6.4}$$

be the ring of the numbers (6.2) with u_k, v_k having denominators prime to p, and let F_p be the automorphism of E with

$$F_p(\zeta) = \zeta^p, \qquad F_p(\sqrt{-d}) = -\sqrt{-d}. \tag{6.5}$$

Then for every $\alpha \in R_p$ and every $h \geq 1$,

$$\alpha^{p^h} - F_p^h(\alpha) \in pR_p. \tag{6.6}$$

In coordinates, for odd h: there is one $\beta \in R_p$ such that, for every residue τ,

$$\alpha(\tau)^{p^h} - \overline{\alpha(-p^h\tau)} = p\,\beta(\tau). \tag{6.7}$$

Proof. Since $p \nmid q$, $\zeta \mapsto \zeta^p$ is an automorphism of $\mathbb{Q}(\zeta)$; and since $\sqrt{-d} \notin \mathbb{Q}(\zeta)$, an automorphism of E may act on ζ and on $\sqrt{-d}$ independently. So F_p exists; it maps R_p to itself. In the ring R_p/pR_p the map $x \mapsto x^p$ is a ring homomorphism, because $(x+y)^p \equiv x^p + y^p \pmod p$. It agrees with F_p on the generators:

- $u^p \equiv u$ for $u \in \mathbb{Z}_{(p)}$ (Fermat);
- $\zeta^p = F_p(\zeta)$;
- $(\sqrt{-d})^p = (-d)^{(p-1)/2}\sqrt{-d} \equiv -\sqrt{-d} = F_p(\sqrt{-d})$, by Euler's criterion.

So $\alpha^p \equiv F_p(\alpha) \pmod{pR_p}$ for all $\alpha \in R_p$. Since F_p maps pR_p to itself, induction gives $\alpha^{p^h} = (\alpha^{p^{h-1}})^p \equiv F_p(\alpha^{p^{h-1}}) \equiv F_p^h(\alpha)$, which is (6.6). Put $\beta = (\alpha^{p^h} - F_p^h(\alpha))/p \in R_p$.

For the coordinates, write $\alpha = \sum_k c_k\zeta^k$ as in (6.2). For odd h, $F_p^h(\alpha) = \sum_k \bar{c}_k\,\zeta^{p^h k}$, where $\bar{c}_k = u_k - v_k\sqrt{-d}$. The embedding τ sends it to

$$\sum_k \bar{c}_k\, e^{2\pi i p^h k\tau/q} = \overline{\sum_k c_k\, e^{-2\pi i p^h k\tau/q}} = \overline{\alpha(-p^h\tau)}, \tag{6.8}$$

since τ sends $\bar{c}_k$ to the complex conjugate of the image of c_k. Applying the embedding τ to $\alpha^{p^h} - F_p^h(\alpha) = p\beta$ gives the claim. □

For odd h we call $-p^h\tau$ the *partner* of τ at p^h. In the literature F_p is the Frobenius automorphism at p, and $\tau \mapsto -p\tau$ is its action on the embeddings; the sign appears because p is inert, so that F_p acts on K as complex conjugation. For even h, F_p^h fixes $\sqrt{-d}$ and the same argument gives $\alpha(\tau)^{p^h} \equiv \alpha(p^h\tau)$, with no complex conjugation. The partners at odd h are the ones the construction uses, matching the odd powers in the weight $1 - \chi_d(p^h)$. The lemma does not apply to the ramified prime $p = d$, whose powers also carry weight in S_d; the proof treats d separately (Section 12).

What is required of q and d. The lemma needs $p \nmid q$: then $\tau \mapsto -p^h\tau$ is a permutation of $(\mathbb{Z}/q)^\times$, so every coordinate has a partner and the couplings of Section 6 are defined. (If $p \mid q$, multiplication by p is not even defined on $(\mathbb{Z}/q)^\times$.) The construction uses the lemma at every inert prime, so the requirement is:

$$\text{every prime dividing } q \text{ splits in } K, \tag{6.9}$$

so that every inert prime is prime to q. This is a condition on d, and it is arranged by the congruence $d \equiv Q/2 - 1 \pmod{Q}$: then $d \equiv 7 \pmod 8$ and $d \equiv -1 \pmod{\ell}$ for every odd prime $\ell \mid Q$, so $-d$ is a square modulo 8 and modulo each such ℓ. The moduli are constrained as follows [*from the Lean*]. In the proof $q = Q \cdot 3^c \cdot 2^s$ with $c \geq 5$ and s large; the extra powers of 2 and 3 make the number of special residues $\tau \equiv 1 \pmod Q$, which is $q/Q = f/\varphi(Q)$, as large as the construction needs. So 2 and 3 always divide q.

- q must be divisible by $5040 = 7! = 2^4 \cdot 3^2 \cdot 5 \cdot 7$, so that E contains the roots of unity of order 5040, which the global part of the construction uses. Since q/Q is a product of powers of 2 and 3, the factors 5 and 7 must come from Q.
- By (6.9) the primes 2 and 3, which divide q, must split in K. With $d \equiv Q/2 - 1 \pmod Q$ this requires $16 \mid Q$ (to get $d \equiv 7 \pmod 8$) and $3 \mid Q$ (to get $d \equiv -1 \pmod 3$).
- The special residues are the classes $\tau \equiv \pm 1 \pmod Q$ (Section 9), and their separation is used at several points of the proof. This is also why the prime sum is restricted to $n \equiv 1 \pmod Q$.

Hence $1680 = 2^4 \cdot 3 \cdot 5 \cdot 7$ must divide Q. The Lean development takes $Q = 18480 = 1680 \cdot 11$, but the factor 11 is unnecessary: we rebuilt both the readable and the compact editions of the Lean proof with 1680 in place of 18480 throughout and the small-prime bound adjusted ($p \geq 13$ becomes $p \geq 11$), and both compile without errors, with the same three standard axioms [*checked*]. With $Q = 18480$ the primes $2, 3, 5, 7, 11$ split, so every inert prime is at least 13; this is a consequence of (6.9), not a separate requirement.

The lemniscate model. We now turn to the functions. *This subsection is an illustration only.* The field $\mathbb{Q}(i)$ is not one of the fields of the proof. The lemniscate sine is the simplest model of the mechanism the proof uses: derivatives of a function with complex multiplication, and hence the determinants built from such derivatives in Section 8, are divisible by the inert primes, here the primes inert in $\mathbb{Q}(i)$. The proof's field $\mathbb{Q}(\sqrt{-d})$ is discussed after the model, in "In the proof" below.

The lemniscate sine sl, the inverse of $t \mapsto \int_0^t du/\sqrt{1-u^4}$, is an elliptic function for a square lattice. It has *complex multiplication*: $\mathrm{sl}(iz) = i\,\mathrm{sl}(z)$, and more generally $\mathrm{sl}((m+ni)z)$ can be expressed through $\mathrm{sl}(z)$ for all integers m, n. This follows from Euler's addition formula

$$\mathrm{sl}(u+v) = \frac{\mathrm{sl}u\,\mathrm{sl}'v + \mathrm{sl}v\,\mathrm{sl}'u}{1 + \mathrm{sl}^2u\,\mathrm{sl}^2v}, \qquad \mathrm{sl}'(u)^2 = 1 - \mathrm{sl}(u)^4. \tag{6.10}$$

The formula involves sl', so in general $\mathrm{sl}((m+ni)z)$ is a rational function of $\mathrm{sl}(z)$ and $\mathrm{sl}'(z)$; for example $\mathrm{sl}(2z) = 2\mathrm{sl}z\,\mathrm{sl}'z/(1+\mathrm{sl}^4 z)$. When $m+n$ is odd, the square roots $\mathrm{sl}' = \pm\sqrt{1-\mathrm{sl}^4}$ cancel and $\mathrm{sl}((m+ni)z)$ is a rational function of $\mathrm{sl}(z)$ alone (Abel). For example,

$$\mathrm{sl}(3z) = \mathrm{sl}z \cdot \frac{3 - 6\mathrm{sl}^4 z - \mathrm{sl}^8 z}{1 + 6\mathrm{sl}^4 z - 3\mathrm{sl}^8 z}. \tag{6.11}$$

The derivatives $A_a = \mathrm{sl}^{(a)}(0)$ of the lemniscate sine at 0 are integers: $1, -12, 3024, -4390848, \ldots$ for $a = 1, 5, 9, 13, \ldots$, and $A_a = 0$ unless $a \equiv 1 \pmod 4$. (This follows from $\mathrm{sl}'' = -2\,\mathrm{sl}^3$ and $\mathrm{sl}'^2 = 1 - \mathrm{sl}^4$; compare Hurwitz [13].)

Theorem 6.2 ([7])**.** *For every prime $p \equiv 3 \pmod 4$ and every a,*

$$p^{\lfloor ap/(p^2-1)\rfloor} \mid A_a. \tag{6.12}$$

The primes $p \equiv 3 \pmod 4$ are exactly those for which -1 is a non-square modulo p, the inert primes of $\mathbb{Q}(i)$; the character χ_d of Section 2 is replaced by the character χ_4 modulo 4. The exponent is rounded once per prime, not once per prime power as in Legendre's formula, and

$$\frac{ap}{p^2-1} = a\Big(\frac{1}{p} + \frac{1}{p^3} + \frac{1}{p^5} + \cdots\Big) \tag{6.13}$$

contains exactly the odd powers of p, matching the weight $1 - \chi_d(p^h)$, which vanishes for even h. The exponent is a lower bound; it is attained in most cases (in 99.6% of the cases with $p \le a \le 6001$ [*computed*]).

Example 6.3 ([*computed*])**.** A_{2001} has 5203 digits. The bound is attained exactly at all 155 primes $p \equiv 3 \pmod 4$ below 2001 (for instance 3^{750}, 7^{291}, 11^{183}), and no prime $p \equiv 1 \pmod 4$ below 6000 divides it.

Proposition 6.4 (A congruence at every odd prime; [*proved*])**.** *For every odd prime p and every $a \ge 1$,*

$$A_{a+p-1} \equiv A_p\, A_a \pmod p, \qquad A_p \equiv \begin{cases} (-1)^{(p-1)/4}\binom{(p-1)/2}{(p-1)/4} & p \equiv 1 \pmod 4, \\ 0 & p \equiv 3 \pmod 4. \end{cases} \tag{6.14}$$

For $p \equiv 1 \pmod 4$ the binomial coefficient is not divisible by p, so whether $p \mid A_a$ depends only on $a \bmod (p-1)$.

Proof. With $x = \mathrm{sl}(z)$, $y = \mathrm{sl}'(z)$ we have $x' = y$, $y' = -2x^3$, $y^2 = 1 - x^4$. So A_n is the value at $(x, y) = (0, 1)$ of $\mathcal{D}^n x$, where $\mathcal{D}$ is the derivation with $\mathcal{D}x = y$, $\mathcal{D}y = -2x^3$; in particular the A_n are integers. Now work modulo p. Let $Y = (1 - x^4)^{1/2} \in \mathbb{F}_p[[x]]$, with constant term 1, and $D = Y\, d/dx$, so that $Dx = Y$ and $DY = -2x^3$. By Leibniz's rule D^p is again a derivation, since the intermediate binomial coefficients vanish modulo p. From

$$D^2(Yx^k) = Y\big(k(k-1)x^{k-2} - (k+2)(k+3)x^{k+2}\big) \tag{6.15}$$

it follows by induction that $D^{2r+1}x = Y\,P_r(x)$ with a polynomial P_r of degree at most $2r$. So $D^p x = Y\,P(x)$ with $\deg P \le p-1$. Applying D^p to $Y^2 = 1 - x^4$ gives $D^p Y = -2x^3 P$. Since D and D^p commute, $D(YP) = D^p Y$, that is, $-2x^3 P + (1-x^4)P' = -2x^3 P$. Hence $P' = 0$, and with $\deg P \le p-1$ this makes P a constant c. So $D^p = cD$ on polynomials in x and Y; evaluating $D^p x$ at $x = 0$ gives $c \equiv A_p$, and $D^{a+p-1}x = D^{a-1}(cDx) = cD^a x$ gives the congruence.

To compute c, write $Y^{-1} = \sum_{j\ge 0} b_j x^j$ and $F(x) = \sum_{j=0}^{p-2} b_j x^{j+1}/(j+1)$. Then $DF = 1 - b_{p-1}x^{p-1} + O(x^p)$, so $(D^pF)(0) = -b_{p-1}(p-1)! = b_{p-1}$ by Wilson's theorem, while $D^pF = cDF$ gives $(D^pF)(0) = c$. Finally b_{p-1}, the coefficient of x^{p-1} in $(1-x^4)^{-1/2}$, is 0 if $p \equiv 3 \pmod 4$, and for $p = 4r+1$ it is $(-1)^r\binom{-1/2}{r} \equiv (-1)^r\binom{(p-1)/2}{r} \pmod p$, which is not divisible by p. □

Remark 6.5 (Split primes; [*computed*]). In this remark $a \equiv 1 \pmod 4$, the indices of the nonzero A_a. The absence of split primes below 6000 in A_{2001} is not a theorem, and it does not persist: the split prime 8101 divides A_{2001}. By Proposition 6.4, a split prime p divides some A_a if and only if it divides one of the finitely many A_r with $1 < r < p$, $r \equiv 1 \pmod 4$. (The binomial coefficient is the Hasse invariant.) For the 147 split primes below 2000 there are 35 such pairs (p, r) in total, against 35.8 expected if each A_r were divisible by p with probability $1/p$. But they are concentrated on few primes: no split prime below 157 divides any A_a, while $157 \mid A_a$ exactly for $a \equiv 101 \pmod{156}$. Split primes divide the A_a sporadically, like irregular primes for the Bernoulli numbers, not systematically like the inert primes.

Why the inert primes divide. Write a function as $f(z) = g(z/\varpi)$, where g has Taylor coefficients g_j that are p-adic integers and ϖ is a number with $v_p(\varpi) = \lambda$. Then

$$f^{(j)}(0) = j!\, g_j\, \varpi^{-j}, \qquad v_p(f^{(j)}(0)) \;\ge\; v_p(j!) - j\lambda. \tag{6.16}$$

So the derivatives are divisible by p at the rate $v_p(j!) \approx j/(p-1)$, minus the cost $j\lambda$ of the rescaling. Everything depends on how small a rescaling makes the function p-adically integral, and that is decided by the multiplication formula.

For an odd prime p the multiplication formula, as in (6.11), gives $\mathrm{sl}(pz) = T\,P(T^4)/R(T^4)$ with $T = \mathrm{sl}(z)$, P and R polynomials with integer coefficients and $R(0) = 1$. So $\mathrm{sl}(pz)$ is a power series in T with integer coefficients, $\mathrm{sl}(pz) = pT + \cdots$.

- $p \equiv 3 \pmod 4$. Modulo p every coefficient of T^k with $k < p^2$ vanishes, and the coefficient of T^{p^2} is a unit [40]. The terms pT and T^{p^2} balance when $v_p(T) = 1/(p^2-1)$, and this is the rescaling: $\lambda = 1/(p^2-1)$. The net rate is

$$\frac{1}{p-1} - \frac{1}{p^2-1} = \frac{p}{p^2-1}. \tag{6.17}$$

- $p \equiv 1 \pmod 4$. Here $p = \pi\bar\pi$ with $\pi = m + ni$, and $\mathrm{sl}(\pi z) \equiv \mathrm{sl}(z)^p \pmod \pi$ for suitably normalized π (Eisenstein's congruence). The terms πT and T^p balance at $v_p(T) = 1/(p-1)$, so $\lambda = 1/(p-1)$, and gain and loss cancel: nothing forces divisibility.

This is why CM derivatives see the inert primes, and only them. (In the language of elliptic curves: the reduction at an inert prime is supersingular, and its formal group has height two [40].) The height explains the leading rate. The simple estimate above loses the digit sum $s_p(j)/(p-1)$ in $v_p(j!) = (j - s_p(j))/(p-1)$; the sharp exponent $\lfloor jp/(p^2-1)\rfloor$ of the theorem needs a finer argument [7].

Two kinds of CM functions. Two different functions attached to a CM lattice Λ occur in this paper, and they should not be confused.

- *An elliptic function*, such as sl, or $t = -2\wp/\wp'$ for a general lattice. It is meromorphic and doubly periodic, with poles at a finite distance ρ from 0. By Cauchy's inequality its j-th derivative at 0 is at most $j!\, r^{-j} \max_{|z|=r} |f|$ for every $r < \rho$; for sl and t, whose poles are simple, the expansion at the nearest poles shows that it has size about $j!\,\rho^{-j}$ along the indices where it does not vanish. It is divisible by the inert primes at the rate $jp/(p^2-1)$. The examples of this section and of Appendix A are of this kind.

- *A theta function.* The function θ that the proof differentiates is entire: it is a translate of a sigma function σ_Λ of the lattice by a half-period t_0, multiplied by an exponential and normalized so that $\theta(0)=1$,

$$\theta(z)=\exp\Big(-\frac{\pi z\,\bar{t}_0}{A_\Lambda}\Big)\frac{\sigma_\Lambda(z+t_0)}{\sigma_\Lambda(t_0)}, \tag{6.18}$$

where A_Λ is the area of a period parallelogram of Λ. Here Λ has complex multiplication by the integers of $K=\mathbb{Q}(\sqrt{-d})$, with $d\equiv 7 \pmod 8$, and σ_Λ is Weierstrass's sigma function multiplied by the factor e^{cz^2} for which

$$\sigma_\Lambda(z+l)=\varepsilon(l)\exp\Big(\frac{\pi\bar{l}}{A_\Lambda}\Big(z+\frac{l}{2}\Big)\Big)\sigma_\Lambda(z)\qquad(l\in\Lambda),$$

with $\varepsilon(l)=1$ for $l\in 2\Lambda$ and $\varepsilon(l)=-1$ otherwise. The point t_0 is the half-period, unique modulo Λ, for which neither $\omega_0 t_0$ nor $\bar{\omega}_0 t_0$ lies in Λ, where $\omega_0=(1+\sqrt{-d})/2$; and Λ and t_0 are scaled together so that $\wp''_\Lambda(t_0)=2$. With this normalization the derivatives of θ at 0 are algebraic numbers [*from the Lean*]. The odd ones vanish, and $(\sqrt{-d})^{n/2}\theta^{(n)}(0)$ is an algebraic integer for even n: the function $u(z)=\wp_\Lambda(z+t_0)-\wp_\Lambda(t_0)$ satisfies $u''=6u^2+a_1^2u+2$, $u(0)=u'(0)=0$, where $a_1^2=12\wp_\Lambda(t_0)$ is an algebraic integer, and $(\log\theta)''=\theta''(0)-u$, where $\sqrt{-d}\,\theta''(0)$ is an algebraic integer [*these two integrality statements from the Lean*]; differentiating the equation for u shows that the derivatives of u at 0 are algebraic integers, the derivatives of an exponential are polynomials with integer coefficients in the derivatives of its exponent, and each term in $\theta^{(n)}(0)$ contains at most $n/2$ factors $\theta''(0)$. The proof differentiates products of translates of θ by torsion points (with the exponential factor that belongs to a translate), composed with linear forms in several variables. Unlike an elliptic function, it has no poles, so bounds for its derivatives come from its Gaussian growth and not from a nearest pole: Cauchy's estimate on a circle of radius about $\sqrt{j}$ gives an upper bound of the form $C_1C_2^j\sqrt{j!}$ (an upper bound, not an asymptotic). The precise bound used in the proof is the size estimate of Theorem A. Its derivatives at 0 are also *less* divisible than those of t: for the sigma function of the lemniscatic lattice the 401st derivative is divisible by 3^{48}, against 3^{150} for the rate above [*computed*].

The link between the two is a normal form at each inert prime p,

$$\theta(\varpi z)=W(z)\exp\Big(\frac{\varpi^{p+1}}{p}\,\Xi(z)\Big), \tag{6.19}$$

with W and Ξ power series with p-integral coefficients and ϖ a rescaling number as in the preceding subsection [*from the Lean*]. So θ is an integral series times an exponential factor. In several variables it is this factor that couples the variable z_τ to the variable of the partner residue, and the proof deals with it at a cost that is accounted for (Section 12). So the statement used in the proof is not "the derivatives of θ are divisible at the rate $jp/(p^2-1)$", but the ramp estimate below, for suitable combinations of derivatives.

In the proof. The same mechanism holds for $K=\mathbb{Q}(\sqrt{-d})$. For the lattice $\omega\mathcal{O}_K$, with ω the period of Section 7 suitably normalized, the derivatives at 0 of $t=-2\wp/\wp'$ are divisible by the inert primes at the rate $np/(p^2-1)$. For $d=9239$, with class number 139, the bound $\lfloor np/(p^2-1)\rfloor$ holds at all ten inert primes $p\le 71$ for odd $n\le 241$, with equality in 96.6% of the cases [*computed*]; Appendix A, A.3, treats $d=23$ in detail. When the class number is larger than one, the class sums of Section 8 involve the lattices of all the ideal classes of K, as in that example. In the proof the basic function is the theta function (6.18) for the lattice $\omega\mathcal{O}_K\subset\mathbb{C}$,

where $\mathcal{O}_K = \mathbb{Z} + \mathbb{Z}\frac{1+\sqrt{-d}}{2}$; the many-variable theta functions of Section 8 are products of θ evaluated at linear forms in the variables z_τ. A derivative of order a in the variable z_τ is linked, by the congruence of Lemma 6.1, to the variable $z_{-p^h\tau}$ of the partner residue: the proof pairs them in monomials $z_\tau^{p^h} z_{-p^h\tau}$ (see (12.13)). Divisibility a/p^h is gained at τ only by spending the order $\kappa(-p^h\tau)$ assigned to the partner, so the guaranteed divisibility contributed by p^h is the ramp $\left(a/p^h - \kappa(-p^h\tau)\right)_+$ of (5.4) [*from the Lean*]; Section 12 derives it, in (12.14).

7. A Gaussian norm and the period

Divisibility alone bounds nothing; it must be compared with size, and for the comparison to give an error of size $x^{-1/2}$ the two must agree in their main terms. A single derivative does not achieve this. In the lemniscate example $\log|A_{2001}| \approx 11980$ while the inert primes account for about 6188: only half (§9).

The proof makes two adjustments.

- *Vectors, with a Gaussian norm.* It works with vectors of derivatives, measured in the Gaussian (Bargmann–Fock) norm

$$\|g\|^2 = \frac{1}{\pi}\int_{\mathbb{C}} |g(z)|^2 e^{-|z|^2}\, dA(z) \tag{7.1}$$

on polynomials g in the variable z of the theta functions (in several variables, one such factor for each variable). The monomials are orthogonal, and $\|z^j\|^2 = j!$: in polar coordinates,

$$\frac{1}{\pi}\int_{\mathbb{C}} z^j \bar{z}^k e^{-|z|^2} dA = \frac{1}{\pi}\int_0^{2\pi} e^{i(j-k)\phi}\, d\phi \int_0^\infty r^{j+k+1} e^{-r^2} dr = \delta_{jk}\, j!. \tag{7.2}$$

Hence

$$\Big\|\sum_j c_j z^j\Big\|^2 = \sum_j |c_j|^2\, j!. \tag{7.3}$$

In this norm the term $c_j z^j$ of a Taylor series, $c_j = f^{(j)}(0)/j!$, has squared length $|f^{(j)}(0)|^2/j!$ [24]. For an elliptic function with complex multiplication, such as sl (Section 6), the two sides of the comparison are then:

- *size:* $|f^{(j)}(0)|$ is about $j!$ times a geometric factor, so the squared length has logarithm $\log j! + O(j) = j\log j + O(j)$;
- *divisibility, counted twice:* $f^{(j)}(0)$ is divisible by $p^{\lfloor jp/(p^2-1)\rfloor}$ at each inert prime, and by Mertens's theorem for the inert primes, $\sum_{p\le j,\ p\text{ inert}} (\log p)/p = \frac{1}{2}\log j + O(1)$, so $2\sum_p \frac{j}{p}\log p = j\log j + O(j)$.

So in the Gaussian norm size and divisibility agree in the term $j\log j$. Measured by its absolute value, a single derivative shows only one half (Section 9). For the lemniscate sine the comparison can be carried through the term linear in j as well, completely explicitly (Appendix A, A.1): with $\mathrm{size}(n) = \log(A_n^2/n!)$ and $\mathrm{ramp}(n)$ the sum of the ramps of Section 5 over the inert primes,

$$\mathrm{size}(n) - \mathrm{ramp}(n) = 2n\log\frac{2}{\pi} + o(n), \tag{7.4}$$

and for $n = 2001$ the two sides, of size about 10^4, differ from this by 4.4. Two cautions. First, this compares the size of a squared length with twice the divisibility of its *numerator*: $|f^{(j)}(0)|^2/j!$ is not an integer, dividing by $j!$ is not free at the finite primes, and for a single derivative no inequality follows. Second, the functions of the proof are theta functions, not elliptic functions (Section 6), and for them the size is not read off from a nearest pole. For

the vectors of the proof the comparison is the content of the estimates of Section 12: a derivative of order j is charged the size $s_\infty(j) = \log j! + j\log(M/\Omega)$ of (12.1) [*from the Lean*].

- *The period.* The next term is linear in the order j, and it is governed by the scale of the lattice.

  The CM period. Let E_0 be an elliptic curve $y^2 = 4x^3 - g_2x - g_3$ with g_2, g_3 algebraic and complex multiplication by $\mathcal{O}_K$, $K = \mathbb{Q}(\sqrt{-d})$. Its lattice of periods $\{\oint dx/y\}$ is $\omega\,\mathfrak{a}$ for a fractional ideal $\mathfrak{a}$ of K and a number $\omega \in \mathbb{C}^\times$; changing the model or the ideal multiplies ω by an algebraic number. This ω is the *CM period* of K, well defined up to algebraic factors. For the lattice $\omega\mathcal{O}_K$ the Weierstrass function satisfies $\wp'^2 = 4\wp^3 - g_2\wp - g_3$ with algebraic g_2, g_3, so the Taylor coefficients of functions such as $t = -2\wp/\wp'$ are algebraic numbers (and integral after a suitable normalization). This is why the functions of the proof are attached to the lattice $\omega\mathcal{O}_K$, not to $\mathcal{O}_K$.

  Why the size involves $j\log(1/|\omega|)$. Scaling the lattice by ω scales the variable. For $t_\Lambda = -2\wp_\Lambda/\wp'_\Lambda$ the exact law is $t_{\omega\Lambda}(z) = \omega\, t_\Lambda(z/\omega)$, so $t^{(j)}_{\omega\Lambda}(0) = \omega^{1-j} t^{(j)}_\Lambda(0)$; a normalized theta function scales as $F(z/\omega)$, with ω^{-j}. In both cases the term linear in j is $j\log(1/|\omega|)$. Equivalently, the nearest singularity of f is at distance proportional to $|\omega|$, and and the j-th derivative of an elliptic function with simple poles, such as t, is at most about $j!\,(\rho|\omega|)^{-j}$, with ρ the distance from 0 to the nearest pole for the lattice $\mathcal{O}_K$ (a pole of order ν would add a factor of size $j^{\nu-1}$). So

$$\log|f^{(j)}(0)| \le \log j! + j\log\frac{1}{\rho\,|\omega|} + O(1), \tag{7.5}$$

  and in a squared length the period contributes $-2j\log|\omega|$.

  The Chowla–Selberg formula [4]. Let $d \equiv 3 \pmod 4$ be a prime, $d > 3$, as in the proof; let $D = -d$, $e_K = |\mu_K| = 2$ the number of roots of unity in K, and h_K the class number. Then ω is an algebraic number times

$$\omega_{\mathrm{CS}} = \sqrt{\pi}\prod_{a=1}^{d-1}\Gamma\Big(\frac{a}{d}\Big)^{e_K\chi_d(a)/(4h_K)}. \tag{7.6}$$

  The classical formula explains why $\frac{L'}{L}(1,\chi_d)$ appears. The construction uses a precise normalization: the area of the lattice of each conjugate CM curve is fixed, up to a number that is a unit at every prime, and a further factor accounts for the different of E. That factor is a unit at the primes not dividing q, so it does not change the valuations at the inert primes. Together these give the exact scale Ω of (7.11) [*from the Lean*].

  Example. The formula holds for every imaginary quadratic field, with χ_d replaced by the character of the field. For $K = \mathbb{Q}(i)$ ($D = -4$, $e_K = 4$, $h_K = 1$) the curve $y^2 = 4x^3 - 4x$ is the lemniscatic curve, and its real period is the lemniscate constant

$$\varpi = 2\int_0^1 \frac{dt}{\sqrt{1-t^4}} = \frac{\Gamma(1/4)^2}{2\sqrt{2\pi}} = 2.62205\ldots, \tag{7.7}$$

  while (7.6) gives $\omega_{\mathrm{CS}} = \sqrt{\pi}\,\Gamma(\frac14)/\Gamma(\frac34) = 2\varpi$, by $\Gamma(\frac14)\Gamma(\frac34) = \pi\sqrt{2}$.

  Lerch's formula [5]. For the Hurwitz zeta function $\zeta(s,x)$,

$$\frac{\partial}{\partial s}\zeta(s,x)\Big|_{s=0} = \log\Gamma(x) - \tfrac12\log(2\pi). \tag{7.8}$$

Since $L(s,\chi_d) = d^{-s}\sum_{a=1}^{d-1}\chi_d(a)\zeta(s,a/d)$ and $\sum_a \chi_d(a) = 0$, this gives

$$\sum_{a=1}^{d-1}\chi_d(a)\log\Gamma\Big(\frac{a}{d}\Big) = L'(0,\chi_d) + L(0,\chi_d)\log d. \tag{7.9}$$

The period and $\frac{L'}{L}(1,\chi_d)$ [*proved, from* (7.6) *and* (7.9)]. Dirichlet's class number formula $h_K = e_K\sqrt{d}\,L(1,\chi_d)/(2\pi)$ [6, Ch. 6, (15)] and the functional equation $L(0,\chi_d) = \sqrt{d}\,L(1,\chi_d)/\pi$ [6, Ch. 9, (13)–(14)] give $L(0,\chi_d) = 2h_K/e_K$. With this, (7.6) and (7.9) give $\log\omega_{\mathrm{CS}} = \frac{1}{2}\log\pi + \frac{1}{2}(\frac{L'}{L}(0,\chi_d) + \log d)$. The functional equation of $\Lambda(s) = (d/\pi)^{(s+1)/2}\Gamma(\frac{s+1}{2})L(s,\chi_d) = \Lambda(1-s)$ gives $\frac{L'}{L}(0,\chi_d) + \frac{L'}{L}(1,\chi_d) = -\log\frac{d}{\pi} + \gamma + \log 2$. Hence

$$|\omega_{\mathrm{CS}}|^2 = 2\pi^2\exp\Big(\gamma - \frac{L'}{L}(1,\chi_d)\Big). \tag{7.10}$$

For $\mathbb{Q}(i)$ this reads $\varpi^2 = \frac{\pi^2}{2}\exp(\gamma - \frac{L'}{L}(1,\chi_4))$, with χ_4 the character modulo 4.

So the factor $|\omega|^{-2j}$ in a squared length contributes $j(\frac{L'}{L}(1,\chi_d) - \gamma)$, plus j times the logarithm of a constant made of 2, π and algebraic numbers. The first part is exactly the principal-character part of $\varphi(Q)\,c_d$ in Proposition 2.2. The Lean development packages the period, together with these constants, the algebraic factor of (7.6) and the normalization of the field E, as

$$\Omega = \exp\Big(\gamma - \mathrm{Re}\,\frac{L'}{L}(1,\chi_d)\Big)\cdot\frac{2\pi}{\mathrm{rd}_E}, \tag{7.11}$$

with $\mathrm{rd}_E = |\mathrm{disc}(E)|^{1/[E:\mathbb{Q}]}$ the root discriminant of the field E. (Since χ_d is real, $\frac{L'}{L}(1,\chi_d)$ is real; the Lean writes its real part.) The remaining terms of c_d, from the characters $\psi \neq \psi_0$, never need to be matched: they cancel in the two-scale identity (Proposition 9.2). The matching itself is carried out in Section 12.

So the period accounts for the constant $\frac{L'}{L}(1,\chi_d) - \gamma$. Other linear terms remain at this point: with $c_1 = \frac{L'}{L}(1,\chi_d) - \gamma - 1$ and $G_0 = \log(M\,\mathrm{rd}_E/2\pi)$ one has $\log(M/\Omega) = 1 + c_1 + G_0$, and the terms coming from the weights and from the level M are compensated by the choice of the auxiliary prime N (Section 12). After that, what remains is the remainder R [*from the Lean*].

It is natural that the period involves $\frac{L'}{L}(1,\chi_d)$: the sum $\sum\chi_d(n)\Lambda(n)/n$ has value $-\frac{L'}{L}(1,\chi_d)$ (Proposition 2.2), and the period is how the arithmetic of K enters the size of CM derivatives.

8. A Gram determinant

We will construct, for each large x, a positive algebraic number H with the properties below. They are stated with the parameters of Section 9: $m \approx \sqrt{x}$ is the number of lifts, $\mathcal{E} = fa(\log b)^4$ is the error unit, and rem is the remainder (9.3), which measures $R(x)$ by the two-scale identity (Proposition 9.2).

(H1) *A Gram determinant of almost D vectors.* $H = \det(\langle v_i, v_j\rangle)_{1\le i,j\le n}$ for vectors $v_1,\dots,v_n$ in a space with a positive definite hermitian inner product, where n is close to a dimension D:

$$\Big(1 - \frac{4}{b^2}\Big)D \;\le\; n \;\le\; D. \tag{8.1}$$

(H2) *Small at every conjugate.* For every automorphism σ of $\mathbb{C}$, let H_σ be the Gram determinant of the vectors σv_i, obtained by applying σ to every entry. Then

$$\log H_\sigma \;\le\; n(h - m\cdot\mathrm{rem} + \mathcal{E}) \qquad \text{for every } \sigma. \tag{8.2}$$

In the language of Section 9, the family has *covolume* at most the right side. Consequently $\log|\sigma(H)|$ satisfies the same bound for every conjugate of H (see (8.25)); the converse implication does not hold. Here $h \geq 0$ is an enormous quantity, the same for every vector, which is never computed.

(H3) *Large, as an algebraic number with few denominators.* The denominators of the vectors are bounded by an index $\delta \leq 10D\log(1+2b^2)\, fa(\log b)^3$, and whenever the family has covolume at most c (Section 9; then $\log|\sigma(H)| \leq c$ at every conjugate),

$$D(h - 3\mathcal{E}) \;\leq\; c + 2\delta. \tag{8.3}$$

Property (H2) plays the role of Chebyshev's upper bound $\binom{2n}{n} \leq 4^n$ after the primes have been divided out, and (H3) the role of "a nonzero integer is at least 1" (Section 5). The same h appears in both: it is the size of a vector less the divisibility it is guaranteed. Because $n \leq D$, it cancels when (H2) and (H3) are combined, and Section 9 shows that what is left is the main inequality $m \cdot \mathrm{rem} \leq fa(\log b)^5$. Section 10 turns this into the bound for $R(x)$. Properties (H2) and (H3) are Theorems (A)–(C) [*from the Lean*]. This section describes the determinant and the vectors, from the general shape down to the entries. Appendix A works out three small examples with actual numbers: a single vector, an 8×8 Gram determinant, and a chain of lifts over three classes of lattices. It also shows where the uncontrolled cofactor sits and recalls how the proof deals with it.

Why a Gram determinant. For vectors $v_1, \ldots, v_n$ the Gram determinant

$$H = \det\big(\langle v_i, v_j\rangle\big)_{1\leq i,j\leq n} > 0 \tag{8.4}$$

is the squared volume of the box they span; it is positive exactly when the v_i are linearly independent. In the proof the v_i are the lifts of independent target forms, and they are independent because the chain of lifts is injective on the target forms, for the untwisted chain and for every twist used [*from the Lean*]. This is an arithmetic non-vanishing statement, not linear algebra; it also makes the determinant ratios below well defined. Three facts about H play the roles of Chebyshev's three facts:

(1) *Size.* Hadamard's inequality

$$H \leq \prod_{i=1}^{n} \|v_i\|^2 \tag{8.5}$$

bounds it from above in terms of the Gaussian norms of Step 7.

(2) *Divisibility.* If every entry of a vector is divisible by a power of p, so is its contribution to H; the inert-prime divisibility of Step 6 passes to H.

(3) *Lower bound.* For a nonzero integer, "≥ 1". For algebraic numbers, the product formula plays this role: for $\alpha \neq 0$ in a number field,

$$\prod_{\sigma} |\sigma(\alpha)| \cdot \prod_{p} \prod_{\mathfrak{p}\mid p} |\alpha|_{\mathfrak{p}} = 1, \tag{8.6}$$

the first product over the embeddings σ of the field into $\mathbb{C}$ and the second over its prime ideals, with $|\alpha|_{\mathfrak{p}} = (\mathrm{N}\mathfrak{p})^{-v_{\mathfrak{p}}(\alpha)}$ (for an algebraic integer this is the norm argument of Section 5). So a nonzero algebraic number that is very divisible by primes must be large at the infinite ones, *provided* its denominators are controlled. In the proof the product formula is applied to one fixed number: the untwisted Gram determinant $\widetilde{H}$ of the lifts of a *basis* of the target space, obtained by completing $y_1, \ldots, y_n$ with $n' = D - n$ auxiliary forms. The first product is then the absolute value of the norm of $\widetilde{H}$, the product of all its conjugates; this is why bounds are needed at *every* conjugate, since a bound at one embedding says

nothing about a norm. The original family enters through its covolume bound: if the auxiliary lifts have logarithmic squared length at most B_{aux} at every conjugate, Fischer's inequality and (8.25) give

$$|\sigma(\widetilde{H})|^2 \;\le\; \widetilde{H}_{\sigma^c}\,\widetilde{H}_{\sigma} \;\le\; H_{\sigma^c}H_{\sigma}\,e^{2n'B_{\mathrm{aux}}} \;\le\; e^{2(c+n'B_{\mathrm{aux}})}. \tag{8.7}$$

So every conjugate of the completed determinant is bounded above in terms of c, and with the local comparisons this gives the lower bound of Theorem C for the covolume parameter c. It is not a separate lower bound for the norm of the $n \times n$ determinant H itself. (The two differ: with $\beta = \sqrt{\sqrt{2}-1}$, $y = (\sqrt{2}, \beta)$ and $z = (0, 1)$, one has $H = \langle y, y\rangle = 1 + \sqrt{2}$ of norm 1, while the determinant of the pair y, z is 2.) The twisted determinants $Z(\xi)$ below serve only as local comparisons for this one $\widetilde{H}$; no single twist has to work at every prime.

These facts dictate what the vectors must be.

- *Divisible at the right rates.* Their entries must be divisible by each inert prime p at the rates of the ramps of Section 5, so that the guaranteed divisibility is the prime sum s_{fin} of (9.1), whose surplus over its average is rem. Derivatives of CM theta functions have exactly this divisibility (Section 6).
- *Of computable length.* Their lengths must be bounded at every conjugate, with the same main terms as the divisibility. This is the role of the Gaussian norm and the period (Section 7).
- *Algebraic, with controlled denominators,* so that the product formula applies.
- *Many.* A lower bound of size Dh is available only for a family that nearly fills a space of dimension D; hence (8.1). One vector, or a few, cannot carry it: products, ratios or small determinants of single CM derivatives do not suffice (Section 9), and it is the high dimension n and the many variables that let size and divisibility agree to within the required error.
- *Amplified.* The remainder must appear with the factor $m \approx \sqrt{x}$ in (8.2): dividing the main inequality by m is what produces the exponent $\frac{1}{2}$ (Section 10).

How the vectors are built. The vectors are built in three stages,

$$\varphi \;\longmapsto\; y = g\,\mathcal{P}\varphi \;\longmapsto\; v = \Theta_1 y, \tag{8.8}$$

from a *source form* φ, through a *target form* y, to its image v under a chain of m theta lifts. Here $\mathcal{P}$ is a Poincaré series and g an element of a finite group G; the precise map, with its normalizing constants, is (11.2). Each stage meets one of the requirements above:

- source forms are where lengths and divisibility can be computed, as finite sums of CM theta derivatives, and where short integral vectors are found (Theorem A);
- target forms are where a family nearly filling a space of known dimension D exists (Theorem B) and where the lower bound (H3) is proved (Theorem C);
- the chain of lifts multiplies the remainder by m, and its last space is definite, so the final vectors are again finite lists of numbers.

We first describe the common ingredients: the field, lattices, forms and theta lifts. Then we describe the three stages. The analogy to keep in mind is the theory of quadratic forms in a fixed genus, with theta series built from positive definite forms.

The field E and hermitian vectors. Divisibility by an inert prime is read one complex embedding at a time, and the ramps come from the congruence linking the embedding τ to its partner $-p^h\tau$ (Section 6). To place different orders of differentiation at different embeddings, and so produce the cutoffs x and x_1, one needs many embeddings, indexed by residues. Let E be the field of numbers (6.2), with q equal to Q times powers of 2 and 3, so that each

number has $f = \varphi(q) \geq 3840$ coordinates $\alpha(\tau)$, $\tau \in (\mathbb{Z}/q)^\times$. Its integers $\mathcal{O}_E$ are the α that are algebraic integers. Write E^l for the l-tuples of elements of E, the hermitian space of rank l. A vector $u \in E^l$ has coordinates $u(\tau) \in \mathbb{C}^l$, one for each residue τ, and its *length* at τ is $|u(\tau)|^2 = \sum_i |u_i(\tau)|^2$.

Lattices and classes. Theta functions of lattices are the source of the CM derivatives, and the classes of a genus make every sum finite. The basic lattice is $\mathcal{O}_E^l$, the vectors with integral entries. A lattice $L \subset E^l$ is in its *genus* if it looks like $\mathcal{O}_E^l$ at every prime: for each prime p there is a length-preserving matrix with p-adic entries carrying L to $\mathcal{O}_E^l$ (a different matrix for each p). Two lattices are in the same *class* if a single length-preserving matrix with entries in E carries one to the other. There are finitely many classes in the genus, $L_1, \dots, L_\nu$. This is the classical theory of hermitian lattices, the unitary counterpart of the genus and class of a quadratic form [9, 10]. In group-theoretic terms the classes are the double cosets $U(E)\backslash U(\mathbb{A}_f)/\mathrm{Stab}(L)$ of the unitary group, their number is finite by Borel's theorem [3], and their weighted count, the *mass*, is given by a mass formula [8]. For quadratic forms, for example, the even unimodular positive definite forms in 16 variables form one genus with two classes, and in 24 variables one genus with 24 classes (Niemeier).

Forms. A vector must be a finite list of algebraic numbers with a positive definite inner product. A *form* (an "algebraic modular form" in the sense of Gross [15]) is a function φ on the classes, $L_i \mapsto \varphi(L_i)$, with values in a fixed finite-dimensional space of polynomials and compatible with the symmetries of each L_i. So a form is just a finite list of vectors of numbers. The classical prototype is the definite quaternion algebra: functions on its classes of ideals, acted on by Brandt matrices, are the modular forms of weight 2 and prime level (Eichler's basis problem).

Theta functions and lifts. This is how CM derivatives become the entries of vectors; the matrix of a lift between definite lattices is a finite array of theta derivatives. For lattices L of rank l and L' of rank l', let $\mathrm{Hom}(L, L')$ be the lattice of $l' \times l$ matrices with entries in E that map L into L'. Its theta function is a Gaussian sum over these matrices, with a matrix variable z_τ for each residue τ. Differentiating at 0 against a polynomial P in the matrix entries gives a number $J(L, L')$: a polynomially weighted count of the matrices mapping L to L', the analogue of the representation numbers of one quadratic form by another. A *theta lift* sends forms on rank-l lattices to forms on rank-l' lattices by the class sum

$$(\mathrm{lift}\,\varphi)(L') = \sum_{i=1}^{\nu} \frac{1}{|\mathrm{Aut}\, L_i|}\, J(L_i, L')\, \varphi(L_i), \tag{8.9}$$

exactly as Brandt matrices act on functions on classes. Since the values are polynomials, $J(L_i, L')$ is a linear map between spaces of polynomials. In coordinates, with bases indexed by α for the values at the source classes L_i and by β for the values at the target classes L'_j, and with $\Gamma_i = \mathrm{Aut}\, L_i$,

$$(\mathrm{lift}\,\varphi)_{j,\beta} = \sum_i \frac{1}{|\Gamma_i|} \sum_\alpha K_{j,\beta;\,i,\alpha}\, \varphi_{i,\alpha}, \tag{8.10}$$

where each entry $K_{j,\beta;\,i,\alpha}$ is a prescribed normalized derivative of a CM theta function, an algebraic number. So a theta lift is a finite block matrix, a chain of lifts is a product of such matrices, and if V is the matrix whose columns are the final vectors, the Gram determinant of Section 8 is $H = \det(V^* B V)$, with B the block-diagonal matrix of the class weights $1/|\Gamma'_j|$ and the metrics on the coefficients. A source form is an arbitrary such list of coefficients; it is the lifted vectors whose entries are sums of theta derivatives. These theta functions factor into

products of the one-variable function θ of Section 6, evaluated at linear forms in the variables, so every entry of J is a polynomial in CM derivatives of the kind in Step 6 [*from the Lean*].

Orders of differentiation. The guaranteed divisibility must reproduce the prime sum with the cutoffs of S_d. The orders are prescribed: order $a = xb$ in the variables z_τ with $\tau \equiv 1 \pmod{Q}$, where $\log b \approx x^{\varepsilon}$; and a weight $\kappa(\sigma)$ at every residue σ, in about k further variables, with $\kappa = b$ at a sparse set Φ_b of residues $\sigma \equiv -1 \pmod{Q}$ and $\kappa = a/x_1$ elsewhere (Section 9). Through the ramps of Step 6 these weights produce the cutoffs x and x_1 (Section 9).

Stages one and two: source forms and target forms. The upper bound (H2) is proved for source forms, whose lengths are computable through definite lattices. The lower bound (H3) needs a family that nearly fills a space of dimension D, and the images of one source space are too few. A single linear map cannot produce more than $\dim S$ independent forms; it is the translates of the image of the Poincaré map under the finite group G that fill the target space. And the Poincaré series itself needs the target space to be indefinite. The form $\mathcal{P}\varphi$ is defined by a *Bessel period*: it is the form whose inner product with any form y on Y_0 is the integral of y, against φ, over a subgroup R of the unitary group of Y_0, the product of a unipotent subgroup (weighted by a generic additive character) and the unitary group of X_0. Such a unipotent subgroup exists only if Y_0 contains *isotropic* vectors, nonzero vectors of length 0, and these are what the hyperbolic planes provide. A space that is positive definite at every residue τ has none, since an isotropic vector would have length 0 at every τ. The indefinite group also carries the analysis of Theorem B: the second moment is computed by unfolding the Poincaré series, with Bessel functions of a discrete series at the infinite places in place of Petersson's J-Bessel kernel. The definite side is recovered twice. Along the chain of lifts the signs change until the last space Y_m is definite (stage three). And the spaces that differ from Y_0 only in their signs, the last of them definite, carry the same Hecke eigenvalues with the same multiplicities among the good forms, which is how the dimension D is compared with a mass in Theorem B [*from the Lean*]. One image of the source space is only a small part of the target space; about $\log b \cdot D/\dim S$ translates fill all but a proportion $4/b^2$ of it (Theorem B). A *source form* is a form on lattices of rank k over E, of order a in the variables z_τ with $\tau \equiv 1 \pmod{Q}$, with a level at a product $M = p_0 p_1$ of two auxiliary primes and at one further prime N. A Poincaré series $\mathcal{P}$ sends source forms to *target forms*: automorphic forms on the unitary group of a larger, indefinite hermitian space of rank r (the source space plus hyperbolic planes and one line of negative length). After a projection, it is followed by translation by an element g of the finite group

$$G = \prod_{i=0,1} \prod_{w \mid p_i} \mathrm{GL}_r(\mathbb{F}_{p_i}), \tag{8.11}$$

where w runs over f chosen prime ideals of E above p_i, each with residue field $\mathbb{F}_{p_i}$. Let D be the dimension of the space of target forms, and S the space of source forms (more precisely, of those satisfying a mild condition at N, the *ramified* source forms of Section 11). The first two stages of (8.8) are $\varphi \mapsto y = g\,\mathcal{P}\varphi$. One chooses n linearly independent target forms $y_1, \dots, y_n$ of this kind, with n in the range (8.1) (Theorem B, Section 13), and puts $v_i = \Theta_1 y_i$.

Stage three: the chain of lifts and the seesaw identity. Each lift contributes the inert-prime surplus once more, so a chain of m lifts multiplies rem by m; the bounds for the accumulated costs, which are quadratic in m, stay within the error budget for $m \asymp \sqrt{x}$ (Section 10). The map Θ_1 in (8.8) is a composite of m theta lifts along a chain of hermitian spaces

$$Y_0 \to Y_1 \to \cdots \to Y_m, \qquad \operatorname{rank} Y_j = r + jt, \quad t = 5r. \tag{8.12}$$

Here Y_0 is the indefinite space of the target forms. Explicitly, Y_j has rank $r + jt$ and exactly $\lfloor (m-j)/2 \rfloor$ negative squares: for $0 \le j < m$, up to a change of basis, its form is a positive

identity matrix, plus $\lfloor (m-1-j)/2 \rfloor$ hyperbolic planes $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, plus one line of length $-N$ if $m-1-j$ is odd. The basis vectors of the hyperbolic planes are the isotropic vectors. The last space Y_m is positive definite [*from the Lean*]. (The arrows are lifts of forms, not inclusions of spaces: the signs of the old coordinates change from one space to the next.) The subscript 1 records that no twist is used. (A *twist* ξ is a character of finite order of the norm-one ideles of E, defined in Section 11. Inserting it in each lift gives a twisted chain Θ_ξ. These twists are used only in the lower bound, Section 14; they are not the Dirichlet characters $\psi\chi_d$ of Sections 3 and 4, and nothing is assumed about the zeros of their L-functions.)

Lifts out of Y_0 cannot be computed by class sums: a form on an indefinite space is a function on an infinite symmetric space, not on finitely many classes. The *seesaw identity* replaces them by class sums. Alongside (8.12) there is a chain of positive definite spaces $X_j = E^{k+j(t+1)}$, $0 \le j \le m$, beginning with the space X_0 of the source forms, and on these every lift is a finite class sum as above. The *composite lift* $\mathcal{L}_1\varphi$ of a source form φ is a finite combination of composites of m lifts $X_0 \to X_1 \to \cdots \to X_m$, restricted to the orthogonal complement in X_m of a vector of length N (a space of rank $k + m(t+1) - 1 = r + mt$, the rank of Y_m), and divided by a power of an element $\bar\pi_N \in \mathcal{O}_E$ with $\pi_N\bar\pi_N = N$. The seesaw identity (11.3) of Section 11 says that $\Theta_1(\mathcal{P}\varphi) = c_{\mathrm{ss}}\,\mathcal{L}_1\varphi$, with a constant $c_{\mathrm{ss}} \neq 0$, for source forms satisfying a mild condition at N. In a diagram, with maps of forms:

$$\begin{array}{ccccccc} X_0 & \longrightarrow & X_1 & \longrightarrow & \cdots & \longrightarrow & X_m \\ \big\downarrow{\scriptstyle \mathcal{P}} & & & & & & \big\downarrow{\scriptstyle \text{restrict, divide by powers of } \bar\pi_N} \\ Y_0 & \longrightarrow & Y_1 & \longrightarrow & \cdots & \longrightarrow & Y_m \end{array} \tag{8.13}$$

The top row is definite, and every arrow in it is a finite class sum. The bottom row starts with the indefinite space of the target forms and ends with a definite one. The seesaw identity says that the square commutes up to the constant c_{ss}.

The name is Kudla's [17]. Let (G_1, G_1') and (G_2, G_2') be dual pairs (pairs of groups, each acting on the other's theta functions through the Weil representation [30]) in the same symplectic group, nested crosswise: $G_2 \subset G_1$ and $G_1' \subset G_2'$. Then for forms f on G_1' and h on G_2,

$$\langle \theta(f)|_{G_2},\, h\rangle_{G_2} = \langle f,\, \theta(h)|_{G_1'}\rangle_{G_1'}, \tag{8.14}$$

because both sides are the integral over $G_2 \times G_1'$ of one theta kernel against $f\,\bar h$, taken in the two orders. Drawn as a cross between the two inclusions, the diagram looks like a seesaw. The proof exchanges the order of integration m times, trading the lift out of the indefinite space Y_0 for lifts between definite lattices.

Finally, the chain commutes with the action of the finite group G [*from the Lean*], and the idempotent e_0 lies in the group algebra of G (Section 11). So for a target form $y = g\mathfrak{p}\varphi$, with $\mathfrak{p}$ as in (11.2),

$$v = \Theta_1 y = c_B\,|G|\; g\,e_0\,\mathcal{L}_1\varphi, \tag{8.15}$$

and the constant c_{ss} cancels. Here c_B is a positive rational number, fixed in Section 12, that divides out the guaranteed divisibility.

The vectors v_i. Each $v_i = \Theta_1 y_i$ is a form on the last space Y_m of the chain of lifts (Section 11), a definite hermitian space of rank

$$R = r + mt = r(1+5m), \qquad 5x^{11/2} \lesssim R \lesssim 30\,x^{11/2}. \tag{8.16}$$

- *Entries.* For each class L of lattices of rank R (at the level of Y_m: the primes above M and the modulus of the twist), the value $v_i(L)$ is a polynomial, that is, a vector of coefficients in the weight representation. So v_i is a list of numbers indexed by pairs (class, coefficient). By

(8.15) each of them is a constant times a finite sum, over classes $L_0, \dots, L_m$ of lattices in $X_0, \dots, X_m$, of products $J(L_0, L_1) \cdots J(L_{m-1}, L_m)$ of theta derivatives with values $\varphi(L_0)$ of a source form. They are algebraic numbers, and after the normalization c_B of Section 12 they are integral up to the index [*from the Lean*].

- *Inner product.* With the Gaussian norm of Section 7 on the polynomial values,

$$\langle v, w \rangle = \sum_L \frac{1}{|\mathrm{Aut}\, L|} \langle v(L), w(L) \rangle. \tag{8.17}$$

Here L runs over the classes of lattices of rank R at the level of Y_m, and $\mathrm{Aut}\, L$ is the group of length-preserving matrices with entries in E that map L to itself and preserve the level structure (finite, since Y_m is definite). How these numbers enter the proof is explained at the end of this section. In the Lean the value $v(L)$ is a vector of coefficients in a fixed integral basis of the weight representation, and its squared length is $\bar{u}\,{}^t C\, u$ for a positive definite integral matrix C, where u is the column of coefficients. This C is not the matrix F of Gaussian pairings of the basis polynomials: the two are related by $CF = h_{\mathrm{hook}} \cdot I$ with a hook number h_{hook}, a duality together with a normalization [*from the Lean*]. The Gaussian pairing explains the factorials in the size estimates; at the finite primes the factorials and hook numbers are accounted for through this normalization.

- *Dimension.* The v_i lie in the space of all forms on Y_m at that level, of dimension $\sum_L \dim V^{\Gamma_L}$, where V is the weight representation and Γ_L the finite group $\mathrm{Aut}\, L$ [15]; this is at most (number of classes) $\times \dim V$. The level has index at most M^{fR^2} [*from the Lean*]. As a rough orientation only, not a proved estimate,

$$\log \dim \;\approx\; f\, R^2 \log M \;\approx\; x^{22+\varepsilon} \tag{8.18}$$

[*heuristic*], more than $\log n$. So the n vectors span a small part of this space. The ambient dimension is never used: the Gram matrix below is $n \times n$, and only the inner products (8.17) enter.

How large D and n are. The parameters satisfy

$$x^5 \le r \le 3x^5, \qquad k \le r \le 2k, \qquad x^{11} \le f \le x^{12}, \qquad b^{20} \le M \le b^{22}, \tag{8.19}$$

and the dimension of the target space is bounded by

$$(1 - 5b^{-4}) \dim S_{\mathrm{full}} \;\le\; D \;\le\; |G| \cdot \deg \cdot \mathrm{mass}(E, r), \qquad |G| \le M^{fr^2}, \tag{8.20}$$

where deg is the dimension of the representation at the infinite places and $\mathrm{mass}(E, r)$ the mass of the genus of $\mathcal{O}_E^r$ [*from the Lean*]. Here S_{full} is the space of all source forms. The lower bound holds because the Poincaré map, followed by the projection e_0, is nearly an isometry in mean square (Section 13); this bounds the dimension of its kernel by $5b^{-4} \dim S_{\mathrm{full}}$, but does not make it injective. These are inequalities, and the proof needs nothing more. As a rough orientation only: if the level factor M^{fr^2} dominates both ends of (8.20), as the mass formula suggests, then

$$\log n \;\asymp\; \log D \;\asymp\; f\, r^2 \log M \;\asymp\; x^{21+\varepsilon} \tag{8.21}$$

up to factors of $\log x$ [*heuristic; not proved here or in the Lean*]. The prime d enters only through the discriminant of E and the mass, which are of lower order. Neither D nor n is computed in the proof: only the ratio n/D in (8.1) is used, and D cancels in Section 9, as h does.

How the inner products are used. The proof never evaluates an entry $\langle v_i, v_j\rangle$ of the Gram matrix. It bounds H from above through the lengths $\|v_i\|$ alone, and from below as a whole.

What (8.17) *says.* Each value $v(L)$ is a vector of coefficients, paired by the integral matrix C described after (8.17). For comparison, and to see where the factorials come from, here is the raw pairing of polynomials in variables $z = (z_1, \dots, z_s)$: the Gaussian inner product of Section 7, with one Gaussian factor per variable, also called the *Fischer inner product.* The monomials $z^\alpha = z_1^{\alpha_1} \cdots z_s^{\alpha_s}$ are orthogonal, with $\|z^\alpha\|^2 = \alpha! = \alpha_1! \cdots \alpha_s!$, so for $P = \sum_\alpha p_\alpha z^\alpha$ and $Q = \sum_\alpha q_\alpha z^\alpha$

$$\langle P, Q\rangle = \sum_\alpha \bar{p}_\alpha \, q_\alpha \, \alpha! = (\bar{P}(\partial)Q)(0), \tag{8.22}$$

conjugate-linear in the first variable, as all inner products here. Here $\bar{P}(\partial)$ is P with conjugated coefficients and each z_i replaced by $\partial/\partial z_i$. The second equality holds because $\partial^\beta z^\alpha$ at 0 is $\alpha!$ if $\beta = \alpha$ and 0 otherwise. So in principle $\langle v_i, v_j\rangle$ is a finite computation: list the classes L, compute each $v_i(L)$ from (8.15) as a class sum of theta derivatives, and pair the coefficient vectors using C. In practice it is out of reach. The number of classes is astronomically large (compare (8.18)), and a single value $v_i(L)$ is a sum over the classes of all the spaces $X_0, \dots, X_m$.

The upper bound uses only the diagonal. Hadamard's inequality needs only the squared lengths $\|v_i\|^2 = \langle v_i, v_i\rangle$, and only upper bounds for them. Theorem A (Section 12) supplies them [*from the Lean*]:

- By (8.15), v_i is a constant times a composite lift: a finite combination of products of the theta derivatives J, that is, of derivatives at 0 of products of the CM theta function θ of Section 6.
- Each such derivative is bounded through the Gaussian norm of Section 7. This is where $\|z^j\|^2 = j!$ and the period Ω enter, through the size $s_\infty(j)$ of (12.1). The number of terms in the class sums contributes only a lower-order error.
- *Every conjugate.* The entries of v_i, and hence H, are algebraic numbers. The product formula involves $|\sigma(H)|$ for every conjugate, so bounds are needed for every automorphism σ of $\mathbb{C}$. For a vector v write σv for the vector with σ applied to every entry, and put

$$\|v\|_\sigma^2 = \langle \sigma v, \sigma v\rangle, \qquad H_\sigma = \det\big(\langle \sigma v_i, \sigma v_j\rangle\big)_{1\le i,j\le n} \ge 0, \tag{8.23}$$

  the Gram determinant of the conjugated vectors. Theorem A bounds $\|v_i\|_\sigma$ for every σ, by the same estimate; this works because the Taylor series of the normalized CM theta functions, and of their torsion translates, are carried by an automorphism fixing K to the corresponding series on conjugate CM lattices. (An automorphism that moves $\sqrt{-d}$ is first composed with complex conjugation.) The linear forms and the finite translation data are carried along, up to a controlled scalar; and the Chowla–Selberg normalization fixes the absolute value of the scale of the derivatives (the remark after (7.6)). So the same Gaussian estimate applies at every conjugate [*from the Lean; compare Shimura* [31]*, although the Lean uses uniformization rather than a reciprocity law*].

A conjugate $\sigma(H)$ is in general *not* the determinant H_σ, and it need not be positive. (For the vector $v = \sqrt{1+\sqrt{2}}$ in dimension one, $H = 1+\sqrt{2}$, and the conjugate $1-\sqrt{2}$ is negative, while $|\sigma v|^2 = \sqrt{2}-1$.) The inner product has rational coefficients in the coordinates used, so

$$\sigma\langle v, w\rangle = \langle \sigma^c v, \sigma w\rangle, \qquad \sigma^c(z) = \overline{\sigma(\bar{z})}, \tag{8.24}$$

a pairing of two different conjugates. (If all entries lie in a field in which every embedding commutes with complex conjugation, as in a CM field, then $\sigma^c = \sigma$ and $\sigma(H) = H_\sigma$; the Lean's

argument does not need this.) The Cauchy–Schwarz inequality for determinants then gives

$$|\sigma(H)|^2 \;\le\; H_{\sigma^c}\, H_\sigma. \tag{8.25}$$

So it is enough to bound H_σ for every σ, and Hadamard's inequality does this: $\log H_\sigma \le \sum_i \log \|v_i\|_\sigma^2$. This is the covolume bound of Section 9, and with (8.25) it gives $\log|\sigma(H)| \le c$ at every conjugate [*from the Lean*].

The lower bound uses the determinant as a whole. Theorem C (Section 14) bounds H from below by the product formula. Its hypotheses, covolume and index (Section 9), are upper bounds: on the determinants H_σ, hence on $\log|\sigma(H)|$ at each conjugate, and on the denominators of the lifts. The individual entries never appear; the proof works with ratios of determinants [*from the Lean*].

- *Completing to a basis, and twisted determinants.* The ratios below concern the lift as a map on the whole target space, not the n chosen vectors. So the family $y_1, \dots, y_n$ is first completed to a basis $w_1, \dots, w_D$ of the target space by a few auxiliary forms, at a small cost (Section 14). For a twist ξ let

$$Z(\xi) = \det(\Theta_\xi^* \Theta_\xi), \tag{8.26}$$

  the determinant of the twisted chain on the whole D-dimensional target space. For a basis w, the Gram determinant of the lifts $\Theta_\xi w_i$ is $Z(\xi)$ times the Gram determinant of w, so in a ratio $Z(1)/Z(\xi)$ the choice of basis cancels. (For the n vectors alone this fails: on a plane with $\Theta_1 = \mathrm{diag}(2,1)$, $\Theta_\xi = I$ and $y_1 = e_2$, the ratio for the one vector is 1 and the ratio for the whole map is 4.) For each prime ℓ and each embedding ι into an algebraic closure of $\mathbb{Q}_\ell$ one takes a twist ξ_ι whose conductor is prime to ℓ; it may depend on ι. The index hypothesis holds uniformly in such twists, so the denominators at ι of the twisted Gram determinant are bounded by the index, and what must be controlled at ι is the ratio $Z(1)/Z(\xi_\iota)$.
- *Exchange identities.* The ratio for the chain of m lifts is compared with ratios $Z_j(1)/Z_j(\xi^j)$, $1 \le j \le m$, for auxiliary one-step lifts $W_j \to V_j$ between spaces of ranks r and $r+t$. These are comparison lifts, not the successive maps $Y_{j-1} \to Y_j$ of the chain; the spaces W_j differ from Y_0 only in their signs and carry the same Hecke eigenvalues with the same multiplicities, which is what gives the same dimension D on both sides. (The maps in the chain are between spaces of different dimensions, so this is not the multiplicativity of determinants.) The tool is an *exchange identity of Siegel–Weil type*. The Siegel–Weil formula [20, 21] grew out of Siegel's mass formula for representation numbers. It says that the integral of a theta function over a unitary (or orthogonal) group is an Eisenstein series; equivalently, so is the weighted average $\sum_L |\mathrm{Aut}\, L|^{-1}\theta_L$ of theta functions over the classes of a genus. The exchange identity compares two such theta integrals over unitary groups of complementary rank. They agree up to a factor c for which $|c|^2$ is a constant times a product of powers of cardinalities of residue fields and of squares of absolute values of partial L-values (L-functions with finitely many Euler factors removed). Applied m^2 times, one square of a grid of spaces at a time, it moves the twist from the whole chain to the one-step lifts (Section 14).
- *Single lifts and L-values.* For a single lift, squared norms are evaluated exactly. A *Hecke operator* at a prime p (in the proof: at prime ideals of E of degree one over a prime that splits, away from M) is the class-sum operator $\varphi \mapsto (L \mapsto \sum_{L'} \varphi(L'))$, the sum over the lattices L' that agree with L away from p and are in a fixed relative position to L at p; it is the analogue of a Brandt matrix. A *Hecke eigenform* is a common eigenvector of these operators. The *Rallis inner product formula* ([18, 28]; for unitary groups in general, [29, 16])

computes the squared norm of the theta lift of an eigenform φ. By the Siegel–Weil formula, $\|\Theta\varphi\|^2$ is an integral of $|\varphi|^2$ against an Eisenstein series. The doubling method turns this integral into a special value of an L-function of φ, times finitely many local factors, times $\|\varphi\|^2$. In the proof the single-lift ratios reduce, through an explicit operator comparing twisted and untwisted lifts, to ratios of values of *Hecke L-functions* $L(z,\chi)=\sum_{\mathfrak{a}}\chi(\mathfrak{a})\,\mathrm{N}\mathfrak{a}^{-z}$ of E (sums over the ideals $\mathfrak{a}$ of $\mathcal{O}_E$, for Hecke characters χ) at the integers $1\le z\le t$ (the series converges in a right half-plane, and the values are those of its analytic continuation where needed). Each single lift returns its share of the main amount $h=h_{\mathrm{main}}$ of (12.2), the cost of its idle directions (the about k further variables differentiated to order $\kappa(\sigma)$). Summed over the lifts and the D dimensions, this is the term Dh of Theorem (C).

So the only exact evaluations of inner products in the proof are squared norms of single lifts of eigenforms, expressed through L-values. The Gram entries of $v_1,\dots,v_n$ are never computed. In summary, the diagonal entries are bounded above at every conjugate (Theorem A, through the Gaussian norm), H is bounded below as a whole (Theorem C, through the product formula), and exact evaluation happens only for single eigenform lifts.

9. One main term, bounded from both sides

The main inequality. The construction attached to a large integer x uses the following data (Section 11 lists all of them) [*from the Lean*]. Fix $0<\varepsilon<\frac{1}{100}$.

- $x_1=\lfloor\exp(x^{\varepsilon/6})\rfloor$; an integer b with $x^{\varepsilon}\le\log b\le x^{\varepsilon}+1$; $a=xb$; and an integer m with $\sqrt{x}\le m\le 2\sqrt{x}$.
- The $f=\varphi(q)$ residues $\tau\in(\mathbb{Z}/q)^{\times}$; among them $\Phi_a=\{\tau\equiv 1 \pmod Q\}$, with $A_0=f/\varphi(Q)$ elements, and a sparse set $\Phi_b\subset\{\tau\equiv -1 \pmod Q\}$ with B_0 elements, where $A_0\le B_0\log b$.
- The *weight* $\kappa(\tau)=a/x=b$ for $\tau\in\Phi_b$, and $\kappa(\tau)=a/x_1$ otherwise.

Let $\lambda(n)=\Lambda(n)(1-\chi_d(n))$, the weight of S_d without the congruence condition. For a residue τ and a real $y>0$ put

$$s_{\mathrm{fin}}(y,\tau)=\sum_{\substack{n\ge 1\\(n,q)=1}}\lambda(n)\Big(\frac{y}{n}-\kappa(-n\tau)\Big)_+,\qquad s_{\mathrm{main}}=a\Big(\log a-\frac{1}{f}\sum_{\sigma}\log\kappa(\sigma)+c_1\Big),\tag{9.1}$$

where $-n\tau$ is the product in $\mathbb{Z}/q$, the sum over σ runs over all residues, and

$$c_1=\frac{L'}{L}(1,\chi_d)-\gamma-1\tag{9.2}$$

is the constant (2.4) for $Q=1$. The condition $(n,q)=1$ makes $\kappa(-n\tau)$ defined, and loses nothing: $\lambda(n)=0$ for the other n, since every prime dividing q splits in K by (6.9); and for $y\le a$ the sum is finite, since the terms with $n>\max(x,x_1)$ vanish. So $s_{\mathrm{fin}}(y,\tau)$ is the sum of the ramps (5.4): the divisibility guaranteed at τ (Section 6). The quantity s_{main} is what the size allows on average (Section 12). The *remainder* is the surplus

$$\mathrm{rem}=\sum_{\tau\in\Phi_a}\big(s_{\mathrm{fin}}(a,\tau)-s_{\mathrm{main}}\big),\tag{9.3}$$

and the *main inequality* (M), derived at the end of this section from Theorems (A), (B) and (C), says that $m\cdot\mathrm{rem}\le fa(\log b)^5$.

Two scales in one identity. Let $R_1(y)=\sum_{n\le y}\lambda(n)(\frac{1}{n}-\frac{1}{y})-\log y-c_1$ be the remainder of the prime sum without the congruence condition, the analogue of R for $Q=1$.

Lemma 9.1 ([*proved*])**.** *For an integer $A \ge 1$ and real $0 < z \le A$,*

$$\sum_{n\equiv 1\,(Q)} \lambda(n)\Big(\frac{A}{n} - \frac{A}{z}\Big)_+ = A\Big(R(z) + \frac{\log z}{\varphi(Q)} + c_d\Big), \tag{9.4}$$

$$\sum_{n\ge 1} \lambda(n)\Big(\frac{A}{n} - \frac{A}{z}\Big)_+ = A(R_1(z) + \log z + c_1). \tag{9.5}$$

Proof. The term is $A(\frac{1}{n} - \frac{1}{z})$ for $n \le z$ and 0 for $n > z$. Compare with (2.3). □

Proposition 9.2 (Two-scale identity; [*proved*])**.**

$$\mathrm{rem} = a\Big[B_0(R(x) - R(x_1)) + A_0 R_1(x_1)\Big]. \tag{9.6}$$

Proof. Fix n prime to q. The map $\tau \mapsto -n\tau$ is a bijection of Φ_a onto the residues $\equiv -n \pmod Q$. This set contains Φ_b if $n \equiv 1 \pmod Q$ and is disjoint from it otherwise. So $-n\tau \in \Phi_b$ for exactly B_0 of the $\tau \in \Phi_a$ in the first case and for none in the second. Since κ is a/x on Φ_b and a/x_1 elsewhere,

$$\sum_{\tau\in\Phi_a} \Big(\frac{a}{n} - \kappa(-n\tau)\Big)_+ = A_0\Big(\frac{a}{n} - \frac{a}{x_1}\Big)_+ + \mathbf{1}_{n\equiv 1\,(Q)}\, B_0\Big[\Big(\frac{a}{n} - \frac{a}{x}\Big)_+ - \Big(\frac{a}{n} - \frac{a}{x_1}\Big)_+\Big]. \tag{9.7}$$

Multiply by $\lambda(n)$ and sum over n. Lemma 9.1, with $A = a$ and $z = x_1$ or x, gives

$$\sum_{\tau\in\Phi_a} s_{\mathrm{fin}}(a,\tau) = a\Big[A_0(R_1(x_1) + \log x_1 + c_1) + B_0\Big(R(x) - R(x_1) + \frac{\log(x/x_1)}{\varphi(Q)}\Big)\Big]. \tag{9.8}$$

For the main term, $\sum_\sigma \log\kappa(\sigma) = f\log(a/x_1) - B_0\log(x/x_1)$, so with $f = A_0\varphi(Q)$

$$A_0\, s_{\mathrm{main}} = a\Big[A_0(\log x_1 + c_1) + B_0\frac{\log(x/x_1)}{\varphi(Q)}\Big]. \tag{9.9}$$

Subtract. □

The identity explains the design. All residues of Φ_a see the primes up to the large cutoff x_1, with the full weight λ. The B_0 special residues lower the cutoff to x, and only for $n \equiv 1 \pmod Q$. So what is measured at the scale x is exactly $R(x)$; the terms at x_1 are removed in Section 10. The constant c_d appears only in the difference $R(x) - R(x_1)$, where it cancels. So the terms of c_d from the characters $\psi \ne \psi_0$ never need to be matched; only c_1, which enters s_{main}, must be matched by the size, and it is, through the period (Sections 7 and 12).

Why a single number fails. Suppose some nonzero rational integer Δ had forced divisibility and size

$$\sum_{p\ \mathrm{inert}} v_p(\Delta)\log p \ \ge\ aS_d(x) - (\mathrm{small}), \qquad \log|\Delta| \ \le\ a\Big(\frac{\log x}{\varphi(Q)} + c_d\Big) + ax^{-1/2+\varepsilon}. \tag{9.10}$$

Then (2.9) would follow at once. But these force Δ to consist almost entirely of its inert-prime factors: everything else must have logarithm at most a proportion $x^{-1/2}$ of the whole. No number we know comes close. For the lemniscate numbers the inert primes carry about half of $\log|A_a|$ (51.7% for $a = 2001$) [*computed*], and Chebyshev-type products, Hankel determinants and ratios of these numbers stay near one half or fail to be integers [*computed*].

The proof's way out. The proof does not make the cofactor small. It compares lengths with a common *main term* $h = h_{\text{main}}$, the explicit quantity (12.2): the size of the idle directions of a vector less their guaranteed divisibility. It is enormous, and its numerical value is never needed. In the lemniscate examples it corresponds to the logarithm of the cofactor, the part of a number not made of inert primes; this is an analogy only. The argument in one paragraph: extra divisibility by the inert primes makes a lifted vector *shorter* than this baseline, by an amount proportional to the positive error in the prime sum. Theorems A and B supply enough independent target forms, with short lifts at every conjugate, to occupy almost all of the D dimensions of the target space. Theorem C bounds the volume of such a family from below relative to the same baseline, allowing for denominators. Comparing the two bounds eliminates h and bounds the positive error of the prime sum, which is the direction Landau's argument needs. A family $y_1, \dots, y_n$ of target forms is measured in two ways.

- *Covolume.* It has covolume at most c if $\log H_\sigma \le c$ for every automorphism σ of $\mathbb{C}$, where H_σ is the Gram determinant (8.23) of the conjugated untwisted lifts σv_i, $v_i = \Theta_1 y_i$. Then $\log|\sigma(H)| \le c$ at every conjugate of H, by (8.25).
- *Index.* It has index at most $\delta = \sum_\ell \delta_\ell \log \ell$, with $\delta_\ell \ge 0$, if for every prime ℓ and every embedding ι of the field of the entries into an algebraic closure of $\mathbb{Q}_\ell$ (equivalently, every prime ideal above ℓ) there is a matrix U_ι with ι-integral entries, with determinant of valuation at most δ_ℓ, such that the lifts Θ_ξ of the family transformed by U_ι are integral at ι, simultaneously for all m-fold twists ξ of conductor prime to ℓ (Section 11). The matrix may depend on ι; the bound δ_ℓ does not. The factor 2 in Theorem C below has an elementary source: a change of the vectors by U changes a Gram matrix to U^*HU, with two determinant factors.

Three theorems are proved in the Lean development [*from the Lean*]; their precise forms and proofs are in Sections 12–14:

(A) *Short forms.* The integral source forms whose lifts, after every translation $g \in G$ and at every conjugate, have squared length at most

$$e^{\,h - m\cdot\mathrm{rem} + fa(\log b)^4} \tag{9.11}$$

span the source space S. Moreover the target forms obtained from integral source forms by a set T of translations have index at most $|T| \cdot \dim S \cdot 2fa(\log b)^3$.

(B) *Enough forms.* Finitely many translates of their images span at least $(1 - 4/b^2)D$ dimensions of the D-dimensional target space.

(C) *Integral forms are long.* If $n \ge (1 - 4/b^2)D$ independent target forms have index at most δ and covolume at most c, then

$$D\big(h - 3fa(\log b)^4\big) \le c + 2\delta. \tag{9.12}$$

The cancellation. Write $\mathcal{E} = fa(\log b)^4$. By (B) the images of S under a set T of translations span at least $(1 - 4/b^2)D$ dimensions, with $|T| \cdot \dim S \le 5D\log(1 + 2b^2)$ (Theorem 13.1). By (A) the short integral forms span S, so their images span the same space, and one can choose $n \ge (1-4/b^2)D$ independent target forms among the images of short integral forms. Their index is at most $\delta \le |T| \cdot \dim S \cdot 2fa(\log b)^3 \le 10D\log(1 + 2b^2)\, fa(\log b)^3$. Hadamard's inequality gives the covolume bound

$$c = n\big(h - m\cdot\mathrm{rem} + \mathcal{E}\big), \tag{9.13}$$

and (C) gives the lower bound

$$D(h - 3\mathcal{E}) \le c + 2\delta, \qquad \text{where } \Big(1 - \frac{4}{b^2}\Big)D \le n \le D, \quad \delta \le 10D\log(1 + 2b^2) fa(\log b)^3. \tag{9.14}$$

Combining,

$$n \cdot m \cdot \mathrm{rem} \;\le\; (n-D)\,h + (3D+n)\,\mathcal{E} + 2\delta. \tag{9.15}$$

Since $n \le D$ and $h \ge 0$, the term $(n-D)h$ is at most 0 and can be dropped: this is the cancellation of the main term h, whose value was never needed. Dividing by $n \ge D/2$,

$$m \cdot \mathrm{rem} \;\le\; 7\,\mathcal{E} + \frac{4\delta}{D} \;\le\; 7fa(\log b)^4 + 40\log(1+2b^2)\,fa(\log b)^3, \tag{9.16}$$

and since $\log(1+2b^2) \le 3\log b$ for large b, the right side is at most $f\,a\,(\log b)^5$:

$$m \cdot \mathrm{rem} \;\le\; f\,a\,(\log b)^5 \tag{M}$$

[*(A), (B), (C) from the Lean; the deduction proved here*]. The hypothesis $n \ge (1-4/b^2)D$ is needed for Theorem C, since only a family that nearly fills the space can have a lower bound of size Dh; the deduction itself uses only $n \ge D/2$.

Theorem C is the analogue of "a nonzero integer is at least 1" and is the hard part; its proof is outlined in Section 14. The product formula by itself gives only a comparison of size with denominators. It does not produce the term Dh: that is the arithmetic content of Theorem C, and it is why the determinant cannot simply be replaced by some other large integer.

10. Iteration and the square root

The construction applies a chain of m theta lifts,

$$X_0 \xrightarrow{T_1} X_1 \xrightarrow{T_2} \cdots \xrightarrow{T_m} X_m, \tag{10.1}$$

and by the seesaw identity (Section 11) the vectors v_i are, up to a constant, composite lifts of source forms. The lifts are applied to each vector before the determinant is formed, so there is still one Gram determinant. Each lift contributes the inert-prime divisibility once more, which is why the left side of (M) is $m \cdot \mathrm{rem}$. The right side $fa(\log b)^5$ is the total allowance after the costs of all m lifts are included. Three of these costs have bounds quadratic in m: in (C), the comparison operators, $O(fm^2b(\log b)^2)$, and the denominators, $O(fm^2b\log b)$; and in (A), the loss at the prime N, where $\log N \asymp m\log b$, which is $O(mfb\log N) = O(fm^2b\log b)$. All fit their allowances, which are fxb times powers of logarithms, for $m \asymp \sqrt{x}$ [*from the Lean*]. (Raising one number to the m-th power would not help: it multiplies the error by m as well.)

From (M) to the bound for R.

Lemma 10.1 ([*proved, given (M)*])**.** *For every sufficiently large integer x, with the data b, m, x_1 of Section 9,*

$$R(x) \le \frac{\varphi(Q)(\log b)^6}{m} + |R(x_1)| + (\log b)\,|R_1(x_1)|. \tag{10.2}$$

Proof. By (M) and (9.6),

$$B_0\big(R(x) - R(x_1)\big) + A_0 R_1(x_1) \;\le\; \frac{f(\log b)^5}{m} = \frac{A_0\,\varphi(Q)(\log b)^5}{m}. \tag{10.3}$$

Divide by B_0 and use $A_0/B_0 \le \log b$. □

Lemma 10.2 (Prime number theorem for a fixed modulus; [6, Ch. 20])**.** *For fixed Q and d and every $B > 0$, $|R(y)| \le (\log y)^{-B}$ and $|R_1(y)| \le (\log y)^{-B}$ for all sufficiently large y.*

This is Proposition 2.2 with de la Vallée Poussin's error term in place of $o(1)$: for each character ξ modulo Q or Qd, $\psi(t,\xi) = \sum_{n\le t} \xi(n)\Lambda(n)$ is $O(t\exp(-c\sqrt{\log t}))$ for ξ non-principal, and since $\frac{1}{n} - \frac{1}{y} = \int_n^y t^{-2}\,dt$, $\Sigma_\xi(y) = \int_1^y \psi(t,\xi)t^{-2}\,dt = -\frac{L'}{L}(1,\xi) - \int_y^\infty \psi(t,\xi)t^{-2}\,dt$. Since the modulus is fixed, a possible exceptional zero changes only the constants and the threshold. No estimate of Riemann-hypothesis quality is used: a bound $(\log x_1)^{-B}$ at $x_1 \approx \exp(x^{\varepsilon/6})$ is already a power of x.

Theorem 10.3 (One-sided bound; [*proved, given (M)*])**.** *For every prime $d \equiv Q/2 - 1 \pmod Q$ and every $\eta > 0$, $R(y) \le y^{-1/2+\eta}$ for all sufficiently large real y. This is (2.9).*

Proof. Fix $\varepsilon = \min(\eta/7, \frac{1}{200})$, so that (M) applies. Let x be a large integer. Since $\log x_1 \ge \frac{1}{2}x^{\varepsilon/6}$, Lemma 10.2 with $B \ge 24/\varepsilon$ gives $|R(x_1)| \le x^{-2}$ and $|R_1(x_1)| \le x^{-2}$. With $m \ge \sqrt{x}$ and $\log b \le 2x^\varepsilon$, Lemma 10.1 gives

$$R(x) \le 64\,\varphi(Q)\,x^{-1/2+6\varepsilon} + x^{-2} + 2x^{\varepsilon-2} \le \big(64\varphi(Q)+3\big)\,x^{-1/2+6\varepsilon}. \tag{10.4}$$

For real $y \ge 2$ and $n_0 = \lfloor y \rfloor$, the sum S_d increases from n_0 to y by

$$\sum_{n\le n_0} \lambda(n)\mathbf{1}_{n\equiv 1\,(Q)}\Big(\frac{1}{n_0} - \frac{1}{y}\Big) \le \frac{1}{y\,n_0}\sum_{n\le n_0} 2\Lambda(n) \le \frac{12}{y}, \tag{10.5}$$

by Chebyshev's bound $\sum_{n\le u}\Lambda(n) \le 6u$ [38, Ch. 2], and the term $-\log y/\varphi(Q)$ does not increase. So $R(y) \le R(n_0) + 12/y$. Since $n_0 \ge y/2$ and the exponent is negative, $n_0^{-1/2+6\varepsilon} \le 2\,y^{-1/2+6\varepsilon}$. Hence $R(y) \le (128\varphi(Q)+18)\,y^{-1/2+6\varepsilon} \le y^{-1/2+\eta}$ for large y, since $7\varepsilon \le \eta$. □

With Theorems 3.2 and 4.1, this is the Riemann hypothesis. The exponent $\frac{1}{2}$ is an optimization at the very end: it is where the linear gain in m meets the quadratic cost.

11. The seesaw identity

Sections 9 and 10 derive the Riemann hypothesis from Theorems (A), (B) and (C). Sections 11–14 outline how these three theorems are proved. This is where the length of the Lean development lies: we describe each step and name its inputs, but we do not prove them. Every statement in Sections 11–14 is [*from the Lean*] unless it is marked otherwise.

The data at scale x. Fix $0 < \varepsilon < \frac{1}{100}$. To every sufficiently large integer x the proof attaches the following numbers; that they exist is part of the Lean development.

	role	size
x_1	second cutoff of the prime sum	$\lfloor \exp(x^{\varepsilon/6}) \rfloor$
b	the weight at the residues of Φ_b; a multiple of $q\,x_1$	$x^\varepsilon \le \log b \le x^\varepsilon + 1$
a	order of differentiation at the residues of Φ_a	xb
m	number of theta lifts; a power of 2	$\sqrt{x} \le m \le 2\sqrt{x}$
r	rank of the target space; a power of 3	$x^5 \le r \le 3x^5$
k	rank of the source lattices	$k = r + 1 - m$
t	each lift raises the rank by $t+1$ (definite side) or t (indefinite side)	$t = 5r$
q	$Q \cdot 3l_0 \cdot 2^s$, with $2^s = m^{22}$ and l_0 a power of 3	$27s < l_0 \le 81s$
f	$\varphi(q)$, the number of residues τ	$x^{11} \le f \le x^{12}$
M	p_0p_1, two primes that split completely in E	$b^{20} \le M \le b^{22}$
N	a prime, $N = \pi_N\bar{\pi}_N$ with $\pi_N \in \mathcal{O}_E$	$m\log b \le \log N \le 30m\log b$

The scales are far apart:

$$(\log b)^{100} \le x, \qquad x^{10^6} \le x_1, \qquad x_1^{10^6} \le b, \tag{11.1}$$

so in increasing order they are $\log b$, x, x_1, b, M, N. The set Φ_b of Section 9 consists of residues $-g^e$, with $g = 1 + 3l_0Q$ and e in a set of Behrend's type [2, 27]. It is chosen so that no residue of Φ_b has two partners $\pm\ell^i\sigma$, $\pm\ell^j\sigma$ ($0 < i < j < l_0$, $\ell \le kt$ prime) that also lie in Φ_b (modulo the part of q prime to ℓ).

Spaces, lifts and twists. All spaces are hermitian spaces over E, written in coordinates.

- *The definite chain.* $X_j = E^{k_j}$ with $k_j = k + j(t+1)$, $0 \le j \le m$, with the form $\sum_i \bar{z}_i w_i$ and the lattice $\mathcal{O}_E^{k_j}$. Source forms live on X_0, and a theta lift $X_{j-1} \to X_j$ is the class-sum matrix of Section 8.
- *The indefinite chain.* Y_0 is X_0 plus hyperbolic planes plus one line of length $-N$, of rank $r = k + m - 1$, and Y_j has rank $r + jt$. Target forms are automorphic forms on the unitary group of Y_0, of a fixed discrete-series type at infinity and with level only at the primes above M.
- *Twists.* A *twist* ξ is a character of finite order of the group of ideles of E of norm one, trivial on the elements of E of norm one and at infinity. Its local data are tame characters at a finite set $\mathcal{T}$ of prime ideals of E, prime to MN, with $\mathcal{T} \cap \bar{\mathcal{T}} = \emptyset$: it contains at most one prime of each pair $w \ne \bar{w}$ of conjugate primes, and no prime equal to its conjugate. (At a prime of the maximal real subfield that splits as $w\bar{w}$, an element u of norm one has $u_{\bar{w}} = \bar{u}_w^{-1}$, so its component at w determines it; this is why one prime of each pair suffices.) The associated Hecke character of E is $x \mapsto \xi(x/\bar{x})$, an *anticyclotomic* character; its conductor is supported on the primes of $\mathcal{T}$ and their conjugates. Each twist gives a chain of lifts $Y_0 \to Y_1 \to \cdots \to Y_m$, written Θ_ξ; Θ_1 is the untwisted chain. A twist is *m-fold* if the local characters of $\xi, \xi^2, \dots, \xi^m$ are nontrivial at every prime of $\mathcal{T}$.

The vectors of Section 8 are the untwisted lifts $v_i = \Theta_1 y_i$ of the target forms y_i.

The Poincaré series and the finite group. A Poincaré series $\mathcal{P}$ sends source forms to automorphic forms on Y_0. The finite group G of (8.11) acts on these through the level at M. A selected family of irreducible characters of each factor $\mathrm{GL}_r(\mathbb{F}_p)$ is indexed by *labels*: monic squarefree polynomials of degree r over $\mathbb{F}_p$ with nonzero constant term [14]. (Not every irreducible character arises this way: $\mathrm{GL}_2(\mathbb{F}_2)$ has three, and only one such polynomial, X^2+X+1.) A family of labels is *good* if no label has a root x with $x^{q_2} = 1$, where $q_2 = (Qd)^{s+6}$ (so in particular 1 is not a root), and if above each of p_0, p_1 at least one label is primitive (irreducible, with a root generating the multiplicative group of the field of degree r). The target forms are the image of the central idempotent e_0 of the good families. Put

$$\mathfrak{p} = c_B\, c_{\mathrm{ss}}^{-1}\, |G|\, e_0 \mathcal{P}, \tag{11.2}$$

where $|G|$ clears the denominators of e_0, c_{ss} is the constant of the seesaw identity below, and c_B is a positive rational number fixed in Section 12. For $g \in G$ the translate $g\mathfrak{p}$ maps source forms to target forms; these are the maps $\varphi \mapsto y = g\,\mathcal{P}\varphi$ of (8.8), up to the constants.

The identity. The *composite lift* $\mathcal{L}_\xi\varphi$ of a source form φ, for a twist ξ, is a finite combination of chains of m lifts $X_0 \to X_1 \to \cdots \to X_m$ built from ξ, restricted to the orthogonal complement in X_m of a vector of length N, and divided by a product of powers of the conjugates of $\bar{\pi}_N$. Since $|\bar{\pi}_N| = N^{1/2}$ at every conjugate, the division lowers the squared length by about N^a at each residue of Φ_a, and homogeneity in the last coordinate fixes the power of $\bar{\pi}_N$ and the saving in size. That the division costs little at the primes above N is a separate local estimate, part of

Theorem A, with an allowed loss of $2mfb$ [*from the Lean*]. A source form is *ramified* at N if its mean under the scalings at each prime dividing π_N is 0; this is a condition of negligible relative codimension, and the small loss is charged in Theorem B. The *seesaw identity* says: for every ramified source form φ and every twist ξ,

$$\Theta_\xi(\mathcal{P}\varphi) = c_{\mathrm{ss}} \cdot \mathcal{L}_\xi \varphi, \tag{11.3}$$

with a constant $c_{\mathrm{ss}} \neq 0$ that depends neither on ξ nor on φ. (It does not assert that the right side is nonzero.)

This is what makes the construction computable. The left side lives on an indefinite unitary group, where there are no finite class sums. The right side is a finite matrix of derivatives of CM theta functions on definite lattices, whose size (Section 7) and divisibility (Section 6) are accessible. The identity is an iterated seesaw identity, in Kudla's sense.

12. Short forms: Theorem A

The statement. Charge a derivative of order j the size

$$s_\infty(j) = \log j! + j \log \frac{M}{\Omega}, \tag{12.1}$$

with Ω the period (7.11) of Section 7. Besides the order a at the residues of Φ_a, each lift differentiates to the order $\kappa(\sigma)$ at every residue σ in about k further *idle directions*. The *main amount* is what these cost, size less guaranteed divisibility:

$$h_{\mathrm{main}} = m\,k \sum_{\sigma \in (\mathbb{Z}/q)^\times} \Big(s_\infty\big(\kappa(\sigma)\big) - s_{\mathrm{fin}}\big(\kappa(\sigma), \sigma\big) \Big) \;\geq\; 0. \tag{12.2}$$

It is enormous and is never computed; this is the h of Section 9. Let S be the space of ramified source forms and $S_{\mathbb{Z}}$ those with integral values.

Theorem 12.1 (A: short forms; [*from the Lean*])**.** (a) *The $\varphi \in S_{\mathbb{Z}}$ with*

$$\big\|\Theta_1(g\mathfrak{p}\varphi)\big\|_\sigma^2 \;\leq\; \exp\big(h_{\mathrm{main}} - m \cdot \mathrm{rem} + fa(\log b)^4\big) \qquad \textit{for all } g \in G \textit{ and all } \sigma \tag{12.3}$$

span S. (b) *For a finite set $T \subset G$, every family of target forms taken from $\bigcup_{g\in T} g\mathfrak{p}(S_{\mathbb{Z}})$ has index at most $|T| \cdot \dim S \cdot 2fa(\log b)^3$.*

Where $h_{\mathrm{main}} - m \cdot \mathrm{rem}$ comes from. Write $\mathcal{E}_0 = fa(\log b)^3$, so that $\mathcal{E} = \mathcal{E}_0 \log b$. A bound of this kind cannot hold for every integral source form: a large integer multiple of a form has a large lift. So the first step is a choice.

- *Short source forms.* There are integral source forms that span S and whose values have squared length at most $e^{\mathcal{E}_0}$ at every conjugate. (Before the condition at N is imposed, forms supported at single classes, with short bases of the polynomial values, do this; imposing the condition multiplies the squared lengths by at most 4^f.) The estimates below are applied to these forms, and this is what gives the spanning statement in (a).

By (11.3) the lift is a composite lift, a combination of theta derivatives. Three facts about it are proved.

(1) *Size.* At every conjugate, the logarithm of its squared length is at most

$$m\,k \sum_\sigma s_\infty\big(\kappa(\sigma)\big) + m \sum_{\tau \in \Phi_a} s_\infty(a) - A_0\, a \log N + 2\mathcal{E}_0, \tag{12.4}$$

where one $\mathcal{E}_0$ comes from the source form and one from the estimate of the kernels. Each of the m lifts differentiates to order $\kappa(\sigma)$ in k idle directions and to order a at the residues of Φ_a, and the final division by $\bar{\pi}_N$ saves $a \log N$ at each residue of Φ_a. The bound comes

from the Gaussian norm of Section 7; the number of terms in the class sums is a lower-order error.

(2) *Divisibility.* At a prime p that does not split in K the composite lift is divisible by p^{gain_p}, up to a bounded loss, where

$$2\sum_p \mathrm{gain}_p \log p = m\Big[\sum_{\tau\in\Phi_a} s_{\mathrm{fin}}(a,\tau) + k\sum_{\sigma} s_{\mathrm{fin}}\big(\kappa(\sigma),\sigma\big)\Big]. \tag{12.5}$$

The ramp $(y/p^h - \kappa(-p^h\tau))_+$ in s_{fin} is the contribution of p^h, as in Section 6; only odd h occur at inert p.

(3) *Normalization.* Multiply the map by the positive rational number $c_B = \prod_p p^{-\lfloor \mathrm{gain}_p - \mathrm{err}_p\rfloor}$, a product over the small primes that do not split in K and over N (where the gain is 0, so the exponent is ≥ 0), with error allowances err_p. The forms stay integral up to the index in (b), and

$$\log c_B \le -\sum_p \mathrm{gain}_p \log p + \mathcal{E}_0 + b. \tag{12.6}$$

The factor $|G|e_0 g$ of (8.15) costs at most x^{99}. Since c_B enters a squared length twice, subtracting the divisibility from the size shows that the logarithm of the squared length of a normalized lift is at most $x^{99} + 4\mathcal{E}_0 + 2b$ plus

$$\underbrace{m\,k\sum_{\sigma}\big(s_\infty(\kappa(\sigma)) - s_{\mathrm{fin}}(\kappa(\sigma),\sigma)\big)}_{h_{\mathrm{main}}} + m\sum_{\tau\in\Phi_a}\big(s_{\mathrm{main}} - s_{\mathrm{fin}}(a,\tau)\big) + A_0\big[m(s_\infty(a) - s_{\mathrm{main}}) - a\log N\big]. \tag{12.7}$$

The middle term is $-m\cdot\mathrm{rem}$ by (9.3). The last term is an error, by the choice of N: the bracket is at most $32a\log b$, so with $A_0 = f/\varphi(Q)$ and $\varphi(Q)\ge 32$ the term is at most $fa\log b$. Altogether

$$\log\big\|\Theta_1(g\mathfrak{p}\varphi)\big\|_\sigma^2 \;\le\; h_{\mathrm{main}} - m\cdot\mathrm{rem} + x^{99} + 4\mathcal{E}_0 + 2b + fa\log b \;\le\; h_{\mathrm{main}} - m\cdot\mathrm{rem} + \mathcal{E} \tag{12.8}$$

for large x, which is part (a). By Stirling's formula and (7.11),

$$s_\infty(a) = a\Big[\log a - 1 - \gamma + \mathrm{Re}\,\frac{L'}{L}(1,\chi_d) + \log\frac{M\,\mathrm{rd}_E}{2\pi}\Big] + O(\log a), \tag{12.9}$$

$$s_{\mathrm{main}} = a\Big[\log a - \log\bar\kappa - 1 - \gamma + \mathrm{Re}\,\frac{L'}{L}(1,\chi_d)\Big], \tag{12.10}$$

where $\log\bar\kappa = \frac{1}{f}\sum_\sigma \log\kappa(\sigma)$. The terms $a\log a$, $-a$, $-\gamma a$ and $a\,\mathrm{Re}\,\frac{L'}{L}(1,\chi_d)$ agree: this is the match through the period announced in Section 7. What is left of one lift is $a(\log\bar\kappa + G_0)$ with $G_0 = \log(M\,\mathrm{rd}_E/2\pi)$, and N is chosen with

$$\log N = m\log\bar\kappa + (m-1)G_0 - \tfrac{15}{2}\log b + O\big(\sqrt{\log b}\big), \tag{12.11}$$

so that $m(s_\infty(a) - s_{\mathrm{main}}) - a\log N = O(a\log b)$. The prime N cannot be taken larger: the number of source forms grows like N^{kf}, and Theorem (B) needs the ratio of the numbers of source and target forms in a fixed range. So $\log N$ is pinned from below by (A) and from above by (B). In the Lean the choice is one-sided, in a window of width about $\sqrt{\log b}$ above the value (12.11).

What goes into the divisibility. At a prime p inert in K the starting point is the normal form (6.19) of Section 6; the digit-sum loss in $v_p(j!)$ and the exponential factor are controlled by local estimates. For products of theta functions the exponential factor couples τ to its partner $-p^h\tau$, and the directions of differentiation are changed, at a determinant cost, so that these couplings vanish to the order needed. In more detail [*from the Lean*]:

- *Where the coupling comes from.* The many-variable theta function is a product over p-integral vectors u_j with entries in E that resolve the identity:

$$\prod_j \theta(\varpi\, u_j^* z)^{w_j}, \qquad \sum_j w_j\, u_{j,i}(\tau)\,\overline{u_{j,i'}(\sigma)} = \delta_{\tau\sigma}\delta_{ii'}, \qquad u_j^* z = \sum_{\tau,i} \overline{u_{j,i}(\tau)}\, z_{\tau,i}, \tag{12.12}$$

  with finitely many vectors u_j and positive integer multiplicities w_j; each $u_j^* z$ is a linear form in the variables. In the normal form (6.19) the exponents add up to $\frac{\varpi^{p+1}}{p}\sum_j w_j\,\Xi(u_j^* z)$. Modulo p the cross terms of a p^h-th power vanish, and Lemma 6.1 turns $\overline{u(\tau)}^{\,p^h}$ into $u(-p^h\tau)$. With the resolution of the identity this gives

$$\sum_j w_j\,(u_j^* z)^{p^h}(u_j^* z) \;\equiv\; \sum_{\tau,i} z_{\tau,i}^{\,p^h}\, z_{-p^h\tau,\,i} \pmod p. \tag{12.13}$$

  So the terms of degree p^h+1 in the exponent pair the variables at τ with those at the partner residue. A multiple of p in the exponent is harmless, because $v_p(\varpi^{p+1}) = (p+1)/(p^2-1) = 1/(p-1)$ is exactly the threshold at which an exponential has p-integral coefficients; it can be absorbed in W.
- *Where the ramp comes from.* At a residue $\tau \in \Phi_a$ the a differentiations are all in one direction. At the partner $\tau' = -p^h\tau$ there are about $\kappa(\tau')$ groups of k differentiations. The coupling between them can be removed in two ways: by a congruence modulo p on each of the $\kappa(\tau')$ groups at τ', at a cost of 1 each in the valuation of a determinant, in all $\kappa(\tau')$; or by a congruence modulo p^{1/p^h} on each of the a differentiations at τ, at a cost of $1/p^h$ each, in all a/p^h. (Why a congruence on a group of k directions costs 1 and not k: for a vector v of k integers one can find an integral matrix U of determinant p with every entry of Uv divisible by p, by clearing, modulo p, all entries but one of least valuation with row operations, and multiplying that row by p. For $p=5$ and $v=(2,3)$, $U = \left(\begin{smallmatrix}5&0\\1&1\end{smallmatrix}\right)$ gives $Uv=(10,5)$.) Taking the cheaper one costs $\min(a/p^h, \kappa(\tau'))$, and the gain a/p^h less this cost is

$$\frac{a}{p^h} - \min\Big(\frac{a}{p^h},\, \kappa(-p^h\tau)\Big) = \Big(\frac{a}{p^h} - \kappa(-p^h\tau)\Big)_+, \tag{12.14}$$

  the ramp. Between two residues outside Φ_a the same choice costs $k\min(\kappa(\sigma)/p^h, \kappa(\sigma'))$, which is where the factor k in h_{main} comes from. This explains the leading term; the actual counts contain shifts of the ranks and corrections that are charged to the local error, and the simultaneous choice over all primes and couplings is the content of the local estimate.

Three kinds of primes need more.

- *Exceptional primes:* inert $p \not\equiv \pm 1 \pmod m$ such that $-p^h\tau \in \Phi_a$ for some $\tau \in \Phi_a$ and some odd h with $p^{h-1} < m$. For these the gain is repaired with Hecke operators whose matrices have small elementary divisors.
- *The ramified prime d.* Since $d \equiv Q/2 - 1 \pmod Q$, a residue and its partner $-d^h\tau$ never lie in the same class modulo Q, so they are never both in Φ_a or both in Φ_b; this bounds the correction at d.

- *The prime N* is treated separately. One restricts the integral source module at colength $O(mfb)$ per source dimension, and on the resulting submodule the composite lift has valuation at least $-2mfb$ at N. The normalization of the lift clears the remaining denominator. The associated loss in the logarithm of the index, per source dimension, is $O(mfb\log N) = O(fm^2b\log b)$.

The two classes ±1 modulo Q are used in two more ways: an exceptional prime cannot carry a residue of Φ_a to one of Φ_b, so the residues marked for the Hecke repair avoid Φ_b; and since Φ_b lies in one class, the residues of Φ_b have no partners in Φ_b at the prime d.

13. Enough forms: Theorem B

Theorem 13.1 (B: enough forms; [*from the Lean*])**.** *$D > 0$, and there is $T \subset G$ with*

$$|T| \cdot \dim S \le 5D\log(1+2b^2) \tag{13.1}$$

such that the images $g\mathfrak{p}(S)$, $g \in T$, together span a subspace of dimension at least $(1-4/b^2)D$.

The proof is a second-moment argument, in four steps.

(1) *The second moment.* The Poincaré map P (the map $\mathcal{P}$, suitably scaled) is defined on the space S_{full} of all source forms. After the projection e_0 it is nearly an isometry into the target forms, and its translates are nearly orthogonal on average:

$$\sum_{g\in G}\|P^*gP\|_{\text{HS}}^2 \;\le\; \Big(1+\frac{9}{b^4}\Big)\frac{(\dim S_{\text{full}})^2}{\deg\cdot\mathrm{mass}(E,r)}, \tag{13.2}$$

with the Hilbert–Schmidt norm. Projecting away the bad labels, and then restricting to the ramified source forms S, loses only a controlled small proportion of the dimensions; this is part of the final $4/b^2$. The left side is computed by unfolding the Poincaré series; the terms are evaluated by repeated Poisson summation and local Bessel functions, and all but the main ones are shown to be small.

(2) *The dimension.* By (8.20), $D \le |G|\cdot\deg\cdot\mathrm{mass}(E,r)$. This uses a comparison with a definite unitary group of the same rank, on which each system of Hecke eigenvalues has the same multiplicity among the good forms (a simple trace formula), and the definition of the mass as a weighted count of classes.

(3) *The factor that is not computed cancels.* Multiplying,

$$D\sum_{g\in G}\|P^*gP\|_{\text{HS}}^2 \le |G|(1+9b^{-4})(\dim S_{\text{full}})^2, \tag{13.3}$$

and the factor $\deg\cdot\mathrm{mass}$ has cancelled.

(4) *Few translates.* By the Cauchy–Schwarz inequality the translates *spread*: every subspace that misses a proportion of the target space is met substantially by the image of some translate. The mechanism is one line: with $A = \sum_g gPP^*g^{-1}$ one has $\sum_g\|\mathrm{pr}_W gP\|_{\text{HS}}^2 = \mathrm{tr}(\mathrm{pr}_W A)$ for a subspace W, and $\|A\|_{\text{HS}}^2 = |G|\sum_g\|P^*gP\|_{\text{HS}}^2$. Choosing translates greedily on the full source space, each step multiplies the missing proportion by at most $1-2\dim S_{\text{full}}/(5D)$, and one stops when $2/b^2$ is missing; this gives the bound for $|T|$. Restricting each of the $|T|$ translates from S_{full} to the ramified forms S, a subspace of relative codimension at most b^{-8}, loses at most a further $2/b^2$ [*from the Lean*].

The number of translates is about $\log b$ times $D/\dim S$; the bound for $|T|\cdot\dim S$ is what enters the index in Section 9.

14. Integral forms are long: Theorem C

Theorem (C) of Section 9 is the lower bound, the analogue of "a nonzero integer is at least 1". Its proof combines the product formula with a twist chosen at each prime, separately for each prime ideal above it.

(1) *Completing to a basis.* Complete $y_1, \dots, y_n$ to a basis of the target space by at most $(4/b^2)D$ *auxiliary* forms. Their untwisted lifts have squared length at most $\exp(2x_1^2b^2)$ at every conjugate, and their twisted lifts become integral after multiplication by a controlled nonzero number $\gamma(\xi)$. Both the size and the cost of this number are absorbed in the error: since $D - n \le 4D/b^2$ and b is far larger than x_1^2, the total is at most $(D-n)(2x_1^2b^2 + b) \le D(8x_1^2 + 4/b)$, much less than Db. This is why the family must fill all but a proportion $4/b^2$ of the target space.

(2) *The product formula.* Apply the product formula to the Gram determinant of the untwisted lifts of this basis. Up to the Gram determinant of the basis itself it is the determinant of the whole chain,

$$Z(\xi) = \det(\Theta_\xi^* \Theta_\xi) \qquad \text{at } \xi = 1, \tag{14.1}$$

as in (8.26). At the conjugates its size is bounded by the covolume. At an embedding ι into an algebraic closure of $\mathbb{Q}_\ell$ its valuation may be read in a *twisted* family, with a twist ξ_ι chosen for ι and of conductor prime to ℓ, because the index hypothesis holds uniformly in such twists. So the lower bound reduces to bounding the ratios $Z(1)/Z(\xi_\iota)$ in the product formula. In the ratio the constant of the chain, which is not computed, cancels.

(3) *From the chain to one-step lifts.* The ratio of chain determinants is expressed, with explicit *exchange* factors, through m ratios for auxiliary one-step lifts,

$$\frac{Z(1)}{Z(\xi)} \longleftrightarrow \prod_{j=1}^{m} \frac{Z_j(1)}{Z_j(\xi^j)}, \tag{14.2}$$

where Z_j is the determinant of a lift $W_j \to V_j$ between spaces of ranks r and $r+t$, on the good forms. These are comparison lifts, not the successive maps of the chain (Section 8). This rests on an *exchange identity* of Siegel–Weil type: two theta integrals over unitary groups of complementary rank agree up to a factor c with $|c|^2$ a constant times a product of powers of residue-field cardinalities and of squared absolute values of partial L-values. Paths through a grid of spaces joined by lifts and transposed lifts are moved, one square at a time, in m^2 steps.

(4) *One lift and Hecke L-values.* Each one-step ratio reduces, through a comparison operator, to ratios of values of Hecke L-functions of E at the points $1 \le z \le t$, for characters of weight between z and $4b$; the L-values appear as squared norms. Each one-step lift returns the main amount of its idle directions: this is where the Dh in Theorem (C) comes from.

(5) *Denominators.* The projection of integral forms onto the lifts can have denominators. They are *congruence numbers* (compare Hida [35]): forms orthogonal to the lifts but with the same Hecke eigenvalues at many primes (*false lifts*) can be congruent to lifts. Through the Galois representations attached to these forms, the congruence numbers are bounded by the lengths of groups of Galois cocycles (Selmer groups). The odd classes are bounded by a finite-level Euler system argument (Stickelberger's theorem and Kolyvagin's method; compare [36, 37]; the main conjecture of Iwasawa theory is not used). The even classes in positive twist need no uniform bound: linear relations at auxiliary primes, of the kind introduced by Taylor and Wiles [34], control them with a bounded loss. The untwisted case uses the class number of E, with $\log h_E \le x^{30}$.

(6) *The choice of twists.* For each prime ℓ, and each embedding ι into an algebraic closure of $\mathbb{Q}_\ell$, there is a family of m-fold twists of conductor prime to ℓ whose denominators are small and whose ratios of L-values have height at most x_1. The twist may depend on ι. The determinant to which the product formula is applied stays fixed; only the auxiliary twisted determinant used to estimate one local valuation changes with the place.

The cost of one lift has two main parts. Per target dimension, the archimedean size of the comparison operator is $O(fmb(\log b)^2)$, and the denominators, dominated by local defects at primes, cost $O(fmb\log b)$. Over the chain these bounds are $O(fm^2b(\log b)^2)$ and $O(fm^2b\log b)$, against allowances of order $fa(\log b)^4$ and $fa(\log b)^3$. With $a = xb$ both fit for $m \asymp \sqrt{x}$, with logarithmic room. This explains the scale chosen in Section 10.

The inputs of this step are the deepest in the proof: Galois representations attached to automorphic forms on unitary groups ([32, 33]; constructed in the Lean through Shimura varieties and p-adic Hodge theory, in a weaker form that suffices: a compatibility of Frobenius with the Hecke polynomials at most primes, to any prescribed ℓ-adic accuracy), class field theory, and Euler systems over cyclotomic fields.

15. Dirichlet L-functions

Theorem 15.1 ([*from the Lean; the reduction sketched here*])**.** *Let χ be a Dirichlet character modulo $q_\chi \geq 1$. Then $L(s,\chi)$ has no zeros in $\operatorname{Re} s > \frac{1}{2}$ (for principal χ, $s = 1$ is a pole).*

The argument is the same with three substitutions.

(1) Replace Q by $Q' = Q\,q_\chi$ and the class $Q/2 - 1$ by $Q'/2 - 1$. A prime d in this class is again $\equiv 7 \pmod 8$ and prime to Q', and $d \equiv -1$ modulo every odd prime dividing Q', so every prime dividing Q' splits in K, as (6.9) requires.
(2) Replace $\varphi(Q)$ by $\varphi(Q')$. The sums of Sections 2–3 run over the characters modulo Q', and among them is the character χ_{ind} modulo Q' induced by χ.
(3) At the end, use $L(s,\chi_{\mathrm{ind}}) = L(s,\chi)\prod_{p\mid Q'}(1-\chi(p)p^{-s})$ in place of the formula for the principal character. For non-principal χ the point $s = 1$ is Dirichlet's theorem.

Nothing else changes: the main inequality (M) is proved for every positive multiple of Q in place of Q, and the constant in the one-sided bound becomes $128\varphi(Q') + 18$.

What is classical and what is new

The analytic reduction (Steps 1–4) is classical in its tools: Landau's lemma, the explicit pole structure of a Laplace transform, and a zero-density estimate proved with theta series and the Pólya–Vinogradov inequality. The mechanism of the local arithmetic (Step 6 and the period in Step 7) is classical too: the lattice $\mathcal{O}_K$, the function $\wp$, complex multiplication, Deuring's supersingular reduction at inert primes, and the Chowla–Selberg formula. The estimate that the proof needs, uniform in many variables and with explicit losses, is not a classical theorem; it is a result of the formal development. Chebyshev's principle (Step 5) is from 1850.

What is new is the global object and its lower bound: Gram determinants of algebraic modular forms on unitary groups, built from iterated theta lifts, and Theorem C, which replaces "a nonzero integer is at least 1" by a lower bound, relative to the same main term as the upper bound, using twists, Hecke L-values, Euler systems and the Galois representations attached to automorphic forms on unitary groups. Whether the proof can be simplified, for instance with classical theta series of quadratic forms or a more transparent proof of Theorem C, is open.

Appendix A. Worked examples

This appendix carries out the construction of Section 8 on three small models, with every ingredient written out: derivatives of a function with complex multiplication, the Gaussian norm, a Gram determinant, its size and its factorization, classes of lattices, class sums, and a chain of lifts. All numbers are [*computed*]: the algebraic arithmetic is exact, and logarithms are evaluated numerically. The models are not the objects of the proof. They use one complex variable, where the proof uses $f \geq 3840$; they have no Poincaré series, no target forms and no twists; and nothing in them corresponds to Theorem C. What they show is the bookkeeping of the upper-bound side: how size, divisibility and the period are compared, and where the uncontrolled cofactor sits.

A.1. One vector: size against the prime sum. Take the lemniscate sine of Section 6, with $A_n = \mathrm{sl}^{(n)}(0)$, and the one-term vector $c_n z^n$, $c_n = A_n/n!$. In the Gaussian norm its squared length is $A_n^2/n!$. Put

$$\mathrm{size}(n) = \log \frac{A_n^2}{n!}, \qquad \mathrm{ramp}(n) = \sum_{k\geq 1} \lambda(k)\Big(\frac{n}{k} - 1\Big)_+, \qquad \lambda(k) = \Lambda(k)(1 - \chi_4(k)), \tag{A.1}$$

where χ_4 is the character modulo 4. So $\lambda(k)$ is $2\log p$ at odd powers of primes $p \equiv 3 \pmod 4$, $\log 2$ at powers of 2, and 0 otherwise; $\mathrm{ramp}(n)$ is the sum of ramps (5.4) with weight 1. Throughout A.1 and A.2, n and a run through the positive integers congruent to 1 modulo 4, since $A_n = 0$ otherwise. Note that $A_n^2/n!$ is not an integer. We compare its size with the divisibility of the numerator A_n^2, counted by $\mathrm{ramp}(n)$; this is not a factorization of $A_n^2/n!$, whose denominator has prime factors of its own. Two classical facts:

$$\mathrm{ramp}(n) = n\log n + n\Big(\frac{L'}{L}(1,\chi_4) - \gamma - 1\Big) + o(n), \tag{A.2}$$

$$\mathrm{size}(n) = n\log n + n\Big(\frac{L'}{L}(1,\chi_4) - \gamma - 1 + 2\log\frac{2}{\pi}\Big) + \tfrac{1}{2}\log n + O(1). \tag{A.3}$$

The first follows by partial summation from $\sum_{k\leq y}\lambda(k)/k = \log y - \gamma + \frac{L'}{L}(1,\chi_4) + o(1)$ and $\sum_{k\leq y}\lambda(k) \sim y$ [6]. For the second, sl has its nearest poles at the four points $(\pm 1 \pm i)\varpi/2$, at distance $R = \varpi/\sqrt{2}$, where $\varpi = \Gamma(\frac{1}{4})^2/(2\sqrt{2\pi}) = 2.62205\ldots$; so $|A_n|/n!$ is R^{-n} up to a bounded factor, and $\mathrm{size}(n) = \log n! - 2n\log R + O(1)$. Lerch's formula (7.9) and the functional equation, for the character χ_4 of $\mathbb{Q}(i)$, give

$$\frac{L'}{L}(1,\chi_4) = \gamma + 2\log 2 + 3\log\pi - 4\log\Gamma(\tfrac{1}{4}) = 0.245609\ldots, \tag{A.4}$$

and therefore $-2\log R = \frac{L'}{L}(1,\chi_4) - \gamma + 2\log\frac{2}{\pi}$. So the transcendental constant of the prime sum is reproduced by the period of the lattice, and

$$\mathrm{size}(n) - \mathrm{ramp}(n) = 2n\log\frac{2}{\pi} + o(n). \tag{A.5}$$

The remaining term is a geometric factor $(2/\pi)^{2n}$. So the terms $n\log n$ agree, and the constant $\frac{L'}{L}(1,\chi) - \gamma - 1$ agrees through the period. This explains the period-dependent constant in the size estimate. It is a statement about sizes only: it does not give an inequality between the divisibility and the size of the rational number $A_n^2/n!$, and the factor $(2/\pi)^{2n}$ cannot simply be scaled away, since $(\pi/2)^n A_n$ is no longer algebraic. In the proof the integral normalization and the control of denominators in Theorems A and C supply what is missing.

n	size(n)	ramp(n)	$2n\log(2/\pi)$	size − ramp − $2n\log(2/\pi)$
101	245.18	333.26	−91.22	3.14
201	621.88	798.85	−181.54	4.57
501	2000.46	2446.69	−452.49	6.26
1001	4684.57	5584.78	−904.07	3.86
2001	10744.89	12547.71	−1807.23	4.41

At $n = 2001$ the numbers 10745, 12548 and −1807 combine to 4.4. What is left is $\frac{1}{2}\log n$, plus a constant, minus the error term of the prime sum (A.2). It is not bounded as n grows: it fluctuates on the scale of that error term (for n up to 10^8 it stays below $0.17\sqrt{n}$ in absolute value); a bound for it from one side is a bound for the error in a sum over the primes $p \equiv 3 \pmod 4$. This is the shape of the proof's main inequality.

The divisibility actually present. For $n = 2001$ the exponent of each prime $p \equiv 3 \pmod 4$ in A_n is exactly $\lfloor np/(p^2-1)\rfloor$, and $v_2(A_n) = (n-1)/2$. Counted twice, these contribute $12376.59 + 1386.29 = 13762.88$, more than $\mathrm{ramp}(n) = 12547.71$: the floors round once per prime, and the ramps are what is guaranteed uniformly.

The cofactor. Write $A_n^2 = P_n\, C_n^2$, with P_n the part made of 2 and the primes $p \equiv 3 \pmod 4$, and C_n the rest. For $n = 2001$,

(A.6)
$$\log A_n^2 = 23959.01, \qquad \log P_n = 13762.88, \qquad \log C_n^2 = 10196.13, \qquad \log n! = 13214.13.$$

No prime $p \equiv 1 \pmod 4$ below 6000 divides C_n; it is a product of large primes. That A_n is a nonzero integer gives only $C_n \geq 1$, hence $\log A_n^2 \geq \log P_n$: an inequality between 23959 and 13763, off by a factor of nearly 2. The comparison (A.5) is sharp because $\log C_n^2$ and $\log n!$ agree on the scale $n\log n$ (10196 against 13214 here), and the integer argument does not give a lower bound for C_n of that strength. This is the uncontrolled cofactor.

A.2. An 8×8 Gram determinant. Now take eight vectors. Fix $a \equiv 1 \pmod 4$ and the degrees $n_j = a + 4j$, $0 \leq j \leq 7$. Let v_i be the part in these degrees of the Taylor series of $\mathrm{sl}^{(4i)}$, $0 \leq i \leq 7$:

$$v_i(z) = \sum_{j=0}^{7} A_{a+4i+4j}\,\frac{z^{n_j}}{n_j!}, \qquad \langle v_i, v_{i'}\rangle = \sum_{j=0}^{7} \frac{A_{a+4i+4j}\,A_{a+4i'+4j}}{n_j!}, \tag{A.7}$$

since $\|z^n/n!\|^2 = 1/n!$. With $X = (A_{a+4i+4j})_{0\leq i,j\leq 7}$, the Gram determinant is

$$H = \det\big(\langle v_i, v_{i'}\rangle\big) = \frac{(\det X)^2}{\prod_j n_j!}, \tag{A.8}$$

a positive rational number. Its divisibility is guaranteed as follows. Each entry A_n is divisible by $p^{e_p(n)}$, $e_p(n) = \lfloor np/(p^2-1)\rfloor$, for $p \equiv 3 \pmod 4$. Expanding the determinant,

$$v_p(\det X) \;\geq\; g_p := \min_{\sigma} \sum_{j=0}^{7} e_p\big(a + 4j + 4\sigma(j)\big), \tag{A.9}$$

the minimum over the permutations σ of $\{0,\dots,7\}$. For $a = 5$ everything can be printed:

$$\det X = 2^{216}\,3^{109}\,5^{51}\,7^{31}\,11^{17}\,13^{14}\,17^{8}\,19^{7}\,23^{5}\,29^{2}\,31, \tag{A.10}$$

$$H = 2^{301}\,3^{155}\,5^{72}\,7^{44}\,11^{24}\,13^{20}\,17^{11}\,19^{10}\,23^{7}\,29^{2}\,31. \tag{A.11}$$

The guaranteed exponents (A.9) are $g_3 = 92$, $g_7 = 31$, $g_{11} = 17$, $g_{19} = 7$, $g_{23} = 5$, $g_{31} = 1$: attained exactly at every prime except 3. That $\det X$ has only small prime factors for $a = 5$ is in the nature of Hankel determinants of the Taylor coefficients of an elliptic function, which

have closed forms (compare Frobenius and Stickelberger [25], Milne [26]); we have not derived this case from them, and for $a = 101$ and $a = 501$ large primes do occur. The budget of $\log H$, for three values of a:

	$a = 5$	$a = 101$	$a = 501$
$\log H$	782.0	3229.2	17684.2
Hadamard bound $\sum_i \log \|v_i\|^2$	1070.7	3715.7	18396.8
$\log(\det X)^2 = \log H + \sum_j \log n_j!$	1115.3	6703.8	39323.1
primes $p \equiv 3 \pmod 4$, guaranteed $2\sum_p g_p \log p$	483.8	3255.8	20223.1
primes $p \equiv 3 \pmod 4$, $p < 3000$, actual	521.1	3304.7	20266.6
the prime 2	299.4	831.8	3049.8
primes $p \equiv 1 \pmod 4$, $p < 3000$	294.8	330.2	330.2
the rest (primes above 3000, of either kind)	0.0	2237.2	15676.5
$\sum_j \log n_j!$	333.3	3474.6	21638.9

Four things can be read off.

- Hadamard's inequality holds with room to spare: the eight vectors are far from orthogonal.
- The guaranteed divisibility (A.9) is nearly all the divisibility there is at these primes.
- In $(\det X)^2$ the primes $p \equiv 3 \pmod 4$ carry 47%, 49% and 52%: about half, as for a single derivative. Taking a determinant does not change the share.
- The Gaussian weights divide by $\prod_j n_j!$, about the other half in size. So $\log H$ is comparable to the divisibility of the numerator $(\det X)^2$: 782 against 521, 3229 against 3305, 17684 against 20267. The difference is of lower order than the main terms. As in A.1 this compares sizes; the factorials are also denominators of H.

A.3. Three classes: $K = \mathbb{Q}(\sqrt{-23})$. The field $\mathbb{Q}(\sqrt{-23})$ has class number 3. Its classes are those of the forms $x^2+xy+6y^2$ and $2x^2 \pm xy+3y^2$, with lattices $[1, \tau_c]$, $\tau_c = (-b_c+\sqrt{-23})/(2a_c)$, $(a_c, b_c) = (1,1), (2,1), (2,-1)$. The three j-invariants are the roots of

$$X^3 + 3491750X^2 - 5151296875X + 12771880859375. \tag{A.12}$$

One of them generates the cubic field $\mathbb{Q}(\alpha)$, $\alpha^3 = \alpha + 1$, and the three embeddings of this field give the three j-invariants. The embeddings $\alpha \mapsto 1.3247$, $\alpha \mapsto -0.6624 \pm 0.5623i$ correspond to the three classes, the real one to the principal class. The curve

$$y^2 = x^3 + Ax + B, \qquad \begin{aligned} A &= -1747440\alpha^2 - 2315520\alpha - 1321920, \\ B &= 1837866240\alpha^2 + 2434707072\alpha + 1387435392, \end{aligned} \tag{A.13}$$

has complex multiplication by the integers of K, and its discriminant has norm $-2^{72}3^{36}23^3$, so it has good reduction at every prime except possibly 2, 3 and 23 (in fact it is bad only at primes above 2 and 23); the primes 2 and 3 split in K. Its three conjugates are the three lattices, each scaled by its period. Let $t = -x/y = -2\wp/\wp'$ and

$$\tau_n = t^{(n)}(0) \in \mathbb{Q}(\alpha), \qquad \tau_n(c) = \text{the value of } \tau_n \text{ at the class } c. \tag{A.14}$$

For example $\tau_5 = 83877120\alpha^2 + 111144960\alpha + 63452160$, of norm $2^{24}3^{12} \cdot 5^3 \cdot 11 \cdot 17 \cdot 23$, with values $3.58 \cdot 10^8$ at the principal class and $(1.14 \pm 0.18i) \cdot 10^5$ at the other two.

Divisibility. The inert primes are $5, 7, 11, 17, 19, 37, \ldots$ Each factors in $\mathbb{Q}(\alpha)$ into two prime ideals. Among the 60 nonzero derivatives τ_n with $1 \le n \le 121$ (n odd, $n \ne 3$), both prime ideals divide τ_n at least $\lfloor np/(p^2-1) \rfloor$ times, and the smaller of the two exponents equals this number in 50, 56, 59, 60, 60, 60 cases for $p = 5, 7, 11, 17, 19, 37$. The bound need not be attained at both prime ideals: for $p = 11$, $n = 97$ the exponents are 9 and 8. At the split primes 13, 29, 31 the exponent never exceeds 1.

Size against the prime sum. Average the sizes over the three classes:

$$S(n)=\frac{1}{3}\sum_{c}\log\frac{|\tau_n(c)|^2}{n!},\qquad \mathrm{ramp}(n)=\sum_{k}\Lambda(k)(1-\chi_{23}(k))\Big(\frac{n}{k}-1\Big)_+. \tag{A.15}$$

Here and below n is odd and $n\neq 3$, since $\tau_n=0$ otherwise. The poles of t nearest to 0 are at half the shortest period, so $S(n)=\log n!-\frac{2}{h}\sum_c n\log(|\omega_c|/2)+O(1)$, with $h=3$ the class number and ω_c the shortest period of the c-th lattice. The periods are given by $\Delta=(2\pi)^{12}\eta(\tau_c)^{24}\omega_c^{-12}$, with η the Dedekind eta function, and Kronecker's limit formula [19], summed over the classes and combined with $\zeta_K=\zeta\cdot L(\cdot,\chi_d)$, in the form

$$\frac{L'}{L}(1,\chi_d)=\gamma-\log 2-\tfrac{1}{2}\log d-\frac{2}{h}\sum_{c}\log\Big(\sqrt{y_c}\,|\eta(\tau_c)|^2\Big),\qquad y_c=\operatorname{Im}\tau_c, \tag{A.16}$$

(equivalent to the Chowla–Selberg formula of Section 7; for $d=4$ it is (A.4)) expresses the average of $\log|\eta(\tau_c)|^2$ through $\frac{L'}{L}(1,\chi_d)$. The result is

$$S(n)-\mathrm{ramp}(n)=n\,E+o(n), \tag{A.17}$$

$$E=\log 2-2\log\pi+\tfrac{1}{2}\log 23+\frac{1}{3}\sum_{c}\log y_c+\frac{1}{18}\log|\mathrm{N}\Delta|=5.8763\ldots, \tag{A.18}$$

where $|\mathrm{N}\Delta|=2^{72}3^{36}23^3$. Again the constant $\frac{L'}{L}(1,\chi_{23})-\gamma-1=-1.99934$ has cancelled, and E is an elementary constant fixed by the scale of the model.

n	$S(n)$	ramp(n)	nE	$S-\mathrm{ramp}-nE$
25	173.82	33.79	146.91	−6.87
49	377.43	95.15	287.94	−5.65
73	593.18	167.36	428.97	−3.15
97	816.91	248.34	570.00	−1.43
121	1046.62	339.48	711.04	−3.90

So the periods of the three lattices, averaged over the classes, reproduce the constant of the sum over the primes inert in $\mathbb{Q}(\sqrt{-23})$. This is Step 7 with a class number larger than one.

A.4. Class sums and a chain of lifts. A *form* is now a function on the three classes $c\in\mathbb{Z}/3$ with polynomial values. The *lift of order* n is the class sum with the matrix

$$M_n=\Big(\tfrac{1}{2}\,\tau_n(c-b)\Big)_{c,b\in\mathbb{Z}/3}, \tag{A.19}$$

where $\frac{1}{2}=1/|\mathrm{Aut}|$ (the units ±1) and $c-b$ is the class of the lattice of maps from the lattice b to the lattice c. Take nine source forms $\varphi_{b,i}$, $b\in\mathbb{Z}/3$, $i\in\{0,1,2\}$: the form supported at the class b with value w^{2i}. The first lift differentiates t against the source polynomial and keeps the degrees $n_1\in\{25,27,29\}$:

$$v^{(1)}_{b,i}(c)(z_1)=\frac{1}{2}\sum_{n_1}\tau_{n_1+2i}(c-b)\,\frac{z_1^{n_1}}{n_1!}. \tag{A.20}$$

Each further lift brings a new variable z_l and keeps the degrees $n_l\in\{25,27\}$:

$$v^{(m)}_{b,i}(c)(z_1,\ldots,z_m)=\sum_{c_1,\ldots,c_{m-1}}\ \sum_{n_1,\ldots,n_m}\frac{\tau_{n_1+2i}(c_1-b)}{2}\prod_{l=2}^{m}\frac{\tau_{n_l}(c_l-c_{l-1})}{2}\prod_{l=1}^{m}\frac{z_l^{n_l}}{n_l!},\qquad c_m=c. \tag{A.21}$$

So every coordinate of $v^{(m)}$ is a finite sum, over chains of classes $b\to c_1\to\cdots\to c_m$, of products of CM derivatives, as in Section 8. The inner product is $\langle v,w\rangle=\sum_c\frac{1}{2}\langle v(c),w(c)\rangle$

with the Gaussian norm, and H_m is the 9×9 Gram determinant of the $v_{b,i}^{(m)}$. The matrices M_n are circulant, so they commute with each other and with their adjoints. Put

$$\Pi = \frac{M_{25}^* M_{25}}{25!} + \frac{M_{27}^* M_{27}}{27!}. \tag{A.22}$$

Summing over the degrees of each new variable replaces the identity on the classes by Π, once for each lift after the first, and the nine vectors of the first lift form a square matrix. Hence

$$H_m = H_1 \, (\det \Pi)^{3(m-1)}. \tag{A.23}$$

H_m is a positive rational number: a cyclic shift of the classes and complex conjugation, which generate the symmetries of the three conjugate fields, fix $\det \Pi$ and H_1.

	$m=1$	$m=2$	$m=3$
number of vectors	9	9	9
dimension of the space they lie in	9	18	36
$\log H_m$	2546.10	4872.23	7198.37
Hadamard bound	2776.09	5102.23	7428.36
$v_5(H_m)$: actual (guaranteed)	48 (42)	84 (78)	120 (114)
$v_7(H_m)$	38 (30)	71 (57)	104 (84)
$v_{11}(H_m)$	18 (18)	36 (36)	54 (54)
$v_{17}(H_m)$, $v_{19}(H_m)$	9 (9)	18 (18)	27 (27)

The guaranteed exponents come from the entries as in (A.9), less the exponents of p in the factorials of the Gaussian weights. What the table shows:

- *Each lift adds the same amount.* By (A.23), $\log H_m$ increases by $3 \log \det \Pi = 2326.14$ at each step, and the exponents of $5, 7, 11, 17, 19$ by $36, 33, 18, 9, 9$. This illustrates the repeated divisibility gain of Step 10: each lift contributes the inert-prime divisibility once more. In this model the initial factor is paid once; the model does not show the accumulated costs of the proof, whose bounds are quadratic in m. The exactness is a feature of this model, where the nine vectors are a full orbit; in the proof the corresponding gains and errors are the content of Theorems A–C.
- *The vectors fill a small part of their space.* At $m = 3$ the nine vectors lie in a space of dimension 36; the Gram matrix is still 9×9, and only inner products enter.
- *Hadamard's bound* exceeds $\log H_m$ by 230.0 for each m (a numerical observation; (A.23) does not imply it).
- *What the model does not show.* At these small orders the scale term nE of (A.17) dominates, and the class sums are governed by the principal class, where τ_n is far larger, not by the average over the classes. So the inert primes account for a small part of $\log H_m$ here. The matching of main terms needs large orders, as in A.1 and A.3, and the proof's normalizations.

A.5. How the proof deals with the cofactor. In A.1 and A.2 the numbers factor as (inert primes) $\times$ (cofactor), and the cofactor is about half of the logarithm. That the cofactor is a nonzero integer gives only that it is at least 1, far from the lower bound that would balance the size. The proof does not try to bound a single cofactor. It proceeds as follows (Sections 9 and 10).

- It measures vectors in the Gaussian norm, so that the size of a vector and twice its divisibility have the same main terms, as in (A.5) and (A.17). What is left per vector, after the guaranteed divisibility is divided out, is an amount h: enormous, and not computed.

- Theorem A bounds each vector from above, at every conjugate: its logarithmic squared length is at most $h - m \cdot \mathrm{rem} + \mathcal{E}$. The prime-sum error rem enters here, as the difference between the divisibility guaranteed by the ramps and its average. With Hadamard's inequality this gives a covolume bound $c \le n(h - m \cdot \mathrm{rem} + \mathcal{E})$.
- Theorem C is the lower bound: for a family that nearly fills the D-dimensional target space, with covolume at most c and denominators controlled by an index δ, it gives $D(h - 3\mathcal{E}) \le c + 2\delta$. This replaces "the cofactor is at least 1". It is the hard part of the proof, and controlling δ is one of its hard inputs.
- Together, $D(h - 3\mathcal{E}) \le n(h - m \cdot \mathrm{rem} + \mathcal{E}) + 2\delta$. Since $n \le D$ and $h \ge 0$, the terms in h drop out and a bound for $m \cdot \mathrm{rem}$ is left. The cofactor is never computed: it is bounded from both sides by the same quantity.

References

[1] L. Alpöge, announcement of a Lean proof of RH and Dirichlet GRH, 20 November 2026, https://alpo.ge/remark.pdf.

[2] F. A. Behrend, On sets of integers which contain no three terms in arithmetical progression, Proc. Nat. Acad. Sci. U.S.A. 32 (1946), 331–332.

[3] A. Borel, Some finiteness properties of adele groups over number fields, Publ. Math. IHÉS 16 (1963), 5–30.

[4] A. Selberg, S. Chowla, On Epstein's zeta-function, J. reine angew. Math. 227 (1967), 86–110.

[5] M. Lerch, Další studie v oboru Malmsténovských řad, Rozpravy České Akad., Třída II, 3 (1894), no. 28, 1–61.

[6] H. Davenport, *Multiplicative Number Theory*, 3rd ed., revised by H. L. Montgomery, Graduate Texts in Mathematics 74, Springer, 2000.

[7] M. O. Rubinstein, *The Alpöge–Claude proof of the Riemann hypothesis explained for the Riemann zeta function*, Hexagon 2611.00832, version 1, 23 November 2026, https://hexagonmath.org/2611.00832v1.

[8] W. T. Gan, J. P. Hanke, J.-K. Yu, On an exact mass formula of Shimura, Duke Math. J. 107 (2001), 103–133.

[9] R. Jacobowitz, Hermitian forms over local fields, Amer. J. Math. 84 (1962), 441–465.

[10] G. Shimura, Arithmetic of unitary groups, Ann. of Math. 79 (1964), 369–409.

[11] D. R. Heath-Brown, A mean value estimate for real character sums, Acta Arith. 72 (1995), 235–275.

[12] M. Jutila, On mean values of Dirichlet polynomials with real characters, Acta Arith. 27 (1975), 191–198.

[13] A. Hurwitz, Über die Entwicklungskoeffizienten der lemniskatischen Funktionen, Math. Ann. 51 (1899), 196–226.

[14] J. A. Green, The characters of the finite general linear groups, Trans. Amer. Math. Soc. 80 (1955), 402–447.

[15] B. H. Gross, Algebraic modular forms, Israel J. Math. 113 (1999), 61–93.

[16] W. T. Gan, Y. Qiu, S. Takeda, The regularized Siegel–Weil formula (the second term identity) and the Rallis inner product formula, Invent. Math. 198 (2014), 739–831.

[17] S. S. Kudla, Seesaw dual reductive pairs, in *Automorphic Forms of Several Variables* (Katata, 1983), Progr. Math. 46, Birkhäuser, 1984, 244–268.

[18] S. Rallis, Injectivity properties of liftings associated to Weil representations, Compositio Math. 52 (1984), 139–169.

[19] C. L. Siegel, *Lectures on Advanced Analytic Number Theory*, Tata Institute Lectures on Mathematics 23, Bombay, 1961 (Chapter I, §1, Theorem 1).

[20] C. L. Siegel, Über die analytische Theorie der quadratischen Formen, Ann. of Math. 36 (1935), 527–606.

[21] A. Weil, Sur la formule de Siegel dans la théorie des groupes classiques, Acta Math. 113 (1965), 1–87.

[22] E. Landau, Über einen Satz von Tschebyschef, Math. Ann. 61 (1905), 527–550.
[23] P. Erdős, Beweis eines Satzes von Tschebyschef, Acta Litt. Sci. Szeged 5 (1932), 194–198.
[24] V. Bargmann, On a Hilbert space of analytic functions and an associated integral transform, Part I, Comm. Pure Appl. Math. 14 (1961), 187–214.
[25] G. Frobenius, L. Stickelberger, Über die Addition und Multiplication der elliptischen Functionen, J. reine angew. Math. 88 (1880), 146–184.
[26] S. C. Milne, Hankel determinants of Eisenstein series, in *Symbolic Computation, Number Theory, Special Functions, Physics and Combinatorics*, Dev. Math. 4, Kluwer, 2001, 171–188.
[27] A. Shapira, Behrend-type constructions for sets of linear equations, Acta Arith. 122 (2006), 17–33.
[28] S. Rallis, *L-functions and the Oscillator Representation*, Lecture Notes in Math. 1245, Springer, 1987.
[29] S. S. Kudla, S. Rallis, A regularized Siegel–Weil formula: the first term identity, Ann. of Math. 140 (1994), 1–80.
[30] A. Weil, Sur certains groupes d'opérateurs unitaires, Acta Math. 111 (1964), 143–211.
[31] G. Shimura, Theta functions with complex multiplication, Duke Math. J. 43 (1976), 673–696.
[32] S. W. Shin, Galois representations arising from some compact Shimura varieties, Ann. of Math. 173 (2011), 1645–1741.
[33] G. Chenevier, M. Harris, Construction of automorphic Galois representations, II, Camb. J. Math. 1 (2013), 53–73.
[34] R. Taylor, A. Wiles, Ring-theoretic properties of certain Hecke algebras, Ann. of Math. 141 (1995), 553–572.
[35] H. Hida, Congruence of cusp forms and special values of their zeta functions, Invent. Math. 63 (1981), 225–261.
[36] K. Rubin, *Euler Systems*, Ann. of Math. Stud. 147, Princeton Univ. Press, 2000.
[37] J. Coates, W. Sinnott, An analogue of Stickelberger's theorem for the higher K-groups, Invent. Math. 24 (1974), 149–161.
[38] H. L. Montgomery, R. C. Vaughan, *Multiplicative Number Theory I*, Cambridge, 2007.
[39] L. Sernau et al., shortened Lean proof of RH and Dirichlet GRH, https://github.com/lsernau/grh-lean, November 2026.
[40] J. H. Silverman, *The Arithmetic of Elliptic Curves*, 2nd ed., Springer, 2009.
[41] A. Sutherland, verification records, https://math.mit.edu/~drew/grh/ and https://math.mit.edu/~drew/shortgrh/.

Department of Pure Mathematics, University of Waterloo, Waterloo, Ontario, Canada