A guide to the shortened Lean development

From Theta Lifts to the Critical Line

The strategy of the Alpöge–Claude proof of the Generalized Riemann Hypothesis

First exposition: an annotated mathematical overview

Based on the compact and readable editions of `grh-lean-main`
November 2026

Purpose. These notes reconstruct the mathematical organization of the shortened Lean proof of RH and Dirichlet GRH. The analytic reduction, principal inequalities, and deduction of the theorem are explained explicitly. The three major arithmetic theorems are stated in a usable form, and their mechanisms are described with pointers to the formal source. Their deep component estimates are inputs to this exposition, not independent proofs supplied here.

Attribution. The underlying proof is due to Levent Alpöge with Claude; the shortened repository identifies its Lean text and documentation as written by Claude with direction from Anthropic colleagues. The present document is an expository reconstruction, not an original proof of GRH or an assertion of independent verification of the Lean development. The shortened Lean code on which this exposition is based is credited to Luke Sernau; see his repository, https://github.com/lsernau/grh-lean.

AI disclaimer. The text of this document was written by AI, ChatGPT-6.2 Sol, and has not been checked by a human.

Version. Reading edition, 26 November 2026. All module references are to the supplied snapshot unless specified otherwise.

Contents

1 What is being proved, and how? 3
1.1 The theorem and the status of these notes . . . 3
1.2 The mathematical idea in a nutshell . . . 3
1.3 The logical organization . . . 4
1.4 How to read the rest . . . 5

2 The analytic problem: a one-sided bound for weighted prime sums 5
2.1 The particular smoothing used in the proof . . . 5
2.2 Why the prime sum contains information about zeros . . . 7
2.3 An elementary form of Landau's one-sided principle . . . 8
2.4 Choosing the auxiliary prime: preventing cancellation . . . 9
2.5 The one-sided zero-free criterion . . . 10
2.6 A good second scale in a long window . . . 11
2.7 From the target bound to GRH . . . 12

3 The auxiliary arithmetic world 13
3.1 Why enlarge the setting? . . . 13
3.2 The scales, and the reason for so many parameters . . . 13
3.3 Existence of the data . . . 14
3.4 Two kinds of automorphic forms . . . 15
3.5 The theta chains and their relation . . . 16
3.6 Norms, squared covolume and index . . . 16
3.7 The common leading term and the prime remainder . . . 17

4 The upper bound: enough integrality for short forms 18
4.1 Why the divisibility of a differential operator counts primes . . . 19
4.2 The collision problem . . . 20
4.3 Elementary divisors and the cost of the repair . . . 21
4.4 Moving to the definite chain . . . 22
4.5 A uniform local estimate . . . 23
4.6 The Archimedean side and the choice of level . . . 23
4.7 Combining the estimates . . . 24

5 Enough forms: a quantitative spanning theorem 25
5.1 The statement and its role . . . 25
5.2 The row space and the good-type projection . . . 26
5.3 From columns to Poincaré series . . . 27
5.4 What the second moment must prove . . . 28
5.5 The rank deduction from the moments . . . 30
5.6 Greedy selection of translates . . . 30

6 The lower bound: why integral families cannot be too short **31**
6.1 Volumes, heights, and the product formula . . . 32
6.2 The first ratio: completing an almost spanning family . . . 33
6.3 The exchange construction . . . 34
6.4 The second ratio: factorization into single lift volumes . . . 35
6.5 Measuring an individual lift without evaluating its norm . . . 35

7 Theorem C, II: congruences and the choice of twists **37**
7.1 What is required of a twist? . . . 37
7.2 The Galois representation of a lift . . . 38
7.3 Shrinking the Selmer modules by one character . . . 39
7.4 Ribet's method and auxiliary primes for the Tate string . . . 40
7.5 Local losses at the primes above ℓ . . . 41
7.6 Summing the local estimates . . . 41

8 Assembly of the proof: why the exponent is one half **42**
8.1 Collecting the arithmetic inputs . . . 42
8.2 From A and B to an almost-basis of short forms . . . 43
8.3 Comparison with the lower bound . . . 44
8.4 Extracting the analytic remainder . . . 44
8.5 Choosing the second cutoff and the power saving . . . 45
8.6 GRH . . . 46

9 A guide to the source and to a longer mathematical treatment **46**
9.1 What should be checked first? . . . 46
9.2 How much has been established in this document? . . . 47
9.3 The chief mathematical blocks for a specialist treatment . . . 47
Primary materials . . . 48

1 What is being proved, and how?

1.1 The theorem and the status of these notes

The formal development establishes the following assertion, in a formulation which is convenient for the proof.

Theorem 1.1 (Dirichlet GRH, half-plane form)**.** *For every integer $q \geq 1$, every Dirichlet character χ modulo q, and every $s \in \mathbb{C}$ with $\Re(s) > 1/2$, one has*

$$L(s, \chi) \neq 0,$$

with the pole $s = 1$ of a principal character excluded. In particular, every zero of a Dirichlet L-function in $0 < \Re(s) < 1$ lies on the critical line. Every nontrivial zero of the Riemann zeta function also lies on that line.

The qualification "Dirichlet" is relevant: the theorem is not a claim about the zeros of every automorphic L-function or about GRH for every number field. The unitary groups and Galois representations introduced below are auxiliary objects. They are used to establish an assertion about degree-one L-functions.

The source starts with standard definitions of L-functions in Mathlib and states the final goal in `src/Statements.lean`. The theorem `New.grh` in the compact edition implies the five exported theorems of `src/Main.lean`. The supplied repository reports successful kernel checking and checks of the axioms; both versions use the standard Lean axioms `propext`, `Classical.choice` and `Quot.sound`. These are reported checks, not a new compilation performed for these notes. The source itself says that it has not yet received a complete human mathematical reading.

Our aim is distinct from checking the executable certificate. We want to know what the mathematical proof does, why the successive ingredients are introduced, and how the estimates interact. In particular, a reader should be able to see which part of a future paper requires the theory of theta correspondence, which part requires an argument in arithmetic geometry, and which part is a comparatively elementary deduction. We will state the main arithmetic theorems and selected supporting results from the formal development and give the final deduction with explicit estimates. The proofs of the highly specialized ingredients of the three arithmetic theorems are not reconstructed in their entirety here. Accordingly, the present notes are a guide to the claimed formal proof, not a self-contained conventional proof of Theorem 1.1.

1.2 The mathematical idea in a nutshell

The starting point is an upper bound on the error in a smoothed sum over nonsplit prime powers. Instead of studying the zeros of $L(s, \chi)$ directly, the argument relates a weighted contribution of primes to the p-adic divisibilities of derivatives of theta functions. It then uses the same derivatives in automorphic representations, not in isolation. On the one hand, the archimedean sizes and finite-place divisibilities of these objects produce families of vectors whose theta lifts are unusually short. On the other hand, a global determinant inequality forces sufficiently many independent integral forms to have long lifts, once their index is taken into account.

Let W be an auxiliary complex space of cusp forms, let $n = \dim_{\mathbb{C}} W > 0$, and let m be the number of theta lifts in an iterated correspondence. Write $\mathcal{R}$ for a particular weighted prime remainder, H for the common leading term in two estimates, B for an error budget, and $L = B \log b$ for a larger budget (here b is a very large auxiliary scale, not an L-function). For an almost spanning family of q independent vectors in W, write V for a bound on the

logarithm of their squared covolume after lifting, and d_{ind} for a bound on their logarithmic index. The two competing estimates have the form

$$V \le q(H - m\mathcal{R} + L),$$
$$n(H - 3L) \le V + 2d_{\mathrm{ind}}.$$

The first inequality concerns the metric on the lifts. The second concerns a height and hence incorporates information from all places. The number q is close to n, and d_{ind}/n is small on the relevant scale. Up to explicitly controlled errors, one obtains

$$H \le H - m\mathcal{R}.$$

It is the cancellation of the huge common term H, not a strong absolute bound on H, that produces a nontrivial estimate for primes. The factor $m \asymp \sqrt{x}$ is designed to give a square-root power saving in the final normalized remainder at x.

This immediately identifies the three arithmetic demands of the proof.

A. Short forms. Construct enough integral source forms whose images have short lifts, and control the denominators used in making their lifts integral for the admissible twists.

B. Enough forms. Prove that only a controlled number of translates of the image of the source is needed to span almost the entire row space. Shortness of a small subspace would not suffice for the determinant argument.

C. Long forms. Prove a lower bound on squared covolume compensated by index. The lower bound must capture exactly the same leading term H as in A. This is the difficult global arithmetic component: theta norm formulas, determinant comparison, twisting and a Selmer-theoretic congruence argument all enter.

The remarkable feature is that these three assertions have rather different mathematical origins yet fit together numerically. In a conventional exposition they should be treated as independently understandable results, connected by the common choice of fields, weights, levels, and norms.

1.3 The logical organization

Here is an informal dependency scheme; arrows mean that the assertion on the right uses the one on the left. The blocks A, B and C are stated more precisely in the following sections.

$$\text{CM periods and theta jet divisibility} \longrightarrow \text{A: short forms},$$
$$\text{Poincaré series and relative second moments} \longrightarrow \text{B: enough forms},$$
$$\text{theta determinant identities and congruences} \longrightarrow \text{C: long forms},$$

$$\text{A + B + C} \longrightarrow m\mathcal{R} \le fa(\log b)^5,$$

$$\text{main inequality + choice of second scale} \longrightarrow R_{N_0}(x) \ll x^{-1/2+\varepsilon},$$

$$\text{one-sided criterion + variation of the quadratic field} \longrightarrow \text{GRH}.$$

The last line uses a Landau-type argument. One-sided estimates are useful here: it is unnecessary to control the absolute value of the remainder for every x. The analytic step does, however, require care about cancellation of poles in the logarithmic derivatives of the relevant Dirichlet L-functions. The auxiliary quadratic field is varied precisely to address that issue.

There are two variants of the shortened proof. The readable edition has about 300,000 lines of code and the compact edition about 76,000; comments and blank lines are excluded from these approximate counts. Both use the same general strategy, but some arguments differ. In particular, we use the compact edition's window argument for choosing a second cutoff, rather than reproducing the large-cutoff prime number theorem in the readable version. For the difficult arithmetic blocks, the readable edition's documentation and index help explain the compact source. When in doubt about numerical constants, the present guide uses the compact edition's statements.

1.4 How to read the rest

Section 2 develops the smoothed prime sum and the analytic reduction. The arithmetic setup following it introduces the two automorphic spaces, the theta chains, and the error parameters. The next four sections describe A, B, and C (with the latter split into a determinant part and a congruence part). The final deduction then puts the inequalities together. A short concordance at the end identifies the source locations of the main results and describes what a full specialist account would still have to prove.

There are two levels of detail throughout. When a displayed assertion is a deep mathematical result from the Lean source, its status as an input is identified. When a conclusion follows by an elementary calculation from those inputs, we give the calculation. This distinction allows the reader to follow the logical spine without confusing an explanation of a proof mechanism with a complete proof of every result used.

2 The analytic problem: a one-sided bound for weighted prime sums

Before constructing the automorphic objects, we isolate the precise problem that they solve. The analytic reduction is useful in its own right: it separates the difficult arithmetic estimate from the comparatively short argument about zeros. In particular, the arithmetic construction need not produce a two-sided error estimate for the ordinary prime counting function.

Throughout this section $M \geq 1$ is a modulus, d is a prime with $d \equiv 7 \pmod 8$ and $(d, M) = 1$, and

$$K = \mathbb{Q}(\sqrt{-d}), \qquad \chi_K(\nu) = \left(\frac{\nu}{d}\right).$$

The discriminant of K is $-d$; quadratic reciprocity identifies the displayed character with its quadratic Dirichlet character. Write $\varphi(M)$ for Euler's totient. The dependence on d in the notation R_M below is suppressed, but it is essential that the auxiliary prime d can eventually be varied.

2.1 The particular smoothing used in the proof

Let $\Lambda(p^j) = \log p$ for prime powers, and let $\Lambda(\nu) = 0$ otherwise. Put

$$\lambda_{d,M}(\nu) = \mathbf{1}_{\nu \equiv 1 \bmod M} \Lambda(\nu)(1 - \chi_K(\nu)). \tag{1}$$

Thus $\lambda_{d,M}$ is non-negative. A prime split in K contributes zero, an inert prime contributes twice its ordinary von Mangoldt weight, and the ramified prime has factor one. The same rule applies to prime powers; for instance an even power of an inert prime has factor zero. Congruence $\nu \equiv 1 \pmod M$ is imposed on the integer ν, not on its underlying prime. In particular, the weighted sum is not literally a sum over primes alone.

For any Dirichlet character ξ, with its modulus understood, denote by

$$\widetilde{L}(s,\xi) = \begin{cases} L(s,\xi), & \xi \text{ non-principal,} \\ (s-1)L(s,\xi), & \xi \text{ principal} \end{cases}$$

the entire function obtained by removing the possible pole at 1 (the second line is interpreted by continuation). The character $\xi\chi_K$ is taken modulo Md; it is non-principal because d is a new odd prime. Define the meromorphic function

$$G_{d,M}(s) = \frac{1}{\varphi(M)} \left\{ \sum_{\xi \bmod M} \left(\frac{\widetilde{L}'}{\widetilde{L}}(s,\xi\chi_K) - \frac{\widetilde{L}'}{\widetilde{L}}(s,\xi) \right) - 1 \right\}. \tag{2}$$

Here and throughout, a sum over characters includes the principal and imprimitive characters. The constant subtracted in the prime error is

$$c_{d,M} = \operatorname{Re} G_{d,M}(1).$$

The classical non-vanishing on $\operatorname{Re} s \geq 1$ of the pole-removed Dirichlet L-functions ensures that this constant exists. In fact $G_{d,M}(1)$ is real, by conjugation of characters.

The error in the compact Lean development, called `primeError`, is exactly

$$R_M(x) = \sum_{1 \leq \nu \leq x} \lambda_{d,M}(\nu) \left(\frac{1}{\nu} - \frac{1}{x} \right) - \frac{\log x}{\varphi(M)} - c_{d,M} \qquad (x > 0). \tag{3}$$

It is important to include the constant as well as the logarithm in the normalization. Omitting it would prevent the Laplace transform used below from being regular at $s = 1$. In modulus one, the constant has a more specific expression. If γ is Euler's constant, write

$$\ell_K = \gamma - \operatorname{Re} \frac{L'}{L}(1,\chi_K).$$

Since $(s-1)\zeta(s) = 1 + \gamma(s-1) + O((s-1)^2)$,

$$c_{d,1} = -1 - \ell_K. \tag{4}$$

The number ℓ_K, called `chowlaSelbergConst` in the code, recurs in the arithmetic estimates: the appearance of the same constant in the norm of a CM theta period is part of the eventual cancellation.

The choice of smoothing in (3) is particularly convenient for the arithmetic proof. For $y > 0$ set $(y)_+ = \max(y, 0)$, and extend this notation to all real numbers. For an integer a and $0 < z \leq a$ one has the exact identity

$$\frac{1}{a} \sum_{\nu=1}^{a} \lambda_{d,M}(\nu) \left(\frac{a}{\nu} - \frac{a}{z} \right)_+ = R_M(z) + \frac{\log z}{\varphi(M)} + c_{d,M}. \tag{5}$$

Indeed the positive part vanishes for $\nu \geq z$ and gives $a(1/\nu - 1/z)$ for $\nu < z$. No asymptotic replacement is involved. The importance of this identity is that the arithmetic construction bounds valuations of jets expressed using precisely such positive parts, with integer order parameter a and scale z. The apparently unusual smoothed prime sum is therefore dictated by a divisibility calculation, not just by analytic convenience.

Our ultimate target is, for every $0 < \varepsilon < 1/100$ and every suitable d,

$$R_{N_0}(x) \leq x^{-1/2+8\varepsilon} \qquad (x \text{ sufficiently large}), \tag{6}$$

where $18480 \mid N_0$ and $d \equiv N_0/2 - 1 \pmod{N_0}$. Only the upper bound occurs in (6). The threshold for x may depend on the fixed parameters N_0, d and ε. We shall explain why such a bound is already strong enough.

Source. The definitions, the constant in modulus one and (5) are in `New/Endgame/PrimeError.lean`, declarations `lambdaClass`, `logDerivMean`, `primeError`, `re_logDerivMean_one`, `primeError_eq_sum`. The function $\widetilde{L}$ is `entireL` in `Lib/General/NumberTheory/QuadField.lean`.

2.2 Why the prime sum contains information about zeros

Set

$$E_M(v) = R_M(e^v), \qquad \Phi_M(s) = \int_0^\infty E_M(v) e^{(1-s)v}\, dv.$$

The variable s is the same complex variable as in the L-functions. The initial integral converges for $\operatorname{Re} s > 2$: for $v \geq 0$, the elementary bound $|\lambda_{d,M}(\nu)| \leq 2\nu$ gives $|E_M(v)| \leq Ce^v$, with C depending on M, d.

Proposition 2.1 (Transform and residues)**.** *The integral Φ_M continues meromorphically, with the expression*

$$\Phi_M(s) = \frac{G_{d,M}(s)/s - c_{d,M}}{s-1}. \tag{7}$$

It is holomorphic in a neighborhood of the closed half-plane $\operatorname{Re} s \geq 1$. For $0 < \operatorname{Re}\rho < 1$ its residue at ρ is

$$\operatorname{res}_{s=\rho} \Phi_M(s) = \frac{\sum_{\xi \bmod M} \left(\operatorname{ord}_\rho \widetilde{L}(\cdot,\xi) - \operatorname{ord}_\rho \widetilde{L}(\cdot,\xi\chi_K)\right)}{\varphi(M)\rho(1-\rho)}. \tag{8}$$

In $\operatorname{Re} s > 0$ any poles are at most simple.

Proof. By orthogonality of characters and the logarithmic derivative of the Euler product, the Dirichlet series of (1), absolutely convergent for $\operatorname{Re} s > 1$, is

$$\begin{aligned} D_{d,M}(s) &= \sum_{\nu \geq 1} \frac{\lambda_{d,M}(\nu)}{\nu^s} \\ &= \frac{1}{\varphi(M)} \sum_{\xi \bmod M} \left(\frac{L'}{L}(s,\xi\chi_K) - \frac{L'}{L}(s,\xi)\right) \\ &= G_{d,M}(s) + \frac{s}{\varphi(M)(s-1)}. \end{aligned}$$

In the last equality the unique principal character modulo M contributes $1/(s-1)$ besides the pole-removed logarithmic derivative, while no twisted character is principal. The additional 1 in the definition of $G_{d,M}$ has now combined with that pole term.

Absolute convergence justifies integrating the sum in (3) term by term for $\operatorname{Re} s > 2$. For an integer $\nu \geq 1$,

$$\int_{\log\nu}^\infty (\nu^{-1} - e^{-v}) e^{(1-s)v}\, dv = \nu^{-s}\left(\frac{1}{s-1} - \frac{1}{s}\right) = \frac{\nu^{-s}}{s(s-1)}.$$

The transform of v is $1/(s-1)^2$, and that of the constant function one is $1/(s-1)$. Consequently

$$\begin{aligned} \Phi_M(s) &= \frac{D_{d,M}(s)}{s(s-1)} - \frac{1}{\varphi(M)(s-1)^2} - \frac{c_{d,M}}{s-1} \\ &= \frac{G_{d,M}(s)/s - c_{d,M}}{s-1}. \end{aligned}$$

This gives the desired continuation. The apparent singularity at 1 is removable because $c_{d,M} = G_{d,M}(1)$. The classical non-vanishing in $\operatorname{Re} s \geq 1$ rules out all other poles there.

If an entire function F has a zero of order j at ρ, then we can write $F(s) = (s-\rho)^j f(s)$ locally, with $f(\rho) \neq 0$. Hence

$$\frac{F'}{F}(s) = \frac{j}{s-\rho} + \frac{f'}{f}(s).$$

This proves (8), including its sign, and shows that the poles are simple. □

The numerator in (8) is a *difference* of zero multiplicities. Thus a bound proving that Φ_M has no poles in a certain half-plane would not, by itself, show that every $L(s,\xi)$ is zero-free there: a pole might cancel between the two sums. The remedy is to choose d so that the finitely many twisted L-functions have no zero at a hypothetical offending zero, and no real zero to the right of the relevant line.

In the code the properties of E_M and its transform are packaged by the predicate `IsPolar`. The assertion `isPolar_primeError` in `New/Endgame/PrimeError.lean` establishes this predicate with the residue function (8) (zero at 1). This abstraction enables the two subsequent analytic arguments to be proved without referring to primes or automorphic forms.

2.3 An elementary form of Landau's one-sided principle

Here is the precise analytic feature that makes an upper bound sufficient. We give the argument, since it explains the sign required in the later choice of the quadratic twist.

Lemma 2.2 (Landau principle)**.** *Let $E : [0,\infty) \to \mathbb{R}$ be measurable with $|E(v)| \leq Ce^v$. Suppose its Laplace transform*

$$\Phi(s) = \int_0^\infty E(v) e^{(1-s)v}\, dv$$

has a meromorphic continuation, holomorphic near $\operatorname{Re} s \geq 1$ and with at most simple poles in $\operatorname{Re} s > 0$. Let $\beta > 0$. Assume that

$$E(v) \leq e^{(\beta-1)v} \quad (v \text{ sufficiently large})$$

and that the residue of Φ at each real $\rho > \beta$ is non-negative (or zero when the point is regular). Then Φ is holomorphic in $\operatorname{Re} s > \beta$.

Proof. Choose $A \geq 1$ large enough that

$$h(v) = Ae^{\beta v} - e^v E(v) \geq 0 \qquad (v \geq 0).$$

The eventual upper bound guarantees the inequality at large v, and the growth assumption gives it on the remaining bounded interval after enlarging A. For sufficiently large $\operatorname{Re} s$ we then have

$$\Psi(s) := \int_0^\infty h(v) e^{-sv}\, dv = \frac{A}{s-\beta} - \Phi(s).$$

Let σ_c be the real abscissa of convergence of the integral for Ψ. Because $h \geq 0$, convergence holds in the right half-plane $\operatorname{Re} s > \sigma_c$, where the integral is holomorphic; the displayed meromorphic expression agrees with it there by analytic continuation.

Suppose $\sigma_c > \beta$. If Ψ were to have a simple pole at the real point σ_c, its residue would necessarily be non-negative: for real $s > \sigma_c$, the integral is non-negative, so $(s-\sigma_c)\Psi(s) \geq 0$. But the meromorphic formula gives that the residue equals minus the residue of Φ and is therefore non-positive. This shows that the residue is zero and the singularity removable.

A Laplace transform of a non-negative function cannot be holomorphic at its finite abscissa of convergence. To see this directly, take a real $c > \sigma_c$ sufficiently close to σ_c that the Taylor series of the continued Ψ about c converges at a real $\sigma < \sigma_c$. Differentiation under the integral at c gives

$$(-1)^j \Psi^{(j)}(c) = \int_0^\infty v^j h(v) e^{-cv}\, dv \geq 0.$$

Taylor convergence and monotone convergence therefore give

$$\begin{aligned} \infty &> \sum_{j\geq 0} \frac{(-1)^j \Psi^{(j)}(c)(c-\sigma)^j}{j!} \\ &= \int_0^\infty h(v) e^{-cv} \sum_{j\geq 0} \frac{((c-\sigma)v)^j}{j!}\, dv \\ &= \int_0^\infty h(v) e^{-\sigma v}\, dv, \end{aligned}$$

contrary to the definition of σ_c. Thus $\sigma_c \leq \beta$, and the formula for Ψ implies the assertion about Φ on $\operatorname{Re} s > \beta$. □

This argument is based on positivity of an auxiliary integrand, not on a hypothesis that E changes sign or has small absolute value. It is the usual Landau principle for Dirichlet integrals with one sign, in the form formalized by `IsPolar.landau` in `Lib/Analytic/Landau/PolarLandau.lean`.

2.4 Choosing the auxiliary prime: preventing cancellation

The following zero-density input is taken from the Lean development. It is an unconditional statement about the family of quadratic twists, not a use of GRH.

Proposition 2.3 (Auxiliary zero-free twists)**.** *Fix a modulus M, a unit class $a_0 \pmod M$ such that every prime $d \equiv a_0 \pmod M$ satisfies $d \equiv 7 \pmod 8$ and $(M, d) = 1$, a number $\beta > 1/2$ and a height $Y \geq 0$. There exists such a prime d with*

$$\widetilde{L}(s, \xi\chi_K) \neq 0 \qquad (\xi \bmod M,\ \operatorname{Re} s \geq \beta,\ |\operatorname{Im} s| \leq Y).$$

Here the height Y is arbitrarily prescribed but fixed before choosing d. In particular, the proposition does not assert that the twists for any one fixed d satisfy GRH, or even that they have no zeros at arbitrary height.

The quantified zero-density estimate proved in the repository is as follows: for any fixed character $\xi \bmod M$, $\eta > 0$ and $Y \geq 0$, there are $O(X^{1-\eta/2})$ primes $d \leq X$ with $d \equiv 7 \pmod 8$, $(M, d) = 1$ and some zero ρ of $\widetilde{L}(s, \xi\chi_K)$ satisfying $\operatorname{Re}\rho \geq 1/2 + 9\eta$ and $|\operatorname{Im}\rho| \leq Y$. Choose η small enough that $1/2 + 9\eta < \beta$. Since there are finitely many ξ, the union of the corresponding bad sets still has count $O(X^{1-\eta/2})$. The reciprocal sum of primes in this union converges, by dyadic summation, whereas the reciprocal sum of all primes in the unit class a_0 diverges by the classical theorem on primes in arithmetic progressions. This proves the proposition from that estimate.

For orientation we describe the mechanism of the estimate. The proof first passes to the primitive character inducing ξ; the finite Euler factors which are thereby discarded cannot vanish in $\operatorname{Re} s > 0$. For a parity $j \in \{0, 1\}$ put $g_j(t) = t^j e^{-\pi t^2}$ and

$$\vartheta_j(c; t) = \sum_{n\geq 1} c(n) g_j(nt), \qquad I_j(c; T) = \int_{1/T}^\infty |\vartheta_j(c; t)|\, \frac{dt}{\sqrt{t}}.$$

The Mellin transform of g_j is $\Gamma_j(s) = \frac{1}{2}\pi^{-(s+j)/2}\Gamma((s+j)/2)$. One uses the Mellin formula for $\vartheta_j(\chi;t)$ and the Gaussian theta transformation for a primitive character χ. For conductors at most T^2, a three-lines argument shows that if $L(s,\chi) = 0$, $1/2 + 9\eta \le \operatorname{Re} s \le 1$ and $I_j(\chi;T) + I_j(\mu\chi;T) \le T^{2\eta}$, then

$$|\Gamma_j(s)|^2 \ll T^{-2\eta},$$

where μ is the Möbius function. The use of $\mu\chi$ provides the inverse of the original Dirichlet series to the right of 1; truncating its theta Mellin integral at $1/T$ provides the small error on the right-hand boundary. On the fixed compact region of possible zeros, $|\Gamma_j(s)|$ is bounded below by a positive constant. For large T, therefore, a zero forces the sum of tail integrals to exceed $T^{2\eta}$.

These large tail integrals are rare in the quadratic-twist family. For bounded coefficients $c(n)$ and $t \ge 1/T$, the repository proves the second-moment estimate

$$\sum_{1\le v\le X}\left|\sum_{n\ge 1} c(n)\Big(\frac{-v}{n}\Big)g_j(nt)\right|^2 \ll_\eta T^2 t^{-1-2\eta}, \qquad X \le T^2,$$

where the symbol denotes the Jacobi symbol with the indicated principal convention at 2. Expanding the square and bounding character sums by Pólya–Vinogradov accounts for the off-diagonal terms; the terms where nm or $2nm$ is a square are bounded separately. Cauchy–Schwarz followed by integrating t from $1/T$ shows that the sum of the corresponding I_j over $v \le X$ is $O(T^{2+\eta})$. This is applied to both theta series, using $\chi_K(n) = (-d/n)$ for the chosen d. Taking T^2 proportional to MX and dividing by the threshold $T^{2\eta}$ gives $O(X^{1-\eta/2})$ bad primes, as asserted. The detailed mean-value and interpolation estimates, rather than their proofs, are the inputs here.

Source. These results appear in `Lib/Analytic/QuadraticTwist/ZeroFreeTwist.lean`: `three_lines_bound_of_LFunction_eq_zero`, `sum_sq_gaussianTheta_jacobiSym_le`, `sum_thetaTailIntegral_le`, `zero_density_quadTwist` and `exists_entireL_quadTwist_ne_zero`. The repository relates this approach to Jutila's mean-value estimates for real characters.

2.5 The one-sided zero-free criterion

Proposition 2.4 (Criterion used in the proof)**.** *Let $\beta > 1/2$, and let M, a_0 satisfy the conditions of Proposition 2.3. Suppose that for* each *prime $d \equiv a_0 \pmod M$ the corresponding prime error (3) satisfies*

$$R_M(x) \le x^{\beta-1} \qquad (x \text{ sufficiently large}).$$

Then all $\widetilde{L}(s,\xi)$ for $\xi \bmod M$ are zero-free in the half-plane $\operatorname{Re} s > \beta$.

Proof. If $\widetilde{L}(\rho,\xi_0) = 0$ for some ρ to the right of β, apply Proposition 2.3 with $Y = |\operatorname{Im}\rho|$. Choose d so that no twisted $\widetilde{L}(s,\xi\chi_K)$ vanishes anywhere in the indicated half-strip. Then, at ρ, the numerator of (8) is a positive integer: the first sum includes the zero of $\widetilde{L}(\cdot,\xi_0)$, while the second sum is zero.

Furthermore at each real point $\sigma \in (\beta,1)$ the numerator of the same residue formula is non-negative and $\sigma(1-\sigma) > 0$. Thus the real residues of Φ_M to the right of β are non-negative; those at and to the right of 1 are zero by classical non-vanishing and normalization. The assumed bound for the chosen d gives $E_M(v) \le e^{(\beta-1)v}$ eventually. The Landau principle (Lemma 2.2) says that Φ_M is holomorphic on $\operatorname{Re} s > \beta$, contradicting its non-zero residue at ρ. □

The logical order matters: the same arithmetic inequality has to be proved for each admissible auxiliary prime d in the given class, while the proof of the criterion chooses d *after* assuming that an offending zero exists. The threshold in the arithmetic estimate need not be uniform in d, and the choice of d need not work for all zeroes at once. This flexibility is what avoids a zero-cancellation problem that an argument involving just one fixed quadratic field would face.

The formal declaration is `entireL_ne_zero_of_primeError_le` in `New/Endgame/OneSidedCriterion.lean`.

2.6 A good second scale in a long window

The compact edition employs a second important analytic argument. The arithmetic construction gives a bound involving two scales x, x_1 rather than one, with a combination of the two prime errors $R_{N_0}(x_1)$ and $R_1(x_1)$. We cannot simply discard those terms: their size relative to the desired saving at x would need justification. Instead, we have the freedom to vary $\log x_1$ in a sufficiently long interval. A high-moment argument produces a good choice of the second scale without requiring a pointwise prime number theorem error estimate there.

Lemma 2.5 (Window lemma)**.** *Let $E, E' : [0,\infty) \to \mathbb{R}$ be measurable with $|E(v)| + |E'(v)| \le Ce^v$. Suppose their Laplace transforms, with kernel $e^{(1-s)v}$, continue holomorphically to a neighborhood of the closed half-plane $\operatorname{Re} s \ge 1$. For every sufficiently large T and every $0 \le X \le e^T$ there exists $u \in [T, T^2]$ such that*

$$E(u) - XE'(u) \le e^{-T}.$$

The threshold for T may depend on the two functions but not on X in this range.

Proof. The closed disc of radius 3 about 4 lies in $\operatorname{Re} s \ge 1$. Choose a fixed $3 < R < 4$ such that both continued transforms are holomorphic on a neighborhood of the closed disc of radius R about 4. Such an R exists by compactness. Differentiation under the original integral near $s = 4$ gives

$$\Phi^{(j)}(4) = \int_0^\infty (-v)^j e^{-3v} E(v)\, dv.$$

The Cauchy estimates for Taylor coefficients yield, for all sufficiently large T, simultaneously for every integer $j \ge 0$,

$$\left| \int_0^\infty v^j e^{-3v} E(v)\, dv \right| \le \frac{j!T}{R^j}.$$

The same holds for E'. We can also require $|E(v)|, |E'(v)| \le Te^v$ for all $v \ge 0$.

Write $k = \lceil T^2/9 \rceil$ and suppose, for a contradiction, that $E(v) - XE'(v) > e^{-T}$ throughout $[T, T^2]$. Consider the function

$$v^k \left(e^{7T - Rv} + e^{2T - T^2 - v} + e^{-3v}(E(v) - XE'(v) - e^{-T}) \right).$$

On the window this is non-negative by hypothesis. Outside the window, the elementary growth bounds give

$$e^{-3v}|E(v) - XE'(v) - e^{-T}| \le e^{2T - 2v} \qquad (v \ge 0),$$

for sufficiently large T, uniformly in $0 \le X \le e^T$. For $0 \le v \le T$, the first exponential in the parentheses dominates this upper bound, since $R < 4$. For $v \ge T^2$, the second exponential dominates it. Thus the integrand is non-negative everywhere.

Integrate over $v \ge 0$, use the moment bounds above, and recall $\int_0^\infty v^k e^{-qv} dv = k!/q^{k+1}$ for $q > 0$. The result, after division by $k!$, is

$$0 \le \frac{e^{7T}}{R^{k+1}} + e^{2T - T^2} + \frac{T(1+X)}{R^k} - \frac{e^{-T}}{3^{k+1}}.$$

This is impossible for large T. Divide all terms by the magnitude $e^{-T}/3^{k+1}$ of the negative term. The three positive ratios are respectively

$$e^{8T}(3/R)^{k+1}, \quad e^{3T-T^2}3^{k+1}, \quad 3e^{T}T(1+X)(3/R)^{k}.$$

The first and third tend to zero as $T \to \infty$, since k is quadratic in T and $3/R < 1$; the convergence is uniform for the third ratio by $X \le e^T$. The second ratio tends to zero because $(\log 3)/9 < 1$. Their sum is less than one for all sufficiently large T, giving the desired contradiction. □

This proof is an explicit example of what a genuine shortening of the formal development looks like when translated into ordinary mathematics. The readable edition uses a pointwise prime number theorem error bound at its much larger second cutoff. The compact edition needs only regularity of the two transforms in $\operatorname{Re} s \ge 1$ and a good point in a window. It is the ability to choose that point, not an assumed strong prime number theorem, that makes the argument work.

Apply the lemma with $E(v) = R_{N_0}(e^v)$ and $E'(v) = R_1(e^v)$ for the same auxiliary prime d. The hypotheses follow from Proposition 2.1. If $T = 10^6 \log x$ and X is the positive multiplier supplied by the arithmetic data, with $X \le e^T$, the lemma permits the choice $x_1 = e^u$ with

$$\log x_1 \in [T, T^2], \qquad R_{N_0}(x_1) - XR_1(x_1) \le x^{-10^6}. \tag{9}$$

This is a one-sided bound for a particular linear combination at a particular point; it should not be confused with a uniform pointwise estimate for either error. In the final argument the change of x_1 has to preserve the admissibility of the automorphic data. The scales are chosen in the arithmetic construction precisely so that this is possible.

Source. The moment and window lemmas are `IsPolar.eventually_moment_le` and `IsPolar.exists_window_sub_mul_le` in `New/Endgame/OneSidedCriterion.lean`; `New/Main.lean` applies them to the remainder in the final bound `eventually_primeError_le_rpow`.

2.7 From the target bound to GRH

To see precisely the quantifiers needed at the end, suppose that (6) has been established by the arithmetic construction. Let $q \ge 1$ be any modulus, $\chi \bmod q$ a Dirichlet character, and ρ a hypothetical zero with $\operatorname{Re}\rho > 1/2$, with the pole of a principal character excepted. Take $N_0 = 18480q$ and the unit class $a_0 = N_0/2 - 1 \pmod{N_0}$. Since $16 \mid N_0$, $a_0 \equiv 7 \pmod 8$ and $a_0^2 \equiv 1 \pmod{N_0}$. Choose $\varepsilon > 0$ with $\varepsilon < 1/100$ and $\beta = 1/2 + 8\varepsilon < \operatorname{Re}\rho$. The estimate (6) holds for every prime $d \equiv a_0 \pmod{N_0}$ and gives $R_{N_0}(x) \le x^{\beta-1}$ eventually. Proposition 2.4 implies that all pole-removed Dirichlet L-functions modulo N_0 are non-zero at ρ.

Changing the level of χ to N_0 multiplies its Dirichlet L-function by a finite product of omitted Euler factors, each non-zero in $\operatorname{Re} s > 0$. Thus a zero of $\widetilde{L}(\cdot, \chi)$ at ρ persists at the new level. This contradicts the preceding conclusion. We have proved, subject only to the arithmetic target estimate,

$$L(s, \chi) \ne 0 \qquad (\operatorname{Re} s > 1/2),$$

apart from the pole of the principal character at $s = 1$.

For a primitive character the functional equation, whose gamma factors are non-zero and finite in the open critical strip, sends a zero at s to a zero for the dual character at $1 - s$. Thus there can be no zero with $0 < \operatorname{Re} s < 1/2$ either. For an imprimitive character the omitted Euler factors again do not vanish in the open strip. Hence every zero there has real part $1/2$. This includes the Riemann zeta function (modulus one); its standard functional equation and trivial-zero classification give the usual full statement of RH.

The rest of the exposition is concerned with the provenance of (6): how a problem about the lengths of automorphic forms and the divisibility of their theta lifts yields exactly the exponent needed for this analytic reduction.

3 The auxiliary arithmetic world

3.1 Why enlarge the setting?

A single Dirichlet L-function contains relatively little geometric structure with which to make estimates of this kind. The argument introduces a CM field, many archimedean places, families of automorphic forms on unitary groups, and a long chain of theta lifts. This enlargement is not a change in the statement to be proved. It is a way to manufacture a comparison between two quantities whose difference remembers a weighted prime sum.

Fix a positive integer N_0 divisible by 18480, and choose a prime $d \equiv N_0/2 - 1 \pmod{N_0}$. This is the class used in the analytic reduction. Put

$$K = \mathbb{Q}(\sqrt{-d}), \qquad E = K(\zeta_{N_1}),$$

where N_1 is a multiple of N_0 depending on the scale of the problem. The prime d is 7 modulo 8 and prime to N_0. With the choices in the development, E is a CM field; the set Φ of its infinite places is identified with the cyclotomic Galois group $(\mathbb{Z}/N_1)^\times$ by choosing the embedding of K. Write $f = |\Phi| = [E : \mathbb{Q}]/2$.

The rest of the data is not fixed as $x \to \infty$: we need a fresh choice of fields, groups and levels for all sufficiently large values of x. Denote a complete collection of this auxiliary data by $\mathcal{D}$. The precise record appears in `New/Data/AdmissibleData.lean`. Let $\mathcal{T} \subseteq \Phi$ consist of the places having cyclotomic class 1 modulo N_0 (the source calls them *tops*). Then

$$f = \varphi(N_0)|\mathcal{T}|.$$

A second subset $\mathcal{S} \subseteq \Phi$ consists of the *thick* places, which have class -1 modulo N_0. It must be sufficiently large and yet avoid certain short multiplicative progressions. It is not just an arbitrary half of the infinite places.

3.2 The scales, and the reason for so many parameters

The principal scale is x; the goal is to estimate $R_{N_0}(x)$. The proof uses another cutoff x_1 and a large order scale b, with

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

These inequalities are intentionally very far apart. They allow polynomial losses in the ranks or field degree to be negligible compared with the principal weights. For each fixed $0 < \varepsilon < 1/100$, the construction starts with $\log b = x^\varepsilon$ and $\log x_1 = x^{\varepsilon/6}$. The second cutoff will later be lowered within an admissible window. In the compact edition it is a real scale; no rounding of x is required.

Parameter	Meaning	Size or relation
m	Number of lifts	$\sqrt{x} \le m \le 2\sqrt{x}$
r	Rank of initial target group	$x^5 \le r \le 3x^5$
k	Rank of initial source group	$k = r + 1 - m$
t	Growth of row ranks per stage	$t = 5r$
h_2	Half-gap in the weights	$h_2 = r^8$
N_1	Cyclotomic order	$N_0 \mid N_1$
$\lvert\mathcal{T}\rvert$	Number of tops	$\asymp x^{11}$ up to log factors
$\lvert\mathcal{S}\rvert$	Number of thick places	$\ge \lvert\mathcal{T}\rvert / \log b$
a	Order of distinguished derivatives	$xb \le a < xb + N_1$
$g = g_0 g_1$	Product of two level primes	$b^{20} \le g \le b^{22}$
N	Extra level prime for source	$m \log b \le \log N \le 30m \log b$

The exact powers 10^6, 100 and 99 which occur in the source are generous margins; one cannot replace them without checking the many subsidiary inequalities, but they are not the source of the exponent $1/2$ in GRH. The significant power is the square-root length of the chain. The choice of m as a power of 2 and r as a power of 3 is compatible with the displayed intervals and serves the discrete combinatorial constructions in the proof.

The effective weights at the infinite places come in two sizes:

$$w_{\mathrm{nom}}(\sigma) = \begin{cases} a/x, & \sigma \in \mathcal{S}, \\ a/x_1, & \sigma \notin \mathcal{S}. \end{cases}$$

The integral weights actually used are rounded upwards to multiples of N_1 and shifted: if $w^-(\sigma)$ is the rounding of $w_{\mathrm{nom}}(\sigma)$ then $w_j(\sigma) = w^-(\sigma) + 2jh_2 + h_2$. Hence neighboring weights have a large gap. This separation permits quantitative archimedean and local estimates while the naive weights still record the two prime cutoffs x and x_1. It is important not to confuse the large weights with the logarithmic scale $\log b$.

3.3 Existence of the data

Proposition 3.1 (Admissible data; input from the formal proof)**.** *For fixed N_0, d as above and $0 < \varepsilon < 1/100$, for all sufficiently large x there exist admissible data $\mathcal{D}$ with the scales and size bounds just described, in particular*

$$\log b = x^{\varepsilon}, \quad \log x_1 = x^{\varepsilon/6} \ge (10^6 \log x)^2, \quad |\mathcal{S}| \ge |\mathcal{T}| / \log b.$$

Moreover the same data can be used with x_1 replaced by e^u for $10^6 \log x \le u \le (10^6 \log x)^2$, interpreting all quantities depending on x_1 with the new value.

The first part is `exists_data` in `New/Data/DataExist.lean`. The last statement follows directly by modifying the $\log x_1$ field of the record: the restrictions involving x_1 are exactly scale inequalities, each preserved by the specified replacement. The auxiliary constructions deriving groups, integral weights, and the prime N from the record are consequently made again for the new value if necessary. We do not fix an already constructed N while changing its defining weight.

Two less familiar ingredients of this proposition are worth explaining. First a Behrend-type large subset is chosen in a cyclotomic coset. An ordinary progression-free set avoids triples u, v, w with $u + w = 2v$. Here one needs a simultaneous version: after projecting into a cyclic factor, a relation $(\tau/\sigma)^j = (\rho/\sigma)^i$ for $0 < i < j \le 324s$ is allowed only in the degenerate case, where s is the cyclotomic tower exponent. The construction uses bounded digit vectors of a fixed Euclidean length, so that a nontrivial convex combination of two distinct vectors on a

sphere cannot remain on that sphere. A pigeonhole bound for the number of vectors gives sufficiently many thick places. This is the role of `exists_three_term_free`. The sparsity condition is used in the estimates for local losses below.

Second, the level primes must have strong splitting and congruence properties. In the chosen windows there are primes congruent to 1 modulo M_ω^2, where $M_\omega = (N_0 d)^{s+6}$. The development proves prime existence in logarithmic windows of length $\sqrt{\log b}$. It chooses two distinct primes g_0, g_1 in the range $[b^{10}, b^{11}]$ whose appropriate split prime ideals in K have principal product. To see the class-group mechanism, take more pairs of conjugate prime ideals than the class number, so that two first ideals represent the same class. The product of one first ideal and the second ideal from the other pair is then principal. An element π_g generating it satisfies $\pi_g \overline{\pi_g} = g$. One member of each conjugate pair of places above g is thereby distinguished. Local determinant and representation-theoretic constructions will be performed at those chosen places.

3.4 Two kinds of automorphic forms

If V is a nondegenerate Hermitian space over E with respect to the involution of E/E^+, $U(V)$ denotes its unitary group over the totally real field E^+. Automorphic forms on such a group are functions (with the specified transformation law at infinity) on an adelic quotient. For a totally definite Hermitian space, the group is compact at infinity, and spaces of algebraic modular forms at a fixed level can be described by values in an algebraic representation on finitely many double cosets. In particular their coefficient integrality and Hermitian pairings can be studied by actual lattice calculations. For the indefinite groups, one instead uses cusp forms of prescribed discrete-series type at infinity. The source imposes suitable finite level and controls their behavior in the cusp.

The proof uses both types. Here are the main spaces; indices on Hermitian spaces should not be confused with indices of vectors.

- The source Hermitian spaces X_j are definite of rank $k + j(t+1)$, $0 \le j \le m$. Let C denote the space of algebraic modular forms on $U(X_0)$ of weight $(a, 0, \dots, 0)$ at tops and zero at the other infinite places, with the prescribed level at g and N. Let $C_{\mathrm{ram}} \subset C$ be the subspace of forms with vanishing averages under the prescribed local scalings at the places of N. These are the "ramified columns."

- The row spaces are forms on Y_j of rank $r + jt$. The starting Y_0 is indefinite, and Y_m is definite. Let W be the selected subspace of forms on $U(Y_0)$, at its prescribed archimedean type and at level g. (We use the letter W only for this vector space; the source also uses W_j for auxiliary rank-r Hermitian spaces.) The spaces selected by a finite-group projector are called good forms.

- For the determinant computation there are also single-lift auxiliary pairs of Hermitian spaces W_j, V_j of ranks $r, r+t$. These are used to replace the norm calculation for a whole chain by the calculations associated to individual lifts and their comparisons.

At the chosen finite places v above g, the principal-level form spaces carry an action of the finite group

$$\mathcal{G} = \prod_v \mathrm{GL}_r(\mathbb{F}_v).$$

The good projector e_0 is the sum of specified idempotents in $\mathbb{C}[\mathcal{G}]$. The good labels are defined through squarefree polynomials over the residue fields, with exclusions at roots of unity and primitive-polynomial conditions. This is more than a harmless way to cut down the dimension: it is designed to guarantee local representation types favorable both to the filling argument

and to the Galois-theoretic analysis of congruences. The precise conditions are discussed in the subsequent sections. We write $n = \dim W$ and $c = \dim C$; neither of these is a Hermitian rank.

3.5 The theta chains and their relation

Let θ be a weight-zero automorphic character of $U(1)$, with tame conductor at a set of places disjoint from its complex conjugate and prime to gN. The successive powers through θ^m must retain the required ramification at that conductor (the trivial character is also allowed). A theta lift is a linear transformation defined by pairing an automorphic form on one group with a theta kernel associated to a Weil representation on a dual pair. Schematic notation for a lift is

$$(\mathcal{L}F)(h) = \int_{[U(V)]} F(u)\,\overline{\vartheta(u,h)}\,du,$$

where the precise conjugations, kernels, measures and characters depend on the convention. The source fixes these choices by its definitions of the spaces and lifts; this schematic integral is a description of the idea, not a replacement for those definitions. For definite groups, such integrals reduce to sums over finitely many classes.

Denote by Θ_θ the composite lift along the row chain beginning in W. The column chain begins in C, runs along the definite X_j, and afterwards restricts from X_m to Y_m. The ranks have been chosen so that

$$\dim_E X_m = k + m(t+1) = r + mt + 1 = \dim_E Y_m + 1.$$

The one-dimensional complement is used in the branching and level normalization. The m extra rank increments on the two sides make the iterated seesaw identity possible. Informally the relationship is

$$\begin{array}{ccccc} C_{\mathrm{ram}} & \longrightarrow & \text{definite chain} & \longrightarrow & X_m\text{-forms} \\ \downarrow P & & & & \downarrow \text{ restriction} \\ \text{initial row forms} & \longrightarrow & \text{row chain} & \longrightarrow & Y_m\text{-forms.} \end{array}$$

Here P is a Poincaré series constructed from a source form by Bessel induction. The actual identity involves a finite combination of definite chains, nonzero normalization factors and finite-place restrictions, not a literal equality between arbitrary individual theta kernels. The projected translates P_s in the upper bound, for $s \in \mathcal{G}$, are obtained from the Poincaré map by translating and projecting; we include the fixed normalizing factors from `projectedPoincare` in this notation.

The two rows of the diagram serve different purposes. The definite chain gives access to coefficient lattices and the sizes and prime divisibilities of derivatives of theta functions. The indefinite row chain gives a large automorphic space, a spanning theorem for Poincaré series, and a determinant calculation. Neither side alone yields the main inequality.

3.6 Norms, squared covolume and index

For each field automorphism σ of $\mathbb{C}$, the target definite space has a conjugated coefficient pairing, with squared norm denoted $\|\cdot\|_\sigma^2$. We mean the pairing obtained by conjugating the coefficient data as specified in `Lib/Adelic/CuspForms/LevelForms.lean`, not the absolute value of an arbitrary conjugate of a real analytic integral. For independent vectors $y_1, \dots, y_q \in W$ define

$$\mathsf{G}_\sigma(y) = \left(\langle \Theta_1 y_i, \Theta_1 y_j\rangle_\sigma\right)_{i,j}, \qquad \mathrm{covol}_\sigma^2(y) = |\det \mathsf{G}_\sigma(y)|.$$

The determinant may be zero for some family before an injectivity statement is available; the upper estimate does not presuppose it is positive. We call V a logarithmic upper bound if $\mathrm{covol}_\sigma^2(y) \le e^V$ for all σ. By the Gram determinant inequality,

$$\mathrm{covol}_\sigma^2(y) \le \prod_{i=1}^{q} \|\Theta_1 y_i\|_\sigma^2.$$

Thus bounds on individual lifts imply a bound on the determinant of their span. The source's definition uses the notation `covolumeSq`.

At a rational prime ℓ, transfer the coefficient model by a field isomorphism $\iota : \mathbb{C} \simeq \overline{\mathbb{Q}}_\ell$. The integral forms for a twist are defined with a lattice in the corresponding weight representation. To measure the integrality of a family, use an integral $q \times q$ change-of-generators matrix U such that the twisted lifts of the resulting q vectors have integral coefficients. If its determinant has ℓ-adic absolute value at least e^{-e_ℓ}, the logarithmic index cost is at most e_ℓ. This is the familiar relationship between determinant and index for lattices over a discretely valued field, stated in the valuation-ring language that is needed in the source.

The assertion $\mathrm{IndexLe}(y, d_{\mathrm{ind}})$ means that one can choose numbers e_ℓ with sum over any finite set of primes at most d_{ind}, for which such a matrix exists at every ι and every admissible twist whose conductor is prime to ℓ. The matrices may depend on the prime, embedding and twist, but the budget is uniform. This is the definition in `New/Endgame/ShortAndEnough.lean`. The square in the Gram determinant explains the coefficient $2d_{\mathrm{ind}}$ in the lower bound: scaling a family changes a squared determinant by two copies of the cost of its change of basis.

3.7 The common leading term and the prime remainder

Here we introduce the quantities that occur in all three theorems. Let $\lambda_{d,1}(\nu) = \Lambda(\nu)(1-\chi_K(\nu))$ be the weight without a congruence restriction, and let s_ν be the place represented by $-\nu$ modulo N_1 when ν is invertible there (with the source's convention of the identity place otherwise). Define the finite prime sum at $\sigma \in \Phi$ by

$$A(y,\sigma) = \sum_{1 \le \nu \le a} \lambda_{d,1}(\nu) \left(\frac{y}{\nu} - w_{\mathrm{nom}}(s_\nu \sigma) \right)_+ .$$

Write $\kappa_K = \gamma - \Re(L'(1,\chi_K)/L(1,\chi_K))$ for the Chowla–Selberg constant used in the proof, and

$$\Omega = \frac{2\pi}{\mathrm{rd}_E} e^{\kappa_K}, \qquad J(j) = \log(j!) - j\log(\Omega/g), \qquad \overline{w} = \frac{1}{f}\sum_{\sigma \in \Phi} \log w_{\mathrm{nom}}(\sigma).$$

The quantity $J(j)$ is the archimedean cost of a derivative of order j in the normalization of the proof. Its dependence on the period of the CM curve is the key to matching archimedean terms with the weighted prime sum. An expected main term for $A(a,\sigma)$, averaged appropriately, is

$$A_0 = a(\log a - 1 - \kappa_K - \overline{w}).$$

The exact quantities in the inequalities are

$$\begin{aligned} \mathcal{R} &= \sum_{\sigma \in \mathcal{T}} \big(A(a,\sigma) - A_0\big), \\ \mathcal{U} &= \sum_{\sigma \in \Phi} \big(J(w_0(\sigma)) - A(w_{\mathrm{nom}}(\sigma),\sigma)\big), \\ H &= mk\mathcal{U}, \qquad B = fa(\log b)^3, \qquad L = B\log b = fa(\log b)^4. \end{aligned}$$

These definitions are `Data.primeSum`, `Data.jetSize`, `Data.dirCost`, `Data.mainHeight`, `Data.mainAsk`, and `Data.remainder`. The quantity $\mathcal{U}$ represents the net cost of an ordinary differentiation direction, obtained by subtracting the finite-place divisibility gain from the archimedean derivative size. Both the short-form argument and the determinant argument are calibrated to this quantity. It is not necessary to determine its precise sign or asymptotic value.

The prime remainder has the following exact interpretation, which will be important in the final deduction:

$$\mathcal{R} = a\{|\mathcal{S}|(R_{N_0}(x) - R_{N_0}(x_1)) + |\mathcal{T}|R_1(x_1)\}. \tag{10}$$

This is `Data.remainder_eq`. We explain its mechanism in Section 8: the two weights replace the finite prime sums by R_{N_0} at two scales and a remainder R_1 without a congruence condition. The role of the analytic choice of x_1 is to control the unwanted second-cutoff expression. Notice that the normalized error $R_M(z)$ of the preceding section is not the same as the unnormalized, large remainder $\mathcal{R}$.

4 The upper bound: enough integrality for short forms

The first arithmetic problem is to produce vectors of W whose lifts are short at the Archimedean places and almost integral at the finite places. There is a substantial difference between arranging either of these conditions separately and arranging both simultaneously. We shall see that the prime sum occurs as the amount by which one can divide out the theta lift while retaining control of its denominators.

Throughout this section $C^{\mathrm{ram}} \subset C$ denotes the subspace of columns ramified at the auxiliary prime N, $C^{\mathrm{ram}}_{\mathrm{int}}$ its everywhere-integral elements, and G the finite group indexing the translates of the projected Poincaré map $P_s : C^{\mathrm{ram}} \longrightarrow W$. The subscript 1 on Θ_1 denotes the trivial twist. Use the squared norm of a lift at a conjugate σ of the coefficient data, denoted by $\|\Theta_1 w\|^2_\sigma$. The finite-place condition on a family is the index bound of the setup: at each prime p, after passing to a p-adic place and for every twist of conductor prime to p, a matrix over the valuation ring makes all the lifted family integral, with a controlled logarithmic determinant loss.

Theorem 4.1 (A: short forms)**.** *There exists a positive rational number q such that the maps $Q_s = qP_s$, $s \in G$, enjoy the following properties.*

1. *There is a set $S \subset C^{\mathrm{ram}}_{\mathrm{int}}$ spanning C^{ram} over $\mathbb{C}$ for which*

$$\|\Theta_1 Q_s u\|^2_\sigma \le \exp(H - m\mathcal{R} + L) \qquad (u \in S,\ s \in G,\ \sigma \in \mathrm{Aut}(\mathbb{C})).$$

2. *For each finite set $J \subset G$, every finite family y in $\bigcup_{s\in J} Q_s(C^{\mathrm{ram}}_{\mathrm{int}})$ has index at most*

$$|J|\ \dim_{\mathbb{C}} C \cdot B.$$

This is the statement of `Data.short_forms`, in `New/Endgame/ShortAndEnough.lean`, with the names of spaces changed to ours. The fact that both assertions use the *same* rational number q is essential. Shortness alone could easily be obtained by multiplying vectors by an arbitrarily small rational number; the index bound prevents this trivial solution.

The rest of the section explains the proof of Theorem 4.1. We state the difficult arithmetic inputs, explain the mechanisms behind them, and give the calculation combining them. It is not a replacement for the detailed proofs of the local theta, seesaw and Hecke estimates.

4.1 Why the divisibility of a differential operator counts primes

Fix the quadratic CM field K and the cyclotomic CM extension E of the setup. Let Φ be its set of infinite places. Write $\mathcal{T} \subset \Phi$ for the tops (the places in the class 1 modulo N_0), and $w(\tau)$ for the nominal weight at a place: $w(\tau) = a/x$ on thick places and $w(\tau) = a/x_1$ elsewhere. If p is a rational prime and χ_K is the quadratic character of K, define, for $0 \le y \le a$,

$$A_p(y,\tau) = \frac{1-\chi_K(p)}{2} \sum_{\substack{j\ge 1 \\ j \text{ odd or } p=d}} \left(\frac{y}{p^j} - w(-p^j\tau)\right)_+, \tag{11}$$

where $u_+ = \max(u,0)$ and $-p^j\tau$ uses the place action of the construction. If the class is undefined because the integer is not a unit modulo the conductor, the place action uses the identity place as its default. The character factor in (11) annuls the contribution at split primes. Inert primes contribute odd powers and the ramified prime d contributes all positive powers with factor $1/2$.

The local ask is exactly the contribution of p to the global prime sum $\mathrm{Ask}(y,\tau)$, not merely an approximation:

$$\mathrm{Ask}(y,\tau) = 2\sum_p A_p(y,\tau)\log p. \tag{12}$$

Indeed the von Mangoldt function is supported on prime powers, and $1-\chi_K(p^j)$ is 2 for odd powers of an inert prime and 0 for the even powers. The equality at the ramified prime follows from $1-\chi_K(d^j) = 1$. The sum is supported below x_1, because $w \ge a/x_1$. In particular, defining

$$F_p = \sum_{\tau\in\Phi} A_p(w(\tau),\tau), \qquad g_p = m\left(\sum_{\tau\in\mathcal{T}} A_p(a,\tau) + kF_p\right),$$

the desired divisibility of the chain is g_p. These definitions are `localAsk`, `forcedPart`, `gain` in `New/Curve/LiveSum.lean`; the identity (12) is `primeSum_eq_sum_localAsk`.

Why should $A_p(y,\tau)$ have anything to do with a theta kernel? For an analytic theta function, the jet along a polynomial Q is

$$J_Q(\vartheta) = Q(\partial)\vartheta|_0\,.$$

If Q is a single monomial, this is a Taylor coefficient multiplied by factorials; for instance $J_{z^h}(\vartheta) = h![z^h]\vartheta(z)$. The factor $h!$ already has a prime-power valuation $v_p(h!) = \sum_{j\ge1}\lfloor h/p^j\rfloor$. The relevant theta functions are not ordinary integral power series in the differential coordinate z; one needs to compare this coordinate with the coordinate of the CM formal group. This is the source of additional scaling costs and interactions between different infinite places.

In the construction the theta kernel uses products of functions of this kind and jets along products of minors (Schur-module matrix coefficients). At a non-split prime of K the change to the formal-group coordinate is controlled by the local CM theory. For instance, the regularized formal parameter S of the compact edition satisfies an equation of the shape

$$S - S^{p^e} = X + \pi E_0(X), \qquad S(0) = 0, quad|\pi|_p < 1,$$

with S, E_0 integral and e the relevant local exponent. The equation is stated in `CMLattice.exists_formal_parameter` in `New/Curve/FormalParameter.lean`. After reduction modulo the small parameter, its solution is obtained by successive applications of Frobenius. It is precisely this phenomenon, rather than an explicit formula for the primes, that produces the prime powers in the valuation estimate.

For inert p, the normal-form theorem for suitable products of theta functions has an especially revealing statement. With ϖ the formal-coordinate scaling, it expresses the rescaled theta series as

$$\vartheta(\varpi z) = \exp(\kappa\Psi(z))V(z), \qquad |\kappa|_p = p^{1-1/(p-1)},$$

where Ψ, V have integral coefficients and $\Psi(0) = 0$. In the reduction of the derivative of Ψ with respect to a variable at place τ, the possible nonzero terms are precisely Frobenius powers of the variables at partner places. More exactly, the coefficient with support a single variable at ν and degree p^j occurs when j is odd and $-p^j\nu = \tau$ (the extra coordinate index is unchanged). This is part of `thetaProducts_normal_form` in `New/Curve/ThetaProducts.lean`; it concerns the specified class of theta products, not arbitrary theta functions.

Thus the gain obtained from a high order of differentiation at one place must be corrected for the order of differentiation at its partner. A partner order w can consume up to w units of the gain predicted by y/p^j. When it is smaller than y/p^j the surplus remains as a divisibility of the jet; when it is larger the surplus vanishes. This describes the positive part in (11). At the matrix level one must account for boxes of weights and pairings of vectors in different columns. One chooses integral tilting matrices that annihilate the relevant Frobenius pairings modulo the small parameter, charging their determinant losses. The result, in `norm_schurJet_le` of `New/Curve/JetDivisibility.lean`, bounds every matrix entry of the Schur jet. The lower bound on valuation includes three kinds of terms: the factorial gain after change of coordinate; the cost of surviving couplings; and explicit truncation and determinant losses. The elementary identity

$$\frac{1-\chi_K(p)}{2} \sum_{\substack{j\ge 1\\ j \text{ odd or } p=d}} p^{-j} = \frac{1}{p-1} - \rho_p$$

relates the factorial's linear term to the formal-group scale $|\varpi|_p = p^{-\rho_p}$.

The integral vacuum is constructed using tight frames of CM lattice vectors, finite translates of a normalized vacuum, and controlled-denominator combinations of these translates. In particular the above matrix estimates apply to the actual kernel data, not just to a freely chosen power series. The compact files `Curve/IntegralVacuum`, `Curve/FrameSections` and `Curve/ConjugatesOfJets` contain the passage from formal power series to normalized theta kernels and their conjugates. The sum of the jet exponents along m successive kernels is the first approximation to g_p; this summation is carried out in `Columns/JetExponentBound`. We discuss next why it does not suffice to apply this estimate successively without modification.

4.2 The collision problem

The extra-long rows of length a occur at the tops. A coupling between two tops can therefore subtract an amount much larger than the ordinary nominal weight. Call a pair (τ, j) a collision at p if τ and $-p^j\tau$ are both tops and j is among the exponents in (11). Denote by C_p the sum of a/p^j over all such pairs. For an untreated chain the direct jet argument gives, up to the precise gauge normalization, the lower bound

$$v_p(\text{chain}) \ge g_p - \ell_p - mC_p$$

on integral columns. This is `gaugeLe_collisionSum` of `New/Columns/LocalStatement.lean`. Here the basic loss is

$$\ell_p = \frac{fa(\log b)^2}{p} \begin{cases} m, & p \equiv \pm 1 \pmod m, \\ 1, & \text{otherwise.} \end{cases}$$

The factor m is permissible for the sparse congruence classes $p \equiv \pm 1 \pmod m$, called short primes in the code. But it is not permissible uniformly in p: the factor $1/m$ in the final argument is the source of the square-root saving. Losing a factor m at every prime would not yield GRH.

For inert p, put $h_p = \lceil \log_p m \rceil$ and let o_p be the order of the class of $-p$ modulo N_0. The particularly delicate case is when the prime is not short and o_p is odd with $o_p \le h_p$. Such primes

are called served primes (`IsServed`). Large-exponent collisions cost little by the geometric decay p^{-j}. At a served prime one marks suitable finite places of E above p. These places split over the maximal totally real subfield of E, even though p is inert in K. One selects a set Λ_p so that the troublesome couplings are covered, with the balance required by the Hecke calculation. The existence and balance assertions are in `Columns/JetExponentBound`. The construction marks places of the extension E rather than changing the quadratic field.

At each marked place λ, let $Y_\lambda = p^{A_\lambda} T_{\lambda,h_p}$ be the scaled spherical Hecke operator on the column lattice. The exponent A_λ is chosen in terms of the weight; the scaling is crucial for integrality. Let

$$Y_p = \prod_{\lambda \in \Lambda_p} Y_\lambda, \qquad j_p = \lceil 2mfa/p \rceil .$$

The Hecke-transfer identities move these operators from the source of a chain to its individual kernels. In the spherical Hecke polynomial of a theta lift, the source polynomial is multiplied by explicit factors corresponding to the added dimensions. Extracting a coefficient uses only controlled denominators. In the compact edition this is `exists_hecke_transfer` and `exists_move_markOperator`.

The local divisibility estimate on forms in $Y_p^m M_p$ is $g_p - 2\ell_p$, where M_p denotes the integral column lattice at the chosen p-adic place. On $p^{j_p} M_p$ there is also enough divisibility, simply because the compensating scalar dominates the raw collision loss. This proves the desired gain on the sum

$$Y_p^m M_p + p^{j_p} M_p.$$

It still does not prove a useful global theorem unless this sum has small colength. A scalar operator which was very divisible would have small images modulo p^{j_p} and make the procedure worthless. This is the second hard aspect of repairing collisions.

4.3 Elementary divisors and the cost of the repair

To understand the estimate needed here, imagine a free integral lattice M over a discrete valuation ring with uniformizer p, and an endomorphism Y with elementary divisor exponents $e_1, \dots, e_s$. The colength of $YM + p^h M$ in M is

$$\sum_{i=1}^{s} \min(e_i, h);$$

we put $e_i = +\infty$ for a zero diagonal entry. It is therefore enough to control truncated sums of elementary divisors, not each slope individually. The formalized version is phrased as `HasType` and `IndexDvd`; the latter supplies a matrix over the valuation ring, with an explicit determinant divisibility bound, giving the required elements of the sublattice. This phrasing does not require a discretely valued coefficient field.

A key result for marked primes in the compact edition is the following bound (`IsMarked.hasType_markOperator`, `New/Columns/MarkedPrimeType.lean`). The operator Y_λ preserves M_p, and for every integer $h \geq 0$ the module $Y_\lambda M_p + p^h M_p$ has index dividing p^c, with

$$c \leq \left(\frac{128 A_\lambda \log b}{mp} + \frac{h}{x^{14}} + x^{40} \right) \dim_{\mathbb{C}} C.$$

The product over the marked places and the m steps can then be estimated uniformly: `hasType_prod_markOperator_pow` bounds the index exponent of $Y_p^m M_p + p^{j_p} M_p$ by $3\ell_p \dim C$.

The proof of this elementary-divisor theorem is a separate piece of arithmetic representation theory, not a formal consequence of the spherical Hecke relation. The compact edition implements it in `New/TwistedHecke`. Here are the principal ideas.

- The local Hecke correspondence is expressed by matrices on algebraic modular forms. In its weight representation the different infinite-place factors are linked into a cycle by Frobenius and, at the appropriate positions, inverse transpose. The Hecke matrix is a sum over parahoric cosets, with weight-depth powers of p.

- One studies the low-valuation terms of polynomials under changes of frame in these factors. A leading term minimizing a Gauss weight in all frames satisfies linear constraints comparing degrees in successive factors. The root-counting mechanism is an elementary multiplicity bound usually called Stepanov's count: a nonzero polynomial of degree D_0 vanishing to order at least e at each of q_0 distinct residue classes satisfies $eq_0 \le D_0$ when the corresponding linear factors are pairwise comaximal. One applies this after varying a Levi matrix entry and using Frobenius in the other factors. In `TwistedHecke/RootLemma` this is the origin of the Frobenius inequalities; the one-variable root bound alone is not the elementary-divisor theorem.

- The allowed monomial exponents occupy a cyclic interlacing cone. Counting its lattice points, together with the number of flags, proves that only a small proportion of the entire weight space can be very deep in every frame. The dimension estimate must include the number of parahoric cosets, not just compare the dimensions of two unrestricted polynomial representations.

- Finally, class means of generic Levi matrices yield an approximate projector onto the part of the space obeying the appropriate inequalities. Integral pairings and this projector allow one to bound the elementary divisors, including those of the complementary part modulo p^h. The resulting statement is `hasType_twisted_hecke` in `TwistedHecke/ElementaryDivisors`.

These steps explain why the local calculation is a dimension-weighted estimate. It controls the total price of modifying the lattice; it is not an assertion of a small bound for the valuation of every Hecke eigenvalue.

4.4 Moving to the definite chain

The original objects whose dimensions and volumes enter the proof are forms on the row chain, beginning on an indefinite unitary group. Direct integral estimates there are not the most convenient. The way to compare these to the CM theta jets is the iterated seesaw identity.

Recall the definite spaces X_j of rank $k + j(t+1)$ and the row spaces Y_j of rank $r + jt$. Denote by T_θ^Ψ the definite chain of m theta lifts with finite datum Ψ followed by restriction to the final row group. The compact formulation of the seesaw theorem is as follows: there is a nonzero scalar c independent of the column and twist such that for every θ there is an admissible good datum Ψ_θ with

$$\Theta_\theta(P(cu)) = T_\theta^{\Psi_\theta} u \qquad (u \in C^{\mathrm{ram}}).$$

Here P is the Poincaré map before the finite translate and projection. This is `Data.seesaw` in `New/Seesaw/SeesawTheorem.lean`. The scalar c is incorporated into P_s, so does not subsequently require a separate height estimate.

The identity is obtained by successive unfolding of a Bessel-type period. At each stage integration against a product of theta kernels can be rewritten with the integration on another member of the seesaw of unitary groups. The final unfolding expands the definite chain as a finite combination of chains with translated theta data. One must track finite-level cuts and the multiplicities of the orbits that occur: their reciprocals are allowed only with controlled

denominators, powers supported by the conductor of the twist. Without that last feature, an equality of complex-valued periods would not transfer the integral estimates at primes away from the twist. In particular, the construction supplies good data simultaneously compatible with the preceding estimates but depending on the twist. It does not simply assert pointwise nonvanishing for every column.

This step explains why the auxiliary source space C has a different and much simpler description from W, yet can be used to construct vectors in the latter. The left-hand side is the quantity with which we will apply a covolume argument, and the right-hand side is amenable to a sum over classes with explicit CM theta data.

4.5 A uniform local estimate

The preceding ingredients yield the precise local estimate used below. Put

$$\epsilon_p = \begin{cases} 3\ell_p, & p < x_1,\ \chi_K(p) \neq 1, \\ 4mfb, & p = N, \\ 0, & \text{otherwise}, \end{cases}$$

for primes p, and set $g_p = 0$ outside the small non-split primes. The level prime N is larger than x_1 and contributes a loss but not a gain. At N a separate ramified-character calculation trades denominators of the theta data for a Gauss-sum-weighted column operator, whose image again has controlled index (`Columns/TwistedHeckeType`).

Proposition 4.2 (Local divisibility and index)**.** *For each prime p and each choice of p-adic place, and every integer $\beta < g_p - \epsilon_p + 1$, there is a submodule $M_p' \subset M_p$ of index dividing p^c with $c \leq \epsilon_p \dim C$ such that for every twist θ prime to p and every good datum Ψ the form $p^{-\beta} T_\theta^\Psi u$ is integral whenever $u \in M_p' \cap C^{\mathrm{ram}}$. Moreover*

$$\sum_p \epsilon_p \log p \leq B.$$

The local assertion is exactly the content (under the present notations) of `Data.exists_indexDvd_Bmu_divisible` in `Columns/LocalStatement`. Notice the quantifier over twists *after* the choice of submodule. Index dividing p^c is an estimate on determinants; no claim of equality with a numerical cardinality is made over a general valuation ring. The last bound is `sum_allowedLoss_mul_log_le` in `Columns/SizeEstimates`. It is a scale calculation: ordinary non-split primes cost essentially $fa(\log b)^2 \sum_{p<x_1} (\log p)/p$, bounded even by comparison with the harmonic sum. For a short prime $p = um \pm 1$ the additional factor m is canceled by the spacing of that progression, giving the same order of upper bound. The loss at N has logarithmic cost $4mfb \log N$. With $a \geq xb$, $m \asymp \sqrt{x}$, $\log N \ll m \log b$ and $(\log x_1)^6 \leq \log b$ there is ample room in $fa(\log b)^3$. The exact constants and numerical slack, important for the theorem as stated, are verified in the cited scale estimate.

4.6 The Archimedean side and the choice of level

For the trivial twist one also has a size bound at every complex conjugate. Let

$$\Omega = \exp(\ell_{\mathrm{CS}}) \frac{2\pi}{\mathrm{rd}_E}, \qquad J(j) = \log(j!) - j \log(\Omega/g),$$

where ℓ_{CS} is the constant from the Chowla–Selberg normalization, $g = g_0 g_1$, and rd_E is the root discriminant. These are `bigOmega` and `jetSize`. Define the direction cost and the preliminary

height per row by

$$\mathcal{U} = \sum_{\tau\in\Phi} \big(J(w_0(\tau)) - \mathrm{Ask}(w(\tau),\tau)\big),$$
$$h_0 = m\sum_{\tau\in\mathcal{T}} \big(J(a) - \mathrm{Ask}(a,\tau)\big) + H - |\mathcal{T}|a\log N,$$

where w_0 is the actual integral weight at stage zero, close to the nominal weight w. The Archimedean estimate for the definite chain (`weylNormSq_Bmu_le`, `Columns/LocalStatement`) says that if the squared Weyl norm of every value of u at the conjugate σ is at most e^{η}, the corresponding bound for $T_1^{\Psi}u$ is

$$\exp\left(\eta + h_0 + 2\sum_p g_p \log p + B\right). \tag{13}$$

This takes account of the sizes of conjugates of individual theta jets, dimension factors, the number of terms of the class sums, and the controlled denominators of the datum. The normalization is arranged so that the translated projected Poincaré maps transfer the required size bound.

Why does one get $H - m\mathcal{R}$ rather than merely H in the Archimedean inequality? Introduce the mean weight logarithm $\overline{w} = f^{-1}\sum_{\tau\in\Phi}\log w(\tau)$ and put $G_0 = \log(g\,\mathrm{rd}_E/(2\pi))$. The expected prime sum at a top is

$$M_0 = a(\log a - 1 - \ell_{\mathrm{CS}} - \overline{w}), \qquad \mathcal{R} = \sum_{\tau\in\mathcal{T}} (\mathrm{Ask}(a,\tau) - M_0).$$

Stirling's bound gives

$$J(a) - M_0 \le a(\overline{w} + G_0) + 1 + \tfrac{1}{2}\log a.$$

The level is chosen at a prime whose logarithm is close to $m(\overline{w}+G_0) - G_0 - \frac{15}{2}\log b$. Consequently

$$\begin{aligned} h_0 &= H - m\mathcal{R} + |\mathcal{T}|\{m(J(a) - M_0) - a\log N\} \\ &\le H - m\mathcal{R} + B. \end{aligned}$$

The cancellation of the large term $m|\mathcal{T}|a(\overline{w}+G_0)$ is why the auxiliary prime cannot be chosen independently of the weight data. The formula and inequality are `heightPerRow_le` in `New/Endgame/ShortAndEnough`. Note that the term H has not disappeared yet; the lower bound of Theorem C must supply the same direction cost.

4.7 Combining the estimates

We finish with the short-form calculation itself. First, the integral column space has sufficiently many independently short elements. The integral structures at finite places and the algebraic modular forms on the definite group reduce this problem to choosing generators at finitely many class representatives with uniformly bounded conjugates. Deep congruence levels make their automorphism groups trivial; applying the ramifying operation selects the subspace of ramified columns with controlled additional size. The result is `span_short_ramifiedColumns` (in `Columns/ChainValuation`): the ramified columns are spanned by everywhere-integral u whose squared Weyl norms at all conjugates and values are at most e^B. Choose S to be this set.

Let

$$e_p = \lfloor \epsilon_p - g_p \rfloor, \qquad q = \prod_p p^{e_p} > 0.$$

There are only finitely many nontrivial factors, and $\beta = -e_p = \lceil g_p - \epsilon_p \rceil$ satisfies the hypothesis of Proposition 4.2. Thus q rescales the lift to one that is integral at p on the local submodule M_p'; all other factors of q are p-adic units. At the Archimedean places the squared norm bound for the rescaled lift of $u \in S$ is, using (13) and the seesaw,

$$\begin{aligned}\log \|\Theta_1 Q_s u\|_\sigma^2 &\le 2\log q + h_0 + 2\sum_p g_p \log p + 2B \\ &\le h_0 + 2B + 2\sum_p \epsilon_p \log p \le h_0 + 4B \le H - m\mathcal{R} + L.\end{aligned}$$

If a norm is zero, the corresponding norm inequality is immediate. The last estimate uses the admissible scale hierarchy. In the nonzero case the equality $\log q = \sum_p e_p \log p$ is the elementary product relation that transforms finite-place divisibility into Archimedean shortness.

It remains to account for the index on families from several translated images. At a fixed p-adic place, the image of the controlled submodule of $M_p \cap C^{\mathrm{ram}}$ has integral twisted lifts under each of the maps Q_s. Pulling back to a family in a union of $|J|$ such images loses at most the product of the respective determinant bounds. Equivalently one adds at most $|J| \dim C\, \epsilon_p \log p$ to the logarithmic index at p. This holds for every choice of p-adic place and every twist prime to p, with the same numerical loss. Summing the losses gives

$$d_{\mathrm{ind}} \le |J| \dim C \sum_p \epsilon_p \log p \le |J| \dim C\, B.$$

This establishes both clauses of Theorem 4.1 from the stated inputs.

The point of the argument is now visible. The factor m amplifies the prime-counting gain in the numerator, while both the local collision repairs and the level choice prevent a proportional undesirable enlargement of the error. The upper bound produces only useful vectors in the images of the maps P_s; it says nothing directly about how much of W these images cover. Establishing that the number of required translates is small and that their images nearly span W is a separate problem, addressed by Theorem B.

5 Enough forms: a quantitative spanning theorem

The upper estimate of Theorem A applies to special vectors obtained by lifting ramified columns. It is not an estimate for every vector of the target space. The point of Theorem B is that one does not need such a universal estimate: one needs a large enough collection of special vectors to form almost a basis. Two competing requirements govern the construction. The family of vectors must have dimension very close to that of the target; on the other hand, the number of translates used to obtain it must be small, since each translate costs some index in the integral estimate.

5.1 The statement and its role

Write $C^{\mathrm{ram}} \subseteq C$ for the ramified columns, $c = \dim_{\mathbb{C}} C$ and $n = \dim_{\mathbb{C}} W$. Let $\mathfrak{G}$ be the finite level group, and let $P_s : C^{\mathrm{ram}} \to W$ ($s \in \mathfrak{G}$) be the projected translates of the Poincaré construction, before the rational rescaling in Theorem A. A common nonzero scalar multiplying these maps has no effect on their images.

Theorem 5.1 (Theorem B: enough forms)**.** *For admissible data $\mathcal{D}$, both c and n are positive. There is a finite set $T \subseteq \mathfrak{G}$ such that*

$$|T| c \le 5n \log(1 + b^2), \qquad \dim_{\mathbb{C}} \left(\sum_{s \in T} P_s(C^{\mathrm{ram}}) \right) \ge (1 - 4b^{-2})n.$$

This is precisely the content of `Data.enough_forms` in `New/Endgame/ShortAndEnough.lean`. Notice in particular the use of the dimension of the *whole* column space in the first inequality, even though the translates used in the second inequality come from the ramified part. The first inequality is a cost estimate and the second a rank estimate. Neither is merely a qualitative spanning assertion.

If $S \subseteq C^{\mathrm{ram}}$ is the spanning set of short integral columns provided by Theorem A, each $P_s(S)$ spans the corresponding image. We can therefore select a linearly independent family

$$y_1, \dots, y_j \in \bigcup_{s \in T} qP_s(S), \qquad j \geq (1 - 4b^{-2})n,$$

where q is the nonzero rational rescaling of Theorem A. The uniform squared-norm estimate for the untwisted composite lifts of the y_i at all the Archimedean conjugates then gives the required upper bound for their Gram determinant by Hadamard's inequality. The integral estimate of Theorem A has cost at most $|T|cB \leq 5n \log(1 + b^2)B$. In particular, no further qualitative or quantitative assertion about the basis is needed before applying the determinant lower bound.

We shall explain why Theorem B holds, treating the substantial estimates of the unfolded second moment as inputs. The last, finite-dimensional argument is elementary and can be given in full.

5.2 The row space and the good-type projection

Let $F = E^+$ be the totally real subfield of the CM field E. The rank r hermitian space Y_0 has a noncompact unitary group at the relevant infinite places. One first constructs a space V of cusp forms on $U(Y_0)$ of the prescribed discrete-series types at infinity and principal level at the chosen split places above each of the two rational primes g_0, g_1. The type at infinity is imposed by a reproducing kernel. In particular V is finite dimensional and endowed with the Petersson inner product. Both the cusp condition and the fixed level matter here; this is not a space of arbitrary smooth functions.

The finite group acting on V is

$$\mathfrak{G} = \mathfrak{G}_0 \times \mathfrak{G}_1, \qquad \mathfrak{G}_i = \prod_{w \mid g_i \text{ chosen}} \mathrm{GL}_r(\mathbb{F}_w).$$

Reduction modulo the chosen places gives its right translation action, denoted ρ. The places above each g_i are chosen so that one works with general linear groups over finite residue fields, despite using global unitary groups. Passing from the full level to the principal level is what makes these nontrivial translates available.

The target W is a direct summand of V, not the whole space. It is defined by a central idempotent $e_0 \in \mathbb{C}[\mathfrak{G}]$ constructed from the finite general linear group characters. For an irreducible representation with character χ of a finite group G, its central idempotent is

$$e_\chi = \frac{\chi(1)}{|G|} \sum_{s \in G} \overline{\chi(s)} s.$$

Thus $e_\chi(1) = \chi(1)^2/|G|$, where the left side denotes the coefficient of the identity. Summing over selected representations amounts to selecting types; the coefficient of the complementary idempotent at the identity is the relative dimension of the complementary summand in the regular representation. This is the Plancherel share of the discarded types, not a priori their proportion in V.

In the development the labels at the chosen places are monic squarefree degree-r polynomials over $\mathbb{F}_w$, with nonzero constant term, parametrizing the selected Green characters. The good

families are subject to two tests: the labels must exclude roots in the specified group of roots of unity, and above *each* of g_0, g_1 at least one label must be primitive, meaning irreducible with a root generating the multiplicative group of the corresponding degree-r extension. In particular the definition is stronger than just excluding a few characters. The idempotent is the sum of the products of label idempotents for all these good families. In the resulting unitary representation, $R = \rho(e_0)$ is an orthogonal projection and

$$W = \operatorname{im} R.$$

The arithmetic reason for these good conditions appears again in the lower bound: the primitive labels yield depth-zero cuspidal types with strong rigidity and trace-transfer properties on the other unitary groups of rank r. For the spanning problem, their crucial property is their small discarded Plancherel share. If e_i projects to good families at g_i, the established estimate is

$$(1 - e_i)(1) \le \tfrac{1}{2} b^{-8}.$$

Indeed the code derives a bound of the form $h\epsilon + \exp(-\lambda)$ for h factors, with $\epsilon \ge (1+M+2r)/q_0$, where q_0 is the residue cardinality, M the excluded torsion order and $2r(2r \log q_0 + 1)\lambda \le h$. This bound uses the degrees of the Green characters and finite-field counts of squarefree and primitive polynomials. Admissibility of the scales makes it sufficiently small. There are many split factors, so one can require the existence of a very particular type at one factor without removing too much of the total Plancherel mass. The relevant concrete definitions and estimates are `goodTupleFamilies`, the definition of e_0 in `New/Seesaw/IndexTypes.lean`, `Data.goodForms` in `New/Trace/GoodIdempotent.lean`, and `bad_share_le` in `New/Trace/BadShare.lean`.

5.3 From columns to Poincaré series

The source permits a particularly explicit parametrization. Its unitary group of rank k is definite, so its adelic classes give a finite coordinate model for the algebraic modular forms of the specified weight. In the Poincaré construction one has indices (h, i): h ranges over these finite classes with the indicated level data, while i ranges over the product of finite sets of representation coordinates at the infinite places. In particular the coordinate space is isomorphic to C; its dimension c is the product of the two cardinalities. In Lean this identification is `Data.columnCoords`. Its coordinate index sets are nonempty,and hence $c > 0$.

The relative geometry comes from a Bessel subgroup of $U(Y_0)$. With $\varrho = m/2 - 1$, the rational space has a distinguished part of rank k, together with ϱ hyperbolic pairs and a middle vector, indeed $r = k + 2\varrho + 1$. A unipotent subgroup U_B and a copy of the smaller definite unitary group make up the Bessel subgroup S_B. A Bessel function is equivariant under the first factor through an additive character and under the second through the chosen representation. This allows a coefficient of a column to be encoded as part of a seed function on the larger group.

For a coordinate (h, i) one forms a seed from three ingredients: a compactly supported finite-place level indicator in the class h, a Bessel function at infinity with coordinate i, and a cutoff adapted to the unipotent quotient. If $\phi_{h,i}$ denotes the seed, the schematic formula for the Poincaré series is

$$(Qz)(g) = \sum_{h,i} z_{h,i} \sum_{\gamma \in U(Y_0)(F)} \phi_{h,i}(\gamma g).$$

The cutoffs in this formula are essential to convergence; their unipotent translates sum to the appropriate identity on the quotient. The result is a cusp form with the correct types and level. The proof fixes all the measures, bases and cutoffs rather than leaving the seed only specified up to a scalar.

The fundamental analytical identity is obtained by unfolding a Petersson product. Write $A(h,s,h')_{i,i'}$ for the sum of amplitudes of rational elements with the level reductions determined by h, s, h'; each amplitude is an integral of two Bessel functions against the relevant character and cutoffs. Then, with a positive constant v_0 coming from the volume of the principal level,

$$\langle \rho(s) Q e_{h',i'}, Q e_{h,i} \rangle = v_0 A(h,s,h')_{i,i'}.$$

The identity is exact. It expresses all matrix coefficients of $Q^* \rho(s) Q$ in terms of explicitly defined relative orbital integrals and arithmetic congruences. The compact formalization defines `Fits.poincareMap` and proves `Fits.petersson_poincare` in `New/Filling/PoincareSeries.lean`.

For the spanning theorem it is convenient to multiply Q by a positive constant, producing a map $P : C \to V$ such that its matrix coefficients equal $A(h,s,h')_{i,i'}/\nu$, where $\nu > 0$ is the unipotent covolume. At $s = 1$ the terms from the Bessel subgroup give precisely the identity matrix in the coordinate representation: their amplitude is $\nu \delta_{h,h'} \delta_{i,i'}$. Thus P is close to an isometry, rather than merely a map with many nonzero values.

5.4 What the second moment must prove

Here is a convenient quantitative formulation of the estimates used in the argument. Put

$$n_0 = |\mathfrak{G}|\, d_\infty\, \mathrm{mass}_r(E),$$

where d_∞ is the product of the formal degrees appearing in the Bessel systems, and $\mathrm{mass}_r(E)$ is the definite genus mass in rank r. Uniform dimension estimates give $n \le n_0$. The mass is a weighted sum over definite adelic classes, with reciprocal automorphism orders. The dimension bound is obtained by transfer to definite unitary groups of the same rank, and counting classes and weight coordinates there. Its use of good types explains why subspaces different from V had to be singled out before analyzing P.

The following estimates, not proved here, summarize the inputs of `New/Filling/Moments.lean`. Write $(u_j)_{1 \le j \le c}$ for the orthonormal coordinate basis of C. Then

$$\sum_j \|(P^*P - I)u_j\| \le b^{-4} c, \quad \frac{1}{|\mathfrak{G}|} \sum_{s \in \mathfrak{G}} \|P^* \rho(s) P\|_{\mathrm{HS}}^2 \le (1 + 9b^{-4}) \frac{c^2}{n_0}, \quad \|(I - R)P\|_{\mathrm{HS}}^2 \le 4b^{-4} c. \tag{14}$$

For the last estimate one applies the projection estimates to each of the two factors $\mathfrak{G}_i$ and uses that the two central projections commute. The second estimate is a nearly sharp upper bound; its leading constant is one. In the ideal situation, the average over translates of the positive operator PP^* would be a multiple of the identity, or a tight frame. Such an average has smallest possible Hilbert–Schmidt energy amongst positive operators of fixed trace and dimension. A coarse estimate of the moment with a large unbounded multiplicative factor would not suffice. The needed conclusion concerns almost *all* target dimensions.

The technical proof of these estimates has three major layers.

The Archimedean estimates. At every infinite place one constructs full Bessel functions using the discrete series with the chosen parameter, and truncated Bessel functions very close to the full ones. For a vector-valued Bessel function W_v write, in the appropriate quotient coordinates,

$$J_v(\gamma) = \int W_v(y) \otimes W_v(\gamma^{-1} y)^*\, dy.$$

These correlations replace the classical Bessel integrals in this large-rank Poincaré expansion. The proof needs bounds on their energy, on derivatives and on their frequency tails. For example the Schur bound for the full functions bounds the squared norm of the operator $\phi \mapsto \int \phi(y) W_v(\gamma y)\, dy$ by $|I_v|/d_v$ times the squared norm of ϕ, with the number of coordinates $|I_v|$ and local formal degree d_v. The truncated functions come with an explicit L^2 approximation and support bound. They admit further estimates on the open cell of the Bessel double coset decomposition, where its frame Gram discriminant does not vanish. These are established using the large gaps in the discrete-series parameter, orthogonality, inequalities for Gram matrices, local coordinate charts and a Laplace localization argument. See `Data.exists_besselSystem` in `New/Filling/FillingData.lean`, and `New/RealPlace/BandLimits.lean` and `New/RealPlace/OpenCellBound.lean`.

Degenerate and incoherent pairs. When the unfolded matrix coefficients are squared and summed over s, pairs of rational elements with the same reduction appear. The Bessel characters must agree on the intersection of two conjugate unipotent groups; incompatible characters make the corresponding orbit sum vanish, by character orthogonality. For a pair with degenerate frame Gram discriminant, character agreement forces intersections of the flags. The proof uses the Eichler flag, normal forms adapted to its successive pieces, lattice counts and bounds for cells where the Gram discriminant is small. The same-congruence requirement forces divisibility, allowing small and incoherent cells to be bounded with room to spare. In particular, if $\overline{S}_B$ denotes the reduced Bessel subgroup, the second-moment contribution from non-Bessel terms over the translations which, at either one of the two groups $\mathfrak{G}_i$, have the reduction of a Bessel element is at most $b^{-12}|\overline{S}_B|c$ in the amplitude energy normalization. This is `energy_compl_besselS_le`; it is one of the key means of keeping the discarded projection and the diagonal error small. The estimates and exact vanishing of disagreeing characters are in `New/Filling/CrudeEnergy.lean`, `BoxPartition.lean`, `LinkPairing.lean` and `Moments.lean` of the same directory.

The near-diagonal main term. The remaining coherent phases are bounded using a system of positive local Bessel kernels. At finite places the orbital integrals and their phases are compared on successive levels of the flag; these local kernels and the Archimedean correlation kernels satisfy a compatible integration relation between consecutive levels. One uses one-sided Poisson summation over the rational coordinates successively, rather than replacing all correlations by their absolute values. This gives the nearly sharp constant required for the second moment. A useful strengthened estimate is the following. If a nonnegative positive-type function $\Xi(s)$ on $\mathfrak{G}$ is dominated in all partial sums by the amplitude energy and is supported on translations fixing specified nonzero frame vectors at the chosen places in a set A above g_0, then

$$\sum_s \Xi(s) \le (1 + 9b^{-4}) \frac{c^2}{d_\infty \operatorname{mass}_r(E)} b^{-6k|A|}.$$

Here positive type means that $(\Xi(s_i^{-1} s_j))_{i,j}$ is positive semidefinite. It matters because the singular-minor conditions that arise in the local decomposition can be controlled using positivity, character sums and the bounds on fixed-vector sets. Taking A empty and $\Xi(s) = \|P^* \rho(s) P\|_{\mathrm{HS}}^2$ yields the desired average bound. The inputs are `energy_fixingSet_le` and `sum_le_of_positive_type`, which use the adelic sieve of `New/Filling/GlobalSieve.lean` and the global-to-local identifications in `New/Filling/GlobalToLocal.lean`.

This three-layer argument is the genuinely difficult part of Theorem B. It is analogous to obtaining an effective trace or large-sieve formula with a sharp main term in a family whose ranks and weights both vary; it does not follow from the usual qualitative spanning theorem for Poincaré series.

5.5 The rank deduction from the moments

We give the deduction to make transparent how all the estimates combine. For a subspace $J \subseteq W$, write $\phi = \dim J/n$ when $n > 0$. Set $A_s = \rho(s)PP^*\rho(s)^*$ and $\overline{A} = |\mathfrak{G}|^{-1}\sum_s A_s$. The second-moment estimate gives

$$\|\overline{A}\|_{\mathrm{HS}}^2 = \frac{1}{|\mathfrak{G}|}\sum_s \|P^*\rho(s)P\|_{\mathrm{HS}}^2 \le (1+9b^{-4})c^2/n_0.$$

The equality results by expanding both averages of translates and reindexing the finite group. Let p_J denote the orthogonal projection onto J in V. For each s, let $U_s = (R\rho(s)P)^{-1}(J)$. Since $R\rho(s)Pu \in J$ for $u \in U_s$, the part of $\rho(s)Pu$ perpendicular to J is in the kernel of R. Summing squared norms over an orthonormal basis of U_s and enlarging the resulting positive sums to all of C gives

$$\dim U_s \le b^{-4}c + \mathrm{tr}(p_J A_s) + \|(I-R)P\|_{\mathrm{HS}}^2.$$

Indeed the trace of $P^*P - I$ against any projection is bounded in absolute value by $\sum_j \|(P^*P - I)u_j\|$. The last term is independent of s, because R is central. Averaging in s and applying Cauchy–Schwarz to the middle term bounds the average of $\dim U_s$ by

$$c\left(5b^{-4} + \sqrt{(1+9b^{-4})\phi}\right) \le \frac{c}{2}(1+\phi+19b^{-4}).$$

Here we used $\dim J/n_0 \le \phi$ and $2\sqrt{AB} \le A + B$. The same argument with $J = 0$ and $R = 0$ would contradict $\dim U_s = c$ by the first and last estimate when b is large. Hence $n > 0$, independently of the subsequent dimension ratios.

Now introduce the ramification operator

$$D_N = \prod_{w|\pi}(1 - z_w)$$

on C, where z_w is translation by a generator of the indicated finite central unit action at a chosen place over N. Its image is C^{ram}. The dimension lost when taking inverse images under D_N is at most c times the number of factors divided by $N - 1$; this is `ramifyingOp_finrank_comap_and_short` in `New/Columns/ChainValuation.lean`. With the admissible scales the additional term together with $(19/2)b^{-4}c$ is at most $b^{-2}c$. Thus for every $J \subseteq W$ there is an $s \in \mathfrak{G}$ with

$$\dim(P_s D_N)^{-1}(J) \le \left(\frac{1+\dim J/n}{2} + b^{-2}\right)c. \tag{B-spread}$$

All relevant normalizations of P_s by nonzero scalars leave this unchanged. This is the filling theorem `Data.ramifiedColumns_fill_rows` obtained from `Fits.exists_translate_finrank_comap_le` in `New/Filling/FillingTheorem.lean` and `Moments.lean`.

5.6 Greedy selection of translates

The remaining argument, formalized as `exists_few_translates` in `New/Heights/OneDeterminant.lean`, has no automorphic content. Let $\eta = b^{-2}$ and start with $J = 0$. Suppose that the current codimension $q = n - \dim J$ is greater than $4\eta n$. From (B-spread), the increase in rank g obtained by adjoining the image of $P_s D_N$ satisfies

$$g \ge c\left(\frac{q}{2n} - \eta\right) \ge \frac{cq}{4n} > 0.$$

Thus that s makes a genuine contribution and was not already selected. Attach to a remaining codimension q the potential

$$\Psi(q) = 5n \log\left(1 + \frac{q}{\eta n}\right).$$

If an image is adjoined while $q > 4\eta n$, the decrease in potential is at least

$$\Psi(q) - \Psi(q-g) = -5n \log\left(1 - \frac{g}{q+\eta n}\right) \geq \frac{5ng}{q+\eta n} \geq c,$$

where the first inequality uses $-\log(1-u) \geq u$ and the second uses the bound for g. This process must terminate since positive integer rank is gained at each step. Since $\Psi(q)$ is nonnegative, the number of additions times c cannot exceed $\Psi(n) = 5n \log(1 + b^2)$. At termination the remaining codimension is at most $4b^{-2}n$, and since D_N maps onto the ramified columns, the chosen images are precisely the images used in Theorem 5.1. This proves the elementary deduction of Theorem B from the filling estimate.

The importance of the particular logarithmic cost can now be seen: a mere assertion that all the translates span would give no useful bound on the index of the family. With this nearly optimal covering argument the total index and missing dimension are both small on the scales of the final inequality.

6 The lower bound: why integral families cannot be too short

Theorem A bounds the lengths of selected lifts from above, whereas Theorem B says that sufficiently many of these lifts can be obtained at a controlled index. We now need a bound in the opposite direction, valid for *every* almost full-dimensional family of small index. This is Theorem C. Its significance is easiest to see before discussing how it is obtained: the lower bound is independent of the prime-counting remainder $\mathcal{R}$, and has the same main term H as the upper bound.

Let W be the row space, $n = \dim_{\mathbb{C}} W$, and let $y_1, \ldots, y_q$ be independent elements of W, where

$$q \geq (1 - 4/b^2)n.$$

For the trivial twist the lifted family has a squared covolume, i.e. a Gram determinant. Suppose the bound for that squared covolume, uniformly over the conjugates specified by the construction, is $\exp V$. Suppose also that the family has index at most d_{ind} in the sense of the twisted integral structures. Thus there is a logarithmic denominator cost e_ℓ for each rational prime ℓ, with $\sum_{\ell \in S} e_\ell \leq d_{\mathrm{ind}}$ for every finite set of primes S; suitable changes of the family with determinants of ℓ-adic size at least e^{-e_ℓ} give integral lifts.

Theorem 6.1 (C: long integral forms)**.** *Under these hypotheses one has*

$$n(H - 3L) \leq V + 2d_{\mathrm{ind}}, \qquad L = fa(\log b)^4. \tag{15}$$

This is the assertion `Data.integral_forms_long` in `New/Endgame/LongForms.lean` (compact edition). The phrase *long* refers here to a bound for the aggregate size of a family of lifted forms, rather than for each individual form. The proof compares the trivial twist with a family of carefully chosen twists. The elementary determinant and product-formula computations below are proved; the deep uniform automorphic and arithmetic assertions are inputs from the Lean development, not independently established in this exposition.

6.1 Volumes, heights, and the product formula

For an allowable m-fold twist θ, write $\Theta_\theta : W \to W_\theta^{\mathrm{end}}$ for the chain of lifts restricted to the rows. The final space is equipped with the prescribed Hermitian pairing. The number

$$Z(\theta) = \det(\Theta_\theta^* \Theta_\theta)$$

is the squared volume of the map, calculated using an orthonormal basis of W. It is nonzero and positive in the construction. If $w_1, \dots, w_n$ is any basis then

$$\det\left(\langle \Theta_\theta w_i, \Theta_\theta w_j \rangle\right) = Z(\theta) \det\left(\langle w_i, w_j \rangle\right). \tag{16}$$

Indeed, writing the basis as columns of an invertible matrix S in orthonormal coordinates, the left side is $\det(S^* \Theta_\theta^* \Theta_\theta S) = |\det S|^2 Z(\theta)$. Consequently ratios of $Z(\theta)$ can be computed with the same, possibly nonorthonormal basis on the two sides. This is important because the vectors given by Theorems A and B need not be orthogonal.

One must obtain information at all primes simultaneously. However, using a *single* twist at every prime would impose needlessly strong conditions. Instead one uses a family indexed by pairs $\iota = (\ell, \iota_\ell)$, where $\iota_\ell : \mathbb{C} \xrightarrow{\sim} \overline{\mathbb{Q}}_\ell$ is a field isomorphism, and chooses a twist θ_ι for each pair. These are abstract field isomorphisms, not continuous identifications with the usual topology on $\mathbb{C}$. Absolute values on the target are normalized by $|\ell|_\ell = \ell^{-1}$.

Here is the particular height formalism used in the compact proof. A family $u = (u_\iota)$ of complex numbers admits a height bound c, denoted $\mathrm{Ht}_{\le c}(u)$, if there exist $Q \in \mathbb{C}^\times$, a real number α, and real numbers h_ℓ such that

$$|\sigma Q| \le e^{\alpha} \qquad (\sigma \in \mathrm{Aut}(\mathbb{C})), \tag{17}$$

$$|\iota_\ell(Q u_\iota)|_\ell \le e^{h_\ell} \qquad (\iota = (\ell, \iota_\ell)), \tag{18}$$

$$\alpha + \sum_{\substack{\ell < M \\ \ell \text{ prime}}} h_\ell \le c \qquad (M \text{ sufficiently large}). \tag{19}$$

In particular, the witness Q is common to the entire family, even though the entries u_ι can vary. This is not the usual nonnegative logarithmic Weil height: a bound c can be negative. The point is to permit separate integral estimates at different primes and join them using one archimedean witness.

Lemma 6.2 (Height calculus)**.** *If* $\mathrm{Ht}_{\le c}(u)$ *and* $\mathrm{Ht}_{\le c'}(u')$*, then* $\mathrm{Ht}_{\le c+c'}(uu')$*. In particular* $\mathrm{Ht}_{\le jc}(u^j)$ *for positive integers* j*. If* $\mathrm{Ht}_{\le c}(1)$ *for the constant family* 1*, then* $c \ge 0$*.*

Proof. Multiplication of the witnesses gives the first assertion, since the archimedean and local logarithmic bounds add. Repeated multiplication gives the second.

For the constant family, the local inequalities bound all ℓ-adic conjugates of Q; this forces Q to be algebraic over $\mathbb{Q}$. Indeed, a transcendental element could be sent by field isomorphisms to transcendental elements of arbitrarily large ℓ-adic absolute value, by extension of maps on transcendence bases. Put $F = \mathbb{Q}(Q)$, $s = [F : \mathbb{Q}]$. All embeddings of F into $\mathbb{C}$ and $\overline{\mathbb{Q}}_\ell$ extend to isomorphisms of the corresponding algebraically closed fields. Hence, for $z = N_{F/\mathbb{Q}} Q \neq 0$,

$$s^{-1} \log |z| \le \alpha, \qquad s^{-1} \log |z|_\ell \le h_\ell.$$

Take M sufficiently large to contain all primes at which $|z|_\ell \neq 1$. The product formula for the rational number z gives

$$0 = s^{-1} \left(\log |z| + \sum_{\ell < M,\ \ell \text{ prime}} \log |z|_\ell \right) \le \alpha + \sum_{\ell < M,\ \ell \text{ prime}} h_\ell \le c.$$

□

The compact file `New/Heights/Heights.lean` implements this definition and the rules `HeightLe.mul`, `HeightLe.pow`, and `HeightLe.nonneg`.

The aim is now to establish two separate bounds for a suitably chosen family of twists:

$$\mathrm{Ht}_{\le V+2d_{\mathrm{ind}}+nb}\left(\frac{Z(\theta_\iota)}{Z(1)}\right), \tag{20}$$

$$\mathrm{Ht}_{\le 2n(2L-H)}\left(\frac{Z(1)^2}{Z(\theta_\iota)^2}\right). \tag{21}$$

These are genuinely different estimates. The first uses the given family's shortness and near integrality, and has no main term $-H$. The second uses the structure of the lifts and is independent of y_i. The first ratio squared times the second equals 1. By Lemma 6.2,

$$0 \le 2(V + 2d_{\mathrm{ind}} + nb) + 2n(2L - H).$$

Admissible scales give $b \le L$, yielding (15). Understanding how (20) and (21) are obtained is thus the main task.

6.2 The first ratio: completing an almost spanning family

The given vectors do not span W in general. Adjoining arbitrary vectors would be useless if their lifts had unbounded conjugates or uncontrolled denominators. One requires a new spanning family of auxiliary rows. They can be much longer than the rows constructed in Theorem A, because they will account for at most $4n/b^2$ of all basis elements.

The needed theorem is `Data.crude_rows`, established in `New/Exchange/CrudeRows.lean`. Before projection onto the good subspace, it supplies a spanning family whose untwisted lifted values have squared norm bounded by $\exp(x_1^2 b^2)$ at every conjugate, and whose twisted lifts, multiplied by one scalar $c(\theta)/c(1)$ common to the family, are integral away from the twist conductor. The scalars are nonzero and their inverse squared norm ratios have small height. After good projection and the passage from value norms to the global pairing, further bounds are required. The final combined cost of the added rows and auxiliary scalars is at most nb; these estimates depend on the precise parameters, including the second cutoff. The relevant calculation is `Data.heightLe_volume_rows` in `New/Endgame/LongForms.lean`.

Here is the determinant argument. Extend $y_1, \dots, y_q$ to a basis w by $k' = n - q$ auxiliary rows. Put

$$G_\theta(w) = \left(\langle \Theta_\theta w_i, \Theta_\theta w_j \rangle\right)_{i,j}.$$

At ι the changes of basis making the original q rows integral extend to changes of the whole basis by a scalar on each of the k' new rows. If M is such a change-of-family matrix, the formal identity

$$\det(M^* G_\theta(w) M) = \overline{\det M}\, \det M\, \det G_\theta(w)$$

expresses the effect of the index on the Gram determinant. When passing to ℓ-adic values, integrality of the two arguments of the Hermitian pairing must be checked at the appropriate embeddings, one composed with complex conjugation. The argument uses changes of family at both embeddings; it does not suppose ι_ℓ commutes with conjugation.

For a scalar γ_ι used on each additional row, the determinant sizes at the two embeddings are bounded below by e^{-e_ℓ} times the corresponding absolute value of $\gamma_\iota^{k'}$. Since the resulting pairings are integral, this implies

$$|\iota_\ell(\det G_{\theta_\iota}(w))|_\ell \le e^{2e_\ell} |\iota_\ell((\gamma_\iota \overline{\gamma_\iota})^{-k'})|_\ell.$$

Use $Q = \det G_1(w)$ as height witness and $u_\iota = Z(\theta_\iota)/Z(1)$. Equation (16) identifies Qu_ι with $\det G_{\theta_\iota}(w)$. The conjugates of Q are bounded by completing the short family, using the bounds for the additional vectors. Conjugation is again handled in the pairing, rather than by assuming all complex field automorphisms preserve the Hermitian norm. The relevant Gram determinant bound is Hadamard's inequality applied after projecting the added vectors orthogonally to the span of the original ones.

If the archimedean cost of the added vectors is at most $k'A$ and the inverse squared scalar ratios admit height bound g, the preceding argument gives

$$V + 2d_{\mathrm{ind}} + k'(A + g).$$

The factor two on the index comes from taking two determinants for a Hermitian pairing, rather than from a loss in an estimate. This step is `one_determinant` in `New/Heights/OneDeterminant.lean`. The scale inequalities and the spanning theorem for auxiliary rows yield (20).

6.3 The exchange construction

Where do the auxiliary rows come from? The upper-bound seesaw is not by itself enough: integrality of auxiliary rows must hold at every twist away from the conductor, after multiplication by a scalar whose variation is controlled. The lower-bound method creates these rows by comparing periods on a grid of unitary spaces. In the compact edition the relevant files lie mainly in `New/Exchange`.

An elementary square represents a situation where a theta lift followed by a descent through one Hermitian space can be exchanged for a descent followed by a lift through another. The two intermediate spaces have ranks differing by an even integer, and their signatures and characters are matched. The exchange is not an equality of arbitrary operators. It is an identity for periods with specified local data, after Fourier transformation and summation over suitable finite residue matrices. Its proportionality factor may depend on the twist through a nonzero factor $c(\theta)$ such that $|c(\theta)|^2$ has an expression in terms of an explicit product of partial Hecke L-values and normalization constants. The comparison factor apart from $c(\theta)$ is independent of the varying twist. See `one_square` in `New/Exchange/OneSquare.lean`; in the general square statement that fixed factor is not yet asserted nonzero.

Regard the construction as a succession of transformations of an integral: express a period using a sequence of given arrows, change the sequence across one square, match the kernels and correct the scalar, then continue. After finitely many such changes the period is expressed at the opposite end of the grid in forms on definite unitary groups where the integral structure can be described by finite class sums. Branching in the algebraic weights, bounds on class sizes, and control of kernel values give the crude size and denominator bounds. In this way one obtains a spanning set on the original indefinite group, not only an identity for an individual eigensystem.

The cost of the scalar comparisons is quantified. A factor c costs s squares when

$$\mathrm{Ht}_{\le 2stx^{999}}\left(|c(\theta_\iota)/c(1)|^{-2}\right)$$

holds for every fit family of twists, a notion explained below. Costs add under multiplication. Factors whose squared norm ratios are powers of the norms of twist conductors have zero cost, since the conductor is prime to the chosen ℓ. This bookkeeping is `Data.CostsSquares` in `New/Exchange/CostOfSquares.lean`. There are two related uses of the exchange construction: moving periods to the definite endpoint for auxiliary rows (route cost $2m^2$), and factoring the squared volume of the entire row chain into individual lift volumes (cost m^2). These costs are small compared with the budget because a contains the very large parameter b.

6.4 The second ratio: factorization into single lift volumes

To see where the height $-H$ in (21) comes from, define $B_0 = B/(mk)$. For $1 \le j \le m$ let

$$u_{j,\vartheta} : W_j \longrightarrow V_{j,\vartheta}$$

denote the individual good-form lift on an auxiliary rank-r unitary group (the notation $W_j, V_{j,\vartheta}$ here refers to spaces of forms). The good source spaces are all of dimension n. Put

$$\lambda_j(\theta) = \det(u_{j,\theta^j}^* u_{j,\theta^j}).$$

These auxiliary lifts are injective. The source of u_{j,θ^j} need not literally be the domain of the jth arrow of the original chain: the proof transfers good forms between different unitary groups and compares their periods and Hecke operators. The next identity is therefore substantive, not an application of multiplicativity of determinants of square matrices. The theorem `Data.volume_iota_eq_prod_liftVolume` in `New/Rows/RowVolume.lean` gives

$$Z(\theta) = |c(\theta)|^n \prod_{j=1}^{m} \lambda_j(\theta). \tag{22}$$

It asserts nonvanishing and controls c with cost m^2 in the above sense. Adjoints of the theta operators give pairings written as Hecke operators by inner-product formulas of Rallis type. Transfer gives the same determinant on good forms with the same multiplicities when the signatures change. The exchange identities compare the composite chain pairing to the product of individual Hecke operators, yielding (22).

The crucial single-lift estimate, `heightLe_liftVolume_ratio` in `New/Rows/LiftVolumeHeight.lean`, is

$$\mathrm{Ht}_{\le nk((\log b)B_0 - \mathcal{U})}\left(\frac{\lambda_j(1)}{\lambda_j(\theta_\iota)}\right), \tag{23}$$

provided the twist family is fit. It pays for k directions in one lift. The height calculus gives the bound $n(L - H)$ for the product of the m ratios. Squaring, multiplying by the scalar ratio from (22), and using its height bound $2m^2 t x^{999} \le 2L$ gives exactly (21). The source of the large negative main term is thus the individual m single-lift estimates.

6.5 Measuring an individual lift without evaluating its norm

To obtain (23), one must avoid bounding unknown periods independently. The proof uses a second family of forms on definite unitary groups to eliminate such periods from a determinant identity.

Suppress the index j. Let T denote the fixed space of coordinate tables, u_ϑ the good lift, and $Z_\vartheta = \operatorname{im} u_\vartheta$ its image. The theorem `exists_comparison_forms` in `New/Rows/ComparisonForms.lean` constructs maps $g_\vartheta : T \to V_{j,\vartheta}$, a single surjection $A : T \to W_j$, and nonzero complex scalars γ_ϑ satisfying

$$u_\vartheta^* g_\vartheta = \gamma_\vartheta A, \qquad |\gamma_\vartheta/\gamma_1|^2 = \Gamma_j(\vartheta)/\Gamma_j(1). \tag{24}$$

Here Γ_j is an explicit product of Hecke L-values with specified Euler factors removed; in the code it is `Data.heckeValueProd`, a product of three strings of values of $U(1)$ characters, including the inverse twist in the string with negative sign. Away from the conductor, g_ϑ sends the indicated integral tables to integral forms, and at the trivial twist one has the uniform estimate

$$h_j \|g_1 \zeta\|_\sigma^2 \le e^{2rB_0} \sum_\alpha \|\sigma \zeta_\alpha\|_{\mathrm{Weyl}}^2$$

for every table and conjugate. Here h_j is the positive hook number of the algebraic weight in the small auxiliary group.

Write p_{Z_ϑ} for orthogonal projection, and consider

$$D_\vartheta = \det\left(h_j p_{Z_\vartheta} g_\vartheta g_\vartheta^*|_{Z_\vartheta}\right).$$

In finite-dimensional Hermitian linear algebra, the identity

$$\det(u_\vartheta^* u_\vartheta) D_\vartheta = h_j^n |\gamma_\vartheta|^{2n} \det(AA^*) \tag{25}$$

follows simply by expressing u_ϑ as an invertible map onto Z_ϑ and taking the determinant of $h_j u_\vartheta^* g_\vartheta g_\vartheta^* u_\vartheta = h_j |\gamma_\vartheta|^2 AA^*$. Surjectivity of A implies all determinants in (25) are nonzero. In particular

$$\frac{\lambda_j(1)}{\lambda_j(\theta_\iota)} = \frac{D_{\theta_\iota^j}}{D_1}\left(\frac{\Gamma_j(1)}{\Gamma_j(\theta_\iota^j)}\right)^n.$$

This is the key cancellation identity for one lift. The comparison maps are constructed using complementary blocks of the definite group, and separation of good forms by products of theta forms proves surjectivity in (24). See `Data.eq_zero_of_integral_fNatural_tableTheta` in `New/Rows/Separation.lean`.

Two further bounds are required. First, for a fit family, the inverse ratio of Hecke values admits height bound

$$10mx^{999} + k(6B_0 - \mathcal{U}) \quad \text{for} \quad \frac{\Gamma_j(1)}{\Gamma_j(\theta_\iota^j)}.$$

See `heightLe_inv_heckeValueProd_div` in `New/Rows/HeckeValueHeight.lean`. This uses rank-one theta lifts to measure individual ratios of Hecke values: integral test forms pair with the twisted theta kernel, producing explicit scalar multiples of the algebraic parts of special values. The local lower bound on the pairing detects forced divisibility of jets; the archimedean bound and product formula convert this to a negative height bound. The long string of ratios accounts for $k\mathcal{U}$, and the short ramified strings have a small total loss. The estimate for an individual rank-one inverse ratio is $6B_0 - \mathcal{U}$ (`heightLe_inv_ratio`).

Second, the determinant ratio $D_{\theta_\iota^j}/D_1$ needs a local bound. Integrality of the comparison maps does *not* imply integrality of projections onto Z_ϑ: distinct characteristic-zero eigenspaces can be congruent at a finite prime. The projection may require division by a congruence ideal. The proof does not ask for an integral projector on the whole target. The relevant comparison forms are tested, namely orthogonal to the false lifts: good forms in $Z_\vartheta^\perp$ satisfying the particular annihilating Hecke-polynomial tests associated with genuine lifts. This ensures the projected comparison forms fall into the case handled by the arithmetic argument.

In `New/Selmer/FitTwists.lean`, `DenominatorLe` with bound β_ℓ is defined by a determinant estimate. If $z = \dim Z_\vartheta$ and v_a, v_b' are integral test families with the specified integrality for the first family at every relevant embedding and the tested condition for the second, then

$$\left|\iota_\ell \det\left(\langle v_a, p_{Z_\vartheta} v_b'\rangle\right)_{a,b=1}^{z}\right|_\ell \le \exp(zr\beta_\ell).$$

This determinant-level control, rather than entrywise control of a full projector, suffices. For the comparison operator it yields $|\iota_\ell D_{\theta_\iota^j}|_\ell \le \exp(nr\beta_\ell)$. At the archimedean places the trivial-twist norm estimate gives $|\sigma D_1| \le \exp(2nrB_0)$ for all σ. These are respectively the assertions `norm_det_compressedGramOp_padic_le` and `norm_det_compressedGramOp_arch_le` in `New/Rows/LiftVolumeHeight.lean`. Using D_1 as the common witness for the height of the determinant ratio and the bound $\sum_{\ell\in S}\beta_\ell \le B_0$ for all finite sets of primes S, we obtain

$$\mathrm{Ht}_{\le 3nrB_0}\left(\frac{D_{\theta_\iota^j}}{D_1}\right).$$

Combining this with the Hecke-value ratio estimate and (25) gives the height bound

$$3nrB_0 + n(10mx^{999} + k(6B_0 - \mathcal{U})) \;\le\; nk((\log b)B_0 - \mathcal{U}).$$

The displayed inequality on the right uses the admissible scales (r/k is bounded, $\log b$ is large, and B_0 grows faster than any fixed power of x). This is the accounting behind the single-lift estimate (23). The existence of twists meeting these denominator bounds simultaneously for all powers θ^j, and also the bounds needed for the rank-one values and exchange scalars, is itself a substantial theorem, the next ingredient of the lower-bound argument.

7 Theorem C, II: congruences and the choice of twists

The determinant comparison in Theorem C requires more than the usual identities for the norms of theta lifts. Even if we know the value of an adjoint composed with a lift over $\mathbb{C}$, the orthogonal projection onto the image of the lift need not preserve any of the integral lattices. The obstruction is a congruence: a form coming from the theta correspondence and a form outside its image may be very close in an ℓ-adic lattice. The lower bound must control this obstruction, simultaneously with the denominators of certain special-value ratios. We explain the mechanism for doing so in the compact Lean development. This is an account of its constituent theorems, not a replacement for their individual proofs.

7.1 What is required of a twist?

Put

$$B_0 = \frac{B}{mk} = \frac{fa(\log b)^3}{mk}.$$

The important point is that the allowance is specified *before* the twists are chosen. One needs a twist for every rational prime ℓ and every identification $\iota : \mathbb{C} \simeq \overline{\mathbb{Q}}_\ell$ as fields, not a single twist that works simultaneously at all primes. Such an identification need not respect the archimedean topology. The twist at ι, denoted θ_ι, is a finite-order character of the norm-one torus $U(1)$. Its conductor is prime to ℓ and to the fixed level. All its powers θ_ι^j, $1 \le j \le m$, must have the required properties. In particular, they are the twists of a single chain, not independently chosen characters for the different arrows of that chain.

Here is a useful precise formulation of the denominator condition. For a subspace Z of the space of forms on the definite member of an auxiliary unitary pair, write P_Z for the orthogonal projection onto Z in the class-sum inner product of that space. Write $q = \dim Z$ and r' for the rank of the source hermitian space of the pair. Given integral forms $u_1, \dots, u_q$ and suitable test forms $v_1, \dots, v_q$, consider

$$A = \left(\langle u_a, P_Z v_c\rangle\right)_{1 \le a,c \le q}.$$

The definition `DenominatorLe` in `New/Selmer/FitTwists.lean` requires that the u_a be integral at all ℓ-adic embeddings and the v_c be integral at the specified embedding. The latter are also tested against the non-lift part: they are orthogonal to the intersections of that part with the chosen idempotent and the kernels of the specified Hecke-polynomial operators. These tests match the forms arising in the auxiliary period argument. The bound in this notation is

$$|\iota(\det A)|_\ell \le \exp(qr'\beta). \tag{26}$$

This is a bound on determinants involving the projection, not an unqualified bound on every matrix entry of the projection on every integral vector. The determinant formulation is particularly well adapted to the covolume comparison of Theorem C.

For the j-th arrow one needs (26) for the image of the good auxiliary lift, and also for the particular rank-one lift lines that express the Hecke values. In the terminology of the Lean code these are respectively the image of `Data.goodLift` and the lines generated by `LValueData.trivialLift`. The same local allowance must work for all the spaces concerned. In addition, for the ramified pairs occurring in the exchange comparison, the inverse ratios of Hecke values must have height at most x^{999}.

Theorem 7.1 (Existence of a fit family of twists)**.** *For admissible data $\mathcal{D}$, there is a family θ_ι as above such that, for every $1 \le j \le m$, there are nonnegative numbers γ_ℓ satisfying*

$$\sum_{\ell \in T} \gamma_\ell \le B_0 \qquad \text{for every finite set of rational primes } T,$$

and (26) holds with $\beta = \gamma_\ell$ for all the stated projection determinants at θ_ι^j. For every index j and every ramified pair and sign parameter of the exchange construction, the family of inverse Hecke-value ratios has height at most x^{999}.

This is the theorem `exists_isFitFamily` in `New/Selmer/FitTwistsExist.lean`, with the definition of `Data.IsFitFamily` in `New/Selmer/FitTwists.lean`. The finite-set formulation of the sum avoids any implicit assumption that only finitely many local allowances are nonzero. It is strong enough for the height argument, since it bounds all finite sums uniformly. We now describe the three main mathematical tasks needed to obtain this result.

7.2 The Galois representation of a lift

The projection problem becomes tractable after passing to simultaneous Hecke eigensystems and their attached ℓ-adic representations. The central statement for an auxiliary theta pair is that the normalized Galois trace of an eigenform in the lift has the form

$$\operatorname{tr} U(g) = \psi(g) \operatorname{tr} R(g) + \sum_{e=0}^{s-1} \chi_\ell(g)^{-e}. \tag{27}$$

Here χ_ℓ is the ℓ-adic cyclotomic character, s is the rank difference of the auxiliary pair, and ψ is the ℓ-adic avatar of the inverse of the twisting character in this normalization. The representation R has rank r' and can be chosen independently of the twist. Its trace and local behavior, not merely its existence, matter. At the level primes the stipulated labels force tame inertia and scalar conditions; in particular its lattices span the full matrix algebra over the coefficient ring. The two chosen rational level primes ensure that at least one source of tame information is available off ℓ, for every ℓ.

The finite strings in (27) account for the change of rank under theta lifting. The trace relation follows from the unramified theta correspondence and the attachment of Galois representations with the relevant local and weight compatibilities. In the compact development see `New/Selmer/RowsExist.lean`, `New/Selmer/LabelledColumns.lean` and the definitions of `IsRow`, `IsColumn` in `New/Selmer/AptTupleDefs.lean`. The latter names refer to Galois representations, not directly to the spaces W and C of automorphic forms. There are only finitely many relevant systems for the fixed data. This will permit simultaneous control of all of them by a twist.

A congruence between a system of the form (27) and another system can be measured by off-diagonal matrix entries in integral lattices, or more invariantly by cocycles extracted from the corresponding trace algebra. This is the content of the Ribet-type method in the proof. Thus the problem of extending the projection integrally is transformed into a problem about the sizes of modules of torsion Galois cocycles. The summands in (27) lead to two distinct problems: the cocycles for the irreducible row block must be made small by twisting, whereas those for the Tate characters in the string need separate treatment.

7.3 Shrinking the Selmer modules by one character

For a lattice T in R, after enlarging the local coefficient ring $\mathcal{O}$ as necessary, consider continuous cocycles with values in $\psi T/\ell^M \psi T$. Their allowed ramification is confined to the places above ℓ, the two level primes, and the places supporting the twist and their conjugates. At places above ℓ impose the relations satisfied by crystalline cocycles. For the normalized modules in the proof there is also an evaluation condition $c(\gamma) = 0$ at a suitable global Galois element γ in the kernel of the cyclotomic character. The row block acts there by a scalar z with $z - 1$ a unit. This is a device for controlling the ambiguity of a cocycle by a coboundary; it is important that the code measures the length of a specified module of cocycles with all these conditions, rather than simply quoting a dimension of an abstract Selmer group.

Let M_g be the order of the product of the finite general linear groups of rank rt at the selected residue-field places above g; this is `Data.chosenGroupOrder`. Set

$$c_\ell = v_\ell(M_g) + 2^{-\ell}, \qquad v_\ell(\ell) = 1.$$

The precise simultaneous twisting target is

$$\mathrm{length}_\ell(\text{normalized cocycle module}) \le (10fr' + 1)c_\ell. \tag{28}$$

If $\mathcal{O}$ is a discrete valuation ring of ramification index $e_\mathcal{O}$ over $\mathbb{Z}_\ell$, the normalized length here is the usual module length divided by $e_\mathcal{O}$. The assertion applies over sufficiently large coefficient rings. In the same choice of twist, the sum of the valuations of the local character values minus one at the auxiliary tame generators is at most c_ℓ, and the inverse ramified-pair special-value ratio Q satisfies

$$|\iota(Q)|_\ell \le \exp(9c_\ell \log \ell). \tag{29}$$

These are the assertions of `exists_mFoldTwist_selmerLengthLe` in `New/Selmer/TwistForRows.lean`, with the precise cocycle module specified by `SelmerLengthLe`. Crucially, they hold for every one of the first m powers of the character.

To explain the proof of the twisting theorem, one starts with auxiliary primes split in the appropriate sense, at which the representations have prescribed Frobenius elements. The intermediate construction uses variable *tied* local conditions: schematically, one requires the value of a cocycle at a tame inertia generator to equal a scalar times its value at Frobenius. The scalar is allowed to vary. Conjugate primes carry compatible conditions. For the balanced representations used here, local and global reciprocity guarantee that when the modules are too large, a new auxiliary prime and a suitable slope can reduce their excess length. A minimality argument on the total module length implements this step repeatedly, with control of the number of primes as their auxiliary depth varies. This is the role of "pencils" of auxiliary primes and the lemma `exists_isPencil` in `New/Selmer/TwistTheorem.lean`.

The next problem is realizing the tied conditions by actual finite-order characters. One presents the modules of twisted cocycles by finitely many coordinates and relations. The relations depend polynomially on the character values at the auxiliary places. Their Fitting ideals give polynomial valuation bounds for the resulting lengths. In a simplified model, a finite torsion module presented by a square matrix $A(z)$ over a discrete valuation ring has normalized length $v_\ell(\det A(z))$ when the determinant is nonzero. In the more general presentations it is appropriate minors that provide the bound, and the same reasoning explains why small polynomial valuations are desired.

The compact proof uses a polynomial selection lemma for tuples of ℓ-power roots of unity. For example, in one variable, suppose a finite family of integral polynomials has Gauss norms of radius $0 < \rho < 1$ bounded below by $y_i > 0$, with $|\ell|_\ell < \rho^{\ell-1}$. At sufficiently large depth, a common ℓ-power root of unity ζ can be chosen with

$$|F_i(\zeta - 1)|_\ell \ge \rho y_i$$

for every i (the precise quantifiers are in `exists_root_of_unity_eval_large`, in `New/Selmer/RootsOfUnityPoints.lean`). The multivariable argument first uses a suitable line of points near 1, then selects roots of unity so as to control the first m powers simultaneously. There are additional polynomial families for the ramified-pair special values. Their behavior under inserting an auxiliary prime is controlled by the value-family lemmas. Class field theory produces the actual anticyclotomic character from the chosen root-of-unity values. This last stage is `exists_anticyclotomic_twist` in `New/Selmer/TwistTheorem.lean`. A bound of the form (28) would be insufficient if the same character forced a large positive valuation in the wrong special-value ratio: this is why (29) is proved at the same time.

7.4 Ribet's method and auxiliary primes for the Tate string

There is still a difficulty with the rank-one Tate blocks of (27). A naive congruence-module estimate could require a bound on the even parts of cyclotomic class-group modules which is not available in the necessary form. In particular, one does not want an additional hypothesis resembling Vandiver's conjecture. The replacement is an auxiliary-level argument using triangular systems of Taylor–Wiles primes.

The basic idea is to choose a detecting collection of points for the relevant cocycles, and then successively select Taylor–Wiles places whose Frobenius elements meet prescribed conditions. Global reciprocity laws relating the even classes to odd Kummer data can be organized so that each new Frobenius kills the earlier reciprocity functions but not the next one. The resulting matrix of restrictions is triangular, up to controlled powers of ℓ. In particular it isolates the even-class contributions without bounding the lengths of the even classes themselves. The proof constructs systems of this kind at every sufficiently deep ℓ-power level, with the number of auxiliary places not increasing with the depth. See `exists_isPrimeTriangle` in `New/Selmer/TrianglePrimes.lean` and `New/Selmer/ReciprocityLaw.lean`.

To use these auxiliary primes on automorphic forms, trace operators from the auxiliary level to the original level are deformed in one parameter X_j for each prime. The trace relations impose a first-order relation between inertia and Frobenius, analogous to the ties used above. This does not simply introduce a new Hecke eigensystem: it yields an operator on the integral form spaces. The full string of roots of the spherical Hecke polynomial makes the local operators applicable to systems having eigenvalue 1 in their normalized Galois representations at every group element. See `New/Selmer/HeckeString.lean` and `New/Patching/OnePrimeTie.lean`.

The algebraic Ribet argument can be described as follows. A determinant of a suitable combination of the matrices $U(g) - 1$ vanishes on the lift family, since its normalized trace decomposition contains the trivial Tate character. Choose the combination so that it suitably separates the remaining eigensystems. The first-order trace-algebra calculation constructs a polynomial multiplier F of this determinant such that the product belongs to the ideal of tie relations. On the distinguished lift systems, coefficients of F at monomials properly dividing $\prod_j X_j$ vanish. The coefficient at $\prod_j X_j$ at the distinguished eigensystem is nonzero and has controlled valuation. This is the precise sort of statement encoded by `IsRibetElement` and `ribet_first_order` in `New/Selmer/RibetFirstOrder.lean`. The triangular conditions are what allow this coefficient to be bounded without an even-class-group term. Taking the corresponding coefficient of the deformed trace operator then supplies a denominator-clearing operator at the original level.

The odd Tate-twisted class modules do enter the bound. The relevant background theorem `exists_length_shaOdd_le` in `Lib/Cyclotomic/OddSha/OddShaBound.lean` supplies a single nonzero integer X whose valuation bounds their normalized lengths at every prime, for the range of twists and levels in the argument. It also bounds the size of X in terms of the cyclotomic conductor, the length of the Tate string and the class number of the CM field. This

is the arithmetic term retained in the final estimate.

7.5 Local losses at the primes above ℓ

An essential subtlety is that the torsion cocycles produced by the trace-algebra argument do not automatically satisfy the crystalline relations used to formulate the Selmer bound. One must bound the discrepancy. Local Tate duality relates it to a defect of the crystalline cocycles on a dual lattice twisted by 1: the quotient by those crystalline classes whose cup products vanish against the classes of the congruent lattices. A general bound based only on the largest Hodge–Tate weight would be wasteful here. The weights of the argument are enormous, with the large weights near a/x or a/x_1, separated from the low ones by large gaps.

Three additional facts explain the architecture of the local estimates. First, the jumps of the row block occur in narrow windows around the prescribed high weights, or in a low range. The fraction of low weights at any embedding is at most m/k, by the definition of `IsRowType`. Secondly, the thick embeddings have been chosen sufficiently sparsely that three of them do not occur in a short Frobenius arc at small primes; this is the spacing lemma `not_thick_of_short_arc`. Thirdly, the primes at which tame ramification can occur form part of the chosen level, and their inertial types satisfy bounds important for controlling descent.

The detailed proofs use Breuil–Kisin module lattices and explicit cokernel estimates. The method constructs suitable submodules of crystalline cocycles, proves that their pairings with the dual congruent lattices vanish after the indicated powers of ℓ, and bounds their colengths. The compact proof separates tame inertial types from crystalline types, using special bounds for the sparse large weights and for the few low weights. The main local packages are in `New/Lift/ExponentTheorems.lean`, `New/Lift/SparseThickLetters.lean` and `New/Lift/FewLowJumps.lean`, assembled in `New/Selmer/CrystallineBound.lean` and `New/Selmer/OnePlaceDefect.lean`. These are substantive integral ℓ-adic Hodge-theoretic estimates, rather than merely the rational crystalline comparison theorem.

Here is the numerical result in the notation of the compact proof. Let $h_2 = r^8$ be half the weight gap, let b_1 be the rounded value of a/x, and set

$$F = (2b_1)! \left(\prod_{\substack{b_1 - 30h_2^3 < u \le b_1 + 30h_2^3 \\ u \ge 1}} u \right)^{h_2^3}, \qquad \lambda_\ell = v_\ell(F).$$

With $g = g_0 g_1$, define

$$\mathrm{loss}_\ell = 10^6 \frac{m}{k} \left(b_1 v_\ell(g) + \lambda_\ell \right). \tag{30}$$

These are the definitions in `New/Selmer/DeterminantBound.lean`. For a row of rank r' and its associated columns of Galois representations, the theorem `sum_blockLoss_le` bounds the total normalized loss, over all their blocks and all places above ℓ, by $4r'f\,\mathrm{loss}_\ell$, independently of the torsion depths. Notice the factor m/k in (30). Its appearance is one reason the choice of weights and their distribution is part of the global proof, not an arbitrary choice of regular weights.

7.6 Summing the local estimates

The previous parts produce the denominator bound with

$$D_\ell = h_2^{27}(c_\ell + v_\ell(X)) + 4f\,\mathrm{loss}_\ell\,. \tag{31}$$

This is `rowBudget` in `New/Selmer/RibetElement.lean`. Specifically, `ribet_columns` bounds the additive valuation of the relevant top coefficient after multiplication by M_g^3 by $r'D_\ell$. The deformation of the trace transfers this bound to the eigenvector-wise denominator-clearing operators, which combine by a determinant estimate to give (26) with $\beta = D_\ell \log \ell$. This last assertion is `denominatorLe_rowBudget` in `New/Selmer/DenominatorBound.lean`. It uses the short Selmer module at depth $\lfloor rD_\ell \rfloor + 1$, so that the depth exceeds the denominator being bounded.

The gain from the explicit formulas is that we can weight the allowances by $\log \ell$ and sum over rational primes. For example, using prime factorization gives exactly

$$\sum_\ell \mathrm{loss}_\ell \log \ell = 10^6 \frac{m}{k}(b_1 \log g + \log F).$$

Likewise $\sum_\ell c_\ell \log \ell = \log M_g + \sum_\ell 2^{-\ell} \log \ell$, with the second term bounded. Stirling-type estimates for the factorial and the size of the short product in F, bounds on $|X|$ and on the finite-group order M_g, and the numerical relations for admissible data now give the inequalities established in `exists_isFitFamily`:

$$\sum_{\ell \in T} D_\ell \log \ell \le \frac{B}{mk}, \qquad \sum_{\ell \in T} 9c_\ell \log \ell \le x^{999} \tag{32}$$

for every finite set of primes T. For perspective, b_1 is of order b, $\log g$ is of order $\log b$, and h_2 is only polynomial in x, while b eventually exceeds every fixed power of x. Thus the factorial loss in (30) is on the scale $(m/k)b \log b$; after multiplication by f it is smaller than $B/(mk)$ by logarithmic savings, since m^2 is of order x and a is of order xb. The exact estimates are uniform in the allowed parameters and do not rely on this informal order-of-magnitude comparison alone.

Equations (26), (29), and (32) give the fit-family theorem. They also explain why the entire congruence argument is needed. A projection cost that grew uncontrolled with the degree of the auxiliary conductor, the number of lifts, or the largest weight could exceed the error allowed in Theorem C. The finite-order twists, the first-order auxiliary-level argument and the crystalline loss estimates remove those obstructions together. The bound is quantitative enough to be multiplied by the ranks and numbers of lifts in the determinant comparison, without overwhelming the cancellation between its main term H and the main term of the upper bound.

8 Assembly of the proof: why the exponent is one half

This section gives the complete calculation from the three stated arithmetic theorems and the analytic criteria. The arguments here are short compared with the ingredients used to justify them. They are also where all the parameter choices become transparent.

8.1 Collecting the arithmetic inputs

With the data of Section 3, set $n = \dim_{\mathbb{C}} W$, $c = \dim_{\mathbb{C}} C$, and retain

$$H = mk\mathcal{U}, \qquad B = fa(\log b)^3, \qquad L = B \log b.$$

For clarity, we give the particular form of the three theorems that is needed in the calculation. The notation in the three statements below is intentionally the same. The proofs of these inputs are the subject of the preceding sections; their deeper constituent results have been taken from the Lean development.

Theorem 8.1 (A: Short forms, recalled)**.** *There are a rational number $q_0 > 0$ and a set $S \subseteq C_{\mathrm{ram}}$ of integral forms spanning C_{ram} over $\mathbb{C}$ such that, putting $T_s = q_0 P_s$, one has*

$$\|\Theta_1 T_s(v)\|_\sigma^2 \le \exp(H - m\mathcal{R} + L) \quad (v \in S,\ s \in \mathcal{G},\ \sigma \in \mathrm{Aut}(\mathbb{C})).$$

For every finite set $F \subseteq \mathcal{G}$ and every finite family y in the images $T_s(C_{\mathrm{ram}}^{\mathrm{int}})$, $s \in F$, one has

$$\mathrm{IndexLe}(y, |F|cB).$$

Theorem 8.2 (B: Enough forms, recalled)**.** *The spaces C and W have positive dimension and there is a finite set $F \subseteq \mathcal{G}$ such that*

$$|F|c \le 5n\log(1+b^2), \qquad \dim\left(\sum_{s\in F}\mathrm{im}(P_s)\right) \ge (1 - 4/b^2)n.$$

Theorem 8.3 (C: Integral forms are long, recalled)**.** *Let $y = (y_1, \dots, y_q)$ be q independent row forms with $q \ge (1 - 4/b^2)n$. If $\mathrm{IndexLe}(y, d_{\mathrm{ind}})$ and for every σ one has $\mathrm{covol}_\sigma^2(y) \le e^V$, then*

$$n(H - 3L) \le V + 2d_{\mathrm{ind}}.$$

Together with the crude bound

$$|H| + m|\mathcal{R}| \le b^2 x^{99}, \tag{33}$$

these are precisely the inputs in `star_inequality` of `New/Main.lean`. Inequality (33) follows from the size parameters and the finite prime sum definitions; the source proves it as `Data.abs_mainHeight_add_remainder_le`.

8.2 From A and B to an almost-basis of short forms

Choose F as in Theorem 8.2 and use the set S from Theorem 8.1. Since q_0 is nonzero and S spans C_{ram}, the set

$$\{T_s(v) : s \in F,\ v \in S\}$$

spans the sum of the images of the P_s, $s \in F$. Choose independent vectors $y_1, \dots, y_q$ from this set which form a basis of its span. In particular,

$$(1 - 4/b^2)n \le q \le n.$$

By A and the bound on the number of translates we may take

$$d_{\mathrm{ind}} \le 5n\log(1+b^2)B.$$

For each conjugated pairing Hadamard's inequality gives

$$\begin{aligned}\mathrm{covol}_\sigma^2(y) &\le \prod_{i=1}^{q} \|\Theta_1 y_i\|_\sigma^2 \\ &\le \exp\left(q(H - m\mathcal{R} + L)\right).\end{aligned}$$

Thus $V = q(H - m\mathcal{R} + L)$ is a valid upper bound. This is the conclusion of `Data.exists_family_covolumeSq_le` in `New/Endgame/LongForms.lean`. The step is only linear algebra, but realizes the crucial purpose of Theorem B: it converts estimates on vectors arising from a particular automorphic construction into an estimate on almost the entire determinant of the row space.

8.3 Comparison with the lower bound

Proposition 8.4 (The main arithmetic inequality)**.** *For admissible data the three arithmetic inputs imply*

$$m\mathcal{R} \le fa(\log b)^5. \tag{$\star$}$$

Proof. Write $h = H - m\mathcal{R} + L$ and apply C to the family just constructed. Then

$$\begin{aligned} n(H - 3L) &\le qh + 2d_{\mathrm{ind}} \\ &= nh + (n-q)(-h) + 2d_{\mathrm{ind}}. \end{aligned}$$

We cannot simply replace q by n in qh, because h may be negative. Instead, use (33) and $L > 0$ to get $-h \le b^2 x^{99}$, and the dimension estimate to get $n - q \le 4n/b^2$. Consequently

$$n(H - 3L) \le n(H - m\mathcal{R} + L) + 4nx^{99} + 10n\log(1+b^2)B.$$

Divide by $n > 0$, cancel the common H, and rearrange:

$$m\mathcal{R} \le 4L + 4x^{99} + 10\log(1+b^2)B. \tag{34}$$

The scale inequalities imply that the right-hand side is at most $(\log b)^2 B$. For the eventual assertion this is immediate from $a \ge xb$, $B = fa(\log b)^3$, and $\log b = x^{\varepsilon}$, since $x^{99}/B \to 0$ and $\log(1+b^2) = O(\log b)$. The Lean version uses its explicit admissibility bounds to make this last comparison for every admissible datum, not just sufficiently large x along this family. Substituting the value of B gives ($\star$). □

It is worth looking closely at the size of the three errors in (34). The $4L$ is the sum of the error per vector in the upper bound and the three units of L in the lower bound. The $4x^{99}$ is the penalty for the missing dimensions, which are so few that even a very crude bound on h suffices. The last term is twice the index. All are smaller than the final budget by a factor of order $\log b$ or more along the asymptotic family. The leading term H has disappeared completely. Without the matching of its exact leading coefficient in A and C, this argument would not give the desired bound.

8.4 Extracting the analytic remainder

Here is the short algebra behind (10). Use the notation of Section 2 and write

$$S_M(z) = \sum_{\nu \le z} \lambda_{d,M}(\nu)(1/\nu - 1/z).$$

For a fixed integer ν, the number of $\sigma \in \mathcal{T}$ such that the place $s_\nu \sigma$ lies in $\mathcal{S}$ is $|\mathcal{S}|$ if $\nu \equiv 1 \pmod{N_0}$ and zero otherwise. Indeed, multiplication by s_ν carries the whole coset $\mathcal{T}$ onto the coset of class $-\nu$ when ν is invertible modulo N_1. The thick places have class -1 modulo N_0. Every $\nu \equiv 1 \pmod{N_0}$ is a unit modulo N_1 in this tower. In the remaining nonunit cases the source's place convention places the product back in $\mathcal{T}$, so it is not thick. It follows that replacing the weight a/x_1 by a/x inside the sum over the tops gives the identity

$$\frac{1}{a}\sum_{\sigma \in \mathcal{T}} A(a,\sigma) = |\mathcal{S}|\big(S_{N_0}(x) - S_{N_0}(x_1)\big) + |\mathcal{T}| S_1(x_1).$$

Here S_M denotes the sum before subtraction of its main terms. By the analytic normalization,

$$S_M(z) = R_M(z) + \frac{\log z}{\varphi(M)} + c_{d,M}.$$

The constants for $M = N_0$ cancel in the difference. For $M = 1$, we have $c_{d,1} = -1 - \kappa_K$. Moreover

$$\log a - \overline{w} = \log x_1 + \frac{|\mathcal{S}|}{f}(\log x - \log x_1).$$

Since $f = \varphi(N_0)|\mathcal{T}|$, we have exactly

$$\frac{|\mathcal{T}|A_0}{a} = |\mathcal{T}|(\log x_1 - 1 - \kappa_K) + \frac{|\mathcal{S}|}{\varphi(N_0)}(\log x - \log x_1).$$

Subtract these main terms from the previous sum identity. This proves (10), including its constants. In particular, no estimate involving the Chowla–Selberg constant is needed at this final stage.

8.5 Choosing the second cutoff and the power saving

Fix $0 < \varepsilon < 1/100$ and choose admissible data by Proposition 3.1 for every sufficiently large x. The quantity $X = |\mathcal{T}|/|\mathcal{S}|$ satisfies $0 < X \le \log b$. Write $T = 10^6 \log x$. For large x we have $X \le e^T$. Apply the window lemma of Section 2 to the two functions $u \mapsto R_{N_0}(e^u)$ and $u \mapsto R_1(e^u)$. It supplies $u \in [T, T^2]$ such that

$$R_{N_0}(e^u) - \frac{|\mathcal{T}|}{|\mathcal{S}|}R_1(e^u) \le e^{-T}.$$

Take $x_1 = e^u$ in the auxiliary data, as permitted by Proposition 3.1. This is the decisive use of the windowed argument: there is no need for a strong pointwise estimate at the initially specified, extremely large, second scale.

Solve (10) for the desired term and use ($\star$):

$$\begin{aligned} R_{N_0}(x) &= \frac{\mathcal{R}}{a|\mathcal{S}|} + R_{N_0}(x_1) - \frac{|\mathcal{T}|}{|\mathcal{S}|}R_1(x_1) \\ &\le \frac{f(\log b)^5}{m|\mathcal{S}|} + e^{-T} \\ &\le \frac{\varphi(N_0)(\log b)^6}{m} + x^{-10^6} \\ &\le \varphi(N_0)x^{-1/2+6\varepsilon} + x^{-10^6}. \end{aligned}$$

Absorbing the fixed constant using 2ε of additional slack gives, for all sufficiently large x,

$$R_{N_0}(x) \le x^{-1/2+8\varepsilon}. \tag{35}$$

This is the precise exponent in the compact development's `eventually_primeError_le_rpow`. Replacing ε by a smaller positive number proves a bound with any fixed exponent greater than $-1/2$.

Remark 8.5 (Where the critical line enters). Everything up to ($\star$) concerns norms, divisibilities, dimensions and heights. The saving $x^{-1/2}$ arises from dividing the error by $m \ge \sqrt{x}$. The large field degree, large weights and rank do not themselves account for it. One cannot hope to improve the analytic conclusion by choosing m arbitrarily larger: the arithmetic theorems and their errors depend crucially on the admissible relationship between the length, the weights, ranks and local losses. A proof of stronger short/long-form estimates on other ranges would be an additional mathematical result.

8.6 GRH

To finish, suppose a Dirichlet character χ modulo q had a zero ρ to the right of $1/2$, excluding the pole convention of Theorem 1.1. Choose $N_0 = 18480q$ and regard χ as a character of level N_0 by induction; the induced L-function differs by finite Euler factors with no zeros in this region. The class $N_0/2 - 1$ is a unit modulo N_0 and forces $d \equiv 7 \pmod 8$. For every prime d in this class, the above arithmetic argument gives (35). Choose ε positive and sufficiently small that $\beta = 1/2 + 8\varepsilon < \Re(\rho)$. The one-sided criterion of Proposition 2.4 forbids the zero ρ. This proves Theorem 1.1, conditional on the arithmetic and analytic inputs used throughout this exposition. The standard functional equations imply the statements about the critical strip. This completes the deduction.

9 A guide to the source and to a longer mathematical treatment

9.1 What should be checked first?

The proof architecture can be read in a much more economical order than the import order of the repository. The following is a concordance of the principal statements, to help a reader switch between the present exposition and the actual code. Module paths are relative to `grh-compact/src/`; identifiers often begin with `Data.`, which expresses the dependence on the common admissible data.

Mathematical task	File	Declaration or topic
Final statements	`Statements.lean`; `Main.lean`	exported RH and Dirichlet GRH
Spine of proof	`New/Main.lean`	`star_inequality`; `New.grh`
Analytic remainder	`New/Endgame/PrimeError.lean`	`primeError`; `isPolar_primeError`
One-sided criterion	`New/Endgame/OneSidedCriterion.lean`	`entireL_ne_zero_of_primeError_le`
Window lemma	same	`IsPolar.exists_window_sub_mul_le`
Admissible data	`New/Data/AdmissibleData.lean`; `New/Data/DataExist.lean`	`Data`; `exists_data`
Main arithmetic inputs	`New/Endgame/ShortAndEnough.lean`	`Data.short_forms`; `Data.enough_forms`; remainder identity
Theorem A	`New/Curve/`; `New/Columns/`; `New/TwistedHecke/`; `New/Seesaw/`	CM jets, Hecke estimates, and seesaw chain
Theorem B	`New/Filling/`; `New/RealPlace/`; `New/DiscreteSeries/`	spreading and spanning, Bessel estimates
Theorem C, formulation	`New/Endgame/LongForms.lean`	`Data.integral_forms_long`
Theorem C, determinants	`New/Rows/RowVolume.lean`; `New/Rows/LiftVolumeHeight.lean`; `New/Exchange/`	chain and individual volume ratios
Theorem C, twists	`New/Selmer/FitTwistsExist.lean`; `New/Selmer/`; `New/Patching/`; `New/Lift/`	`exists_isFitFamily`; Selmer bounds and local losses

In particular, the best first file is `New/Main.lean`, whose detailed introductory comment lays out the major inputs and especially the terms that cancel. The files `LongForms.lean` and `ShortAndEnough.lean` then explain the precise interface between the elementary endgame and the major arithmetic constructions. This is a natural interface for a human paper as well: the

analytic deduction can be finished before reading the proof of any of A, B or C, provided they are explicitly identified as input theorems.

The readable edition is worth using even when one wishes to present the compact proof. Its `INDEX.md` gives a theorem-level map of the major components and places them in relation to the background literature; `CONTENTS.md` is a file-level index. Its `GUIDE.md` describes the geometry and ranks of the various Hermitian spaces, notation for periods and twists, and the quantitative scales. The readable edition's final reduction is in `New/Reduction/`, with the main inequality in `MainInequality.lean`; its chapters on derivatives, upper bound, spanning, determinant, one lift, and congruences shadow the major parts of these notes. One must not assume that identically purposed statements in the two variants have literally identical hypotheses or proofs.

9.2 How much has been established in this document?

The analytic calculations in Section 2, the comparison of the three bounds, the prime-sum identity and the final exponent in Section 8 are intended to be read as mathematical arguments with explicit premises. For the intermediate chapters, we have separated the stated technical inputs from deductions that can be presented with little formalism. In particular, the explanation of an iterated seesaw, a relative second moment, a theta exchange identity or a Selmer argument is not in itself a proof of the required uniform quantitative theorem. A specialist research article would have to replace each such imported assertion by either a rigorous reference applicable in precisely these conventions or an actual proof.

The formal source divides its results between `Lib`, the background reconstructed because it is absent from Mathlib in the required form, and `New`, the peculiar constructions and estimates of this proof. In the compact edition there are 573 background files and 219 new files; in the readable edition 495 and 306, respectively. The corresponding approximate numbers of lines of code are 76,000 and 300,000 for the entire source, rather than the number of lines of a conventional mathematical proof.

These counts explain an opportunity for an ordinary written exposition. A mathematical paper can cite the standard theory and write a single conceptual lemma where the Lean proof needs many definitions and tactical lemmas. On the other hand, the classification as background is not itself evidence that a particular assertion can be taken unchanged from a paper. The compact edition's `REFERENCES.md` states that of 196 citations checked to varying degrees, 57 were compared directly with their source and located, 30 were located at second hand, 102 identify a work without a precise location, and 7 were not compared. Citations in additional files were outside that check. In particular, the reference list is a useful starting point rather than an audited bibliography guaranteeing each exact result.

9.3 The chief mathematical blocks for a specialist treatment

The following issues seem to us to have the highest priority in expanding this exposition. Each is central enough that the relevant theorem should have a clean, standalone mathematical statement with all uniformities made explicit.

1. For A, isolate the divisibility theorem for jets of the CM theta kernel and explain why it survives the indefinite-to-definite seesaw. In particular, the loss and repair at exceptional inert primes must be controlled on the actual lattices of source forms, not just for formal polynomials.

2. For B, identify the precise relative second moment, the contributions to its diagonal or main term, and the uniform upper bound for non-main terms. State the finite-group type

restrictions in a way which makes the dimension estimate transparent.

3. For C, prove the determinant comparison along the entire chain, including the factorization into single-lift volumes and the cancellation of the auxiliary-period factors. Supply the exact paired coefficient-lattice structures to which the determinant arguments apply.

4. For the fit-family theorem, check the passage from congruences of automorphic forms to cohomology classes, the simultaneous choice of twists, the uniform Selmer and crystalline local-loss bounds, and their summation over rational primes. This is one of the places where standard background theorems stated without quantitative refinements would not be enough.

5. Audit the published inputs of the reconstructed libraries, particularly the integral specializations of theta identities, the attaching of Galois representations, comparison theorems and the relevant local p-adic Hodge theory. Each citation must imply the actual statement, including its dependence on the growing ranks and weights.

A well organized full account should proceed by theorem packages rather than translate line-by-line. For instance, the proof of the main arithmetic inequality in Section 8 requires no internal knowledge of a CM formal group once the theorem about short forms has been proved. Conversely, the local jet argument should have a formulation which can be understood without following the final numerical proof of GRH. Keeping the interfaces narrow and precise should make the larger exposition easier to check and easier to shorten later.

Primary materials

- L. Alpöge, with Claude, *Remark* (announcement of the proof), announcement available at `https://alpo.ge/remark.pdf`. The proof attribution and link are given in the supplied repository.

- *GRH for Dirichlet L-functions, in Lean 4*, supplied repository snapshot `grh-lean-main`, containing `grh/` and `grh-compact/`. See the root `README.md`, and each edition's `README.md`, `GUIDE.md`, and `REFERENCES.md`. The readable edition also contains `INDEX.md` and `CONTENTS.md`. Apache 2.0, copyright 2026 Anthropic, PBC.

- The corresponding Mathlib definitions and results are imported by `Statements.lean`; the snapshot pins Lean 4.35.0 and Mathlib commit `56a005855ea17e0b0958f73ec801ae140b0a1679`. The repository's `scripts/PrintAxioms.lean` records the intended axioms check.

The material in the mathematical chapters is grounded in the supplied snapshot. This source list is not a substitute for the comprehensive standard-literature bibliography required by a complete specialist proof.