



Valentin Blomer · Gergely Harcos · Péter Maga · Djordje Milićević

The sup-norm problem for $GL(2)$ over number fields

Received August 12, 2016

Abstract. We solve the sup-norm problem for spherical Hecke–Maaß newforms of square-free level for the group $GL(2)$ over a number field, with a power saving over the local geometric bound simultaneously in the eigenvalue and the level aspect. Our bounds feature a Weyl-type exponent in the level aspect, they reproduce or improve upon all known special cases, and over totally real fields they are as strong as the best known hybrid result over the rationals.

Keywords. Sup-norm, automorphic form, arithmetic manifold, amplification, pre-trace formula, diophantine analysis, geometry of numbers

1. Introduction

1.1. The sup-norm problem

The sup-norm problem has taken a prominent position in recent years at the interface of automorphic forms and analytic number theory. It is inspired by a classical question in analysis about comparing two norms on an infinite-dimensional Hilbert space: given an eigenfunction ϕ on a locally symmetric space X with a large Laplace eigenvalue λ and $\|\phi\|_2 = 1$, what can be said about its sup-norm $\|\phi\|_\infty$?

This question is closely connected to the multiplicity of eigenvalues [41], and it is motivated by the correspondence principle of quantum mechanics, where the high energy limit $\lambda \rightarrow \infty$ provides a connection between classical and quantum mechanics. The sup-norm of an eigenform with large eigenvalue gives some information on the distribution

V. Blomer: Mathematisches Institut, Bunsenstr. 3-5, D-37073 Göttingen, Germany, and Mathematisches Institut, Endenicher Allee 60, D-53177 Bonn, Germany; e-mail: blomer@math.uni-bonn.de

G. Harcos: Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences, POB 127, Budapest H-1364, Hungary, and MTA Rényi Intézet Lendület Automorphic Research Group, and Central European University, Nador u. 9, Budapest H-1051, Hungary; e-mail: gharcos@renyi.hu, harcosg@ceu.edu

P. Maga: Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences, POB 127, Budapest H-1364, Hungary, and MTA Rényi Intézet Lendület Automorphic Research Group; e-mail: magapeter@gmail.com

D. Milićević: Department of Mathematics, Bryn Mawr College, 101 North Merion Avenue, Bryn Mawr, PA 19010, USA; e-mail: dmilicevic@brynmawr.edu

Mathematics Subject Classification (2010): Primary 11F72, 11F55, 11J25

of its mass on X , which sheds light on the question to what extent these eigenstates can localize (“scarring”). Despite a lot of work from different points of view, in the case of a classically chaotic Hamiltonian (for instance when X is a compact hyperbolic manifold), the relation between the classical mechanics and the quantum mechanics in the semi-classical limit is currently not well-understood. This goes by the name of *quantum chaos*. We refer the reader to the excellent surveys [39, 40] and the references therein for an introduction to this topic and further details.

Purely analytic techniques can be used to give a best-possible solution to the sup-norm problem on a general compact locally symmetric space X of dimension d and rank r [41]:

$$\|\phi\|_\infty \ll_X \lambda^{(d-r)/4}, \quad (1.1)$$

and this bound is sharp as it is attained, for instance, for the round sphere. (The symbol \ll is introduced formally at the end of Section 2.) The bound is local in nature, in that its proof is insensitive to the global geometry of X , and in general it still allows for significant concentration of mass at individual points. In many cases, in particular for compact hyperbolic manifolds, a stronger bound is expected. The *sup-norm problem* aims at decreasing the exponent in (1.1) or in a refined version thereof.

The beauty of the sup-norm problem lies in particular in the fact that it is amenable to *arithmetic* techniques when the manifold is equipped with additional arithmetic structure. Two classical examples in dimension 2 are the round sphere $X = \mathcal{S}^2 = \mathrm{SO}_3(\mathbb{R})/\mathrm{SO}_2(\mathbb{R})$, realized as a quotient of the projective group of units in the Hamilton quaternions, and the modular surface $X = \mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}^2$, where \mathcal{H}^2 denotes the Poincaré upper half-plane of complex numbers on which $\mathrm{SL}_2(\mathbb{R})$ acts by hyperbolic isometries. In both cases, there is an arithmetically defined family of Hecke operators commuting with the Laplacian, so that it makes sense to consider joint eigenfunctions. A combination of analytic and arithmetic techniques led to a significant improvement of (1.1) for joint Hecke–Laplace eigenfunctions on these two arithmetic surfaces [24, 47]:

$$\|\phi\|_\infty \ll_\varepsilon \lambda^{5/24+\varepsilon} \quad \text{for every } \varepsilon > 0.$$

For applications, for instance in connection with Faltings’ delta function [25–27], it is also important to consider the dependence on X of the implied constant in (1.1), in particular as X varies through a sequence of covers. A typical situation is the case of the congruence covers

$$\Gamma_0(N)\backslash\mathcal{H}^2 \rightarrow \mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}^2, \quad (1.2)$$

where $\Gamma_0(N)$ is the usual Hecke congruence subgroup consisting of 2×2 integral unimodular matrices with lower left entry divisible by N .

Following the original breakthrough of Iwaniec and Sarnak [24], a lot of effort went into proving good upper bounds (and also lower bounds, but this is not the focus of the present paper) for joint eigenfunctions, in a great variety of situations and with various applications in mind: see for instance [3–7, 10, 18–20, 28, 31, 38, 45, 46, 51]. The results fall roughly into two categories. On the one hand, one can try to establish bounds as strong as possible. Somewhat reminiscent of the *subconvexity problem* in the theory of L -functions, this often leads eventually to a “natural” exponent that marks the limit of

techniques of analytic number theory. For example, in the case of the congruence cover (1.2) for N a square-free integer, Templier [46] proved the important benchmark result

$$\|\phi\|_\infty \ll_\varepsilon \lambda^{5/24+\varepsilon} N^{1/3+\varepsilon}, \quad (1.3)$$

improving simultaneously¹ in both aspects on the generic local bound $\lambda^{1/4} N^{1/2}$. On the other hand, one can also confine oneself to some small numerical improvement over the trivial bound, but use techniques that work on very general spaces X . Here the most general available result is due to Marshall [31] for semisimple split Lie groups over totally real fields and their totally imaginary quadratic extensions (CM-fields).

1.2. General number fields

In this paper, we address both points of view, and for the first time we address the sup-norm problem for the group GL_2 over a general number field F of degree $n = r_1 + 2r_2$ over \mathbb{Q} , with r_1 real embeddings and r_2 conjugate pairs of complex embeddings. From the perspective of automorphic forms, this is certainly the natural framework, and there is little reason to treat the ground field \mathbb{Q} separately. The underlying manifold is then a quotient of the product of r_1 copies of the upper half-plane \mathcal{H}^2 and r_2 copies of the upper half-space \mathcal{H}^3 , so it has dimension $d = 2r_1 + 3r_2$ and rank $r = r_1 + r_2$ (cf. (1.1)). As is well-known, the passage from \mathbb{Q} to a general number field introduces two abelian groups, the finite class group and the (except for the imaginary quadratic case) infinite unit group. As has been observed in many contexts (e.g. in the context of cubic hypersurfaces [8] and the Ramanujan conjecture [2]), these groups cause considerable technical difficulties for arguments of analytic number theory; the general strategy is always to use an adelic treatment to deal with issues of the class group and to use carefully chosen units in order to work with algebraic integers whose size is comparable in all archimedean embeddings. Our paper provides a general adelic counting scheme for such situations (see Section 6).

However, in our case the difficulties go much deeper than dealing with the class group and the unit group. As soon as F has a complex place, the formalism of the amplified pre-trace formula leads to counting integral matrices $\gamma \in M_2(\mathfrak{o}_F)$, which lie suitably close to a certain maximal compact subgroup of $GL_2(F_\infty)$, and whose entries are described by conditions involving real and imaginary parts at each complex place separately. If F is not a CM-field, there is no global complex conjugation (see e.g. [36]), and hence the global counting techniques that work over number fields like \mathbb{Q} or $\mathbb{Q}(i)$ break down in the general situation. In fact, the maximal compact subgroups of $GL_2(F_\infty)$ cannot be defined over F unless F is a totally real field or a CM-field.

Another difficulty is signified by a fundamental difference between $PSL_2(\mathbb{R})$ and $PSL_2(\mathbb{C})$. On the one hand, every arithmetic Fuchsian subgroup of $PSL_2(\mathbb{R})$ is commensurable with $SO^+(L)$ for a suitable lattice L in a quadratic space V of signature $(2, 1)$, upon identifying $SO^+(V)$ with $PSL_2(\mathbb{R})$. On the other hand, an arithmetic Kleinian subgroup of $PSL_2(\mathbb{C})$ is commensurable with $SO^+(M)$ for a suitable lattice M in a quadratic space W of signature $(3, 1)$, upon identifying $SO^+(W)$ with $PSL_2(\mathbb{C})$, if and

¹ See also Remark 7 in Subsection 10.3.

only if it contains a nonelementary Fuchsian subgroup [29, Theorem 10.2.3]. These special Kleinian subgroups are already known to behave distinctively for the sup-norm problem [33], and they can be described in terms of the invariant trace fields and quaternion algebras; in particular, their trace field is a quadratic extension of the maximal totally real subfield.

For a general number field F , these structural features make the sup-norm problem in many ways a very different problem. Therefore, we introduce a number of new devices into the argument to leverage the specific interplay between the maximal compact subgroups of $\mathrm{GL}_2(F_\infty)$ and the arithmetic of F . In the hardest situation in our counting problem, F is not totally real, and the field element $\xi := \mathrm{tr}(\gamma)^2/\mathrm{det}(\gamma)$ is bounded in F_∞ and very close to being totally real. In this case, we combine two observations that appear to be novel in this context. On the one hand, we exploit a certain *rigidity of number fields* (see Section 7) to show that ξ lies in a *proper* subfield of F . However, the denominator of ξ is arithmetically controlled by our specific *amplifier* (see Section 9), so $\xi \in F$ must be an algebraic integer. This is already a very strong conclusion when coupled with the boundedness of ξ in F_∞ ; however, except for special number fields F , we do not know how to deal with the non-parabolic cases $\xi \neq 4$. On the other hand, by artificially extending the spectrum, we can improve the performance of the pre-trace formula on the geometric side so that $\gamma \in \mathrm{M}_2(\mathfrak{o}_F)$ is also localized modulo some auxiliary ideal \mathfrak{q} . Specifically, we can ensure that γ is *locally parabolic* modulo \mathfrak{q} . As a result, $\xi \in 4 + \mathfrak{q}$, which forces $\xi = 4$ when the norm of \mathfrak{q} is large. In conclusion, in the hardest situation we can eliminate all but parabolic matrices, which are relatively simple to count. We refer the reader to Lemma 17 for a precise version of this argument, as well as to Lemma 16 for another application of the realness rigidity of number fields.

The precise setup of extending the spectrum and hence localizing γ modulo \mathfrak{q} is described in Sections 2 and 3, with a special view toward treating the units in \mathfrak{o}_F efficiently. Indeed, there is a natural ambiguity of $\mathrm{det}(\gamma)$ by units modulo squared units, while our congruence conditions force the units that appear here to be quadratic residues modulo \mathfrak{q} ; we can choose \mathfrak{q} in such a way that these units are automatically squared units. Thus the success of our method rests on three pillars: passage to a suitably chosen congruence subgroup, a carefully designed amplifier equipped with arithmetic features as described in Subsection 9.3, and the rigidity results for number fields mentioned above. At the technical level, we rely heavily on Atkin–Lehner operators (see Section 4) and the geometry of numbers (see Section 5), which allow an efficient counting of the matrices γ in Section 10.

In retrospect, the general idea of extending the spectrum to thin out the geometric side of the pre-trace formula is not unprecedented, the most spectacular example being Iwaniec’s approach [21] to the Ramanujan conjecture for the metaplectic group (see also [44] for another example). We believe that our variation of it, based on arithmetic properties of a certain congruence subgroup and the underlying number field, introduces a novel and flexible tool into the machinery of the sup-norm problem that may be useful in other situations.

As another useful feature, our argument also uses positivity more strongly than the previous treatments. Rather than carrying out an exact spectral average, we use positivity

of our operators to establish a *pre-trace inequality*. This streamlines the argument substantially, e.g. we do not even have to mention Eisenstein series and oldforms. A similar idea in the context of infinite volume subgroups was used by Gamburd [15].

Finally, we mention that in Section 8 we develop a uniform Fourier bound for spherical Hecke–Maaß newforms for the group GL_2 over a number field, which might be of independent interest.

1.3. Main results

Our main result is a solution of the sup-norm problem for GL_2 over any number field simultaneously in the eigenvalue and the level aspect, provided the level is square-free. In certain cases, we recover a Weyl-type saving, the strongest bound one can expect with the current technology. To formulate our results, we introduce the tuple

$$\lambda := (\lambda_1, \dots, \lambda_{r_1}, \lambda_{r_1+1}, \dots, \lambda_{r_1+r_2})$$

of Laplace eigenvalues at the r_1 real places and the r_2 complex places, and we write

$$|\lambda|_\infty := |\lambda|_{\mathbb{R}} \cdot |\lambda|_{\mathbb{C}}, \quad |\lambda|_{\mathbb{R}} := \prod_{j=1}^{r_1} \lambda_j, \quad |\lambda|_{\mathbb{C}} := \prod_{j=r_1+1}^{r_1+r_2} \lambda_j^2. \quad (1.4)$$

As usual, empty products are defined to be 1. We also denote by $\mathcal{N}\mathfrak{n}$ the norm of an integral ideal \mathfrak{n} (see Section 2 for further notation).

In classical language, we are looking at a cusp form ϕ on a congruence manifold X (see Section 2 for precise definitions). The connected components of X correspond to the ideal classes of F : they are left quotients of $(\mathcal{H}^2)^{r_1} \times (\mathcal{H}^3)^{r_2}$ by $\Gamma_0(\mathfrak{n})$ and related level \mathfrak{n} subgroups (cf. [42]). Assuming that $\|\phi\|_2 = 1$ with respect to the *probability measure* coming from invariant measures on \mathcal{H}^2 and \mathcal{H}^3 , the generic local bound reads

$$\|\phi\|_\infty \ll_{F,\varepsilon} |\lambda|_\infty^{1/4+\varepsilon} (\mathcal{N}\mathfrak{n})^{1/2+\varepsilon}.$$

Theorem 1. *Let ϕ be an L^2 -normalized Hecke–Maaß cuspidal newform on GL_2 over F of square-free level \mathfrak{n} and trivial central character. Suppose that ϕ is spherical at the archimedean places. Then for any $\varepsilon > 0$ we have*

$$\|\phi\|_\infty \ll_{F,\varepsilon} |\lambda|_\infty^{5/24+\varepsilon} (\mathcal{N}\mathfrak{n})^{1/3+\varepsilon} + |\lambda|_{\mathbb{R}}^{1/8+\varepsilon} |\lambda|_{\mathbb{C}}^{1/4+\varepsilon} (\mathcal{N}\mathfrak{n})^{1/4+\varepsilon}.$$

We emphasize that this result is new with any exponent less than $1/2$ over $\mathcal{N}\mathfrak{n}$, any exponent less than $1/4$ over $|\lambda|_{\mathbb{R}}$, and for any number field F other than \mathbb{Q} and $\mathbb{Q}(i)$ (cf. [3, 46]). In particular, for totally real fields this is the proper analogue of (1.3). In view of the above remarks on the difficulties with general number fields, it is remarkable that the methods in the level aspect—which historically appeared to be the harder parameter—are flexible enough to produce a Weyl-type exponent in a general setup.

For a general number field F , the strength of Theorem 1 in the eigenvalue aspect $|\lambda|_\infty$ depends on the relative sizes of $|\lambda|_{\mathbb{R}}$ and $|\lambda|_{\mathbb{C}}$. It is particularly strong for totally real fields; for other fields, it fails to solve the sup-norm problem when $|\lambda|_{\mathbb{C}}$ gets large relative to $|\lambda|_{\mathbb{R}}$ and $\mathcal{N}\mathfrak{n}$. The next theorem, in which F_0 denotes the maximal totally real subfield of F , fixes this issue by saving in all aspects for any number field other than a totally real field.

Theorem 2. *Suppose that $[F : F_0] \geq 2$. Then, under the same assumptions as in Theorem 1, we have*

$$\|\phi\|_\infty \ll_{F,\varepsilon} (|\lambda|_\infty^{1/2} \mathcal{N}\mathfrak{n})^{\frac{1}{2} - \frac{1}{8[F:F_0]-4} + \varepsilon}.$$

In the special case $[F : F_0] = 2$ this bound reads

$$\|\phi\|_\infty \ll_{F,\varepsilon} |\lambda|_\infty^{5/24+\varepsilon} (\mathcal{N}\mathfrak{n})^{5/12+\varepsilon},$$

so Theorem 2 improves on [3, Theorems 2–3] even in the case $F = \mathbb{Q}(i)$, and the proof differs substantially in several aspects. Further, Theorem 2 with any exponent less than $1/4$ over $|\lambda|_\infty$ is new for any non-CM-field. For a sequence of fields with $[F : F_0] \rightarrow \infty$, the exponents of $|\lambda|_\infty$ and $\mathcal{N}\mathfrak{n}$ degenerate to $1/4$ and $1/2$, respectively, but this defect only impacts the $|\lambda|_\mathbb{C}$ -aspect due to the uniform exponents in Theorem 1. It would be desirable to treat all number fields on equal footing (as was accomplished in other contexts such as [2, 8]), but for that a new idea (or a completely new method) would be needed to handle more efficiently the difficulties described in the previous subsection.

Recently, Assing [1] extended Theorems 1 and 2 to arbitrary level and central character by combining the ideas of the present paper with the methods of Saha [38]. In Assing's results, the dependence on the Laplace eigenvalues and the square-free part of the level is the same as ours, and this is coupled with a rather good dependence on the (remaining) square part of the level and the conductor of the central character.

Convention. In this paper, we regard the number field F as being fixed, and we allow all implied constants to depend on it (unless we emphasize the opposite).

2. Basic setup and notation

Let F be a number field of degree $n = r_1 + 2r_2$ over \mathbb{Q} with ring of integers \mathfrak{o} and different ideal \mathfrak{d} . The completions F_v at the various places v are equipped with canonical norms (or modules) as in [50]. In particular, at an archimedean place v we have $|x|_v = |x|^{[F_v:\mathbb{R}]}$, where $|\cdot|$ denotes the usual absolute value. We reserve the symbol \mathfrak{p} for prime ideals of \mathfrak{o} , and we use it to label nonarchimedean places of F in the usual way. For each prime \mathfrak{p} , we fix a uniformizer $\varpi_{\mathfrak{p}} \in \mathfrak{o}_{\mathfrak{p}}$ of $\mathfrak{p}\mathfrak{o}_{\mathfrak{p}}$. As usual, we define the adèle ring of F as a restricted direct product

$$\mathbb{A} := F_\infty \times \mathbb{A}_{\text{fin}}, \quad F_\infty := \prod_{v|\infty} F_v, \quad \mathbb{A}_{\text{fin}} := \prod'_{\mathfrak{p}} F_{\mathfrak{p}},$$

and we write accordingly

$$|x|_{\mathbb{A}} := |x_\infty|_\infty \cdot |x_{\text{fin}}|_{\text{fin}}, \quad |x_\infty|_\infty := \prod_{v|\infty} |x_v|_v, \quad |x_{\text{fin}}|_{\text{fin}} := \prod_{\mathfrak{p}} |x_{\mathfrak{p}}|_{\mathfrak{p}}$$

for the module of an idele $x \in \mathbb{A}^\times$. We further decompose the archimedean module as

$$|x_\infty|_\infty = |x_\infty|_{\mathbb{R}} \cdot |x_\infty|_{\mathbb{C}}, \quad |x_\infty|_{\mathbb{R}} := \prod_{v \text{ real}} |x_v|, \quad |x_\infty|_{\mathbb{C}} := \prod_{v \text{ complex}} |x_v|^2. \quad (2.1)$$

This is consistent with (1.4). We introduce the following notation for the closure of \mathfrak{o} in \mathbb{A}_{fin} :

$$\hat{\mathfrak{o}} := \prod_{\mathfrak{p}} \mathfrak{o}_{\mathfrak{p}}. \quad (2.2)$$

We call a field element $x \in F^\times$ *totally positive* if $x_v > 0$ for every real place v . We denote the group of totally positive field elements by F_+^\times , and the group of totally positive units by \mathfrak{o}_+^\times . We choose a set of representatives $\theta_1, \dots, \theta_h \in \mathbb{A}_{\text{fin}}^\times$ for the ideal classes of F ; without loss of generality, they lie in $\hat{\mathfrak{o}}$.

As mentioned in the introduction, we can fix a square-free ideal $\mathfrak{q} \subseteq \mathfrak{o}$ in such a way that the only elements of \mathfrak{o}^\times that are quadratic residues modulo \mathfrak{q} are the elements of $(\mathfrak{o}^\times)^2$. Indeed, if u is a *nonsquare* unit, then $F(\sqrt{u})/F$ is one of the finitely many quadratic extensions corresponding to the square classes in $\mathfrak{o}^\times/(\mathfrak{o}^\times)^2$. Moreover, for any prime \mathfrak{p} that is inert in this extension, u is a quadratic nonresidue modulo \mathfrak{p} . So if we choose an inert prime for each of the above-mentioned extensions, and \mathfrak{q} is divisible by all these primes, then \mathfrak{q} has the required property. We fix such an ideal \mathfrak{q} once and for all, with the additional requirement that

$$\mathcal{N}\mathfrak{q} \geq 300^n. \quad (2.3)$$

We can clearly think of \mathfrak{q} as a function of F . (For concreteness, we could pick \mathfrak{q} so that its norm is minimal, and with additional constraints we could even pin down \mathfrak{q} uniquely.)

We fix a square-free ideal $\mathfrak{n} \subseteq \mathfrak{o}$, and we consider the corresponding global Hecke congruence subgroup

$$K := \prod_v K_v \quad \text{with} \quad K_v := \begin{cases} \text{O}_2(\mathbb{R}) & \text{for } v \text{ real,} \\ \text{U}_2(\mathbb{C}) & \text{for } v \text{ complex,} \\ \text{GL}_2(\mathfrak{o}_{\mathfrak{p}}) & \text{for } \mathfrak{p} \nmid \mathfrak{n}, \\ K_0(\mathfrak{p}\mathfrak{o}_{\mathfrak{p}}) & \text{for } \mathfrak{p} \mid \mathfrak{n}, \end{cases} \quad (2.4)$$

where

$$K_0(\mathfrak{p}\mathfrak{o}_{\mathfrak{p}}) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{GL}_2(\mathfrak{o}_{\mathfrak{p}}) : c \in \mathfrak{p}\mathfrak{o}_{\mathfrak{p}} \right\}$$

is the subgroup of $\text{GL}_2(\mathfrak{o}_{\mathfrak{p}})$ consisting of the matrices whose lower left entry is divisible by $\mathfrak{p}\mathfrak{o}_{\mathfrak{p}}$. As explained in the introduction, we need to enlarge our spectrum a bit. With this in mind, we introduce

$$K^b := \prod_v K_v^b \quad \text{with} \quad K_v^b := \begin{cases} K_v & \text{for } v \nmid \mathfrak{q}, \\ K_1(\mathfrak{p}\mathfrak{o}_{\mathfrak{p}}) & \text{for } \mathfrak{p} \mid \mathfrak{q}, \end{cases} \quad (2.5)$$

where

$$K_1(\mathfrak{p}\mathfrak{o}_{\mathfrak{p}}) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{GL}_2(\mathfrak{o}_{\mathfrak{p}}) : a - d \in \mathfrak{p}\mathfrak{o}_{\mathfrak{p}}, c \in \mathfrak{p}\mathfrak{o}_{\mathfrak{p}} \right\}$$

is the subgroup of $K_0(\mathfrak{p}\mathfrak{o}_{\mathfrak{p}})$ consisting of the matrices whose diagonal entries are congruent to each other modulo $\mathfrak{p}\mathfrak{o}_{\mathfrak{p}}$.

We fix a Haar measure on $\mathrm{GL}_2(\mathbb{A})$, and we use it to define the Hilbert space $L^2(\tilde{X})$, where \tilde{X} is the finite volume coset space²

$$\tilde{X} := \mathrm{GL}_2(F) \backslash \mathrm{GL}_2(\mathbb{A}) / Z(F_\infty). \quad (2.6)$$

All the other L^2 -spaces in this paper will be regarded or defined as Hilbert subspaces of $L^2(\tilde{X})$. We consider a spherical Hecke–Maaß newform ϕ on GL_2 over F of level \mathfrak{n} and trivial central character. By definition, $\phi : \mathrm{GL}_2(\mathbb{A}) \rightarrow \mathbb{C}$ is a left $\mathrm{GL}_2(F)$ -invariant and right $Z(\mathbb{A})K$ -invariant function that generates an irreducible cuspidal representation π of GL_2 over F of conductor \mathfrak{n} . It spans the one-dimensional newspace π^K , and it corresponds to a pure tensor $\otimes_v \phi_v$ of local newvectors $\phi_v \in \pi_v^{K_v}$ (cf. [34, Cor. 2], [13, Th. 4], [9, Th. 1]). In particular, ϕ is a cuspidal eigenfunction of the Hecke algebra for K (as defined in Section 3). Inspired by Venkatesh [48, Subsection 2.3], we regard ϕ as a square-integrable function on the coset spaces

$$X := \tilde{X}/K \quad \text{and} \quad X^b := \tilde{X}/K^b. \quad (2.7)$$

More precisely, we identify $L^2(X)$ and $L^2(X^b)$ with the right K -invariant and right K^b -invariant subspaces of $L^2(\tilde{X})$ defined above.

In adelic treatments, one usually divides by $Z(\mathbb{A})$ instead of $Z(F_\infty)$, especially if the central character is assumed to be trivial. Dividing by the smaller group $Z(F_\infty)$ in (2.6) makes the spaces in (2.7) larger and separates the infinite part and the finite part nicely. The cost to pay is that one has to deal with a bigger automorphic spectrum: instead of the trivial central character, one needs to consider all ideal class characters as central characters.³ Introducing the ideal \mathfrak{q} , i.e. switching from X to X^b , allows one to work with Hecke operators of smaller support (see Section 3), which is immensely beneficial for our matrix counting scheme (see Section 10). However, this has a similar effect (already for $F = \mathbb{Q}$) of enlarging the automorphic spectrum. Indeed, the resulting quotients X and X^b are orbifolds with finitely many connected components. The connected components of X correspond to the ideal classes of F , while those of X^b correspond to certain cosets of the ray class group modulo \mathfrak{q} . More concretely, reduction modulo \mathfrak{q} embeds $U := \mathfrak{o}^\times / (\mathfrak{o}^\times)^2$ into $V := (\mathfrak{o}/\mathfrak{q})^\times / (\mathfrak{o}/\mathfrak{q})^{\times 2}$ by our choice of \mathfrak{q} , and each connected component of X is covered by exactly $[V : U]$ connected components of X^b under the natural covering map $X^b \rightarrow X$.

For a ramified place v (i.e. for $v = \mathfrak{p}$ dividing the level \mathfrak{n}), the matrix $A_v := \begin{pmatrix} \varpi_v & 1 \\ 0 & 1 \end{pmatrix} \in \mathrm{GL}_2(F_v)$ normalizes K_v . The group K_v^* generated by K_v and A_v contains the center $Z(F_v)$, and $K_v^*/Z(F_v)K_v$ has order 2. By multiplicity one and the assumption that the central character of ϕ is trivial, we infer that the right action of K_v^* on ϕ is given by a character $K_v^* \rightarrow \{\pm 1\}$. It follows that $|\phi|$ is right invariant by the global Atkin–Lehner group $K^* := \prod_v K_v^*$, where we put $K_v^* := Z(F_v)K_v$ for all unramified places v , including the archimedean ones. That is, for the purpose of studying the sup-norm $\|\phi\|_\infty$, we

² For any ring R , we denote by $Z(R)$ the matrix group $\left\{ \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} : a \in R^\times \right\}$.

³ This subtlety enters in (3.13), where we assume that $\mathrm{gcd}(l, m)$ is a product of principal prime ideals.

can regard $|\phi|$ as a square-integrable function on the coset space

$$X^* := \mathrm{GL}_2(F) \backslash \mathrm{GL}_2(\mathbb{A}) / K^*. \quad (2.8)$$

We emphasize again that we regard $L^2(X^*)$, $L^2(X)$, $L^2(X^b)$ as Hilbert subspaces of $L^2(\tilde{X})$, and we assume that $\|\phi\|_2 = 1$ in all these spaces.

We allow all implied constants depend on the number field F , hence also on the auxiliary ideal \mathfrak{q} chosen above for F . Accordingly, $A \ll B$ means that $|A| \leq C|B|$ holds for a constant $C = C(F) > 0$ depending on F , while $A \ll_S B$ means the same for a constant $C = C(F, S) > 0$ depending on F and S . If S is a list of quantities including ε , then it is implicitly meant that the bound holds for any sufficiently small $\varepsilon > 0$. The relation $A \asymp B$ means that $A \ll B$ and $B \ll A$ hold simultaneously, while $A \asymp_S B$ means that $A \ll_S B$ and $B \ll_S A$ hold simultaneously. Finally, inspired by [3, 19], we shall use the notation

$$A \preceq B \stackrel{\text{def}}{\iff} A \ll_\varepsilon B |\lambda|_\infty^\varepsilon (\mathcal{N}\mathfrak{n})^\varepsilon, \quad (2.9)$$

which will be in force for the rest of the paper.

3. Hecke algebras and the idea of amplification

While our newform ϕ lives naturally on the space X , and in fact $|\phi|$ is well-defined even on the space X^* , it is convenient to view ϕ as a function on X^b which is equipped with more suitable operators for the purpose of amplification.

The groups $Z(F_\infty)K$ and $Z(F_\infty)K^b$ contain the central subgroup (cf. (2.2), (2.4), (2.5))

$$Z := Z(F_\infty \hat{\mathfrak{o}}) = \prod_{v|\infty} Z(F_v) \prod_{\mathfrak{p}} Z(\mathfrak{o}_{\mathfrak{p}}), \quad (3.1)$$

hence we can identify X with $\Gamma \backslash G / K$, and X^b with $\Gamma \backslash G / K^b$, where

$$G := \mathrm{GL}_2(\mathbb{A}) / Z, \quad \Gamma := \mathrm{GL}_2(F)Z / Z \cong \mathrm{GL}_2(F) / Z(\mathfrak{o}). \quad (3.2)$$

In particular, we can identify the functions on X (resp. X^b) with those functions on G that are left Γ -invariant and right K -invariant (resp. right K^b -invariant). Accordingly, we have an inclusion of Hilbert spaces, each defined via the Haar measure that we fixed on $\mathrm{GL}_2(\mathbb{A})$,

$$L^2(X) \leq L^2(X^b) \leq L^2(\Gamma \backslash G) \leq L^2(\tilde{X}). \quad (3.3)$$

We define the $Z(F_\infty)$ -invariant norm

$$\|g_\infty\| := \prod_{v|\infty} \frac{|a_v|^2 + |b_v|^2 + |c_v|^2 + |d_v|^2}{2|a_v d_v - b_v c_v|}, \quad g_\infty = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(F_\infty),$$

and we say that $f : G \rightarrow \mathbb{C}$ is a *rapidly decaying smooth* function if the following properties hold for $g = g_\infty g_{\mathrm{fin}}$, where $g_\infty \in \mathrm{GL}_2(F_\infty)$ and $g_{\mathrm{fin}} \in \mathrm{GL}_2(\mathbb{A}_{\mathrm{fin}})$:

- f is compactly supported in g_{fin} , and it is locally constant in g_{fin} for any fixed g_∞ ;
- $\|g_\infty\|^N f(g)$ is bounded for any $N > 0$, and f is C^∞ in g_∞ for any fixed g_{fin} .

We denote by $C(G)$ the convolution algebra of these rapidly decaying smooth functions on G . For $f \in C(G)$ and $\psi \in L^2(\Gamma \backslash G)$, we consider the function $R(f)\psi \in L^2(\Gamma \backslash G)$ given by

$$\begin{aligned} (R(f)\psi)(x) &:= \int_G f(y)\psi(xy) dy = \int_G f(x^{-1}y)\psi(y) dy \\ &= \int_{\Gamma \backslash G} \left(\sum_{\gamma \in \Gamma} f(x^{-1}\gamma y) \right) \psi(y) dy. \end{aligned} \quad (3.4)$$

That is, $R(f)$ is an integral operator on $\Gamma \backslash G$ with kernel

$$k_f(x, y) := \sum_{\gamma \in \Gamma} f(x^{-1}\gamma y). \quad (3.5)$$

Then $R(f_1 * f_2) = R(f_1)R(f_2)$ for $f_1, f_2 \in C(G)$, and the adjoint of $R(f)$ equals $R(\check{f})$ with

$$\check{f}(g) := \overline{f(g^{-1})}, \quad g \in G.$$

We shall define convenient subalgebras of $C(G)$ in terms of the restricted product decomposition

$$G = \prod'_v G_v, \quad G_v := \begin{cases} \mathrm{GL}_2(F_v)/Z(F_v) & \text{for } v \mid \infty, \\ \mathrm{GL}_2(F_{\mathfrak{p}})/Z(\mathfrak{o}_{\mathfrak{p}}) & \text{for } v \nmid \infty. \end{cases} \quad (3.6)$$

We choose a Haar measure on each of the groups $\mathrm{GL}_2(F_v)$ so that their product is the Haar measure we fixed on $\mathrm{GL}_2(\mathbb{A})$ earlier, and the measure of $\mathrm{GL}_2(\mathfrak{o}_{\mathfrak{p}})$ within $\mathrm{GL}_2(F_{\mathfrak{p}})$ is 1. We define $C(G_v)$ and its action $f_v \mapsto R(f_v)$ on $L^2(\Gamma \backslash G)$ just as for $C(G)$, but with integration over G_v instead of G . The restricted tensor product of these algebras is the \mathbb{C} -span of pure tensors $\otimes_v f_v$ such that $f_v \in C(G_v)$ for all places v and $f_{\mathfrak{p}}$ is the characteristic function of $\mathrm{GL}_2(\mathfrak{o}_{\mathfrak{p}})$ for all but finitely many primes \mathfrak{p} . We regard this product as a subalgebra of $C(G)$ in the usual way, namely by identifying $\otimes_v f_v$ with the function $x \mapsto \prod_v f_v(x_v)$ so that also $R(\otimes_v f_v) = \prod_v R(f_v)$; the products are finite in the sense that the factors equal the identity for all but finitely many v 's.

We write $\hat{\mathrm{GL}}_2(\mathfrak{o}_{\mathfrak{p}})$ for $\mathrm{M}_2(\mathfrak{o}_{\mathfrak{p}}) \cap \mathrm{GL}_2(F_{\mathfrak{p}})$, and we define $\hat{K}_0(\mathfrak{p}\mathfrak{o}_{\mathfrak{p}})$ as the subsemigroup of $\hat{\mathrm{GL}}_2(\mathfrak{o}_{\mathfrak{p}})$ consisting of the matrices with lower left entry divisible by \mathfrak{p} and upper left entry coprime to \mathfrak{p} . In accordance with (2.4) and (2.5), we consider the following two open subsemigroups of G :

$$M := \prod_v M_v \quad \text{with} \quad M_v := \begin{cases} G_v & \text{for } v \mid \infty, \\ \hat{\mathrm{GL}}_2(\mathfrak{o}_{\mathfrak{p}})/Z(\mathfrak{o}_{\mathfrak{p}}) & \text{for } \mathfrak{p} \nmid \mathfrak{n}, \\ \hat{K}_0(\mathfrak{p}\mathfrak{o}_{\mathfrak{p}})/Z(\mathfrak{o}_{\mathfrak{p}}) & \text{for } \mathfrak{p} \mid \mathfrak{n}; \end{cases} \quad (3.7)$$

$$M^b := \prod_v M_v^b \quad \text{with} \quad M_v^b := \begin{cases} M_v & \text{for } v \nmid \mathfrak{q}, \\ K_1(\mathfrak{p}\mathfrak{o}_{\mathfrak{p}})/Z(\mathfrak{o}_{\mathfrak{p}}) & \text{for } \mathfrak{p} \mid \mathfrak{q}. \end{cases} \quad (3.8)$$

Note that M (resp. M^b) is left and right invariant by K (resp. K^b). Finally, we define the *Hecke algebra* for K , and the *unramified Hecke algebra* at \mathfrak{q} for K^b , as the restricted tensor products

$$\mathcal{H} := \bigotimes'_v \mathcal{H}_v \quad \text{with} \quad \mathcal{H}_v := C(K_v \backslash M_v / K_v),$$

$$\mathcal{H}^b := \bigotimes'_v \mathcal{H}_v^b \quad \text{with} \quad \mathcal{H}_v^b := C(K_v^b \backslash M_v^b / K_v^b).$$

These algebras have a unity element, unlike $C(G)$ or $C(G_v)$.

Note that \mathcal{H} acts on $L^2(X)$, and \mathcal{H}^b acts on $L^2(X^b)$, through $f \mapsto R(f)$. There is a \mathbb{C} -algebra embedding $\iota : \mathcal{H}^b \hookrightarrow \mathcal{H}$ such that any $f \in \mathcal{H}^b$ acts on the subspace $L^2(X)$ of $L^2(X^b)$ exactly as $\iota(f) \in \mathcal{H}$ does. We define this embedding as $\iota := \otimes_v \iota_v$, by choosing an appropriate \mathbb{C} -algebra embedding $\iota_v : \mathcal{H}_v^b \hookrightarrow \mathcal{H}_v$ at each place v . For $v \nmid \mathfrak{q}$, the local factors \mathcal{H}_v^b and \mathcal{H}_v are equal, so we choose ι_v to be the identity map. For $v \mid \mathfrak{q}$, the local factor \mathcal{H}_v^b is isomorphic to \mathbb{C} , so there is a unique choice for ι_v . From now on, we use the usual convention that the subscript “ ∞ ” collects the local factors at $v \mid \infty$, while the subscript “ fin ” collects the local factors at $v \nmid \infty$. Then, in particular, we can talk about the \mathbb{C} -algebra embedding $\iota_{\text{fin}} : \mathcal{H}_{\text{fin}}^b \hookrightarrow \mathcal{H}_{\text{fin}}$. Under this embedding, thinking of \mathcal{H}_{fin} (resp. $\mathcal{H}_{\text{fin}}^b$) as a subalgebra of $C(K_{\text{fin}} \backslash G_{\text{fin}} / K_{\text{fin}})$ (resp. $C(K_{\text{fin}}^b \backslash G_{\text{fin}} / K_{\text{fin}}^b)$), the constant function $\text{vol}(K_{\text{fin}}^b)^{-1}$ on a double coset $K_{\text{fin}}^b g K_{\text{fin}}^b \subset M_{\text{fin}}^b$ becomes the constant function $\text{vol}(K_{\text{fin}})^{-1}$ on the double coset $K_{\text{fin}} g K_{\text{fin}} \subset M_{\text{fin}}$. In the next two paragraphs, we introduce the Hecke operators for X in terms of \mathcal{H}_{fin} , and the (unramified at \mathfrak{q}) Hecke operators for X^b in terms of $\mathcal{H}_{\text{fin}}^b$. Our presentation is based to some extent on [42, Section 2] and [43, Ch. 3].

For any nonzero ideal $\mathfrak{m} \subseteq \mathfrak{o}$, we consider the Hecke operator $T_{\mathfrak{m}} := R(t_{\mathfrak{m}})$ on $L^2(X)$, where $t_{\mathfrak{m}} \in \mathcal{H}_{\text{fin}}$ is given by

$$t_{\mathfrak{m}}(x) := \begin{cases} (\mathcal{N}\mathfrak{m})^{-1/2} \text{vol}(K_{\text{fin}})^{-1} & \text{for } x \in M_{\text{fin}} \text{ and } (\det x)\mathfrak{o} = \mathfrak{m}, \\ 0 & \text{otherwise.} \end{cases} \quad (3.9)$$

We note that here the determinant of $x \in M_2(\hat{\mathfrak{o}})/Z(\hat{\mathfrak{o}})$ is an element of $\hat{\mathfrak{o}}/(\hat{\mathfrak{o}}^\times)^2$ rather than an element of $\hat{\mathfrak{o}}$, but still it determines a unique ideal in \mathfrak{o} that we denoted by $(\det x)\mathfrak{o}$. We also need the supplementary operator $S_{\mathfrak{m}} := R(s_{\mathfrak{m}})$ on $L^2(X)$, with $s_{\mathfrak{m}} \in \mathcal{H}_{\text{fin}}$ defined as follows. First we assume that \mathfrak{m} is coprime to the level \mathfrak{n} . We take any finite idele $\mu \in \mathbb{A}_{\text{fin}}^\times$ representing \mathfrak{m} , i.e. $\mathfrak{m} = \mu\mathfrak{o}$, and then we put

$$s_{\mathfrak{m}}(x) := \begin{cases} \text{vol}(K_{\text{fin}})^{-1} & \text{for } x \in \begin{pmatrix} \mu & \\ & \mu \end{pmatrix} K_{\text{fin}} / Z(\hat{\mathfrak{o}}), \\ 0 & \text{otherwise.} \end{cases} \quad (3.10)$$

The function $s_{\mathfrak{m}} \in \mathcal{H}_{\text{fin}}$ is independent of the representative μ , because K_{fin} contains $Z(\hat{\mathfrak{o}})$, and $S_{\mathfrak{m}}$ agrees with the right action of $\begin{pmatrix} \mu & \\ & \mu \end{pmatrix}$. In particular, $S_{\mathfrak{m}}$ commutes with the Hecke operators. Using the fact that $L^2(X)$ consists of functions invariant under $Z(F F_\infty \hat{\mathfrak{o}})$, we

see that S_m only depends on the ideal class of m , and it is the identity map whenever m is principal. If m and n are not coprime, then we define s_m (hence also S_m) to be zero. The Hecke operators commute with each other as a consequence of the following multiplicativity relation, valid for all nonzero ideals $l, m \subseteq \mathfrak{o}$ (cf. [42, (2.12)]):

$$t_l * t_m = \sum_{\mathfrak{f}|\gcd(l,m)} s_{\mathfrak{f}} * t_{lm/\mathfrak{f}^2}, \quad \text{therefore} \quad T_l T_m = \sum_{\mathfrak{f}|\gcd(l,m)} S_{\mathfrak{f}} T_{lm/\mathfrak{f}^2}. \quad (3.11)$$

In addition, if m is coprime to the level n , then $t_m = s_m * \check{t}_m$, whence T_m is a normal operator, and it is even self-adjoint when m is principal.

For any ideal $m \subseteq \mathfrak{o}$ coprime to q , we define the functions $t_m^b, s_m^b \in \mathcal{H}_{\text{fin}}^b$ in the same way as $t_m, s_m \in \mathcal{H}_{\text{fin}}$, but with K_{fin} and M_{fin} replaced by K_{fin}^b and M_{fin}^b (cf. (3.9)–(3.10)). In particular,

$$t_m^b(x) := \begin{cases} (\mathcal{N}m)^{-1/2} \text{vol}(K_{\text{fin}}^b)^{-1} & \text{for } x \in M_{\text{fin}}^b \text{ and } (\det x)\mathfrak{o} = m, \\ 0 & \text{otherwise.} \end{cases} \quad (3.12)$$

The corresponding operators on $L^2(X^b)$ are $T_m^b := R(t_m^b)$ and $S_m^b := R(s_m^b)$. Then in fact $\iota_{\text{fin}}(t_m^b) = t_m$ and $\iota_{\text{fin}}(s_m^b) = s_m$ under the \mathbb{C} -algebra embedding $\iota_{\text{fin}} : \mathcal{H}_{\text{fin}}^b \hookrightarrow \mathcal{H}_{\text{fin}}$, hence (3.11) implies the analogous relations

$$t_l^b * t_m^b = \sum_{\mathfrak{f}|\gcd(l,m)} s_{\mathfrak{f}}^b * t_{lm/\mathfrak{f}^2}^b, \quad \text{therefore} \quad T_l^b T_m^b = \sum_{\mathfrak{f}|\gcd(l,m)} S_{\mathfrak{f}}^b T_{lm/\mathfrak{f}^2}^b.$$

An important special case is when $\gcd(l, m)$ is a product of *principal prime ideals* not dividing nq . In this case, the above relations simplify to

$$t_l^b * t_m^b = \sum_{\mathfrak{f}|\gcd(l,m)} t_{lm/\mathfrak{f}^2}^b, \quad \text{therefore} \quad T_l^b T_m^b = \sum_{\mathfrak{f}|\gcd(l,m)} T_{lm/\mathfrak{f}^2}^b. \quad (3.13)$$

In addition, if m is coprime to nq , then $t_m^b = s_m^b * \check{t}_m^b$, whence T_m^b is a normal operator, and it is even self-adjoint when m is principal.

Let $f \in \mathcal{H}$ be arbitrary. As $\phi \in L^2(X)$ is a newform of level n , we have $R(f)\phi = c(f)\phi$ for some $c(f) \in \mathbb{C}$. The same also holds for $f \in \mathcal{H}^b$, because in this case $\iota(f) \in \mathcal{H}$, and $R(f)\phi = R(\iota(f))\phi$. Moreover, if $f = \otimes_v f_v$ is a pure tensor from \mathcal{H}^b , then $R(f_v)\phi = c(f_v)\phi$ for some $c(f_v) \in \mathbb{C}$, and $c(f) = \prod_v c(f_v)$; there is a similar decomposition $c(f) = c(f_{\infty})c(f_{\text{fin}})$ for partial tensors $f = f_{\infty} \otimes f_{\text{fin}} \in \mathcal{H}_{\infty}^b \otimes \mathcal{H}_{\text{fin}}^b$. In particular, ϕ is an eigenfunction of each Hecke operator T_m with eigenvalue

$$\lambda(m) := c(t_m), \quad (3.14)$$

and for m coprime q it is also an eigenfunction of T_m^b with the same eigenvalue.

Now we describe along these lines the idea of *amplification*, a technique pioneered by Duke, Friedlander, Iwaniec, and Sarnak [11, 14, 24] to prove efficient bounds for automorphic L -functions on the critical line, and also for $|\phi(g)|$ at a given $g \in \text{GL}_2(\mathbb{A})$. Assume that $f \in \mathcal{H}^b$ is such that the operator $R(f)$ on $L^2(X^b)$ is positive. Then the

eigenvalue $c(f)$ is nonnegative, and the orthogonal decomposition $L^2(X^b) = (\mathbb{C}\phi) + (\mathbb{C}\phi)^\perp$ is $R(f)$ -invariant (because $R(f)$ is self-adjoint). Any $\psi \in L^2(X^b)$ decomposes uniquely as $\psi = \psi_1 + \psi_2$, where $\psi_1 \in \mathbb{C}\phi$ and $\psi_2 \in (\mathbb{C}\phi)^\perp$, and therefore

$$\langle R(f)\psi, \psi \rangle = \langle R(f)\psi_1, \psi_1 \rangle + \langle R(f)\psi_2, \psi_2 \rangle \geq \langle R(f)\psi_1, \psi_1 \rangle.$$

On the right hand side, we have explicitly $\psi_1 = \langle \psi, \phi \rangle \phi$, hence flipping the two sides we obtain

$$c(f)|\langle \psi, \phi \rangle|^2 \leq \langle R(f)\psi, \psi \rangle.$$

The inner products and also $R(f)\psi$ can be expressed as integrals over X^b (cf. (3.3)–(3.5)), yielding

$$c(f) \int_{X^b \times X^b} \phi(x) \overline{\phi(y)} \psi(y) \overline{\psi(x)} dx dy \leq \int_{X^b \times X^b} k_f(x, y) \psi(y) \overline{\psi(x)} dx dy.$$

We can use this inequality to estimate the value $|\phi(g)|$ at the given point $g \in GL_2(\mathbb{A})$ as follows. Note that the integrals are over a rather concrete space: an orbifold with finitely many connected components. We take a basis of open neighborhoods $\{V\} \subset X^b$ of the point $\Gamma g Z K^b \in X^b$ (the image of the coset $gZ \in G$), and we let $\psi = \psi_V \in L^2(X^b)$ run through the corresponding characteristic functions. Then by continuity, as V approaches the point $\Gamma g Z K^b \in X^b$,

$$c(f)(|\phi(g)|^2 + o(1)) \text{vol}(V \times V) \leq (k_f(g, g) + o(1)) \text{vol}(V \times V).$$

We conclude

$$c(f)|\phi(g)|^2 \leq k_f(g, g) = \sum_{\gamma \in \Gamma} f(g^{-1}\gamma g). \quad (3.15)$$

This is the pre-trace inequality mentioned in the introduction. The idea of amplification is to find, in terms of ϕ , a positive operator $R(f)$ as above such that $c(f)$ is relatively large, while the right hand side is relatively small. By dividing the last inequality by $c(f)$, we see that such an operator gives rise to an upper bound for $|\phi(g)|$. We note that the above argument goes back to Mercer [32]; see especially the end of Section 6 in his paper, and see also [37, Section 98] for a modern account.

4. Iwasawa decomposition modulo Atkin–Lehner operators

In the next two sections, we establish a nice fundamental domain for the space X^* (cf. (2.8)), which is the natural habitat of $|\phi|$. We start by developing a variant of the usual Iwasawa decomposition for $GL_2(F_v)$. The results are probably known to experts.

First we recall the action of $GL_2(\mathbb{R})$ on the hyperbolic plane \mathcal{H}^2 , and the action of $GL_2(\mathbb{C})$ on the hyperbolic 3-space \mathcal{H}^3 . We identify \mathcal{H}^2 with a half-plane in the set of complex numbers $\mathbb{C} = \mathbb{R} + \mathbb{R}i$,

$$\mathcal{H}^2 := \{x + yi : x \in \mathbb{R}, y > 0\} \subset \mathbb{C}. \quad (4.1)$$

A matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2^+(\mathbb{R})$ of positive determinant maps $P \in \mathcal{H}^2$ to $(aP + b)(cP + d)^{-1} \in \mathcal{H}^2$, while $\begin{pmatrix} -1 & \\ & 1 \end{pmatrix} \in \mathrm{GL}_2(\mathbb{R})$ maps it to $-\overline{P} \in \mathcal{H}^2$. This determines a transitive action of $\mathrm{GL}_2(\mathbb{R})$ on \mathcal{H}^2 , and by examining the stabilizer of the point $i \in \mathcal{H}^2$, we see that

$$\mathcal{H}^2 \cong \mathrm{GL}_2(\mathbb{R}) / \mathrm{Z}(\mathbb{R}) \mathrm{O}_2(\mathbb{R}). \quad (4.2)$$

Similarly, we identify \mathcal{H}^3 with a half-space in the set of Hamilton quaternions $\mathbb{H} = \mathbb{R} + \mathbb{R}i + \mathbb{R}j + \mathbb{R}k$,

$$\mathcal{H}^3 := \{x + yj : x \in \mathbb{C}, y > 0\} \subset \mathbb{H}. \quad (4.3)$$

A matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2^+(\mathbb{C})$ of positive real determinant maps a point $P \in \mathcal{H}^3$ to $(aP + b)(cP + d)^{-1} \in \mathcal{H}^3$, while any central element $\begin{pmatrix} a & \\ & a \end{pmatrix} \in \mathrm{GL}_2(\mathbb{C})$ fixes it. This determines a transitive action of $\mathrm{GL}_2(\mathbb{C})$ on \mathcal{H}^3 , and by examining the stabilizer of the point $j \in \mathcal{H}^3$, we see that

$$\mathcal{H}^3 \cong \mathrm{GL}_2(\mathbb{C}) / \mathrm{Z}(\mathbb{C}) \mathrm{U}_2(\mathbb{C}). \quad (4.4)$$

The following two lemmas provide explicit local Iwasawa decompositions; in particular, Lemma 1 (with $y > 0$) explicates the isomorphisms (4.2) and (4.4). Recall that $K_v^* = \mathrm{Z}(F_v)K_v$ for $v \mid \infty$.

Lemma 1. *Let $v \mid \infty$ be an archimedean place. Any matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(F_v)$ can be decomposed as*

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} y & x \\ & 1 \end{pmatrix} k, \quad (4.5)$$

where⁴ $\begin{pmatrix} y & x \\ & 1 \end{pmatrix} \in \mathrm{P}(F_v)$ and $k \in K_v^*$. Moreover, the absolute value of y is uniquely determined by

$$|y| = \frac{|ad - bc|}{|c|^2 + |d|^2}. \quad (4.6)$$

Proof. Existence with a unique $y > 0$ is clear from our remarks above, especially from (4.2) and (4.4). Equation (4.6) is well-known: we multiply both sides of (4.5) with its conjugate transpose. We have $k \in \begin{pmatrix} u & \\ & u \end{pmatrix} K_v$ for some $u \in F_v^\times$, and then we get

$$\begin{pmatrix} |a|^2 + |b|^2 & * \\ * & |c|^2 + |d|^2 \end{pmatrix} = \begin{pmatrix} |u|^2 & \\ & |u|^2 \end{pmatrix} \begin{pmatrix} |y|^2 + |x|^2 & x \\ \bar{x} & 1 \end{pmatrix}.$$

It follows that $|u|^2 = |c|^2 + |d|^2$, while taking the determinant of both sides reveals that $|ad - bc|^2 = |u|^4 |y|^2$, and the claim follows. \square

⁴ For any ring R , we denote by $\mathrm{P}(R)$ the matrix group $\left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} : a, d \in R^\times, b \in R \right\}$.

Lemma 2. *Let $v = \mathfrak{p}$ be a nonarchimedean place. Any matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL_2(F_v)$ can be decomposed as*

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} y & x \\ & 1 \end{pmatrix} k, \quad (4.7)$$

where $\begin{pmatrix} y & x \\ & 1 \end{pmatrix} \in P(F_v)$ and $k \in K_v^*$. Moreover, the \mathfrak{p} -adic absolute value of y is uniquely determined by

$$|y|_v = \begin{cases} |(ad - bc)/\gcd(c, d)^2|_v & \text{when } |c|_v < |d|_v \text{ or } \mathfrak{p} \nmid \mathfrak{n}, \\ |\varpi_{\mathfrak{p}}(ad - bc)/\gcd(c, d)^2|_v & \text{when } |c|_v \geq |d|_v \text{ and } \mathfrak{p} \mid \mathfrak{n}. \end{cases} \quad (4.8)$$

Here, $\gcd(c, d)$ stands for any generator of the $\mathfrak{o}_{\mathfrak{p}}$ -ideal $c\mathfrak{o}_{\mathfrak{p}} + d\mathfrak{o}_{\mathfrak{p}}$. Similarly, the image of k in the group $K_v^*/Z(F_v)K_v$ is uniquely determined, namely

$$k \in \begin{cases} Z(F_v)K_v & \text{when } |c|_v < |d|_v \text{ or } \mathfrak{p} \nmid \mathfrak{n}, \\ Z(F_v)K_v A_v & \text{when } |c|_v \geq |d|_v \text{ and } \mathfrak{p} \mid \mathfrak{n}. \end{cases} \quad (4.9)$$

Proof. First we show that a decomposition of the form (4.7) exists with a y -coordinate satisfying (4.8) and a k -component satisfying (4.9). We start with the decomposition, valid for $d \neq 0$,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} (ad - bc)/d^2 & b/d \\ & 1 \end{pmatrix} \begin{pmatrix} d & \\ & c \end{pmatrix}.$$

This is of the form (4.7) as long as $|c|_v \leq |d|_v$ and $\mathfrak{p} \nmid \mathfrak{n}$, or $|c|_v < |d|_v$ and $\mathfrak{p} \mid \mathfrak{n}$ (because \mathfrak{n} is square-free). Moreover, the y -coordinate equals here $(ad - bc)/d^2$, and also $c\mathfrak{o}_{\mathfrak{p}} + d\mathfrak{o}_{\mathfrak{p}} = d\mathfrak{o}_{\mathfrak{p}}$ by $|c|_v \leq |d|_v$, which verifies (4.8) for this particular case. Similarly, the k -component equals here $\begin{pmatrix} d & \\ & c \end{pmatrix} \in Z(F_v)K_v$, so that (4.9) holds as well. The two cases in which we have established (4.7) can be summarized as the case of $|c/w|_v \leq |d|_v$, where we put

$$w := \begin{cases} 1 & \text{for } \mathfrak{p} \nmid \mathfrak{n}, \\ \varpi_{\mathfrak{p}} & \text{for } \mathfrak{p} \mid \mathfrak{n}. \end{cases}$$

Assume now that we are in the complementary case $|c/w|_v > |d|_v$ (including the case $d = 0$), so that in particular $|c|_v \geq |d|_v$. We consider the decomposition

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} b & a/w \\ d & c/w \end{pmatrix} \begin{pmatrix} & 1 \\ w & \end{pmatrix}.$$

By our initial case, the first factor on the right hand side has a suitable decomposition

$$\begin{pmatrix} b & a/w \\ d & c/w \end{pmatrix} = \begin{pmatrix} y & x \\ & 1 \end{pmatrix} k$$

with $k \in Z(F_v)K_v$, $x \in F_v$, and

$$y := \frac{(ad - bc)/w}{(c/w)^2} = \frac{w(ad - bc)}{c^2}. \quad (4.10)$$

Therefore,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} y & x \\ & 1 \end{pmatrix} k \begin{pmatrix} & \\ w & 1 \end{pmatrix},$$

and this is a suitable decomposition upon regarding the product of the last two factors as a single element $\tilde{k} \in K_v^*$. Moreover, $c\mathfrak{o}_p + d\mathfrak{o}_p = c\mathfrak{o}_p$ by $|c|_v \geq |d|_v$, hence (4.10) satisfies (4.8) for this particular case. Similarly, $k \in Z(F_v)K_v$ shows that $\tilde{k} \in Z(F_v)K_v$ or $\tilde{k} \in Z(F_v)K_vA_v$ depending on whether $p \nmid n$ or $p \mid n$, and so (4.9) holds as well.

Now we prove that the p -adic absolute value $|y|_v$ and the image of k in the group $K_v^*/Z(F_v)K_v$ are constant along all decompositions (4.7) of a given matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ in $\mathrm{GL}_2(F_v)$. In other words,

$$\begin{pmatrix} y_1 & x_1 \\ & 1 \end{pmatrix} k_1 = \begin{pmatrix} y_2 & x_2 \\ & 1 \end{pmatrix} k_2$$

for $y_1, y_2 \in F_v^\times$, $x_1, x_2 \in F_v$, $k_1, k_2 \in K_v^*$ implies that

$$y_1/y_2 \in \mathfrak{o}_p^\times \quad \text{and} \quad k_2 k_1^{-1} \in Z(F_v)K_v.$$

Rearranging, we get

$$\begin{pmatrix} y_1/y_2 & (x_1 - x_2)/y_2 \\ & 1 \end{pmatrix} = \begin{pmatrix} y_2 & x_2 \\ & 1 \end{pmatrix}^{-1} \begin{pmatrix} y_1 & x_1 \\ & 1 \end{pmatrix} = k_2 k_1^{-1}.$$

In particular, both sides lie in $P(F_v) \cap K_v^*$ which, by inspection, equals $Z(F_v)P(\mathfrak{o}_p)$. It follows that $y_1/y_2 \in \mathfrak{o}_p^\times$ and $k_2 k_1^{-1} \in Z(F_v)P(\mathfrak{o}_p) \subset Z(F_v)K_v$. \square

5. Geometry of numbers and the fundamental domain

We start this section with a simple but useful observation about balancing infinite ideles with units.

Lemma 3. *Let $y, z \in F_\infty^\times$ be two infinite ideles such that $|y|_\infty = |z|_\infty$. Then for any positive integer m , there is an m -th powered unit $t \in (\mathfrak{o}^\times)^m$ such that*

$$|ty_v|_v \asymp_m |z_v|_v, \quad v \mid \infty.$$

Proof. We fix m , and we look for $t \in (\mathfrak{o}^\times)^m$ in the form $t = u^m$ with $u \in \mathfrak{o}^\times$. The infinite idele $s := z/y \in F_\infty^\times$ satisfies $|s|_\infty = 1$, and the conclusion can be rewritten as

$$m \log(|u|_v) = \log(|s_v|_v) + O_m(1), \quad v \mid \infty.$$

Let us introduce the notation

$$l(x) := (\log(|x_v|_v))_{v \mid \infty} \in \prod_{v \mid \infty} \mathbb{R}, \quad x \in F_\infty^\times. \quad (5.1)$$

Then $\{l(u) : u \in \mathfrak{o}^\times\}$ is a lattice in the hyperplane

$$W := \left\{ w \in \prod_{v|\infty} \mathbb{R} : \sum_{v|\infty} w_v = 0 \right\} \quad (5.2)$$

by Dirichlet's unit theorem (cf. [50, p. 93]). As the vector $l(s)/m$ lies in W , there exists a lattice point $l(u)$ within $O(1)$ distance from it. Multiplying by m , we get the required conclusion in the stronger form

$$m \log(|u|_v) = \log(|s|_v) + O(m), \quad v|\infty.$$

The proof is complete. \square

Remark 1. Applying Lemma 3 with $m = 1$ (or its proof if more geometric features are needed), we see that $F_\infty^\times/\mathfrak{o}^\times$ has a fundamental domain lying in $\{y \in F_\infty^\times : |y_v| \asymp |y_\infty|^{1/n} \text{ for all } v|\infty\}$. We recall here that $n = [F : \mathbb{Q}]$, and $|y_v|_v = |y_v|$ when v is a real place, but $|y_v|_v = |y_v|^2$ when v is a complex place. Switching to $m = 2$, we see the same for $F_\infty^\times/(\mathfrak{o}^\times)^2$, or (mutatis mutandis) for $F_\infty^\times/\mathfrak{o}_+^\times$ and $F_+^\times/\mathfrak{o}_+^\times$ (cf. Section 2).

By Lemmata 1 and 2, any global matrix $g \in GL_2(\mathbb{A})$ can be decomposed (non-uniquely) as

$$g = \begin{pmatrix} y & x \\ & 1 \end{pmatrix} k, \quad (5.3)$$

where $\begin{pmatrix} y & x \\ & 1 \end{pmatrix} \in P(\mathbb{A})$ and $k \in K^*$. Moreover, its height

$$\text{ht}(g) := |y|_{\mathbb{A}} = \prod_v |y_v|_v$$

and the image of k in the group $K^*/Z(\mathbb{A})K \cong \prod_{p|n} \{\pm 1\}$ are uniquely determined. By using the ideal class representatives $\theta_1, \dots, \theta_h \in \mathbb{A}_{\text{fin}}^\times$ introduced in Section 2, we can refine (5.3) and obtain a convenient fundamental domain for X^* (cf. (2.8)).

Lemma 4. Any $g \in GL_2(\mathbb{A})$ can be decomposed as

$$g = \begin{pmatrix} t & s \\ & 1 \end{pmatrix} \begin{pmatrix} y & x \\ & 1 \end{pmatrix} \begin{pmatrix} \theta_i & \\ & 1 \end{pmatrix} k, \quad (5.4)$$

where $\begin{pmatrix} t & s \\ & 1 \end{pmatrix} \in P(F)$, $\begin{pmatrix} y & x \\ & 1 \end{pmatrix} \in P(F_\infty)$, and $k \in K^*$. Moreover, the possible modules $|y|_\infty$ that occur for a given $g \in GL_2(\mathbb{A})$ are essentially constant, namely $\text{ht}(g) \asymp |y|_\infty$.

Proof. By (5.3), we have a decomposition

$$g = \begin{pmatrix} \tilde{y} & \tilde{x} \\ & 1 \end{pmatrix} \tilde{k},$$

where $\begin{pmatrix} \tilde{y} & \tilde{x} \\ & 1 \end{pmatrix} \in P(\mathbb{A})$ and $\tilde{k} \in K^*$. We can write $\tilde{y} = t\theta_i y y'$ with some $t \in F^\times$, $y \in F_\infty^\times$, $y' \in \hat{\mathfrak{o}}^\times$, and a unique index $i \in \{1, \dots, h\}$ depending on the fractional ideal $\tilde{y}_{\text{fin}}\mathfrak{o}$. Here

\hat{o} is as in (2.2). In addition, as $F + F_\infty$ is dense in \mathbb{A} (see [50, Cor. 2 to Th. 3 in Ch. IV-2]), we can write $\tilde{x} = s + t\theta_i(x + x')$ with some $s \in F$, $x \in F_\infty$, and $x' \in \hat{o}$. Therefore,

$$\begin{pmatrix} \tilde{y} & \tilde{x} \\ & 1 \end{pmatrix} = \begin{pmatrix} t\theta_i & s \\ & 1 \end{pmatrix} \begin{pmatrix} yy' & x + x' \\ & 1 \end{pmatrix} = \begin{pmatrix} t & s \\ & 1 \end{pmatrix} \begin{pmatrix} \theta_i & \\ & 1 \end{pmatrix} \begin{pmatrix} y & x \\ & 1 \end{pmatrix} \begin{pmatrix} y' & x' \\ & 1 \end{pmatrix}.$$

On the right hand side, the second and third factors commute, while the fourth factor (call it k') lies in K^* , hence (5.4) follows with $k := k'\tilde{k} \in K^*$.

The second statement is straightforward by the fact that (5.4) is an instance of the Iwasawa decomposition (5.3). Indeed,

$$\text{ht}(g) = |ty\theta_i|_{\mathbb{A}} = |y\theta_i|_{\mathbb{A}} = |y|_{\infty}|\theta_i|_{\mathbb{A}} \asymp |y|_{\infty}.$$

Here we have used $|t|_{\mathbb{A}} = 1$ due to $t \in F^\times$ (see [50, Th. 5 in Ch. IV-4]). \square

In contrast, the module $|y|_{\infty}$ can vary considerably if we allow any matrix $\gamma \in \text{GL}_2(F)$ in place of $\begin{pmatrix} t & s \\ & 1 \end{pmatrix} \in \text{P}(F)$, and it will be useful for us to take $|y|_{\infty}$ as large as possible in this more general context.

Lemma 5. *Any $g \in \text{GL}_2(\mathbb{A})$ can be decomposed as*

$$g = \gamma \begin{pmatrix} y & x \\ & 1 \end{pmatrix} \begin{pmatrix} \theta_i & \\ & 1 \end{pmatrix} k, \quad (5.5)$$

where $\gamma \in \text{GL}_2(F)$, $\begin{pmatrix} y & x \\ & 1 \end{pmatrix} \in \text{P}(F_\infty)$, and $k \in K^*$. Moreover, among the possible modules $|y|_{\infty}$ that occur for a given $g \in \text{GL}_2(\mathbb{A})$, there is a maximal one.

Proof. By Lemma 4, or alternatively by Lemma 1 and strong approximation for $\text{SL}_2(\mathbb{A})$, a decomposition of the form (5.5) certainly exists. Let us now fix any decomposition as in (5.5), and consider the alternative decompositions

$$g = \tilde{\gamma} \begin{pmatrix} \tilde{y} & \tilde{x} \\ & 1 \end{pmatrix} \begin{pmatrix} \theta_j & \\ & 1 \end{pmatrix} \tilde{k},$$

where $\tilde{\gamma} \in \text{GL}_2(F)$, $\begin{pmatrix} \tilde{y} & \tilde{x} \\ & 1 \end{pmatrix} \in \text{P}(F_\infty)$, and $\tilde{k} \in K^*$. It suffices to show that there are only finitely many possible values of $|\tilde{y}|_{\infty}$ that exceed $|y|_{\infty}$. So we assume that $|\tilde{y}|_{\infty} > |y|_{\infty}$, and rearrange the terms to get

$$\tilde{\gamma}^{-1} \gamma \begin{pmatrix} y & x \\ & 1 \end{pmatrix} \begin{pmatrix} \theta_i & \\ & 1 \end{pmatrix} = \begin{pmatrix} \tilde{y} & \tilde{x} \\ & 1 \end{pmatrix} \begin{pmatrix} \theta_j & \\ & 1 \end{pmatrix} (\tilde{k}k^{-1}).$$

Then, with the notation $\begin{pmatrix} a & b \\ c & d \end{pmatrix} := \tilde{\gamma}^{-1} \gamma \in \text{GL}_2(F)$ and $k' := \tilde{k}k^{-1} \in K^*$, we have

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} y & x \\ & 1 \end{pmatrix} \begin{pmatrix} \theta_i & \\ & 1 \end{pmatrix} = \begin{pmatrix} \tilde{y} & \tilde{x} \\ & 1 \end{pmatrix} \begin{pmatrix} \theta_j & \\ & 1 \end{pmatrix} k'.$$

Multiplying both sides by a suitable matrix in $Z(F)$, we can further achieve that the greatest common divisor $c\mathfrak{o} + d\mathfrak{o}$ equals $\theta_m\mathfrak{o}$ for some m . Now we calculate the height of both sides, using Lemmata 1–4. On the right hand side, we get $\asymp |\tilde{y}|_{\infty}$ by Lemma 4. On the

left hand side, the product of the local factors $|ad - bc|_v$ coming from (4.6) and (4.8) equals 1 by $ad - bc \in F^\times$ (cf. [50, Th. 5 in Ch. IV-4]). The product of the other factors coming from (4.8) is $\asymp_n 1$ due to finitely many possibilities for the pair (θ_i, θ_m) and the fact that $\begin{pmatrix} y & x \\ & 1 \end{pmatrix}$ has no finite components. So we can focus on the remaining factors coming from (4.6), and we conclude that

$$\frac{|y|_\infty}{\prod_{v|\infty} (|c_v y_v|^2 + |c_v x_v + d_v|^2)^{[F_v:\mathbb{R}]}} \asymp_n |\tilde{y}|_\infty.$$

Along these lines, we also see that the fractional ideals $c\mathfrak{o}$ and $d\mathfrak{o}$ together with the denominator on the left hand side determine $|\tilde{y}|_\infty$ up to $\ll_n 1$ choices, so it suffices to show that these quantities only take on finitely many different values. At any rate, the right hand side exceeds $|y|_\infty$ by assumption, hence we immediately get

$$\prod_{v|\infty} (|c_v y_v|^2 + |c_v x_v + d_v|^2)^{[F_v:\mathbb{R}]} \ll_n 1. \quad (5.6)$$

If $c = 0$, then (5.6) yields $|d|_\infty \ll_n 1$, so in this case there are $\ll_n 1$ choices for the fractional ideal $d\mathfrak{o} \subseteq \theta_m \mathfrak{o}$ and its norm $|d|_\infty$, whose square is apparently the left hand side of (5.6). If $c \neq 0$, then (5.6) yields $|cy|_\infty \ll_n 1$, so in this case there are $\ll_{n,y} 1$ choices for the fractional ideal $c\mathfrak{o} \subseteq \theta_m \mathfrak{o}$. Let us fix a nonzero choice for $c\mathfrak{o}$ (including an arbitrary choice of the generator c ; however, none of the implied constants below will depend on this choice). Dividing (5.6) by the squared norm $|c|_\infty^2$ (which is $\gg 1$), we get

$$\prod_{v|\infty} (|y_v|^2 + |x_v + d_v/c_v|^2)^{[F_v:\mathbb{R}]} \ll_n 1. \quad (5.7)$$

The factors are $\gg_y 1$, hence $x_v + d_v/c_v \ll_{n,y} 1$ for all $v|\infty$. Using also $d/c \in (\theta_m \mathfrak{o})(c\mathfrak{o})^{-1}$, we see that d/c as an element of \mathbb{A} lies in a fixed compact set, so there are finitely many choices for d/c and consequently for d as well. In the end, the left hand side of (5.7) takes on finitely many different values, and the same is true of the left hand side of (5.6). The proof is complete. \square

By Lemma 5, any double coset in X^* can be represented by a matrix of the form $\begin{pmatrix} y & x \\ & 1 \end{pmatrix} \begin{pmatrix} \theta_i & \\ & 1 \end{pmatrix}$, where $\begin{pmatrix} y & x \\ & 1 \end{pmatrix} \in P(F_\infty)$ and $|y|_\infty$ is maximal. We shall call such matrices *special*. We can further specify these representatives by observing that, for any unit $t \in \mathfrak{o}^\times$ and any field element $s \in \theta_i \mathfrak{o}$,

$$\begin{pmatrix} t & s \\ & 1 \end{pmatrix} \begin{pmatrix} y & x \\ & 1 \end{pmatrix} \begin{pmatrix} \theta_i & \\ & 1 \end{pmatrix} = \begin{pmatrix} t_\infty y & t_\infty x + s_\infty \\ & 1 \end{pmatrix} \begin{pmatrix} \theta_i & \\ & 1 \end{pmatrix} \begin{pmatrix} t_{\text{fin}} & \theta_i^{-1} s_{\text{fin}} \\ & 1 \end{pmatrix}.$$

Indeed, the leftmost matrix lies in $GL_2(F)$ and the rightmost matrix lies in K^* , hence we can replace y by $t_\infty y$ and x by $t_\infty x + s_\infty$ without changing $|y|_\infty$ and the double coset represented. In particular, we can restrict $y \in F_\infty^\times$ to a fixed fundamental domain for $F_\infty^\times/\mathfrak{o}^\times$ and $x \in F_\infty$ to a fixed fundamental domain for $F_\infty/\theta_i \mathfrak{o}$. We can further tweak $y \in F_\infty^\times$ by replacing each component y_v by its absolute value $|y_v|$, thanks to the observation that

$$\begin{pmatrix} y_v/|y_v| & \\ & 1 \end{pmatrix} \in K_v^*, \quad v|\infty. \quad (5.8)$$

Again, $|y|_\infty$ is invariant under such a replacement. In this way, using appropriate fundamental domains for $F_\infty^\times/\mathfrak{o}^\times$ and $F_\infty/\theta_i\mathfrak{o}$, we obtain a set of representatives $\mathcal{F}(\mathfrak{n}) \subset \mathrm{GL}_2(\mathbb{A})$ for X^* consisting of special matrices $\begin{pmatrix} y & x \\ & 1 \end{pmatrix}^{(\theta_i)}$ such that the components of $y \in F_\infty^\times$ and $x \in F_\infty$ satisfy (cf. Remark 1)

$$0 < y_v \asymp |y|_\infty^{1/n} \quad \text{and} \quad x_v \ll 1, \quad v \mid \infty. \quad (5.9)$$

By possibly shrinking $\mathcal{F}(\mathfrak{n})$ further, we could get a true fundamental domain for X^* (i.e. a nice subset of $\mathrm{GL}_2(\mathbb{A})$ representing each double coset in X^* exactly once), but the above construction is sufficient for our purposes.

Now we consider the $2n$ -dimensional \mathbb{R} -algebra

$$\mathbb{M} := \prod_{v \text{ real}} \mathbb{C} \prod_{v \text{ complex}} \mathbb{H}, \quad (5.10)$$

which becomes a Euclidean space if we postulate that the standard basis (the union of the bases $\{1, i\}$ and $\{1, i, j, k\}$ of the various factors \mathbb{C} and \mathbb{H} embedded into \mathbb{M} as subspaces) is orthonormal. Within this space, we consider the generalized upper half-space (cf. (4.1) and (4.3))

$$\mathcal{H} := \prod_{v \text{ real}} \mathcal{H}^2 \prod_{v \text{ complex}} \mathcal{H}^3, \quad (5.11)$$

and to each special matrix $\begin{pmatrix} y & x \\ & 1 \end{pmatrix}^{(\theta_i)} \in \mathcal{F}(\mathfrak{n})$, we associate the point (cf. (5.9))

$$P = P(x, y) := \prod_{v \text{ real}} \{x_v + y_v i\} \times \prod_{v \text{ complex}} \{x_v + y_v j\} \in \mathcal{H} \quad (5.12)$$

and the $2n$ -dimensional \mathbb{Z} -lattice

$$\Lambda(P) := \{cP + d : c, d \in \mathfrak{o}\} \subset \mathbb{M}. \quad (5.13)$$

Note that a given point $P \in \mathcal{H}$ corresponds to at most h elements of $\mathcal{F}(\mathfrak{n})$ in the above sense.

The next lemma, a generalization of [3, Lemma 2], shows in terms of these lattices that $\mathcal{F}(\mathfrak{n})$ has good diophantine properties.

Lemma 6. *Let $\begin{pmatrix} y & x \\ & 1 \end{pmatrix}^{(\theta_i)} \in \mathcal{F}(\mathfrak{n})$ with corresponding point $P = P(x, y) \in \mathcal{H}$. Then the lattice $\Lambda(P) \subset \mathbb{M}$ and its successive minima $m_1 \leq \dots \leq m_{2n}$ satisfy:*

- (a) $m_1 \cdots m_{2n} \asymp |y|_\infty$;
- (b) $|y|_\infty \gg m_1^{2n} \gg (\mathcal{N}\mathfrak{n})^{-1}$;
- (c) $m_1 \asymp \dots \asymp m_n$ and $m_{n+1} \asymp \dots \asymp m_{2n}$;
- (d) *in any ball of radius R the number of lattice points is $\ll 1 + R^n (\mathcal{N}\mathfrak{n})^{1/2} + R^{2n} |y|_\infty^{-1}$.*

Proof. (a) Let $\{u_1, \dots, u_n\}$ be a \mathbb{Z} -basis of \mathfrak{o} . Then $\{u_1, \dots, u_n, u_1 P, \dots, u_n P\}$ is a \mathbb{Z} -basis of $\Lambda(P)$, and calculating its exterior product coupled with Minkowski's theorem [16, Th. 3 on p. 124] shows that

$$m_1 \cdots m_{2n} \asymp \mathrm{vol}(\mathbb{M}/\Lambda(P)) = |y|_\infty \mathrm{vol}(F_\infty/\mathfrak{o})^2 \asymp |y|_\infty.$$

(b) The bound $|y|_\infty \gg m_1^{2n}$ is clear from (a), so it suffices to show $m_1^{2n} \gg (\mathcal{N}\mathfrak{n})^{-1}$. Let $(c, d) \in \mathfrak{o}^2$ be any pair distinct from $(0, 0)$. We choose any pair $(a, b) \in F^2$ such that $ad - bc \neq 0$, and we consider the translated matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} y & x \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \theta_i \\ 1 \end{pmatrix} \in GL_2(\mathbb{A})$. According to Lemma 4, we have a decomposition

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} y & x \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \theta_i \\ 1 \end{pmatrix} = \begin{pmatrix} t & s \\ & 1 \end{pmatrix} \begin{pmatrix} \tilde{y} & \tilde{x} \\ & 1 \end{pmatrix} \begin{pmatrix} \theta_j \\ 1 \end{pmatrix} k,$$

where $\begin{pmatrix} t & s \\ & 1 \end{pmatrix} \in P(F)$, $\begin{pmatrix} \tilde{y} & \tilde{x} \\ & 1 \end{pmatrix} \in P(F_\infty)$, and $k \in K^*$. The key observation here is that $\begin{pmatrix} y & x \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \theta_i \\ 1 \end{pmatrix}$ and $\begin{pmatrix} \tilde{y} & \tilde{x} \\ & 1 \end{pmatrix} \begin{pmatrix} \theta_j \\ 1 \end{pmatrix}$ represent the same double coset in X^* , hence $|\tilde{y}|_\infty \leq |y|_\infty$ by the very assumption $\begin{pmatrix} y & x \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \theta_i \\ 1 \end{pmatrix} \in \mathcal{F}(\mathfrak{n})$. Now we proceed somewhat similarly to the proof of Lemma 5. That is, we calculate the height of both sides, using Lemmata 1–4. On the right hand side, we get $\asymp |\tilde{y}|_\infty$ by Lemma 4. On the left hand side, the product of the local factors $|ad - bc|_v$ coming from (4.6) and (4.8) equals 1 by $ad - bc \in F^\times$ (cf. [50, Th. 5 in Ch. IV-4]). The product of the other factors coming from (4.8) is $\gg (\mathcal{N}\mathfrak{n})^{-1}$ due to the facts that $c, d \in \mathfrak{o}$ are integers and $\begin{pmatrix} y & x \\ 1 & 1 \end{pmatrix}$ has no finite components. So, we can focus on the remaining factors coming from (4.6), and we conclude that

$$\frac{|y|_\infty (\mathcal{N}\mathfrak{n})^{-1}}{\prod_{v|\infty} (|c_v y_v|^2 + |c_v x_v + d_v|^2)^{[F_v:\mathbb{R}]}} \ll |\tilde{y}|_\infty \leq |y|_\infty.$$

Comparing the two sides gives

$$\prod_{v|\infty} (|c_v y_v|^2 + |c_v x_v + d_v|^2)^{[F_v:\mathbb{R}]} \gg (\mathcal{N}\mathfrak{n})^{-1}.$$

If we regard the left hand side as a product of $\sum_{v|\infty} [F_v:\mathbb{R}] = n$ factors, the inequality between the arithmetic and geometric mean readily yields

$$\left(\sum_{v|\infty} (|c_v y_v|^2 + |c_v x_v + d_v|^2) \right)^n \gg (\mathcal{N}\mathfrak{n})^{-1}.$$

The sum on the left hand side equals the squared Euclidean norm $\|cP + d\|^2$, hence

$$\|cP + d\|^{2n} \gg (\mathcal{N}\mathfrak{n})^{-1}.$$

This holds for all pairs $(c, d) \in \mathfrak{o}^2$ distinct from $(0, 0)$, so $m_1^{2n} \gg (\mathcal{N}\mathfrak{n})^{-1}$ as stated.

(c) It suffices to show that $m_n \ll m_1$ and $m_{2n} \ll m_{n+1}$, and for this we utilize the left \mathfrak{o} -invariance of $\Lambda(P)$. To prove the first relation, we take a nonzero lattice point Q in $\Lambda(P)$ of Euclidean norm at most m_1 . Then $\mathfrak{o}Q$ is an n -dimensional sublattice of $\Lambda(P)$ with \mathbb{Z} -basis $\{u_1 Q, \dots, u_n Q\}$, where $\{u_1, \dots, u_n\}$ is a \mathbb{Z} -basis of \mathfrak{o} as before. Therefore,

$$m_n \leq \max\{\|u_1 Q\|, \dots, \|u_n Q\|\} \ll \|Q\| \leq m_1.$$

To prove the second relation, we take \mathbb{Z} -independent lattice points $Q_1, \dots, Q_{n+1} \in \Lambda(P)$ of Euclidean norms at most m_{n+1} . At least two of these, say Q and Q' , must be \mathfrak{o} -independent, and then $\mathfrak{o}Q + \mathfrak{o}Q'$ is a $2n$ -dimensional sublattice of $\Lambda(P)$ with \mathbb{Z} -basis $\{u_1Q, \dots, u_nQ, u_1Q', \dots, u_nQ'\}$. Therefore,

$$m_{2n} \leq \max\{\|u_1Q\|, \dots, \|u_nQ\|, \|u_1Q'\|, \dots, \|u_nQ'\|\} \ll \max\{\|Q\|, \|Q'\|\} \leq m_{n+1}.$$

(d) Let $B \subset \mathbb{M}$ be any ball of radius $R > 0$ (with respect to the Euclidean norm). Then, by a lattice point counting argument using successive minima (see e.g. [3, Lemma 1]⁵),

$$\#(\Lambda(P) \cap B) \ll 1 + \frac{R}{m_1} + \frac{R^2}{m_1 m_2} + \dots + \frac{R^{2n}}{m_1 \cdots m_{2n}}.$$

Denoting by t_k the degree k term, we can estimate, by part (c),

$$t_k \ll \begin{cases} t_n^{k/n} \leq \max(1, t_n) & \text{when } 0 \leq k \leq n, \\ t_n (t_{2n}/t_n)^{(k-n)/n} \leq \max(t_n, t_{2n}) & \text{when } n \leq k \leq 2n, \end{cases}$$

hence our bound simplifies to

$$\#(\Lambda(P) \cap B) \ll 1 + \frac{R^n}{m_1 \cdots m_n} + \frac{R^{2n}}{m_1 \cdots m_{2n}}.$$

The first denominator is $\gg (\mathcal{N}\mathfrak{n})^{-1/2}$ by (b), and the second is $\asymp |y|_\infty$ by (a), so the stated bound follows. \square

6. Counting the field elements

In this section, we collect some convenient counting bounds in the number field F . These bounds are rather standard but indispensable for our goals, and we present them in a consistent adelic language.

Lemma 7. *Let $t \in \mathbb{A}^\times$ be an arbitrary idele. Then*

- (a) $\#\{x \in F^\times : |x|_v \leq |t_v|_v \text{ for all places } v\} \ll |t|_{\mathbb{A}};$
- (b) $\#\{x \in F^\times : |x|_v \leq |t_v|_v \text{ for all } v \mid \infty \text{ and } |x|_v = |t_v|_v \text{ for all } v \nmid \infty\} \ll_\varepsilon |t|_{\mathbb{A}}^\varepsilon$ for every $\varepsilon > 0$.

⁵ We use this opportunity to note that, in the proof of [3, Lemma 1], kB' should be k^2B' , while $(k+1)B'$ should be $(k^2+1)B'$, as follows from an observation of Mahler [16, p. 68]. The statement and proof of [3, Lemma 1] are otherwise unchanged.

Proof. (a) Let S be the set in part (a). If S is empty, we are done. Otherwise, we fix any element $\tilde{x} \in S$. By assumption, the idele $s := t/\tilde{x} \in \mathbb{A}^\times$ satisfies $1 \leq |s_v|_v$ for all places v . Moreover, any $x \in S$ is determined by the field element $u := x/\tilde{x} \in F^\times$, which in turn satisfies $|u|_v \leq |s_v|_v$ for all places v . Now we consider a fundamental parallelotope P for F_∞/\mathfrak{o} . Then $\mathbf{P} := P \times \hat{\mathfrak{o}}$ (cf. (2.2)) is a fundamental domain for \mathbb{A}/F (see [50, Prop. 6 in Ch. V-4]). The translates $u + \mathbf{P}$ (with u as before) have pairwise disjoint interiors and, for a suitable idele $y \in \mathbb{A}^\times$ depending only on P , they all lie in the adelic box

$$\mathbf{B} := \{z \in \mathbb{A} : |z_v|_v \leq |y_v s_v|_v \text{ for all places } v\}.$$

Here we have used $|s_v|_v \geq 1$ for all places v , and we remark that we can take $y_v = 1$ at the nonarchimedean places v . An adelic volume calculation now gives

$$\#S \ll_P \#S \text{vol}(\mathbf{P}) \leq \text{vol}(\mathbf{B}) = |y s|_{\mathbb{A}} \ll_P |s|_{\mathbb{A}} = |t|_{\mathbb{A}},$$

and the bound (a) follows.

(b) Let S be the set in (b). If S is empty, we are done. Otherwise, we fix any $\tilde{x} \in S$. By assumption, the idele $s := t/\tilde{x} \in \mathbb{A}^\times$ satisfies $1 \leq |s_v|_v$ for all $v \mid \infty$ and $1 = |s_v|_v$ for all $v \nmid \infty$. Moreover, any $x \in S$ is determined by $u := x/\tilde{x} \in F^\times$, which in turn satisfies $|u|_v \leq |s_v|_v$ for all $v \mid \infty$ and $|u|_v = |s_v|_v = 1$ for all $v \nmid \infty$. The second relation implies that $u \in \mathfrak{o}^\times$ is a unit, while the first relation implies the lower bounds

$$|u|_v = \prod_{\substack{w \mid \infty \\ w \neq v}} |u|_w^{-1} \geq \prod_{\substack{w \mid \infty \\ w \neq v}} |s_w|_w^{-1} = |s_v|_v \cdot |s|_{\infty}^{-1}, \quad v \mid \infty.$$

In short,

$$\log(|s_v|_v) - \log(|s|_{\infty}) \leq \log(|u|_v) \leq \log(|s_v|_v), \quad v \mid \infty.$$

With the notations (5.1)–(5.2), the vector $l(u)$ lies in a fixed ball of radius $\ll \log(|s|_{\infty})$ intersected with a fixed lattice in W by Dirichlet's unit theorem (cf. [50, p. 93]), hence the usual volume argument yields

$$\#l(u) \ll 1 + (\log |s|_{\infty})^{\dim W} \ll_{\varepsilon} |s|_{\infty}^{\varepsilon} = |s|_{\mathbb{A}}^{\varepsilon} = |t|_{\mathbb{A}}^{\varepsilon}.$$

However, u is determined by $l(u)$ up to a root of unity (cf. [50, Th. 8 in Ch. IV-4]), so the bound (b) follows. \square

We shall use Lemma 7 in the following classical formulation.

Corollary 1. *Let $y \in F_\infty^\times$ be an infinite idele, and let $\mathfrak{m} \subseteq F$ be a nonzero fractional ideal. Then*

- (a) $\#\{x \in F^\times : |x|_v \leq |y_v|_v \text{ for all } v \mid \infty \text{ and } x\mathfrak{o} \subseteq \mathfrak{m}\} \ll |y|_{\infty}/\mathcal{N}\mathfrak{m}$;
- (b) $\#\{x \in F^\times : |x|_v \leq |y_v|_v \text{ for all } v \mid \infty \text{ and } x\mathfrak{o} = \mathfrak{m}\} \ll_{\varepsilon} (|y|_{\infty}/\mathcal{N}\mathfrak{m})^{\varepsilon}$ for every $\varepsilon > 0$.

Proof. We take any finite idele $z \in \mathbb{A}_{\text{fin}}^\times$ such that $\mathfrak{m} = z\mathfrak{o}$, and we write $t := yz \in \mathbb{A}^\times$. Clearly, $|y_v|_v = |t_v|_v$ for all $v \mid \infty$, while $x\mathfrak{o} \subseteq \mathfrak{m}$ (resp. $x\mathfrak{o} = \mathfrak{m}$) is equivalent to $|x|_v \leq |t_v|_v$ (resp. $|x|_v = |t_v|_v$) for all $v \nmid \infty$. Moreover,

$$|t|_{\mathbb{A}} = |y|_{\infty} \prod_{v \nmid \infty} |z_v|_v = |y|_{\infty} / \mathcal{N}(z\mathfrak{o}) = |y|_{\infty} / \mathcal{N}\mathfrak{m}.$$

Hence (a) and (b) are immediate from the corresponding parts of Lemma 7. \square

Remark 2. It is a valuable feature of Lemma 7 that the bounds only depend on the module $|t|_{\mathbb{A}}$. This type of result proved to be useful in earlier investigations of the sup-norm problem; for instance, [6, (4.1)–(4.2)] constitute a special case of Lemma 7 and Corollary 1.

7. Realness rigidity of number fields

This section is devoted to the proof of the following result which will be used in the proof of Theorem 2, whose germ is the argument in [3, beginning of Subsection 11.2].

Lemma 8. *Let F_0 be the maximal totally real subfield of F , and put $m := [F : F_0]$. Let $\xi \in F$ be such that $F = F_0(\xi)$. Suppose that $\xi\mathfrak{o} = \mathfrak{k}/\mathfrak{l}$, where $\mathfrak{k}, \mathfrak{l} \subseteq \mathfrak{o}$ are integral ideals. Suppose that*

$$|\xi_v| \leq A \quad \text{and} \quad |\text{Im } \xi_v| \leq A\sqrt{\delta_v}, \quad v \mid \infty, \quad (7.1)$$

where $A \geq 1$ and $\delta_v > 0$ are arbitrary. Then

$$1 \leq ((2A)^n (\mathcal{N}\mathfrak{l})^2)^{2(m-1)} |\delta|_{\mathbb{C}}. \quad (7.2)$$

Remark 3. The assumption $F = F_0(\xi)$ serves convenience. Without it, (7.2) still holds with m being the degree of ξ over F_0 , and with $|\delta|_{\mathbb{C}}$ replaced by the relevant subproduct restricted to the places of F that lie over the *complex* places of $F_0(\xi)$. In this more general formulation, the result says that if $\xi \in F$ is “close to being totally real”, then it is totally real. Thus the result expresses a certain “rigidity of realness” in number fields.

Remark 4. If $F = \mathbb{Q}(i)$, and $\xi = \kappa/\lambda$ with $\kappa, \lambda \in \mathbb{Z}[i]$ satisfying $|\kappa| \ll |\lambda| \asymp \Lambda$, then the above statement says that if $|\text{Im } \xi| \ll \Lambda^{-2}$ with a sufficiently small implied constant, then $\xi \in \mathbb{Q}$. This is precisely the crucial input in [3, Subsection 11.2], which in this special case is completely elementary.

Proof of Lemma 8. We shall assume that $m \geq 2$, because (7.2) is trivially true for $m = 1$ (in this case the right hand side equals 1). Let \tilde{F} be the Galois closure of F with ring of integers $\tilde{\mathfrak{o}}$, and consider the Galois groups

$$G := \text{Gal}(\tilde{F}/\mathbb{Q}), \quad H := \text{Gal}(\tilde{F}/F_0), \quad I := \text{Gal}(\tilde{F}/F).$$

Note that $I \leq H \leq G$, and

$$[G : I] = [F : \mathbb{Q}] = n, \quad [G : H] = [F_0 : \mathbb{Q}] = n/m, \quad [H : I] = [F : F_0] = m. \quad (7.3)$$

Indeed, the elements of $I \backslash G$ correspond to the n embeddings $F \hookrightarrow \tilde{F}$, the elements of $H \backslash G$ correspond to the n/m embeddings $F_0 \hookrightarrow \tilde{F}$, and the elements of $I \backslash H$ correspond to the m conjugates of ξ over F_0 via the right action $\sigma \mapsto \xi^\sigma$ of H on ξ . In particular,

$$p(X) := \prod_{\alpha \in I \backslash H} (X - \xi^\alpha)$$

is the minimal polynomial of ξ over F_0 (a separable polynomial of degree m), and

$$\Delta := \prod_{\substack{\alpha, \beta \in I \backslash H \\ \alpha \neq \beta}} (\xi^\alpha - \xi^\beta)$$

is the discriminant of $p(X)$. Note that right multiplication by any $\sigma \in H$ permutes the elements of $I \backslash H$, hence H fixes Δ , and so $\Delta \in F_0^\times$. The key to deriving (7.2) is to examine the absolute norm $|N_{F_0/\mathbb{Q}}(\Delta)| \in \mathbb{Q}^\times$, which is a *positive rational* number.

Fixing an embedding $\tilde{F} \hookrightarrow \mathbb{C}$, we get (cf. [50, Cor. 3 to Prop. 4 in Ch. III-3])

$$|N_{F_0/\mathbb{Q}}(\Delta)| = \prod_{\gamma \in H \backslash G} |\Delta^\gamma| = \prod_{\substack{\alpha, \beta \in I \backslash H \\ \alpha \neq \beta}} \prod_{\gamma \in H \backslash G} |\xi^{\alpha\gamma} - \xi^{\beta\gamma}|.$$

The products $\alpha\gamma$ and $\beta\gamma$ run through distinct representatives of $I \backslash G$ such that $H\alpha\gamma = H\beta\gamma$ (namely $H\alpha\gamma$ and $H\beta\gamma$ are equal to $H\gamma$), hence they correspond to distinct embeddings $\sigma, \tau : F \hookrightarrow \mathbb{C}$ which agree on F_0 . We infer

$$|N_{F_0/\mathbb{Q}}(\Delta)| = \prod_{\substack{\sigma, \tau : F \hookrightarrow \mathbb{C} \\ \sigma \neq \tau \\ \sigma|_{F_0} = \tau|_{F_0}}} |\xi^\sigma - \xi^\tau|. \quad (7.4)$$

On the right hand side, there are $(n/m)m(m-1) = n(m-1)$ factors by (7.3), all of which are at most $2A$ by the first part of (7.1). By definition, each complex place v of F can be identified with an *unordered* complex conjugate pair $\{\sigma, \tau\}$ of distinct embeddings $\sigma, \tau : F \hookrightarrow \mathbb{C}$. For such a pair $\{\sigma, \tau\}$, we have $\{\xi^\sigma, \xi^\tau\} = \{\xi_v, \bar{\xi}_v\}$, while the condition $\sigma|_{F_0} = \tau|_{F_0}$ is automatic as F_0 is totally real (the restricted embedding $F_0 \hookrightarrow \mathbb{R}$ corresponds to the real place of F_0 below v). Hence each complex place v of F contributes *two* factors $|\xi_v - \bar{\xi}_v|$ to the product in (7.4), and these are at most $2A\sqrt{\delta_v}$ by the second part of (7.1). Bounding the other factors $|\xi^\sigma - \xi^\tau|$ in (7.4) by $2A$ as remarked initially, we obtain in the end

$$|N_{F_0/\mathbb{Q}}(\Delta)| \leq (2A)^{n(m-1)} |\delta|_{\mathbb{C}}^{1/2}. \quad (7.5)$$

On the other hand, with $\tilde{\mathfrak{k}} := \mathfrak{k}\tilde{\mathfrak{o}}$ and $\tilde{\mathfrak{l}} := \mathfrak{l}\tilde{\mathfrak{o}}$, the nonzero fractional ideal in \tilde{F} given by

$$\Delta \prod_{\substack{\alpha, \beta \in I \backslash H \\ \alpha \neq \beta}} (\tilde{\mathfrak{l}}^\alpha \tilde{\mathfrak{l}}^\beta) = \prod_{\substack{\alpha, \beta \in I \backslash H \\ \alpha \neq \beta}} (\xi^\alpha - \xi^\beta) (\tilde{\mathfrak{l}}^\alpha \tilde{\mathfrak{l}}^\beta) \subseteq \prod_{\substack{\alpha, \beta \in I \backslash H \\ \alpha \neq \beta}} (\tilde{\mathfrak{k}}^\alpha \tilde{\mathfrak{l}}^\beta + \tilde{\mathfrak{l}}^\alpha \tilde{\mathfrak{k}}^\beta) \subseteq \tilde{\mathfrak{o}}$$

is integral, hence its norm is at least 1. The norm of each factor $\tilde{l}^\alpha \tilde{l}^\beta \subseteq \tilde{\delta}$ on the left hand side equals

$$\mathcal{N}(\tilde{l}^\alpha) \mathcal{N}(\tilde{l}^\beta) = (\mathcal{N}\tilde{l})^2 = (\mathcal{N}\mathfrak{l})^{2[\tilde{F}:F]},$$

hence we get

$$1 \leq |N_{\tilde{F}/\mathbb{Q}}(\Delta)| (\mathcal{N}\mathfrak{l})^{2m(m-1)[\tilde{F}:F]} = |N_{\tilde{F}/\mathbb{Q}}(\Delta)| (\mathcal{N}\mathfrak{l})^{2(m-1)[\tilde{F}:F_0]}.$$

Raising the two sides to the power $[\tilde{F}:F_0]^{-1}$, and combining with (7.5), we conclude

$$1 \leq |N_{F_0/\mathbb{Q}}(\Delta)| (\mathcal{N}\mathfrak{l})^{2(m-1)} \leq ((2A)^n (\mathcal{N}\mathfrak{l})^2)^{m-1} |\delta|_{\mathbb{C}}^{1/2}.$$

Finally, by squaring the two sides, (7.2) follows. \square

8. The Fourier bound

We denote by $t_v \in \mathbb{R} \cup \mathbb{R}i$ the *spectral parameter* of ϕ at the archimedean place v , so that the Laplace eigenvalue of ϕ at v equals

$$\lambda_v = \begin{cases} 1/4 + t_v^2 & \text{when } v \text{ is real,} \\ 1 + 4t_v^2 & \text{when } v \text{ is complex.} \end{cases} \quad (8.1)$$

The archimedean Ramanujan conjecture states that $t_v \in \mathbb{R}$, while it is known [2, Th. 1] that

$$t_v \in \mathbb{R} \cup [-7/64, 7/64]i.$$

For convenience, we also introduce the tuple $T := (T_v)_{v|\infty}$, where

$$T_v := \max(1/2, |t_v|) \asymp \sqrt{\lambda_v}. \quad (8.2)$$

The aim of this section is to prove the following bound.

Lemma 9. *Let $\begin{pmatrix} y & x \\ 1 & 1 \end{pmatrix} \in P(F_\infty)$ and $i \in \{1, \dots, h\}$. Then for any $\varepsilon > 0$ we have*

$$\phi\left(\begin{pmatrix} y & x \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \theta_i & \\ & 1 \end{pmatrix}\right) \ll_\varepsilon (|T|_\infty^{1/6} + |T/y|_\infty^{1/2})^{1+\varepsilon} (\mathcal{N}\mathfrak{n})^\varepsilon. \quad (8.3)$$

Proof. Our starting point is the Fourier–Whittaker decomposition (cf. [30, Subsection 2.3.2])

$$\phi\left(\begin{pmatrix} y & x \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \theta_i & \\ & 1 \end{pmatrix}\right) = \rho(\mathfrak{o}) \sum_{0 \neq n \in \theta_i^{-1} \mathfrak{d}^{-1}} \frac{\lambda(n\theta_i \mathfrak{d})}{\sqrt{\mathcal{N}(n\theta_i \mathfrak{d})}} W(ny) \psi(nx). \quad (8.4)$$

Here, \mathfrak{d} is the different ideal of F , $\rho(\mathfrak{o})$ is a nonzero constant (“the first Fourier coefficient”), $\lambda(\mathfrak{m})$ are the Hecke eigenvalues introduced in (3.14),

$$W(\xi) := \prod_{v|\infty} W_v(\xi_v), \quad W_v(\xi_v) := \begin{cases} \frac{|\xi_v|^{1/2} K_{it_v}(2\pi|\xi_v|)}{|\Gamma(1/2+it_v)\Gamma(1/2-it_v)|^{1/2}}, & v \text{ real,} \\ \frac{|\xi_v|^{1/2} K_{2it_v}(4\pi|\xi_v|)}{|\Gamma(1+2it_v)\Gamma(1-2it_v)|^{1/2}}, & v \text{ complex,} \end{cases}$$

is an L^2 -normalized (weight zero) Whittaker function for an appropriate Haar measure on F_∞^\times , and

$$\psi(\xi) := \prod_{v \text{ real}} e^{2\pi i \xi_v} \prod_{v \text{ complex}} e^{2\pi i (\xi_v + \bar{\xi}_v)}$$

is the corresponding (appropriately normalized) additive character on F_∞ .

The local Whittaker function W_v satisfies (even for $T_v = 1/2$)

$$W_v(\xi_v) \ll \begin{cases} \min(T_v^{1/6}, T_v^{1/4} |2\pi |\xi_v| - T_v|^{-1/4}), & |\xi_v| \leq T_v, \\ e^{-\pi |\xi_v|}, & |\xi_v| > T_v, \end{cases} \quad (8.5)$$

as follows from [4, p. 679] and [17, Prop. 9], upon noting that $|\text{Im}(t_v)| \leq 1/2$ since $\lambda_v \geq 0$ (cf. (8.1)). We fix an $\varepsilon > 0$ for the rest of this section. Then, by [30, Prop. 2.2, (2.16), (3.5), Prop. 3.2] and the fact that ϕ is a newform, we also have

$$|T|_\infty^{-\varepsilon} (\mathcal{N}\mathfrak{n})^{-\varepsilon} \ll_\varepsilon |\rho(\mathfrak{o})| \ll_\varepsilon |T|_\infty^\varepsilon (\mathcal{N}\mathfrak{n})^\varepsilon. \quad (8.6)$$

Multiplying $y \in F_\infty^\times$ and $x \in F_\infty$ by a given unit from \mathfrak{o}^\times leaves the bound (8.3) unchanged, hence we may assume that

$$|y_v|_v \asymp |T_v|_v^{\log |y|_\infty / \log |T|_\infty}, \quad v | \infty. \quad (8.7)$$

To see this, we apply Lemma 3 with $m = 1$ and observe that the product of the right hand side over all $v | \infty$ equals $|y|_\infty$. We may further assume that $y_v > 0$ for all $v | \infty$, thanks to (5.8).

We estimate the sum on the right hand side of (8.4) by the Cauchy–Schwarz inequality as

$$\begin{aligned} &\leq \left(\sum_{0 \neq n \in \theta_i^{-1} \mathfrak{d}^{-1}} \frac{|\lambda(n\theta_i \mathfrak{d})|^2}{|n|_\infty \prod_{v | \infty} (1 + |n_v y_v|_v)^{2\varepsilon}} \right)^{1/2} \\ &\quad \times \left(\sum_{0 \neq n \in \theta_i^{-1} \mathfrak{d}^{-1}} |W(ny)|^2 \prod_{v | \infty} (1 + |n_v y_v|_v)^{2\varepsilon} \right)^{1/2}, \end{aligned} \quad (8.8)$$

and we claim that the first factor satisfies

$$\sum_{0 \neq n \in \theta_i^{-1} \mathfrak{d}^{-1}} \frac{|\lambda(n\theta_i \mathfrak{d})|^2}{|n|_\infty \prod_{v | \infty} (1 + |n_v y_v|_v)^{2\varepsilon}} \ll_\varepsilon |T/y|_\infty^\varepsilon (\mathcal{N}\mathfrak{n})^\varepsilon. \quad (8.9)$$

To see this, we fix a nonzero principal fractional ideal $\mathfrak{m} \subseteq \theta_i^{-1} \mathfrak{d}^{-1}$ along with a nonnegative integral vector $\mathbf{l} = (l_v) \in \mathbb{N}^{\{v | \infty\}}$, and we estimate first the contribution of the n 's lying in

$$\{n \in \theta_i^{-1} \mathfrak{d}^{-1} : 2^{l_v} \leq 1 + |n_v y_v|_v < 2^{l_v+1} \text{ for all } v | \infty \text{ and } n\mathfrak{o} = \mathfrak{m}\}.$$

By Corollary 1(b), this contribution is

$$\ll_\varepsilon \left(\frac{|\lambda(\mathfrak{m}\theta_i \mathfrak{d})|^2}{\mathcal{N}\mathfrak{m}} \prod_{v | \infty} 2^{-2\varepsilon l_v} \right) \times \left((\mathcal{N}\mathfrak{m})^{-\varepsilon} |y|_\infty^{-\varepsilon} \prod_{v | \infty} 2^{\varepsilon l_v} \right).$$

Summing over all pairs (\mathbf{m}, \mathbf{l}) , we infer that the left hand side of (8.9) is indeed

$$\ll_{\varepsilon} |y|_{\infty}^{-\varepsilon} \sum_{0 \neq \mathbf{m} \subseteq \theta_i^{-1} \mathfrak{d}^{-1}} \frac{|\lambda(\mathbf{m} \theta_i \mathfrak{d})|^2}{(\mathcal{N} \mathbf{m})^{1+\varepsilon}} \prod_{v|\infty} \left(\sum_{l_v=1}^{\infty} 2^{-\varepsilon l_v} \right) \ll_{\varepsilon} |T/y|_{\infty}^{\varepsilon} (\mathcal{N} \mathbf{n})^{\varepsilon},$$

since the \mathbf{m} -sum is $\ll_{\varepsilon} |T|_{\infty}^{\varepsilon} (\mathcal{N} \mathbf{n})^{\varepsilon}$ by Iwaniec's trick [22, pp. 72–73].

To estimate the second factor in (8.8), we decompose F_{∞}^{\times} into generalized boxes

$$F_{\infty}^{\times} = \bigcup_{\mathbf{k}} I(\mathbf{k}), \quad I(\mathbf{k}) := \prod_{v|\infty} I_v(k_v),$$

where $\mathbf{k} = (k_v) \in \mathbb{Z}^{[v|\infty]}$ runs through integral vectors, and the local components are defined as

$$I_v(k_v) := \begin{cases} \left\{ \xi_v \in F_v^{\times} : k_v \frac{T_v}{y_v} < |\xi_v| \leq (k_v + 1) \frac{T_v}{y_v} \right\}, & k_v \geq 1, \\ \left\{ \xi_v \in F_v^{\times} : |\xi_v| \leq \frac{T_v}{y_v} \text{ and } -k_v \leq \left| |\xi_v| - \frac{T_v}{2\pi y_v} \right| < -k_v + 1 \right\}, & k_v \leq 0. \end{cases}$$

It is easy to see that $I(\mathbf{k}) = \emptyset$ unless $k_v \geq -\lfloor T_v/y_v \rfloor$ for each $v|\infty$. Correspondingly,

$$\begin{aligned} \sum_{0 \neq \mathbf{n} \in \theta_i^{-1} \mathfrak{d}^{-1}} |W(\mathbf{n}y)|^2 \prod_{v|\infty} (1 + |n_v y_v|_v)^{2\varepsilon} \\ = \sum_{\substack{\mathbf{k}=(k_v) \\ k_v \geq -\lfloor T_v/y_v \rfloor}} \sum_{\mathbf{n} \in \theta_i^{-1} \mathfrak{d}^{-1} \cap I(\mathbf{k})} |W(\mathbf{n}y)|^2 \prod_{v|\infty} (1 + |n_v y_v|_v)^{2\varepsilon}. \end{aligned} \quad (8.10)$$

Then the inner sum (the contribution of a given \mathbf{k}) is

$$\leq \#(\theta_i^{-1} \mathfrak{d}^{-1} \cap I(\mathbf{k})) \times \sup_{\mathbf{n} \in \theta_i^{-1} \mathfrak{d}^{-1} \cap I(\mathbf{k})} |W(\mathbf{n}y)|^2 \prod_{v|\infty} (1 + |n_v y_v|_v)^{2\varepsilon}.$$

We shall estimate both factors on the right hand side by a product of local factors over $v|\infty$. To estimate the lattice point count, we consider a fundamental parallelootope P_i for the fractional ideal $\theta_i^{-1} \mathfrak{d}^{-1}$. We can and will assume that P_i contains the origin and has diameter $D_i \ll 1$. Then we observe that

$$\bigcup_{\mathbf{n} \in \theta_i^{-1} \mathfrak{d}^{-1} \cap I(\mathbf{k})} (\mathbf{n} + P_i) \subseteq J(\mathbf{k}), \quad J(\mathbf{k}) := \prod_{v|\infty} J_v(k_v),$$

where the union on the left hand side is disjoint, and the local components are defined as

$$J_v(k_v) := \begin{cases} \left\{ \xi_v \in F_v : k_v \frac{T_v}{y_v} - D_i < |\xi_v| \leq (k_v + 1) \frac{T_v}{y_v} + D_i \right\}, & k_v \geq 1, \\ \left\{ \xi_v \in F_v : -k_v - D_i \leq \left| |\xi_v| - \frac{T_v}{2\pi y_v} \right| < -k_v + 1 + D_i \right\}, & k_v \leq 0. \end{cases}$$

Using $\text{vol}(P_i) \gg 1$ and $D_i \ll 1$, it is now clear that

$$\#(\theta_i^{-1} \mathfrak{d}^{-1} \cap I(\mathbf{k})) \ll \text{vol}(J(\mathbf{k})) = \prod_{v|\infty} \text{vol}(J_v(k_v)) \ll \prod_{v|\infty} f_v(k_v),$$

where

$$f_v(k_v) := \begin{cases} 1 + T_v/y_v, & v \text{ real and } k_v \geq 1, \\ 1, & v \text{ real and } k_v \leq 0, \\ k_v(1 + T_v/y_v)^2, & v \text{ complex and } k_v \geq 1, \\ 1 + T_v/y_v, & v \text{ complex and } k_v \leq 0. \end{cases}$$

By (8.5), we also have

$$\sup_{n \in \theta_i^{-1} \mathfrak{d}^{-1} \cap I(\mathbf{k})} |W(ny)|^2 \prod_{v|\infty} (1 + |n_v y_v|_v)^{2\varepsilon} \ll \prod_{v|\infty} g_v(k_v),$$

where

$$g_v(k_v) := \begin{cases} |k_v T_v|_v^{2\varepsilon} e^{-2\pi k_v T_v}, & k_v \geq 1, \\ |T_v|_v^{2\varepsilon} \min(T_v^{1/3}, T_v^{1/2} |k_v y_v|^{-1/2}), & k_v \leq 0. \end{cases}$$

By distributivity, we infer that the right hand side of (8.10) is

$$\ll \sum_{\substack{\mathbf{k}=(k_v) \\ k_v \geq -\lfloor T_v/y_v \rfloor}} \left(\prod_{v|\infty} f_v(k_v) g_v(k_v) \right) = \prod_{v|\infty} \left(\sum_{k_v=-\lfloor T_v/y_v \rfloor}^{\infty} f_v(k_v) g_v(k_v) \right). \quad (8.11)$$

The local factor at a real place v is

$$\begin{aligned} &\ll \sum_{k_v=1}^{\infty} (k_v T_v)^{2\varepsilon} (1 + T_v/y_v) e^{-2\pi k_v T_v} + \sum_{k_v=-\lfloor T_v/y_v \rfloor}^0 T_v^{2\varepsilon} \min(T_v^{1/3}, T_v^{1/2} |k_v y_v|^{-1/2}) \\ &\ll_{\varepsilon} T_v^{2\varepsilon} (T_v^{1/3} + T_v/y_v), \end{aligned}$$

as follows by estimating the minimum in the second sum by $T_v^{1/3}$ for $k_v = 0$ and by $T_v^{1/2} |k_v y_v|^{-1/2}$ for $k_v < 0$. Similarly, the local factor at a complex place v is

$$\begin{aligned} &\ll \sum_{k_v=1}^{\infty} (k_v T_v)^{4\varepsilon} k_v (1 + T_v/y_v)^2 e^{-2\pi k_v T_v} \\ &\quad + \sum_{k_v=-\lfloor T_v/y_v \rfloor}^0 T_v^{4\varepsilon} (1 + T_v/y_v) \min(T_v^{1/3}, T_v^{1/2} |k_v y_v|^{-1/2}) \\ &\ll_{\varepsilon} T_v^{4\varepsilon} (1 + T_v/y_v) (T_v^{1/3} + T_v/y_v) \ll T_v^{4\varepsilon} (T_v^{1/3} + T_v/y_v)^2 \ll T_v^{4\varepsilon} (T_v^{2/3} + T_v^2/y_v^2). \end{aligned}$$

All in all, the local factor at each archimedean place v is

$$\ll_{\varepsilon} |T_v|_v^{2\varepsilon} (|T_v|_v^{1/3} + |T_v/y_v|_v).$$

We conclude, by (8.10) and (8.11), that

$$\begin{aligned} \sum_{0 \neq n \in \theta_i^{-1} \mathfrak{o}^{-1}} |W(ny)|^2 \prod_{v|\infty} (1 + |n_v y_v|_v)^{2\varepsilon} &\ll_\varepsilon \prod_{v|\infty} |T_v|_v^{2\varepsilon} (|T_v|_v^{1/3} + |T_v/y_v|_v) \\ &\ll |T|_\infty^{2\varepsilon} (|T|_\infty^{1/3} + |T/y|_\infty), \end{aligned} \quad (8.12)$$

where we have used the balancing assumption (8.7) in the last step as follows:

$$\begin{aligned} |y|_\infty \leq |T|_\infty^{2/3} &\implies |T_v|_v^{1/3} \ll |T_v/y_v|_v \text{ for all } v|\infty, \\ |y|_\infty \geq |T|_\infty^{2/3} &\implies |T_v|_v^{1/3} \gg |T_v/y_v|_v \text{ for all } v|\infty. \end{aligned}$$

Substituting (8.12) into (8.8), and also using (8.9) and (8.6), we obtain finally

$$\begin{aligned} \phi \left(\begin{pmatrix} y & x \\ & 1 \end{pmatrix} \begin{pmatrix} \theta_i & \\ & 1 \end{pmatrix} \right) &\ll_\varepsilon (|T|_\infty^{1/6} + |T/y|_\infty^{1/2}) |T|_\infty^{2\varepsilon} |T/y|_\infty^{\varepsilon/2} (\mathcal{N}\mathfrak{n})^{2\varepsilon} \\ &\ll_\varepsilon (|T|_\infty^{1/6} + |T/y|_\infty^{1/2})^{1+12\varepsilon+\varepsilon} (\mathcal{N}\mathfrak{n})^{2\varepsilon}. \end{aligned}$$

Replacing ε by $\varepsilon/13$ completes the proof of (8.3). \square

Remark 5. Lemma 9 is close to optimal when $|y|_\infty$ is around $|T|_\infty$, but it does not capture the exponential decay of ϕ in the cusps. This result, however, suffices for our purposes, and in combination with Lemma 5, it yields a bound for $|\phi(g)|$ at any $g \in \mathrm{GL}_2(\mathbb{A})$.

9. The amplifier

In Section 3, we have seen that any function $f \in \mathcal{H}^b$ such that the operator $R(f)$ is positive can be used to bound $|\phi(g)|$ at a given point $g \in \mathrm{GL}_2(\mathbb{A})$ (cf. (3.15)), complementing the Fourier bound of the previous section (cf. Remark 5). In this section, we construct this *amplifier* as a pure tensor $f = f_\infty \otimes f_{\mathrm{fin}}$, where $f_\infty \in \mathcal{H}_\infty^b$ and $f_{\mathrm{fin}} \in \mathcal{H}_{\mathrm{fin}}^b$. Then $R(f)$ is the product of the commuting operators $R(f_\infty)$ and $R(f_{\mathrm{fin}})$, hence it is positive as long as both $R(f_\infty)$ and $R(f_{\mathrm{fin}})$ are (cf. [37, Section 104]).

9.1. The archimedean part of the amplifier

In Section 5, we introduced the generalized upper half-space \mathcal{H} (cf. (5.11)) as a subset of the Euclidean space \mathbb{M} (cf. (5.10)), and we identify it with G_∞/K_∞ via (4.2), (4.4), (2.4), (3.6). In particular, the left action of G_∞ on \mathcal{H} is given by generalized fractional linear transformations. For a point $P = (P_v)_{v|\infty} \in \mathcal{H}$ as in (5.12), we write $\mathfrak{S}(P_v)$ for y_v , and we define

$$u_v(P_v, Q_v) := \frac{\|P_v - Q_v\|^2}{2\mathfrak{S}(P_v)\mathfrak{S}(Q_v)}, \quad P, Q \in \mathcal{H}, \quad (9.1)$$

where $\|\cdot\|$ stands for the Euclidean norm (length) in the corresponding \mathbb{C} or \mathbb{H} component of \mathbb{M} . Then $u = (u_v)_{v|\infty}$ is a point-pair invariant on $\mathcal{H} \times \mathcal{H}$, i.e. it is invariant under the diagonal left action of G_∞ .

We define $f_\infty(g)$ in terms of the nonnegative vector $u(g\mathbf{i}, \mathbf{i})$, where $\mathbf{i} := P(0, 1) \in \mathcal{H}$ corresponds to the identity element of G_∞ (cf. (5.12)). Specifically, in the next subsection, we shall choose⁶ a smooth, compactly supported function $k_v : [0, \infty) \rightarrow \mathbb{R}$ for each $v | \infty$, and we put, for $g \in G_\infty$,

$$f_\infty(g) := \prod_{v|\infty} f_v(g_v), \quad f_v(g_v) := k_v(u_v(g_v \mathbf{i}_v, \mathbf{i}_v)). \quad (9.2)$$

That is, $R(f_\infty) = R(\otimes_{v|\infty} f_v) = \prod_{v|\infty} R(f_v)$ is the product of the commuting operators $R(f_v)$, hence it is positive as long as the factors $R(f_v)$ are (cf. [37, Section 104]).

We determine $k_v(u)$ in terms of its Selberg/Harish-Chandra transform $h_v(t)$ on \mathcal{H}^2 or \mathcal{H}^3 (depending on the type of v), which is necessarily holomorphic, even, and rapidly decaying in every strip $|\operatorname{Im} t| < A$, and it is real-valued on $\mathbb{R} \cup \mathbb{R}i$. We shall focus on the strip $|\operatorname{Im} t| < 1$, because for the positivity of $R(f_v)$ it suffices that $h_v(t)$ has positive real part there. Indeed, in this case $h_v(t)$ is the square of an even, holomorphic, rapidly decaying function $\tilde{h}_v(t)$ in the strip $|\operatorname{Im} t| < 1$, which is real-valued on $\mathbb{R} \cup (-1, 1)i$, therefore $R(f_v)$ is the square of the corresponding self-adjoint operator $R(\tilde{f}_v)$.

9.2. Selberg/Harish-Chandra pairs

We fix once and for all a holomorphic, rapidly decaying function

$$j : \{t \in \mathbb{C} : |\operatorname{Im} t| < 1\} \rightarrow \{z \in \mathbb{C} : \operatorname{Re} z > 0\} \quad (9.3)$$

satisfying the symmetries

$$\overline{j(\bar{t})} = j(t) = j(-t). \quad (9.4)$$

We assume further that its *inverse Fourier transform*

$$\hat{j}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} j(t) e^{itx} dt, \quad x \in \mathbb{R},$$

is smooth and supported in $[-1, 1]$. Such a function $j(t)$ certainly exists, and we provide an example based on the original construction of Iwaniec–Sarnak [24]. We take a smooth, even, not identically zero function $m : \mathbb{R} \rightarrow \mathbb{R}$ supported in $[-1/2, 1/2]$ with *Fourier transform*

$$\check{m}(t) := \int_{-\infty}^{\infty} m(x) e^{-itx} dx, \quad t \in \mathbb{C},$$

and we define $j(t)$ as the convolution

$$j(t) := \int_{-\infty}^{\infty} \operatorname{sech}\left(\frac{\pi(t-s)}{2}\right) \check{m}(s)^2 ds, \quad |\operatorname{Im} t| < 1.$$

⁶ The notations f_v, g_v, k_v are independent of those in the previous section. Hopefully, this does not cause confusion.

Then $\hat{j}(x)$ is the product of 2 $\operatorname{sech}(x)$ and the self-convolution $(m*m)(x)$, and the required properties of $j(t)$ and $\hat{j}(x)$ are straightforward to verify.

After this preparation, for each archimedean place $v \mid \infty$ we consider the even holomorphic function

$$h_v(t) := j(t - T_v) + j(t + T_v), \quad (9.5)$$

where T_v is defined by (8.2). Then, in accordance with [23, (1.64)] and Lemma 5.5 in [12, Ch. 3], we obtain the inverse Selberg/Harish-Chandra transform $k_v(u)$ of $h_v(t)$ in three steps:

$$g_v(x) := \hat{h}_v(x) = 2 \cos(T_v x) \hat{j}(x), \quad q_v(w) := \frac{1}{2} g_v(|2|_v \log(\sqrt{w+1} + \sqrt{w})), \quad (9.6)$$

$$k_v(u) := \int_{u/2}^{\infty} \frac{-q'_v(w) dw}{\pi \sqrt{w-u/2}} \quad \text{for } v \text{ real}, \quad k_v(u) := \frac{-q'_v(u/2)}{\pi} \quad \text{for } v \text{ complex}. \quad (9.7)$$

In checking these formulae, it is good to keep in mind the following. For v real, u in [23, (1.64)] is $u/2$ here. For v complex, $h(1+t^2)$ and $g(x)$ in [12, Ch. 3, (5.32)] are $h_v(t/2)$ and $2g_v(2x)$ here, while $k(x)$ and $Q(\cosh(x))$ in [12, Ch. 3, (5.35)] are $k_v(x-1)$ and $4q_v(w)$ here, upon writing $\cosh(x)$ as $1+2w$, i.e. $x = 2 \log(\sqrt{w+1} + \sqrt{w})$. The next lemma summarizes the properties that we need of these functions.

Lemma 10. *The function $h_v(t)$ is holomorphic, even, and rapidly decaying in the strip $|\operatorname{Im} t| < 1$. It is positive on $\mathbb{R} \cup (-1, 1)i$, and it has positive real part in the strip $|\operatorname{Im} t| < 1$. Moreover, at the spectral parameter t_v of ϕ (cf. (8.1)) it satisfies the bound*

$$h_v(t_v) \gg 1. \quad (9.8)$$

The inverse Selberg/Harish-Chandra transform $k_v(u)$ of $h_v(t)$ is smooth, real-valued, and supported in $[0, 1]$. Moreover, it satisfies the bound

$$k_v(u) \ll \min(|T_v|_v, |T_v|_v^{1/2} |u|_v^{-1/4}), \quad u \geq 0. \quad (9.9)$$

Proof. It is clear by our construction that $h_v(t)$ is even, holomorphic, and rapidly decaying in $|\operatorname{Im} t| < 1$. It is also clear by (9.3) and (9.5) that $h_v(t)$ has positive real part in $|\operatorname{Im} t| < 1$. From (9.4) it follows that $h_v(t)$ is real on $\mathbb{R} \cup (-1, 1)i$, hence in fact it is positive there. In particular, $h_v(t_v)$ is positive. Now we prove (9.8). If $|t_v| \leq 1/2$, then $T_v = 1/2$ by (8.2), hence $h_v(t_v)$ lies in a fixed compact subset of \mathbb{C} , and (9.8) follows. If $|t_v| > 1/2$, then $t_v \in \mathbb{R}$ and $T_v = |t_v|$ by (8.2), hence $h_v(t_v) > j(0)$, and (9.8) follows again.

It is clear by our construction that $k_v(u)$ is smooth and real-valued. It is also clear by (9.6) that $g_v(x)$ is supported in $[-1, 1]$, hence $k_v(u)$ vanishes when $u \geq 2 \sinh^2(1/2)$. This shows that $k_v(u)$ is supported in $[0, 1]$. Finally, we prove (9.9). For a real place v , we follow closely the proof of [24, Lemma 1.1], so we shall be brief. By (9.6) we have

$$q_v(w) \ll 1 \quad \text{and} \quad q'_v(w) \ll \min(T_v^2, T_v w^{-1/2}),$$

hence by (9.7) we have, for any $u, \eta > 0$,

$$k_v(u) = \int_0^{\infty} \frac{-q'_v(t+u/2)}{\pi \sqrt{t}} dt = \int_0^{\eta} + \int_{\eta}^{\infty} \ll \eta^{1/2} \min(T_v^2, T_v u^{-1/2}) + \eta^{-1/2}.$$

Choosing $\eta := \min(T_v^2, T_v u^{-1/2})^{-1}$ yields the bound (9.9). For a complex place v , the bound (9.9) is immediate from (9.6) and (9.7), since in this case

$$q'_v(u/2) \ll \min(T_v^2, T_v u^{-1/2}) = \min(|T_v|_v, |T_v|_v^{1/2} |u|_v^{-1/4}).$$

The proof is complete. \square

9.3. The nonarchimedean part of the amplifier

Let $C_0 \geq 2$ be a sufficiently large constant depending only on the number field F , and let

$$C_0 \leq L \leq |T|_\infty(\mathcal{N}\mathfrak{n}) \quad (9.10)$$

be a parameter to be optimized later.⁷ We consider all the *totally split principal prime ideals* in \mathfrak{o} coprime to $\mathfrak{n}\mathfrak{q}$ which are generated by an element from $F_+^\times \cap (1 + \mathfrak{q})$ and whose norm lies in $[L, 2L]$, and for each of them we choose a totally positive generator

$$l \in F_+^\times \cap (1 + \mathfrak{q}). \quad (9.11)$$

For each rational prime that occurs as a norm here, we keep exactly one of the l 's above it; furthermore, by restricting l to an appropriate fundamental domain for $F_+^\times/\mathfrak{o}_+^\times$, we can and will assume that $l_v \asymp L^{1/n}$ for each archimedean place v (cf. Remark 1). We denote the set of these prime elements $l \in \mathfrak{o}$ by $P(L)$. If the constant $C_0 \geq 2$ in (9.10) is sufficiently large, then $P(L)$ is nonempty and its cardinality satisfies

$$\#P(L) \asymp L/\log L \quad (9.12)$$

by the extension of Dirichlet's theorem⁸ to narrow ray class groups [35, §13 in Ch. VII], since \mathfrak{q} is a fixed ideal depending only on F .

Inspired by [49, (4.11)], and using the notation of (3.12) and (3.14), we set

$$x_m := \text{sgn}(\lambda(m\mathfrak{o})) \quad \text{for } m \in \mathfrak{o} \text{ coprime to } \mathfrak{n}\mathfrak{q}, \quad (9.13)$$

$$f_{\text{fin}} := \left(\sum_{l \in P(L)} x_l t_{l\mathfrak{o}}^b \right) * \left(\sum_{l \in P(L)} x_l t_{l\mathfrak{o}}^b \right) + \left(\sum_{l \in P(L)} x_{l^2} t_{l^2\mathfrak{o}}^b \right) * \left(\sum_{l \in P(L)} x_{l^2} t_{l^2\mathfrak{o}}^b \right). \quad (9.14)$$

Clearly, $R(f_{\text{fin}})$ is positive, because it is a sum of squares of self-adjoint operators (cf. Section 3):

$$R(f_{\text{fin}}) = \left(\sum_{l \in P(L)} x_l T_{l\mathfrak{o}}^b \right)^2 + \left(\sum_{l \in P(L)} x_{l^2} T_{l^2\mathfrak{o}}^b \right)^2. \quad (9.15)$$

⁷ See also the first sentence in Section 11.

⁸ For the purposes of this paper, the following weaker version of (9.11) would suffice: for each real $v \mid \infty$ the sign of $l_v \in F_v^\times$ is constant, and for each $\mathfrak{p} \mid \mathfrak{q}$ the square class of $l_{\mathfrak{p}} \in \mathfrak{o}_{\mathfrak{p}}^\times$ is constant. Hence, by the pigeonhole principle, we could do with Dirichlet's theorem for the class group only, and we could even avoid this variant by allowing a more general amplifier.

In addition, by the multiplicativity relation (3.13), we can linearize the quadratic expression (9.14) as

$$f_{\text{fin}} = \sum_{0 \neq m \in \mathfrak{o}} w_m t_{m\mathfrak{o}}^{\flat},$$

$$w_m := \begin{cases} \sum_{l \in P(L)} (x_l^2 + x_{l^2}^2), & m = 1, \\ x_{l_1} x_{l_2} + \delta_{l_1=l_2} x_{l_1^2} x_{l_2^2}, & m = l_1 l_2 \text{ for some } l_1, l_2 \in P(L), \\ x_{l_1^2} x_{l_2^2}, & m = l_1^2 l_2^2 \text{ for some } l_1, l_2 \in P(L), \\ 0, & \text{otherwise.} \end{cases} \quad (9.16)$$

It follows from (9.13), (9.15), and the discussion of Section 3, that $R(f_{\text{fin}})\phi = c(f_{\text{fin}})\phi$ with

$$c(f_{\text{fin}}) = \left(\sum_{l \in P(L)} |\lambda(l\mathfrak{o})| \right)^2 + \left(\sum_{l \in P(L)} |\lambda(l^2\mathfrak{o})| \right)^2$$

$$\geq \frac{1}{2} \left(\sum_{l \in P(L)} |\lambda(l\mathfrak{o})| + |\lambda(l^2\mathfrak{o})| \right)^2 > \frac{1}{8} (\#P(L))^2. \quad (9.17)$$

In the last step we have used the bound $|\lambda(l\mathfrak{o})| + |\lambda(l^2\mathfrak{o})| > 1/2$ following from (3.13).

9.4. Reduction to a counting problem

Combining (8.1), (8.2), (9.8), (9.17), we see that

$$c(f) = c(f_{\text{fin}})c(f_{\infty}) = c(f_{\text{fin}}) \prod_{v|\infty} h_v(t_v) \gg c(f_{\text{fin}}) \gg (\#P(L))^2,$$

hence (3.15), (9.10), (9.12) imply for the amplifier $f \in \mathcal{H}^{\flat}$ constructed above (cf. (2.9), (8.2)) that

$$L^2 |\phi(g)|^2 \ll \sum_{\gamma \in \Gamma} f(g^{-1}\gamma g).$$

We recall that γ runs through the elements of $\text{GL}_2(F)$ modulo $Z(\mathfrak{o})$ (cf. (3.2)). Now we fix a special matrix

$$g := \begin{pmatrix} y & x \\ & 1 \end{pmatrix} \begin{pmatrix} \theta_i & \\ & 1 \end{pmatrix} \in \mathcal{F}(\mathfrak{n}) \quad (9.18)$$

and the corresponding point $P = P(x, y) \in \mathcal{H}$ as in Lemma 6. Then, by (9.2) and (9.16), the previous inequality becomes

$$L^2 |\phi(g)|^2 \ll \sum_{0 \neq m \in \mathfrak{o}} |w_m| \sum_{\gamma \in \Gamma} t_{m\mathfrak{o}}^{\flat} \left(\begin{pmatrix} \theta_i^{-1} & \\ & 1 \end{pmatrix} \gamma_{\text{fin}} \begin{pmatrix} \theta_i & \\ & 1 \end{pmatrix} \right) |k(u(\gamma P, P))|,$$

where

$$k(u(\gamma P, P)) := \prod_{v|\infty} k_v(u_v(\gamma_v P_v, P_v)). \quad (9.19)$$

Assume that $\gamma \in \Gamma$ contributes to the inner sum for a given nonzero integer $m \in \mathfrak{o}$. By (9.16) and (9.11), m is totally positive and congruent to 1 modulo \mathfrak{q} , and by (3.7), (3.8), (3.12), γ is represented by a matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL_2(F)$ satisfying

$$a, d \in \mathfrak{o}, \quad a-d \in \mathfrak{q}, \quad b \in \theta_i \mathfrak{o}, \quad c \in \theta_i^{-1}(\mathfrak{n} \cap \mathfrak{q}), \quad ad - bc = mu \text{ for some } u \in \mathfrak{o}^\times.$$

Here $bc \in \mathfrak{q}$, hence $ad - bc \in a^2 + \mathfrak{q}$, and so the unit $u \in \mathfrak{o}^\times$ is a quadratic residue modulo \mathfrak{q} . By the choice of \mathfrak{q} (cf. Section 2), we see that $u = v^2$ for some $v \in \mathfrak{o}^\times$, and our relations become

$$a, d \in \mathfrak{o}, \quad a-d \in \mathfrak{q}, \quad b \in \theta_i \mathfrak{o}, \quad c \in \theta_i^{-1}(\mathfrak{n} \cap \mathfrak{q}), \quad ad - bc = mv^2 \text{ for some } v \in \mathfrak{o}^\times.$$

Multiplying $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL_2(F)$ by the inverse of $\begin{pmatrix} v & \\ & v \end{pmatrix} \in Z(\mathfrak{o})$ does not change the class $\gamma \in \Gamma$ represented, nor does it change any of the congruence conditions on the entries a, b, c, d , so we can and will assume that $v = 1$. Looking at the range of $t_{m\mathfrak{o}}^b$ (cf. (3.12)), we arrive at

$$L^2 |\phi(g)|^2 \ll \text{vol}(K_{\text{fin}}^b)^{-1} \sum_{0 \neq m \in \mathfrak{o}} \frac{|w_m|}{\sqrt{N(m\mathfrak{o})}} \sum_{\gamma \in \Gamma(i, m)} |k(u(\gamma P, P))|, \quad (9.20)$$

where

$$\Gamma(i, m) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL_2(F) : a, d \in \mathfrak{o}, a-d \in \mathfrak{q}, b \in \theta_i \mathfrak{o}, \right. \\ \left. c \in \theta_i^{-1}(\mathfrak{n} \cap \mathfrak{q}), ad - bc = m \right\}. \quad (9.21)$$

We note that the entry b here is always integral, because our ideal class representative $\theta_i \in \mathbb{A}_{\text{fin}}^\times$ lies in $\hat{\mathfrak{o}}$ (cf. Section 2). Most of the time we shall ignore the condition $a-d \in \mathfrak{q}$. It will only become important in Lemma 17, which is the cornerstone for the proof of Theorem 2.

We now associate to each $\gamma \in \Gamma(i, m)$ a certain dyadic

$$\delta = (\delta_v)_{v|\infty} = (2^{k_v})_{v|\infty} \in (0, 4]^{v|\infty}$$

with the property that the contribution $|k(u(\gamma P, P))|$ to the right hand side of (9.20) can be estimated in terms of $|\delta|_\infty$. Pick any $\gamma \in \Gamma(i, m)$, and for each $v|\infty$ consider the smallest $k_v \in \mathbb{Z}$ such that

$$\max(T_v^{-2}, u_v(\gamma_v P_v, P_v)) \leq 2^{k_v}.$$

We can restrict to the case $k_v \leq 2$, because otherwise $u_v(\gamma_v P_v, P_v) > 4$ (noting that $T_v^{-2} \leq 4$), and hence $k(u(\gamma P, P)) = 0$ by (9.19) and Lemma 10. Denoting $\delta_v := 2^{k_v}$ for a moment, we have either $T_v^{-2} > \delta_v/2$ or $u_v(\gamma_v P_v, P_v) > \delta_v/2$, hence in both cases we get, by (9.9),

$$k_v(u_v(\gamma_v P_v, P_v)) \ll |T_v|_v^{1/2} |\delta_v|_v^{-1/4}.$$

This in turn implies, by (9.19),

$$k(u(\gamma P, P)) \ll |T|_{\infty}^{1/2} |\delta|_{\infty}^{-1/4}. \quad (9.22)$$

In (9.20), we rearrange the matrices $\gamma \in \Gamma(i, m)$ according to their integral vectors $\mathbf{k} \in \mathbb{Z}^{\{v|\infty\}}$, and we group together the nonzero integers $m \in \mathfrak{o}$ with the same number of prime factors (0 or 2 or 4 prime factors in $P(L)$ as in (9.16)). To summarize our findings, we define for any $j \in \{0, 1, 2\}$ and for any nonnegative vector $\delta = (\delta_v)_{v|\infty}$,

$$M(L, j, \delta) := \# \bigcup_{l_1, l_2 \in P(L)} \left\{ \gamma \in \Gamma(i, l_1^j l_2^j) : u_v(\gamma P, P) \leq \delta_v \text{ for all } v|\infty \right\}. \quad (9.23)$$

We observe that in (9.20) we have, by (2.4), (2.5), (2.9),

$$\text{vol}(K_{\text{fin}}^b)^{-1} = \text{vol}(K_{\text{fin}})^{-1} \prod_{\mathfrak{p}|q} [K_{\mathfrak{p}} : K_{\mathfrak{p}}^b] \ll \text{vol}(K_{\text{fin}})^{-1} \asymp \mathcal{N} \mathfrak{n},$$

while also $w_1 \ll L$ and $w_m \ll 1$ for $m \neq 1$ by (9.13) and (9.16). We conclude, also using (9.22),

$$|\phi(g)|^2 \asymp (\mathcal{N} \mathfrak{n}) \sum_{\substack{\mathbf{k} \in \mathbb{Z}^{\{v|\infty\}} \\ T_v^{-2} \leq \delta_v = 2^{k_v} \leq 4}} \frac{|T|_{\infty}^{1/2}}{|\delta|_{\infty}^{1/4}} \left(\frac{M(L, 0, \delta)}{L} + \frac{M(L, 1, \delta)}{L^3} + \frac{M(L, 2, \delta)}{L^4} \right). \quad (9.24)$$

We stress here that $g \in \text{GL}_2(\mathbb{A})$ is a special matrix of the form (9.18). It remains to bound the matrix counts $M(L, j, \delta)$ for the special 2-adic vectors $\delta = (\delta_v)_{v|\infty}$ as above. This is the subject of the next section.

10. Counting matrices

We shall analyze in depth the matrices $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{GL}_2(F)$ counted by $M(L, j, \delta)$ for some

$$\delta = (\delta_v)_{v|\infty} \quad \text{with} \quad 0 < \delta_v \leq 4, \quad (10.1)$$

which we assume for the rest of the paper. We recall the definitions (9.21) and (9.23). In particular, the determinant $l := ad - bc$ is totally positive: it is of the form $l = l_1^j l_2^j$ with $l_1, l_2 \in P(L)$, hence it satisfies⁹ $l_v \asymp L^{2j/n}$ for each archimedean place $v|\infty$.

⁹ Hopefully, the typographical similarity of l_v and l_1, l_2 will cause no confusion.

10.1. Estimates on matrix entries

We claim that the following inequalities hold true at any archimedean place $v \mid \infty$, with absolute implied constants (which are independent of the number field F , the auxiliary ideal \mathfrak{q} , etc.):

$$|c_v P_v + d_v| = |l_v|^{1/2}(1 + O(\sqrt{\delta_v})), \quad (10.2)$$

$$|c_v P_v - a_v| = |l_v|^{1/2}(1 + O(\sqrt{\delta_v})), \quad (10.3)$$

$$c_v y_v \ll |l_v|^{1/2}, \quad (10.4)$$

$$2c_v x_v - a_v + d_v \ll |l_v|^{1/2}, \quad (10.5)$$

$$a_v + d_v \ll |l_v|^{1/2}, \quad (10.6)$$

$$\frac{c_v}{|l_v|} y_v^2 - c_v x_v^2 + (a_v - d_v)x_v + b_v \ll y_v |l_v|^{1/2} \sqrt{\delta_v}, \quad (10.7)$$

$$-c_v x_v^2 + (a_v - d_v)x_v + b_v \ll y_v |l_v|^{1/2}, \quad (10.8)$$

$$\operatorname{Re}\left(\frac{2c_v x_v - a_v + d_v}{\sqrt{l_v}}\right) \ll \sqrt{\delta_v}, \quad (10.9)$$

$$\operatorname{Im}\left(\frac{a_v + d_v}{\sqrt{l_v}}\right) \ll \sqrt{\delta_v}. \quad (10.10)$$

Here $\sqrt{l_v}$ denotes either of the two square-roots of l_v . At the complex places, these bounds follow from [3, Section 6] upon noting that $\delta_v \ll 1$. We show below that these bounds also hold at the real places.

Let v be a real place. We restrict ourselves to $l_v > 0$, which is the only case needed for this paper, but we stress that the bounds also hold when $l_v < 0$; we omit the proof for this case in order to save space.

Starting from $\gamma_v P_v = (a_v P_v + b_v)/(c_v P_v + d_v)$, it follows that

$$\frac{\Im(\gamma_v P_v)}{\Im(P_v)} = \frac{l_v}{\|c_v P_v + d_v\|^2},$$

and so

$$\delta_v \geq u_v(\gamma_v P_v, P_v) \geq \frac{|\Im(\gamma_v P_v) - \Im(P_v)|^2}{2\Im(\gamma_v P_v)\Im(P_v)} = \frac{1}{2} \left| \frac{|l_v|^{1/2}}{\|c_v P_v + d_v\|} - \frac{\|c_v P_v + d_v\|}{|l_v|^{1/2}} \right|^2.$$

This gives (10.2) immediately, and also (10.3) upon noting that $u_v(\gamma_v P_v, P_v) = u_v(P_v, \gamma_v^{-1} P_v)$. By considering $\operatorname{Im}(c_v P_v + d_v)$, we obtain (10.4) from (10.2) upon noting that $\delta_v \ll 1$. By considering $\operatorname{Re}(c_v P_v + d_v) \pm \operatorname{Re}(c_v P_v - a_v)$, we obtain (10.5) and (10.6) from (10.2) and (10.3) upon noting that $\delta_v \ll 1$. Using also

$$\delta_v \geq u_v(\gamma_v P_v, P_v) = \frac{\|a_v P_v + b_v - c_v P_v^2 - d_v P_v\|^2}{2l_v y_v^2},$$

we obtain (10.7) and (10.9) by considering the real and imaginary parts of the complex number in the numerator. Finally, (10.8) is a consequence of (10.7) coupled with (10.4) and $\delta_v \ll 1$, while (10.10) is trivial as its left hand side vanishes.

Remark 6. As noted at the beginning of this subsection, all implied constants are absolute here. In particular, our arguments yield the following explicit version of (10.6):

$$|a_v + d_v| < 17|l_v|^{1/2}. \quad (10.11)$$

10.2. Preliminary bounds

We shall use Corollary 1(a) several times below, whenever we need to count lattice points in a box.

Lemma 11. *We have*

$$M(L, 0, \delta) \ll 1 + |y|_\infty |\delta|_\infty^{1/2}.$$

Proof. Combining (10.4) with Lemma 6(b), we see that the number of possibilities for c is

$$\#c \ll 1 + |y|_\infty^{-1} (\mathcal{N}\mathfrak{n})^{-1} \ll 1.$$

For a fixed c , the number of possibilities for the difference $a - d$ can be bounded similarly, using (10.5), namely

$$\#(a - d) \ll 1.$$

For a fixed pair $(c, a - d)$ and $l = \det \gamma = 1$, the number of possibilities for b is, by (10.7),

$$\#b \ll 1 + |y|_\infty |\delta|_\infty^{1/2}.$$

Finally, the quadruple $(c, a - d, l, b)$ determines $(c, a - d, ad, b)$, and the latter determines (c, a, d, b) up to two choices. Hence we conclude the proof by multiplying the above bounds. \square

Let $M_0(L, j, \delta)$ and $M_1(L, j, \delta)$ be the number of matrices $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(F)$ counted by $M(L, j, \delta)$ with $c = 0$ and $c \neq 0$, respectively.

Lemma 12. *For $j \in \{1, 2\}$ we have*

$$M_0(L, j, \delta) \preceq L^2 + L^{2+j} |y|_\infty |\delta|_\infty^{1/2}.$$

Proof. For the determinant $l = l_1^j l_2^j$ with $l_1, l_2 \in P(L)$, there are $\ll L^2$ choices. Let us fix the determinant; then since $l = ad$ there are $\preceq 1$ possibilities for the ideals $a\mathfrak{o}$ and $d\mathfrak{o}$. On the other hand, (10.2) and (10.3) imply that $a_v, d_v \ll |l_v|^{1/2}$, hence by Corollary 1(b) there are $\preceq 1$ choices for (a, d) . If we fix this pair, the number of possibilities for b is $\ll 1 + L^j |y|_\infty |\delta|_\infty^{1/2}$ by (10.7), and we conclude the proof by multiplying the above bounds. \square

10.3. Parabolic matrices

Let us write

$$M_1(L, j, \delta) = M_2(L, j, \delta) + M_3(L, j, \delta), \quad j \in \{1, 2\},$$

where $M_2(L, j, \delta)$ and $M_3(L, j, \delta)$ stand for the number of parabolic and nonparabolic matrices counted by $M_1(L, j, \delta)$, respectively. Here we call a matrix γ *parabolic* if $(\operatorname{tr} \gamma)^2 = 4 \det \gamma$. The main result in this subsection is the bound for $M_2(L, j, \delta)$ in Lemma 14 below. Estimates for $M_3(L, j, \delta)$ are the subject of the next subsection.

Lemma 13. *Let $j \in \{1, 2\}$. Then $M_2(L, j, \delta) = 0$ unless $|\delta|_\infty \gg L^{-2j} |y|_\infty^{-2}$.*

Proof. Assume that $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL_2(F)$ is a matrix counted by $M_2(L, j, \delta)$, so that $c \neq 0$ and $(a+d)^2 = 4(ad-bc)$. Then γ has a unique fixed point in $\mathcal{H} \cup F_\infty$, namely the field element $(a-d)/(2c) \in F$ embedded in F_∞ . For convenience we write $p := a-d$ and $q := 2c$, and we extend these elements to an invertible matrix $\begin{pmatrix} s & -r \\ -q & p \end{pmatrix} \in GL_2(F)$.

By Lemma 4, and recalling again (9.18), we have a decomposition

$$\begin{pmatrix} s & -r \\ -q & p \end{pmatrix} \begin{pmatrix} y & x \\ & 1 \end{pmatrix} \begin{pmatrix} \theta_i & \\ & 1 \end{pmatrix} = \begin{pmatrix} \tilde{t} & \tilde{s} \\ & 1 \end{pmatrix} \begin{pmatrix} \tilde{y} & \tilde{x} \\ & 1 \end{pmatrix} \begin{pmatrix} \theta_j & \\ & 1 \end{pmatrix} k,$$

where $\begin{pmatrix} \tilde{t} & \tilde{s} \\ & 1 \end{pmatrix} \in P(F)$, $\begin{pmatrix} \tilde{y} & \tilde{x} \\ & 1 \end{pmatrix} \in P(F_\infty)$, and $k \in K^*$. Multiplication on the left by $\begin{pmatrix} \tilde{t} & \tilde{s} \\ & 1 \end{pmatrix}^{-1}$ does not affect the bottom row of $\begin{pmatrix} s & -r \\ -q & p \end{pmatrix}$, and so we may assume that $s, r \in F$ have been chosen so that

$$\begin{pmatrix} s & -r \\ -q & p \end{pmatrix} \begin{pmatrix} y & x \\ & 1 \end{pmatrix} \begin{pmatrix} \theta_i & \\ & 1 \end{pmatrix} = \begin{pmatrix} \tilde{y} & \tilde{x} \\ & 1 \end{pmatrix} \begin{pmatrix} \theta_j & \\ & 1 \end{pmatrix} k. \quad (10.12)$$

Furthermore, by changing k if necessary, we may assume that $\tilde{y}_v > 0$ for any archimedean place $v | \infty$ (cf. (5.8)). Note that $|y|_\infty \geq |\tilde{y}|_\infty$ by $\begin{pmatrix} y & x \\ & 1 \end{pmatrix} \begin{pmatrix} \theta_i & \\ & 1 \end{pmatrix} \in \mathcal{F}(\mathfrak{n})$.

With such a choice of $\begin{pmatrix} s & -r \\ -q & p \end{pmatrix}$ and k , we set $\sigma := \frac{1}{ps-qr} \begin{pmatrix} p & r \\ q & s \end{pmatrix} \in GL_2(F)$ for the corresponding inverse, and we examine the parabolic matrix $\sigma^{-1} \gamma \sigma \in GL_2(F)$ by looking at its infinite and finite component separately. The infinite component has (unique) fixed point ∞ , which implies readily that

$$\sigma^{-1} \gamma \sigma = \begin{pmatrix} \lambda & b' \\ & \lambda \end{pmatrix} \in GL_2(F) \quad (10.13)$$

with $\lambda^2 = l$. In particular, $\lambda \in F$ implies $\lambda \in \mathfrak{o}$ (i.e. l is a square). We claim that $b' \in \theta_j \mathfrak{o}$. Note that $b' \neq 0$, because $c \neq 0$; therefore, this will furnish the useful bound $|b'|_\infty \gg 1$.

To justify the claim, we start by observing that (10.12) yields

$$\begin{aligned} \sigma_{\text{fin}}^{-1} &= \begin{pmatrix} s & -r \\ -q & p \end{pmatrix}_{\text{fin}} = \begin{pmatrix} \theta_j & \\ & 1 \end{pmatrix} k_{\text{fin}} \begin{pmatrix} \theta_i^{-1} & \\ & 1 \end{pmatrix}, \\ (\sigma^{-1} \gamma \sigma)_{\text{fin}} &= \begin{pmatrix} \theta_j & \\ & 1 \end{pmatrix} k_{\text{fin}} \begin{pmatrix} \theta_i^{-1} & \\ & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix}_{\text{fin}} \begin{pmatrix} \theta_i & \\ & 1 \end{pmatrix} k_{\text{fin}}^{-1} \begin{pmatrix} \theta_j^{-1} & \\ & 1 \end{pmatrix}. \end{aligned}$$

As the determinant $ad - bc$ is coprime to \mathfrak{n} , our assumptions (cf. (9.21)) yield for the middle part

$$\begin{pmatrix} \theta_i^{-1} & \\ & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix}_{\text{fin}} \begin{pmatrix} \theta_i & \\ & 1 \end{pmatrix} \in \prod_{\mathfrak{p} \nmid \mathfrak{n}} M_2(\mathfrak{o}_{\mathfrak{p}}) \prod_{\mathfrak{p} | \mathfrak{n}} K_{\mathfrak{p}}.$$

The set on the right hand side is normalized by the Atkin–Lehner group $K_{\text{fin}}^* := \prod_{\mathfrak{p}} K_{\mathfrak{p}}^*$, hence we infer

$$(\sigma^{-1} \gamma \sigma)_{\text{fin}} \in \begin{pmatrix} \theta_j & \\ & 1 \end{pmatrix} M_2(\hat{\mathfrak{o}}) \begin{pmatrix} \theta_j & \\ & 1 \end{pmatrix}^{-1}.$$

By (10.13), this implies $b' \in \theta_j \mathfrak{o}$ as claimed.

If we look at the infinite component again, (10.12) shows that $\sigma^{-1} P = \tilde{P}$, where $P = P(x, y) \in \mathcal{H}$ is as before, and $\tilde{P} := P(\tilde{x}, \tilde{y}) \in \mathcal{H}$. With this notation, we find for any archimedean place $v \mid \infty$,

$$u_v(\gamma P_v, P_v) = u_v(\sigma^{-1} \gamma P_v, \sigma^{-1} P_v) = u_v(\sigma^{-1} \gamma \sigma \tilde{P}_v, \tilde{P}_v) = u_v \left(\begin{pmatrix} \lambda & b' \\ & \lambda \end{pmatrix} \tilde{P}_v, \tilde{P}_v \right).$$

Bounding the left hand side by δ_v and evaluating the right hand side via (9.1), we see that $\delta_v \gg |b'_v|^2 |\lambda_v|^{-2} \tilde{y}_v^{-2}$ with the usual absolute value in each archimedean completion F_v . In particular, $|\delta|_{\infty} \gg |b'|_{\infty}^2 L^{-2j} |\tilde{y}|_{\infty}^{-2}$. However, as we have remarked earlier, $|b'|_{\infty} \gg 1$ and $|y|_{\infty} \geq |\tilde{y}|_{\infty}$, and therefore $|\delta|_{\infty} \gg L^{-2j} |y|_{\infty}^{-2}$. \square

We now recall the notation

$$|\delta|_{\infty} = |\delta|_{\mathbb{R}} \cdot |\delta|_{\mathbb{C}}, \quad |\delta|_{\mathbb{R}} = \prod_{v \text{ real}} \delta_v, \quad |\delta|_{\mathbb{C}} = \prod_{v \text{ complex}} \delta_v^2,$$

which is a special case of (2.1).

Lemma 14. *For $j \in \{1, 2\}$ we have*

$$M_2(L, j, \delta) \ll L^{3j} |\delta|_{\mathbb{R}}^{3/4} |\delta|_{\mathbb{C}}^{1/4} (\mathcal{N}\mathfrak{n})^{-1}.$$

Remark 7. This result is a number field version of [46, Lemma 4.4]. Unfortunately, we were unable to reconstruct the proof of Templier’s lemma. It remains unclear to us how the referenced argument in [18] would generalize to produce the bound of [46, Lemma 4.4], as there does not appear to be an obvious reason why the number of scaling matrices $\sigma_{\mathfrak{a}}$ is uniformly bounded. Here we give a simple but robust proof of an alternative bound that suffices for our purposes and would also suffice for [46].

Proof of Lemma 14. Assume that $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{GL}_2(F)$ is a matrix counted by $M_2(L, j, \delta)$, so that $c \neq 0$ and $(a + d)^2 = 4(ad - bc)$. Then, as in the previous proof, the determinant $l = ad - bc$ is a square, and $a + d = 2\lambda$ for one the two square-roots λ of l . At a real place v , we can estimate c_v by (10.2) and (10.9):

$$(c_v y_v)^2 = |c_v P_v + d_v|^2 - (c_v x_v + d_v)^2 = \lambda_v^2 (1 + O(\sqrt{\delta_v}))^2 - \lambda_v^2 (1 + O(\sqrt{\delta_v}))^2 \ll \lambda_v^2 \sqrt{\delta_v},$$

and hence

$$c_v y_v \ll L^{j/n} \delta_v^{1/4}. \quad (10.14)$$

At a complex place v , we record the simpler and weaker bound (10.4):

$$c_v y_v \ll L^{j/n}. \quad (10.15)$$

For a fixed $c \neq 0$, the identity $(a-d)^2 + 4bc = 0$ shows that the integer $a-d$ is divisible by a fixed ideal of norm at least $(\mathcal{N}(co))^{1/2}$, hence by (10.9) and (10.5) combined with Corollary 1(a), there are

$$\#(a-d) \ll 1 + L^j |\delta|_{\mathbb{R}}^{1/2} (\mathcal{N}(co))^{-1/2}$$

choices for this integer. We estimate the total number of pairs $(c, a-d)$ by summing these bounds over all nonzero elements $c \in \theta_i^{-1} \mathfrak{n}$ that satisfy (10.14) and (10.15), collecting first the pairs corresponding to a given fractional ideal co , and then applying Corollary 1(b) for each such subsum. In this way we get, for any $\varepsilon > 0$,

$$\#(c, a-d) \ll \frac{L^j |\delta|_{\mathbb{R}}^{1/4}}{|y|_{\infty}^{1+\varepsilon} (\mathcal{N}\mathfrak{n})} + \frac{L^{3j/2} |\delta|_{\mathbb{R}}^{5/8}}{|y|_{\infty}^{1/2+\varepsilon} (\mathcal{N}\mathfrak{n})} \ll \frac{L^{2j} |\delta|_{\mathbb{R}}^{3/4} |\delta|_{\mathbb{C}}^{1/4}}{\mathcal{N}\mathfrak{n}},$$

where in the last step we have bounded $|y|_{\infty}$ from below by invoking Lemma 13. Finally, the trace $a+d = 2\lambda$ can be chosen in $\ll L^j$ ways, so by $(a-d)^2 + 4bc = 0$ the total number of possibilities for the parabolic matrix γ is

$$\#(c, a-d, a+d) \ll L^{3j} |\delta|_{\mathbb{R}}^{3/4} |\delta|_{\mathbb{C}}^{1/4} (\mathcal{N}\mathfrak{n})^{-1}.$$

The proof is complete. \square

10.4. Generic matrices

Again, we shall use Corollary 1(a) several times below.

Lemma 15. *We have*

$$M_3(L, 1, \delta) \ll L^2 + \frac{L^{5/2} |\delta|_{\mathbb{R}}^{1/4}}{(\mathcal{N}\mathfrak{n})^{1/4}} + \frac{L^4 |\delta|_{\mathbb{R}} |\delta|_{\mathbb{C}}^{3/4}}{\mathcal{N}\mathfrak{n}}, \quad (10.16)$$

$$M_3(L, 2, \delta) \ll L^2 + \frac{L^4 |\delta|_{\mathbb{R}}^{1/2}}{(\mathcal{N}\mathfrak{n})^{1/2}} + \frac{L^6 |\delta|_{\mathbb{R}} |\delta|_{\mathbb{C}}^{1/2}}{\mathcal{N}\mathfrak{n}}. \quad (10.17)$$

Remark 8. Our proof actually shows that (10.16) holds for $M_1(L, 1, \delta)$ in place of $M_3(L, 1, \delta)$, but we preferred the current formulation for harmony.

Proof of Lemma 15. Let $j \in \{1, 2\}$. If $M_3(L, j, \delta)$ vanishes, then the bound stated for it holds trivially. Otherwise, we fix some $c \neq 0$ that occurs in $M_3(L, j, \delta)$. By (10.4), the number of possibilities for c satisfies

$$\#c \ll \frac{L^j}{|y|_\infty(\mathcal{N}\mathbf{n})}. \quad (10.18)$$

We denote by $M_3(L, j, \delta, c)$ the subcount of $M_3(L, j, \delta)$ with the given c , and we shall subdivide it as

$$M_3(L, j, \delta, c) = M_3(*) = \sum_{\mathbf{n}} M_3(*, \mathbf{n}) = \sum_{\mathbf{n}, \mathbf{p}} M_3(*, \mathbf{n}, \mathbf{p}) = \sum_{\mathbf{n}, \mathbf{p}, \mathbf{q}} M_3(*, \mathbf{n}, \mathbf{p}, \mathbf{q}), \quad (10.19)$$

where $*$ abbreviates “ L, j, δ, c ”, and $\mathbf{n} = (n_v)$, $\mathbf{p} = (p_v)$, $\mathbf{q} = (q_v)$ are vectors in $\mathbb{Z}^{\{v \text{ complex}\}}$ satisfying

$$n_v, p_v, q_v \ll \delta_v^{-1/2}, \quad v \text{ complex}. \quad (10.20)$$

The role of the parameters $\mathbf{n}, \mathbf{p}, \mathbf{q}$ is to partially localize $l, a - d, a + d$ at the various complex places, so that we can make the most of the bounds collected in Subsection 10.1.

The components of $\mathbf{n} \in \mathbb{Z}^{\{v \text{ complex}\}}$ satisfy

$$0 \leq n_v < 2\pi/\sqrt{\delta_v}, \quad v \text{ complex},$$

and we denote by $M_3(*, \mathbf{n})$ the subcount of $M_3(*)$ with the additional condition

$$n_v \sqrt{\delta_v} \leq \arg(l_v) < (n_v + 1)\sqrt{\delta_v}, \quad v \text{ complex}, \quad (10.21)$$

for the determinant $l = ad - bc$. If $M_3(*, \mathbf{n})$ vanishes, then we subdivide it trivially as

$$M_3(*, \mathbf{n}) = M_3(*, \mathbf{n}, \mathbf{0}) = M_3(*, \mathbf{n}, \mathbf{0}, \mathbf{0}) = 0.$$

Otherwise, we fix a matrix $\gamma_{\mathbf{n}} = \begin{pmatrix} a_{\mathbf{n}} & b_{\mathbf{n}} \\ c & d_{\mathbf{n}} \end{pmatrix}$ counted by $M_3(*, \mathbf{n})$. Any matrix $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ counted by $M_3(*, \mathbf{n})$ is determined by the differences

$$a' := a - a_{\mathbf{n}}, \quad d' := d - d_{\mathbf{n}}, \quad b' := b - b_{\mathbf{n}}. \quad (10.22)$$

By (10.21), the determinants $l := \det \gamma$ and $l_{\mathbf{n}} := \det \gamma_{\mathbf{n}}$ satisfy

$$\frac{l_v}{|l_v|} - \frac{l_{\mathbf{n},v}}{|l_{\mathbf{n},v}|} \ll \sqrt{\delta_v} \quad \text{and} \quad \frac{\sqrt{l_v}}{\sqrt{l_{\mathbf{n},v}}} = \left| \frac{l_v}{l_{\mathbf{n},v}} \right|^{1/2} + O(\sqrt{\delta_v}), \quad v | \infty, \quad (10.23)$$

with a suitable choice of the square-roots, upon noting that our determinants are totally positive. By using $l_v, l_{\mathbf{n},v} \asymp L^{2j/n}$ and $y_v \asymp |y|_\infty^{1/n}$ (cf. (5.9)), the first part of (10.23) combined with (10.4) and (10.7) yields

$$(a'_v - d'_v)x_v + b'_v \ll L^{j/n}|y|_\infty^{1/n}\sqrt{\delta_v}, \quad v | \infty, \quad (10.24)$$

while the second part of (10.23) combined with (10.9) and (10.10) yields

$$\operatorname{Re}\left(\frac{a'_v - d'_v}{\sqrt{l_{\mathbf{n},v}}}\right), \operatorname{Im}\left(\frac{a'_v + d'_v}{\sqrt{l_{\mathbf{n},v}}}\right) \ll \sqrt{\delta_v}, \quad v \mid \infty. \quad (10.25)$$

We complement these bounds with the simpler relations that follow from (10.5) and (10.6),

$$\operatorname{Im}\left(\frac{a'_v - d'_v}{\sqrt{l_{\mathbf{n},v}}}\right), \operatorname{Re}\left(\frac{a'_v + d'_v}{\sqrt{l_{\mathbf{n},v}}}\right) \ll 1, \quad v \mid \infty. \quad (10.26)$$

Of course, the above bounds for imaginary parts are trivial at the real places.

Now we denote by $M_3(*, \mathbf{n}, \mathbf{p}, \mathbf{q})$ the subcount of $M_3(*, \mathbf{n})$ with the additional conditions

$$p_v \sqrt{\delta_v} \leq \operatorname{Im}\left(\frac{a'_v - d'_v}{\sqrt{l_{\mathbf{n},v}}}\right) < (p_v + 1) \sqrt{\delta_v}, \quad v \text{ complex}, \quad (10.27)$$

$$q_v \sqrt{\delta_v} \leq \operatorname{Re}\left(\frac{a'_v + d'_v}{\sqrt{l_{\mathbf{n},v}}}\right) < (q_v + 1) \sqrt{\delta_v}, \quad v \text{ complex}. \quad (10.28)$$

By (10.26), $M_3(*, \mathbf{n}, \mathbf{p}, \mathbf{q})$ vanishes unless $p_v, q_v \ll \delta_v^{-1/2}$, so we shall restrict to $\mathbf{p}, \mathbf{q} \in \mathbb{Z}^{\{v \text{ complex}\}}$ satisfying this condition. If $M_3(*, \mathbf{n}, \mathbf{p}, \mathbf{q}) \neq 0$, then we fix a matrix $\gamma_{\mathbf{n}, \mathbf{p}, \mathbf{q}} = \begin{pmatrix} a_{\mathbf{n}, \mathbf{p}, \mathbf{q}} & b_{\mathbf{n}, \mathbf{p}, \mathbf{q}} \\ c & d_{\mathbf{n}, \mathbf{p}, \mathbf{q}} \end{pmatrix}$ counted by $M_3(*, \mathbf{n}, \mathbf{p}, \mathbf{q})$. Any matrix $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ counted by $M_3(*, \mathbf{n}, \mathbf{p}, \mathbf{q})$ is determined by the differences

$$a'' := a - a_{\mathbf{n}, \mathbf{p}, \mathbf{q}}, \quad d'' := d - d_{\mathbf{n}, \mathbf{p}, \mathbf{q}}, \quad b'' := b - b_{\mathbf{n}, \mathbf{p}, \mathbf{q}}.$$

We remark that with (10.22) and the analogous notation

$$a'_{\mathbf{n}, \mathbf{p}, \mathbf{q}} := a_{\mathbf{n}, \mathbf{p}, \mathbf{q}} - a_{\mathbf{n}}, \quad d'_{\mathbf{n}, \mathbf{p}, \mathbf{q}} := d_{\mathbf{n}, \mathbf{p}, \mathbf{q}} - d_{\mathbf{n}}, \quad b'_{\mathbf{n}, \mathbf{p}, \mathbf{q}} := b_{\mathbf{n}, \mathbf{p}, \mathbf{q}} - b_{\mathbf{n}},$$

we can also write

$$a'' = a' - a'_{\mathbf{n}, \mathbf{p}, \mathbf{q}}, \quad d'' = d' - d'_{\mathbf{n}, \mathbf{p}, \mathbf{q}}, \quad b'' = b' - b'_{\mathbf{n}, \mathbf{p}, \mathbf{q}}.$$

The point is that (10.24)–(10.28) also hold with $(a'_{\mathbf{n}, \mathbf{p}, \mathbf{q}}, d'_{\mathbf{n}, \mathbf{p}, \mathbf{q}}, b'_{\mathbf{n}, \mathbf{p}, \mathbf{q}})$ in place of (a', d', b') . In order to balance out the different sizes of δ_v at the various archimedean places, we fix a totally positive unit $s \in \mathfrak{o}_+^\times$ such that (cf. Remark 1)

$$s_v \sqrt{\delta_v} \asymp |\delta|_\infty^{1/(2n)}, \quad v \mid \infty,$$

and we switch to the variables

$$\tilde{a} := sa'', \quad \tilde{d} := sd'', \quad \tilde{b} := sb''.$$

By (10.24), these scaled differences satisfy

$$(\tilde{a}_v - \tilde{d}_v)x_v + \tilde{b}_v \ll L^{j/n} |y|_\infty^{1/n} |\delta|_\infty^{1/(2n)}, \quad v \mid \infty, \quad (10.29)$$

and by (10.25)–(10.28) they also satisfy

$$\tilde{a}_v - \tilde{d}_v \ll L^{j/n} |\delta|_\infty^{1/(2n)} \quad \text{and} \quad \tilde{a}_v + \tilde{d}_v \ll L^{j/n} s_v, \quad v \text{ real}, \quad (10.30)$$

$$\tilde{a}_v - \tilde{d}_v \ll L^{j/n} |\delta|_\infty^{1/(2n)} \quad \text{and} \quad \tilde{a}_v + \tilde{d}_v \ll L^{j/n} s_v \sqrt{\delta_v}, \quad v \text{ complex}. \quad (10.31)$$

By (10.29) and the first parts of (10.30)–(10.31), the Euclidean norm of the lattice point $(\tilde{a} - \tilde{d})P + \tilde{b} \in \Lambda(P)$ (cf. (5.13)) can be bounded as

$$\|(\tilde{a} - \tilde{d})P + \tilde{b}\| \ll L^{j/n} |y|_\infty^{1/n} |\delta|_\infty^{1/(2n)},$$

and hence by Lemma 6(d),

$$\#(\tilde{a} - \tilde{d}, \tilde{b}) \ll 1 + L^j |y|_\infty |\delta|_\infty^{1/2} (\mathcal{N}\mathfrak{n})^{1/2} + L^{2j} |y|_\infty |\delta|_\infty.$$

In addition, by the second parts of (10.30)–(10.31),

$$\#(\tilde{a} + \tilde{d}) \ll 1 + L^j |\delta|_\mathbb{C}^{1/2}. \quad (10.32)$$

Let us assume $|\delta|_\mathbb{C} \gg L^{-2j}$ for a moment. Then, as any matrix $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ counted by $M_3(*, \mathbf{n}, \mathbf{p}, \mathbf{q})$ is determined by the triple $(\tilde{a} - \tilde{d}, \tilde{b}, \tilde{a} + \tilde{d})$, we see that

$$M_3(*, \mathbf{n}, \mathbf{p}, \mathbf{q}) \ll L^j |\delta|_\mathbb{C}^{1/2} (1 + L^j |y|_\infty |\delta|_\infty^{1/2} (\mathcal{N}\mathfrak{n})^{1/2} + L^{2j} |y|_\infty |\delta|_\infty).$$

We combine this bound with (10.18)–(10.20). Using also Lemma 6(b), we obtain

$$\begin{aligned} M_3(L, j, \delta) &\ll \frac{L^j}{|y|_\infty (\mathcal{N}\mathfrak{n})} \frac{L^j |\delta|_\mathbb{C}^{1/2}}{|\delta|_\mathbb{C}^{3/4}} (1 + L^j |y|_\infty |\delta|_\infty^{1/2} (\mathcal{N}\mathfrak{n})^{1/2} + L^{2j} |y|_\infty |\delta|_\infty) \\ &\ll \frac{L^{2j}}{|\delta|_\mathbb{C}^{1/4}} + \frac{L^{3j} |\delta|_\mathbb{R}^{1/2} |\delta|_\mathbb{C}^{1/4}}{(\mathcal{N}\mathfrak{n})^{1/2}} + \frac{L^{4j} |\delta|_\mathbb{R} |\delta|_\mathbb{C}^{3/4}}{\mathcal{N}\mathfrak{n}} \\ &\ll \frac{L^{2j}}{|\delta|_\mathbb{C}^{1/4}} + \frac{L^{4j} |\delta|_\mathbb{R} |\delta|_\mathbb{C}^{3/4}}{\mathcal{N}\mathfrak{n}}. \end{aligned} \quad (10.33)$$

In the last step, we dropped the middle term, because it is the geometric mean of the other two terms. The bound obtained is valid under the assumptions (10.1) and $|\delta|_\mathbb{C} \gg L^{-2j}$.

We shall use (10.33) for $j = 1$ only, because for $j = 2$ we can do without (10.32), hence also without the second parts of (10.30)–(10.31), by the following observation. For any matrix $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ counted by $M_3(*, \mathbf{n})$, the determinant $l = l_1^2 l_2^2$ is a square and $(a - d)^2 + 4bc \neq 0$. The identity

$$0 \neq (a - d)^2 + 4bc = (a + d)^2 - 4l = (a + d - 2l_1 l_2)(a + d + 2l_1 l_2)$$

combined with (10.6) implies that each pair $(a - d, b)$ gives rise to $\ll 1$ choices for $a + d$. Indeed, there are $\ll 1$ choices for the ideals $(a + d \pm 2l_1 l_2)\mathfrak{o}$ as their product is a fixed nonzero ideal of norm $\ll L^4$. Using Corollary 1(b), and again keeping in mind (10.6), we see that for each choice of ideals there are $\ll 1$ possibilities for their

generators $a + d \pm 2l_1 l_2 \in \mathfrak{o}$, which in turn determine $a + d$. So, in the case of $j = 2$, we only need the first parts of (10.30)–(10.31), along with (10.29). Hence, instead of $M_3(*, \mathbf{n}, \mathbf{p}, \mathbf{q})$, we consider $M_3(*, \mathbf{n}, \mathbf{p})$ defined as the subcount of $M_3(*, \mathbf{n})$ with the additional condition (10.27), and we obtain an improved version of (10.33) even without the assumption $|\delta|_{\mathbb{C}} \gg L^{-2j}$:

$$\begin{aligned} M_3(L, 2, \delta) &\asymp \frac{L^2}{|y|_{\infty}(\mathcal{N}\mathfrak{n})} \frac{1}{|\delta|_{\mathbb{C}}^{1/2}} (1 + L^2 |y|_{\infty} |\delta|_{\infty}^{1/2} (\mathcal{N}\mathfrak{n})^{1/2} + L^4 |y|_{\infty} |\delta|_{\infty}) \\ &\asymp \frac{L^2}{|\delta|_{\mathbb{C}}^{1/2}} + \frac{L^4 |\delta|_{\mathbb{R}}^{1/2}}{(\mathcal{N}\mathfrak{n})^{1/2}} + \frac{L^6 |\delta|_{\mathbb{R}} |\delta|_{\mathbb{C}}^{1/2}}{\mathcal{N}\mathfrak{n}} \\ &\asymp \frac{L^2}{|\delta|_{\mathbb{C}}^{1/2}} + \frac{L^6 |\delta|_{\mathbb{R}} |\delta|_{\mathbb{C}}^{1/2}}{\mathcal{N}\mathfrak{n}}. \end{aligned} \quad (10.34)$$

In the last step, we dropped the middle term, because it is the geometric mean of the other two terms. The bound obtained is valid under the assumption (10.1).

Now we derive (10.16) from (10.33) specialized to $j = 1$. If $|\delta|_{\mathbb{C}} > L^{-2} |\delta|_{\mathbb{R}}^{-1} (\mathcal{N}\mathfrak{n})$, then (10.33) is applicable, and the second term dominates in it, so (10.16) follows. If $|\delta|_{\mathbb{C}} \leq L^{-2} |\delta|_{\mathbb{R}}^{-1} (\mathcal{N}\mathfrak{n})$, then we can replace each δ_v by some $\delta_v \leq \tilde{\delta}_v \leq 4$ so that

$$|\tilde{\delta}|_{\mathbb{R}} = |\delta|_{\mathbb{R}} \quad \text{and} \quad |\tilde{\delta}|_{\mathbb{C}} = \min(16^{j^2}, L^{-2} |\delta|_{\mathbb{R}}^{-1} (\mathcal{N}\mathfrak{n})).$$

As (10.33) is applicable with $\tilde{\delta}$ in place of δ , we obtain

$$M_3(L, 1, \delta) \leq M_3(L, 1, \tilde{\delta}) \ll \frac{L^2}{|\tilde{\delta}|_{\mathbb{C}}^{1/4}} + \frac{L^4 |\delta|_{\mathbb{R}} |\tilde{\delta}|_{\mathbb{C}}^{3/4}}{\mathcal{N}\mathfrak{n}} \ll \frac{L^2}{|\tilde{\delta}|_{\mathbb{C}}^{1/4}} \ll L^2 + \frac{L^{5/2} |\delta|_{\mathbb{R}}^{1/4}}{(\mathcal{N}\mathfrak{n})^{1/4}},$$

and (10.16) follows again.

Finally, we derive (10.17) from (10.34). If $|\delta|_{\mathbb{C}} > L^{-4} |\delta|_{\mathbb{R}}^{-1} (\mathcal{N}\mathfrak{n})$, then the second term dominates in (10.34), so (10.17) follows. If $|\delta|_{\mathbb{C}} \leq L^{-4} |\delta|_{\mathbb{R}}^{-1} (\mathcal{N}\mathfrak{n})$, then we can replace each δ_v by some $\delta_v \leq \tilde{\delta}_v \leq 4$ so that

$$|\tilde{\delta}|_{\mathbb{R}} = |\delta|_{\mathbb{R}} \quad \text{and} \quad |\tilde{\delta}|_{\mathbb{C}} = \min(16^{j^2}, L^{-4} |\delta|_{\mathbb{R}}^{-1} (\mathcal{N}\mathfrak{n})).$$

Applying (10.34) with $\tilde{\delta}$ in place of δ , we obtain

$$M_3(L, 2, \delta) \leq M_3(L, 2, \tilde{\delta}) \asymp \frac{L^2}{|\tilde{\delta}|_{\mathbb{C}}^{1/2}} + \frac{L^6 |\delta|_{\mathbb{R}} |\tilde{\delta}|_{\mathbb{C}}^{1/2}}{\mathcal{N}\mathfrak{n}} \ll \frac{L^2}{|\tilde{\delta}|_{\mathbb{C}}^{1/2}} \ll L^2 + \frac{L^4 |\delta|_{\mathbb{R}}^{1/2}}{(\mathcal{N}\mathfrak{n})^{1/2}},$$

and (10.17) follows again. \square

Our final two lemmas are the main applications of the rigidity Lemma 8. Lemma 17 below is the crucial input for the proof of Theorem 2 and the only point where the congruences in the matrix count imposed by the ideal \mathfrak{q} become relevant. Let F_0 be the maximal totally real subfield of F , and put $m := [F : F_0]$.

Lemma 16. *Suppose that F is not totally real, i.e. $m \geq 2$. Then*

$$M_3(L, 1, \delta) \ll L^2 + L^{2m} |\delta|_{\mathbb{R}}^{1/2} |\delta|_{\mathbb{C}}^{1/4} + \frac{L^{2m+1} |\delta|_{\mathbb{R}} |\delta|_{\mathbb{C}}^{3/4}}{\mathcal{N}\mathfrak{n}}.$$

Remark 9. Our proof actually shows that this bound holds for $M_1(L, 1, \delta)$ in place of $M_3(L, 1, \delta)$, but we preferred the current formulation for harmony. The result is the precise analogue of [3, Subsection 11.1], but the present proof is very different and applies to all number fields without any special treatment of CM-fields. The bound is sharp for very small distances and is responsible for a strong exponent of $|T|_{\infty}$ in Theorem 2. In fact, by arguing similarly to the proof of Lemma 17 with $F_1 := F_0((a+d)^2/l)$, we can prove that $M_3(L, 1, \delta) = 0$ unless $1 \ll L^{8(m-1)} |\delta|_{\mathbb{C}}$. We do not see how to exploit this conclusion in the context of estimating $M_3(L, 1, \delta)$ in the endgame in Section 11, so we omit this statement and its proof.

Proof of Lemma 16. The proof shares several common elements with the proof of Lemma 15, so we shall be brief at certain points. If $M_3(L, 1, \delta)$ vanishes, then the stated bound holds trivially. Otherwise, we fix some $c \neq 0$ that occurs in $M_3(L, 1, \delta)$. By (10.4), the number of possibilities for c satisfies (10.18), where we set $j := 1$. We denote by $M_3(L, 1, \delta, c)$ the subcount of $M_3(L, 1, \delta)$ with the given c , and we split it as

$$M_3(L, 1, \delta, c) = M_3(*) = M_3(*, 0) + M_3(*, 1), \quad (10.35)$$

where $*$ abbreviates “ $L, 1, \delta, c$ ”, and the subsums $M_3(*, 0)$ and $M_3(*, 1)$ refer to two complementary ranges described in (10.36) and (10.39) below.

Specifically, $M_3(*, 0)$ denotes the subcount of $M_3(*)$ satisfying the additional condition

$$|2cx - a + d|_{\infty} \leq 1. \quad (10.36)$$

In order to estimate $M_3(*, 0)$, we shall subdivide it as

$$M_3(*, 0) = \sum_{\mathbf{n}, \mathbf{q}} M_3(*, 0, \mathbf{n}, \mathbf{q}),$$

where $\mathbf{n}, \mathbf{q} \in \mathbb{Z}^{\{v \text{ complex}\}}$ are as in the proof of Lemma 15, and $M_3(*, 0, \mathbf{n}, \mathbf{q})$ denotes the subcount of $M_3(*, 0)$ satisfying (10.21) and (10.28). By (10.36), the number of choices for $a - d$ is $\ll 1$. For fixed $c, a - d, \mathbf{n}, \mathbf{q}$, (10.29) and (10.32) give

$$\#b \ll 1 + L|y|_{\infty} |\delta|_{\infty}^{1/2} \quad \text{and} \quad \#(a+d) \ll 1 + L|\delta|_{\mathbb{C}}^{1/2}.$$

Let us assume $|\delta|_{\mathbb{C}} \geq L^{-2}$ for a moment. As the triple $(a - d, b, a + d)$ determines the matrix $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ counted by $M_3(*, 0, \mathbf{n}, \mathbf{q})$, we see that

$$M_3(*, 0) = \sum_{\mathbf{n}, \mathbf{q}} M_3(*, 0, \mathbf{n}, \mathbf{q}) \ll \frac{L|\delta|_{\mathbb{C}}^{1/2}}{|\delta|_{\mathbb{C}}^{1/2}} (1 + L|y|_{\infty} |\delta|_{\infty}^{1/2}) = L + L^2 |y|_{\infty} |\delta|_{\infty}^{1/2}. \quad (10.37)$$

For a general $\delta = (\delta_v)_{v|\infty}$, we obtain

$$M_3(*, 0) \ll L + L^2|y|_\infty|\delta|_\infty^{1/2} + L|y|_\infty|\delta|_\mathbb{R}^{1/2}. \quad (10.38)$$

Indeed, if $|\delta|_\mathbb{C} \geq L^{-2}$, then this is obvious by (10.37). If $|\delta|_\mathbb{C} < L^{-2}$, then we can replace each δ_v by some $\tilde{\delta}_v \leq \tilde{\delta}_v \leq 4$ so that $|\tilde{\delta}|_\mathbb{R} = |\delta|_\mathbb{R}$ and $|\tilde{\delta}|_\mathbb{C} = L^{-2}$. As (10.37) is applicable with $\tilde{\delta}$ in place of δ , and the left hand side of (10.37) is nondecreasing in $|\delta|_\mathbb{C}$, the bound (10.38) follows again.

We turn to the analysis of $M_3(*, 1)$, which we define as the subcount of $M_3(*)$ satisfying the additional condition

$$|2cx - a + d|_\infty > 1. \quad (10.39)$$

We subdivide $M_3(*, 1)$ in terms of *dyadic vectors* $\mathbf{z} = (z_v) \in \mathbb{N}^{\{v|\infty\}}$,

$$M_3(*, 1) = \sum_{\mathbf{z}} M_3(*, 1, \mathbf{z}), \quad (10.40)$$

where $M_3(*, 1, \mathbf{z})$ denotes the subcount of $M_3(*, 1)$ with the additional condition

$$\begin{aligned} \frac{1}{2} < \frac{z_v|2c_v x_v - a_v + d_v|}{C_1|l_v|^{1/2}\sqrt{\delta_v}} &\leq 1 && \text{for } v \text{ real,} \\ \frac{1}{2} < \frac{z_v|2c_v x_v - a_v + d_v|}{C_1|l_v|^{1/2}} &\leq 1 && \text{for } v \text{ complex.} \end{aligned} \quad (10.41)$$

Here, $C_1 > 0$ is the maximum of the two implied constants in (10.5) and (10.9), so that (10.40) holds with dyadic vectors $\mathbf{z} \in \mathbb{N}^{\{v|\infty\}}$. By (10.39) and (10.41), either $M_3(*, 1, \mathbf{z})$ is empty or $|\mathbf{z}|_\infty \ll L|\delta|_\mathbb{R}^{1/2}$, and hence by $l_v \asymp L^{2/n}$ the number of choices for $a - d$ is

$$\#(a - d) \ll L|\delta|_\mathbb{R}^{1/2}|\mathbf{z}|_\infty^{-1}. \quad (10.42)$$

Fixing $a - d$ and combining (10.9) with (10.41), a moment's thought gives

$$\arg(l_v) \equiv \pi + 2 \arg(2c_v x_v - a_v + d_v) + \mathcal{O}(z_v \sqrt{\delta_v}) \pmod{2\pi}, \quad v \text{ complex.} \quad (10.43)$$

At this point the stage is set for the application of realness rigidity, Lemma 8. We write

$$N := \lceil (C_2 L^{2(m-1)} |\delta|_\mathbb{C}^{1/4})^{1/r_2} \rceil,$$

where $C_2 > 0$ is a sufficiently large constant (depending only on F) to be chosen shortly. Inequality (10.43) states that $\arg(l_v)$ belongs to a fixed interval of length $\ll z_v \sqrt{\delta_v}$ for every complex place v . We split each of these intervals into subintervals $I(v, j_v)$ of length at most $\sqrt{\delta_v}/N$, where $j_v \ll z_v N$ is a positive integer. Then, with $\mathbf{j} := (j_v)$, the total number of combinations of these subintervals at the various complex places can be bounded as

$$\#\mathbf{j} \ll |\mathbf{z}|_\mathbb{C}^{1/2} N^{r_2} \ll |\mathbf{z}|_\mathbb{C}^{1/2} (1 + L^{2(m-1)} |\delta|_\mathbb{C}^{1/4}). \quad (10.44)$$

We claim that for each combination \mathbf{j} , there exists at most one determinant $l = l_1 l_2$ with $l_1, l_2 \in P(L)$ such that $\arg(l_v) \in I(v, j_v)$ for all complex places v . Indeed, let $l = l_1 l_2$

and $l' = l'_1 l'_2$ be any two such determinants, and let $\xi := l'/l$ be their ratio. Then $\arg(l_v)$ and $\arg(l'_v)$ differ by at most $\sqrt{\delta_v}/N$ at every complex place v , and hence by the definition of $P(L)$ in Subsection 9.3,

$$\xi_v \ll 1 \quad \text{and} \quad \text{Im } \xi_v \ll \sqrt{\delta_v}/N, \quad v \mid \infty.$$

Consider the number field $F_1 := F_0(\xi)$. If $F = F_1$, then Lemma 8 is applicable, and we obtain

$$1 \ll L^{8(m-1)} |\delta|_{\mathbb{C}} / N^{4r_2} \leq 1/C_2^4.$$

By choosing the constant $C_2 > 0$ to be sufficiently large, this inequality is impossible, hence F_1 is a *proper* subfield of F . We claim that $\xi \in \mathfrak{o}$. We can write $\xi \mathfrak{o}_{F_1}$ as a ratio $\mathfrak{a}/\mathfrak{b}$ of nonzero coprime ideals $\mathfrak{a}, \mathfrak{b} \subseteq \mathfrak{o}_{F_1}$ in F_1 . Then $\xi \mathfrak{o}$ is the ratio $(\mathfrak{a}\mathfrak{o})/(\mathfrak{b}\mathfrak{o})$ of the nonzero coprime ideals $\mathfrak{a}\mathfrak{o}, \mathfrak{b}\mathfrak{o} \subseteq \mathfrak{o}$ in F , and therefore $\mathfrak{b}\mathfrak{o}$ divides $l\mathfrak{o}$. By the definition of $P(L)$, $l\mathfrak{o}$ is a product of totally split prime ideals such that distinct prime ideal factors lie above distinct rational primes, hence we infer that $\mathfrak{b}\mathfrak{o} = \mathfrak{o}$. As a result, $\xi \in \mathfrak{a}\mathfrak{o} \subseteq \mathfrak{o}$ is an integer as claimed. By switching the roles l and l' , we also see that $\xi^{-1} \in \mathfrak{o}$, i.e. $\xi \in \mathfrak{o}^\times$ is a unit. However, again by the definition of $P(L)$, this forces $\xi = 1$, i.e. $l = l'$.

By the previous paragraph, we have (cf. (10.44))

$$\#l \leq \#\mathbf{j} \ll |\mathbf{z}|_{\mathbb{C}}^{1/2} (1 + L^{2(m-1)} |\delta|_{\mathbb{C}}^{1/4}).$$

Moreover, by (10.7), for each fixed pair $(a - d, l)$ we have (noting that c is also fixed)

$$\#b \ll 1 + L|y|_{\infty} |\delta|_{\infty}^{1/2}.$$

As the triple $(a - d, l, b)$ determines the matrix $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ counted by $M_3(*, 1, \mathbf{z})$ up to two choices, we see by (10.42) and our last two bounds that

$$M_3(*, 1, \mathbf{z}) \ll \frac{L|\delta|_{\mathbb{R}}^{1/2}}{|\mathbf{z}|_{\mathbb{R}} |\mathbf{z}|_{\mathbb{C}}^{1/2}} (1 + L^{2(m-1)} |\delta|_{\mathbb{C}}^{1/4}) (1 + L|y|_{\infty} |\delta|_{\infty}^{1/2}).$$

Summing up over all dyadic vectors $\mathbf{z} \in \mathbb{N}^{\{v \mid \infty\}}$, we infer

$$M_3(*, 1) \ll (L + L^{2m-1} |\delta|_{\mathbb{R}}^{1/2} |\delta|_{\mathbb{C}}^{1/4}) (1 + L|y|_{\infty} |\delta|_{\infty}^{1/2}). \quad (10.45)$$

Finally, combining (10.18), (10.35), (10.38), (10.45), and using also Lemma 6(b), we obtain

$$\begin{aligned} M_3(L, 1, \delta) &\ll \frac{L}{|y|_{\infty} (\mathcal{N}\mathfrak{n})} (L + L^2 |y|_{\infty} |\delta|_{\infty}^{1/2} + L|y|_{\infty} |\delta|_{\mathbb{R}}^{1/2}) \\ &\quad + \frac{L}{|y|_{\infty} (\mathcal{N}\mathfrak{n})} (L + L^{2m-1} |\delta|_{\mathbb{R}}^{1/2} |\delta|_{\mathbb{C}}^{1/4}) (1 + L|y|_{\infty} |\delta|_{\infty}^{1/2}) \\ &\ll L^2 + L^{2m} |\delta|_{\mathbb{R}}^{1/2} |\delta|_{\mathbb{C}}^{1/4} + \frac{L^{2m+1} |\delta|_{\mathbb{R}} |\delta|_{\mathbb{C}}^{3/4}}{\mathcal{N}\mathfrak{n}}. \end{aligned}$$

The proof is complete. \square

Lemma 17. *We have $M_3(L, 2, \delta) = 0$ unless*

$$1 \ll L^{8(m-1)} |\delta|_{\mathbb{C}}. \quad (10.46)$$

Proof. Assume that $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL_2(F)$ is a matrix counted by $M_3(L, 2, \delta)$. Recall that the determinant is of the form $l = l_1^2 l_2^2$ with $l_1, l_2 \in P(L)$. We observe that

$$\xi := \frac{a+d}{l_1 l_2} \in F \quad (10.47)$$

satisfies, by (10.11) and (10.10),

$$|\xi_v| < 17 \quad \text{and} \quad \text{Im } \xi_v \ll \sqrt{\delta_v}, \quad v | \infty. \quad (10.48)$$

Consider the number field $F_1 := F_0(\xi)$. If $F = F_1$, then Lemma 8 is applicable, and we obtain (10.46) from (10.47) and (10.48). Otherwise, F_1 is a *proper* subfield of F . We claim that $\xi \in \mathfrak{o}$. If $\xi = 0$, then the claim is trivial. If $\xi \neq 0$, then we can write $\xi \mathfrak{o}_{F_1}$ as a ratio $\mathfrak{a}/\mathfrak{b}$ of nonzero coprime ideals $\mathfrak{a}, \mathfrak{b} \subseteq \mathfrak{o}_{F_1}$ in F_1 . Then $\xi \mathfrak{o}$ is the ratio $(\mathfrak{a}\mathfrak{o})/(\mathfrak{b}\mathfrak{o})$ of the nonzero coprime ideals $\mathfrak{a}\mathfrak{o}, \mathfrak{b}\mathfrak{o} \subseteq \mathfrak{o}$ in F , and therefore $\mathfrak{b}\mathfrak{o}$ divides $l_1 l_2 \mathfrak{o}$. Using the definition of $P(L)$ in Subsection 9.3, $l_1 l_2 \mathfrak{o}$ is a product of totally split prime ideals such that distinct prime ideal factors lie above distinct rational primes, hence we infer that $\mathfrak{b}\mathfrak{o} = \mathfrak{o}$. As a result, $\xi \in \mathfrak{a}\mathfrak{o} \subseteq \mathfrak{o}$ is an integer as claimed. Using $a-d \in \mathfrak{q}$ and $bc \in \mathfrak{q}$ (cf. (9.21)), we even see that $\xi^2 - 4 = ((a-d)^2 + 4bc)/l \in \mathfrak{q}$. Now $\xi^2 - 4 \in \mathfrak{q}$ is nonzero (because γ is nonparabolic), hence its norm $|\xi^2 - 4|_{\infty}$ is at least $\mathcal{N}\mathfrak{q}$. However, this is impossible by (2.3) and (10.48) as the following short calculation shows:

$$300^n \leq \mathcal{N}\mathfrak{q} \leq |\xi^2 - 4|_{\infty} = \prod_{v|\infty} |\xi_v^2 - 4|_v < \prod_{v|\infty} |17^2 + 4|_v = 293^n.$$

The proof is complete. \square

11. The endgame

Theorems 1 and 2 are trivial when $|T|_{\infty}(\mathcal{N}\mathfrak{n})$ is bounded, hence we can assume that $|T|_{\infty}(\mathcal{N}\mathfrak{n})$ is sufficiently large in terms of the number field F . We recall (9.24) in the form

$$|\phi(g)|^2 \ll (\mathcal{N}\mathfrak{n}) \sup_{\substack{\delta = (\delta_v)_{v|\infty} \\ T_v^{-2} \leq \delta_v \leq 4 \text{ for all } v|\infty}} \frac{|T|_{\infty}^{1/2}}{|\delta|_{\infty}^{1/4}} \left(\frac{M(L, 0, \delta)}{L} + \frac{M(L, 1, \delta)}{L^3} + \frac{M(L, 2, \delta)}{L^4} \right), \quad (11.1)$$

where L is an arbitrary amplifier length satisfying (9.10). We note that the vector δ satisfies

$$|T|_{\mathbb{R}}^{-2} \leq |\delta|_{\mathbb{R}}, \quad |T|_{\mathbb{C}}^{-2} \leq |\delta|_{\mathbb{C}}. \quad (11.2)$$

First we prove Theorem 1. For $j \in \{1, 2\}$ we have:

$$\begin{aligned}
M(L, 0, \delta) &\ll 1 + |y|_\infty |\delta|_\infty^{1/2} && \text{by Lemma 11,} \\
M_0(L, j, \delta) &\asymp L^2 + L^{2+j} |y|_\infty |\delta|_\infty^{1/2} && \text{by Lemma 12,} \\
M_2(L, j, \delta) &\asymp L^{3j} |\delta|_{\mathbb{R}}^{3/4} |\delta|_{\mathbb{C}}^{1/4} (\mathcal{N}\mathfrak{n})^{-1} && \text{by Lemma 14,} \\
M_3(L, 1, \delta) &\ll L^2 + L^{5/2} |\delta|_{\mathbb{R}}^{1/4} (\mathcal{N}\mathfrak{n})^{-1/4} + L^4 |\delta|_{\mathbb{R}} |\delta|_{\mathbb{C}}^{3/4} (\mathcal{N}\mathfrak{n})^{-1} && \text{by Lemma 15,} \\
M_3(L, 2, \delta) &\asymp L^2 + L^4 |\delta|_{\mathbb{R}}^{1/2} (\mathcal{N}\mathfrak{n})^{-1/2} + L^6 |\delta|_{\mathbb{R}} |\delta|_{\mathbb{C}}^{1/2} (\mathcal{N}\mathfrak{n})^{-1} && \text{by Lemma 15.}
\end{aligned}$$

From (11.1) and (11.2) we infer

$$|\phi(g)|^2 \asymp (\mathcal{N}\mathfrak{n}) \left(\frac{|T|_\infty}{L} + |T|_\infty^{1/2} |y|_\infty + \frac{L^2 |T|_\infty^{1/2}}{\mathcal{N}\mathfrak{n}} + \frac{|T|_{\mathbb{R}}^{1/2} |T|_{\mathbb{C}}}{(\mathcal{N}\mathfrak{n})^{1/2}} \right).$$

We equate the first and third term by choosing $L := |T|_\infty^{1/6} (\mathcal{N}\mathfrak{n})^{1/3}$, and then our bound becomes

$$\phi(g) \asymp |T|_\infty^{5/12} (\mathcal{N}\mathfrak{n})^{1/3} + |T|_{\mathbb{R}}^{1/4} |T|_{\mathbb{C}}^{1/2} (\mathcal{N}\mathfrak{n})^{1/4} \quad (11.3)$$

as long as $|y|_\infty \leq |T|_\infty^{1/3} (\mathcal{N}\mathfrak{n})^{-1/3}$, but (11.3) remains true in the opposite case by Lemma 9. This completes the proof of Theorem 1.

We prove Theorem 2 following a similar strategy. For $j \in \{1, 2\}$ we have:

$$\begin{aligned}
M(L, 0, \delta) &\ll 1 + |y|_\infty |\delta|_\infty^{1/2} && \text{by Lemma 11,} \\
M_0(L, j, \delta) &\asymp L^2 + L^{2+j} |y|_\infty |\delta|_\infty^{1/2} && \text{by Lemma 12,} \\
M_2(L, j, \delta) &\asymp L^{3j} |\delta|_{\mathbb{R}}^{3/4} |\delta|_{\mathbb{C}}^{1/4} (\mathcal{N}\mathfrak{n})^{-1} && \text{by Lemma 14,} \\
M_3(L, 1, \delta) &\ll L^2 + L^{2m} |\delta|_{\mathbb{R}}^{1/2} |\delta|_{\mathbb{C}}^{1/4} + L^{2m+1} |\delta|_{\mathbb{R}} |\delta|_{\mathbb{C}}^{3/4} (\mathcal{N}\mathfrak{n})^{-1} && \text{by Lemma 16,} \\
M_3(L, 2, \delta) &\asymp L^2 + L^{2m+2} |\delta|_{\mathbb{R}}^{1/2} |\delta|_{\mathbb{C}}^{1/4} (\mathcal{N}\mathfrak{n})^{-1/2} + L^6 |\delta|_{\mathbb{R}} |\delta|_{\mathbb{C}}^{1/2} (\mathcal{N}\mathfrak{n})^{-1} && \\
&&& \text{by Lemmata 15 and 17,}
\end{aligned}$$

where $m = [F : F_0] \geq 2$, and we inserted the factor $L^{2(m-1)} |\delta|_{\mathbb{C}}^{1/4} \gg 1$ artificially (cf. (10.46)) in the second term on the last line. From (11.1) and (11.2) we infer

$$|\phi(g)|^2 \asymp (\mathcal{N}\mathfrak{n}) \left(\frac{|T|_\infty}{L} + |T|_\infty^{1/2} \left(|y|_\infty + L^{2m-3} + \frac{L^{2m-2}}{(\mathcal{N}\mathfrak{n})^{1/2}} \right) \right).$$

Introducing the constant $C_3 := 2^n C_0$ (cf. (9.10)) and choosing

$$L := \min \left((|T|_\infty \mathcal{N}\mathfrak{n})^{\frac{1}{4m-2}}, C_3 |T|_\infty^{\frac{1}{4m-4}} \right),$$

we obtain

$$\phi(g) \asymp (|T|_\infty \mathcal{N}\mathfrak{n})^{1/2} \left((|T|_\infty \mathcal{N}\mathfrak{n})^{-\frac{1}{8m-4}} + |T|_\infty^{-\frac{1}{8m-8}} \right) \quad (11.4)$$

as long as $|y|_\infty \leq |T|_\infty^{1/4}$, but (11.4) remains true in the opposite case by Lemma 9. Using also (11.3), we can strengthen (11.4) to

$$\phi(g) \ll (|T|_\infty \mathcal{N}\mathfrak{n})^{1/2} \left((|T|_\infty \mathcal{N}\mathfrak{n})^{-\frac{1}{8m-4}} + \min\left(|T|_\infty^{-\frac{1}{8m-8}}, (\mathcal{N}\mathfrak{n})^{-1/4}\right) \right).$$

However,

$$\min\left(|T|_\infty^{-\frac{1}{8m-8}}, (\mathcal{N}\mathfrak{n})^{-1/4}\right) \leq \left(|T|_\infty^{-\frac{1}{8m-8}}\right)^{\frac{2m-2}{2m-1}} \left((\mathcal{N}\mathfrak{n})^{-1/4}\right)^{\frac{1}{2m-1}} = (|T|_\infty \mathcal{N}\mathfrak{n})^{-\frac{1}{8m-4}},$$

hence in fact our bound simplifies to

$$\phi(g) \ll (|T|_\infty \mathcal{N}\mathfrak{n})^{\frac{1}{2} - \frac{1}{8m-4}}.$$

This completes the proof of Theorem 2.

Acknowledgments. We thank the referee for a thorough reading and several constructive remarks.

The first author was supported in part by the Volkswagen Foundation, by DFG-SNF lead agency program grant BL 915/2-1, and by NSF grant DMS-1128155 while enjoying the hospitality of the Institute for Advanced Study. The second and third authors were supported by NKFIH (National Research, Development and Innovation Office) grants NK 104183, ERC_HU_15 118946, K 119528, and by the MTA Rényi Intézet Lendület Automorphic Research Group. The second author was also supported by NKFIH grant K 101855 and ERC grant AdG-321104. The third author was also supported by the Premium Postdoctoral Fellowship of the Hungarian Academy of Sciences. The fourth author was supported by NSF grant DMS-1503629 and ARC grant DP130100674, and he thanks the Max Planck Institute for Mathematics for their support and exceptional research infrastructure.

References

- [1] Assing, E.: On sup-norm bounds part I: ramified Maaß newforms over number fields. [arXiv:1710.00362](https://arxiv.org/abs/1710.00362) (2017)
- [2] Blomer, V., Brumley, F.: On the Ramanujan conjecture over number fields. *Ann. of Math. (2)* **174**, 581–605 (2011) [Zbl 1322.11039](https://zbmath.org/journal/1322.11039) [MR 2811610](https://doi.org/10.2307/2811610)
- [3] Blomer, V., Harcos, G., Milićević, D.: Bounds for eigenforms on arithmetic hyperbolic 3-manifolds. *Duke Math. J.* **165**, 625–659 (2016) [Zbl 1339.11062](https://zbmath.org/journal/1339.11062) [MR 3474814](https://doi.org/10.2307/3474814)
- [4] Blomer, V., Holowinsky, R.: Bounding sup-norms of cusp forms of large level. *Invent. Math.* **179**, 645–681 (2010) [Zbl 1243.11059](https://zbmath.org/journal/1243.11059) [MR 2587342](https://doi.org/10.2307/2587342)
- [5] Blomer, V., Maga, P.: Subconvexity for sup-norms of cusp forms on $PGL(n)$. *Selecta Math.* **22**, 1269–1287 (2016) [Zbl 1352.11054](https://zbmath.org/journal/1352.11054) [MR 3518551](https://doi.org/10.2307/3518551)
- [6] Blomer, V., Michel, P.: Hybrid bounds for automorphic forms on ellipsoids over number fields. *J. Inst. Math. Jussieu* **12**, 727–758 (2013) [Zbl 1355.11062](https://zbmath.org/journal/1355.11062) [MR 3103131](https://doi.org/10.2307/3103131)
- [7] Blomer, V., Pohl, A.: The sup-norm problem on the Siegel modular space of rank two. *Amer. J. Math.* **138**, 999–1027 (2016) [Zbl 1350.11061](https://zbmath.org/journal/1350.11061) [MR 3538149](https://doi.org/10.2307/3538149)
- [8] Browning, T., Vishe, P.: Cubic hypersurfaces and a version of the circle method for number fields. *Duke Math. J.* **163**, 1825–1883 (2014) [Zbl 1298.11098](https://zbmath.org/journal/1298.11098) [MR 3229043](https://doi.org/10.2307/3229043)
- [9] Casselman, W.: On some results of Atkin and Lehner. *Math. Ann.* **201**, 301–314 (1973) [Zbl 0239.10015](https://zbmath.org/journal/0239.10015) [MR 0337789](https://doi.org/10.2307/0337789)

- [10] Das, S., Sengupta, J.: L^∞ norms of holomorphic modular forms in the case of compact quotient. *Forum Math.* **27**, 1987–2001 (2015) [Zbl 1394.11035](#) [MR 3365786](#)
- [11] Duke, W., Friedlander, J. B., Iwaniec, H.: Bounds for automorphic L -functions. *Invent. Math.* **112**, 1–8 (1993); II, *ibid.* **115**, 219–239 (1994); Erratum to II, *ibid.* **140**, 227–242 (2000); III, *ibid.* **143**, 221–248 (2001) [Zbl 0765.11038\(I\)](#) [MR 1207474\(I\)](#) [Zbl 0812.11032\(II\)](#) [MR 1258904\(II\)](#) [Zbl 1056.11504\(Err.\)](#) [MR 1779802\(Err.\)](#) [Zbl 1163.11325\(III\)](#) [MR 1835388\(III\)](#)
- [12] Elstrodt, J., Grunewald, F., Mennicke, J.: *Groups Acting on Hyperbolic Space*. Springer Monogr. Math., Springer, Berlin (1998) [Zbl 0888.11001](#) [MR 1483315](#)
- [13] Flath, D.: Decomposition of representations into tensor products. In: *Automorphic Forms, Representations, and L -functions*, A. Borel and W. Casselman (eds.), Part 1, Proc. Sympos. Pure Math. 33, Amer. Math. Soc., Providence, RI, 179–183 (1979) [Zbl 0414.22019](#) [MR 0546596](#)
- [14] Friedlander, J., Iwaniec, H.: A mean-value theorem for character sums. *Michigan Math. J.* **39**, 153–159 (1992) [Zbl 0765.11037](#) [MR 1137896](#)
- [15] Gamburd, A.: On the spectral gap for infinite index “congruence” subgroups of $SL_2(\mathbf{Z})$. *Israel J. Math.* **127**, 157–200 (2002) [Zbl 1009.22018](#) [MR 1900698](#)
- [16] Gruber, P. M., Lekkerkerker, C. G.: *Geometry of Numbers*. 2nd ed., North-Holland Math. Library 37, North-Holland, Amsterdam (1987) [Zbl 0611.10017](#) [MR 0893813](#)
- [17] Harcos, G., Michel, P.: The subconvexity problem for Rankin–Selberg L -functions and equidistribution of Heegner points. II. *Invent. Math.* **163**, 581–655 (2006) [Zbl 1111.11027](#) [MR 2207235](#)
- [18] Harcos, G., Templier, N.: On the sup-norm of Maass cusp forms of large level: II. *Int. Math. Res. Notices* **2012**, 4764–4774 [Zbl 1332.11048](#) [MR 2989618](#)
- [19] Harcos, G., Templier, N.: On the sup-norm of Maass cusp forms of large level. III. *Math. Ann.* **356**, 209–216 (2013) [Zbl 1332.11049](#) [MR 3038127](#)
- [20] Holowinsky, R., Ricotta, G., Royer, E.: On the sup-norm of SL_3 Hecke–Maass cusp forms. [arXiv:1404.3622](#) (2014)
- [21] Iwaniec, H.: Fourier coefficients of modular forms of half-integral weight. *Invent. Math.* **87**, 385–401 (1987) [Zbl 0606.10017](#) [MR 0870736](#)
- [22] Iwaniec, H.: Small eigenvalues of Laplacian for $\Gamma_0(N)$. *Acta Arith.* **56**, 65–82 (1990) [Zbl 0702.11034](#) [MR 1067982](#)
- [23] Iwaniec, H.: *Spectral Methods of Automorphic Forms*. 2nd ed., Grad. Stud. Math. 53, Amer. Math. Soc., Providence, RI (2002) [Zbl 1006.11024](#) [MR 1942691](#)
- [24] Iwaniec, H., Sarnak, P.: L^∞ norms of eigenfunctions of arithmetic surfaces. *Ann. of Math.* (2) **141**, 301–320 (1995) [Zbl 0833.11019](#) [MR 1324136](#)
- [25] Jorgenson, J., Kramer, J.: Bounding the sup-norm of automorphic forms. *Geom. Funct. Anal.* **14**, 1267–1277 (2004) [Zbl 1078.11027](#) [MR 2135167](#)
- [26] Jorgenson, J., Kramer, J.: Bounds on Faltings’s delta function through covers. *Ann. of Math.* (2) **170**, 1–43 (2009) [Zbl 1169.14020](#) [MR 2521110](#)
- [27] Jorgenson, J., Kramer, J.: Effective bounds for Faltings’s delta function. *Ann. Fac. Sci. Toulouse Math.* (6) **23**, 665–698 (2014) [Zbl 1327.14127](#) [MR 3266709](#)
- [28] Kırıl, E. M.: Bounds on sup-norms of half-integral weight modular forms. *Acta Arith.* **165**, 385–399 (2014) [Zbl 1329.11032](#) [MR 3268710](#)
- [29] Maclachlan, C., Reid, A.: *The Arithmetic of Hyperbolic 3-Manifolds*. Grad. Texts in Math. 219, Springer, New York (2003) [Zbl 1025.57001](#) [MR 1937957](#)
- [30] Maga, P.: Subconvexity and shifted convolution sums over number fields. Ph.D. thesis, Central European Univ. (2013); http://www.etd.ceu.hu/2014/maga_peter.pdf
- [31] Marshall, S.: Sup norms of Maass forms on semisimple groups. [arXiv:1405.7033](#) (2014)

- [32] Mercer, J.: Functions of positive and negative type, and their connection with the theory of integral equations. *Philos. Trans. Roy. Soc. London Ser. A* **209**, 415–446 (1909) [Zbl 40.0408.02](#)
- [33] Milićević, D.: Large values of eigenfunctions on arithmetic hyperbolic 3-manifolds. *Geom. Funct. Anal.* **21**, 1375–1418 (2011) [Zbl 1287.11072](#) [MR 2860192](#)
- [34] Miyake, T.: On automorphic forms on GL_2 and Hecke operators. *Ann. of Math.* **94**, 174–189 (1971) [Zbl 0204.54201](#) [MR 0299559](#)
- [35] Neukirch, J.: *Algebraic Number Theory*. Grundlehren Math. Wiss. 322, Springer, Berlin (1999) [Zbl 0956.11021](#) [MR 1697859](#)
- [36] Okazaki, R.: Inclusion of CM-fields and divisibility of relative class numbers. *Acta Arith.* **92**, 319–338 (2000) [Zbl 0952.11023](#) [MR 1760241](#)
- [37] Riesz, F., Sz.-Nagy, B.: *Functional Analysis*. Ungar, New York (1955) [Zbl 0070.10902](#) [MR 1068530](#)
- [38] Saha, A.: Hybrid sup-norm bounds for Maass newforms of powerful level. *Algebra Number Theory* **11**, 1009–1045 (2017) [Zbl 06748165](#) [MR 3671430](#)
- [39] Sarnak, P.: Arithmetic quantum chaos. In: *The Schur Lectures* (Tel Aviv, 1992), Israel Math. Conf. Proc. 8, Bar-Ilan Univ., Ramat Gan, 183–236 (1995); <http://publications.ias.edu/sarnak> [Zbl 0831.58045](#) [MR 1321639](#)
- [40] Sarnak, P.: Spectra of hyperbolic surfaces. *Bull. Amer. Math. Soc. (N.S.)* **40**, 441–478 (2003) [Zbl 1045.11033](#) [MR 1997348](#)
- [41] Sarnak, P.: Letter to Cathleen Morawetz. August 2004, <http://publications.ias.edu/sarnak>
- [42] Shimura, G.: The special values of the zeta functions associated with Hilbert modular forms. *Duke Math. J.* **45**, 637–679 (1978); Corrections, *ibid.* **48**, 697 (1981) [Zbl 0394.10015](#) [MR 0630592](#)
- [43] Shimura, G.: *Introduction to the Arithmetic Theory of Automorphic Functions*. Reprint of the 1971 original, Princeton Univ. Press, Princeton, NJ (1994) [Zbl 0872.11023](#) [MR 1291394](#)
- [44] Spinu, F.: The L^4 norm of the Eisenstein series. Ph.D. thesis, Princeton Univ. (2003); <http://www.math.jhu.edu/~fspinu/math/thesis.pdf>
- [45] Steiner, R. S.: Supnorm of modular forms of half-integral weight in the weight aspect. *Acta Arith.* **177**, 201–218 (2017) [Zbl 06693133](#) [MR 3612872](#)
- [46] Templier, N.: Hybrid sup-norm bounds for Hecke–Maass cusp forms. *J. Eur. Math. Soc.* **17**, 2069–2082 (2015) [Zbl 1376.11030](#) [MR 3372076](#)
- [47] VanderKam, J.: L^∞ norms and quantum ergodicity on the sphere. *Int. Math. Res. Notices* **1997**, 329–347; Correction, *ibid.* **1998**, 65 [Zbl 0877.58056](#) [MR 1440572](#)
- [48] Venkatesh, A.: “Beyond Endoscopy” and special forms on $GL(2)$. *J. Reine Angew. Math.* **577**, 23–80 (2004) [Zbl 1061.22019](#) [MR 2108212](#)
- [49] Venkatesh, A.: Sparse equidistribution problems, period bounds, and subconvexity. *Ann. of Math. (2)* **172**, 989–1094 (2010) [Zbl 1214.11051](#) [MR 2680486](#)
- [50] Weil, A.: *Basic Number Theory*, 3rd ed., Grundlehren Math. Wiss. 144, Springer, New York (1974) [Zbl 0326.12001](#) [MR 0427267](#)
- [51] Xia, H.: On L^∞ norms of holomorphic cusp forms. *J. Number Theory* **124**, 325–327 (2007) [Zbl 1121.11038](#) [MR 2321365](#)