Computer methods for Hilbert modular forms

John Voight
University of Vermont

Workshop on Computer Methods for L-functions and Automorphic Forms
Centre de Récherche Mathématiques (CRM)
22 March 2010

Computer methods

The history of computing modular forms goes back to some of the earliest days of machine computation for number theory. The first tables of (modular) elliptic curves over $\mathbb{Q}$ appeared in the proceedings of the Antwerp conference in 1972 and these were tremendously influential on the development of the subject.

We are here (according to the "brochure") as part of an effort

> *to generate challenges and new questions for people working both on the theoretical and the experimental side of the subject, as well as [to gather] valuable data that will be precious in suggesting conjectures or revealing new lines of enquiry.*

As we turn from GL_2 over $\mathbb{Q}$ to more general settings, experiment and computation will remain an essential tool.

Modular forms

The group $\mathrm{GL}_2^+(\mathbb{Q}) = \{\gamma \in \mathrm{GL}_2(\mathbb{Q}) : \det \gamma > 0\}$ acts on the upper half-plane $\mathcal{H}$ by linear fractional transformations. For $N \in \mathbb{Z}_{>0}$, let

$$\Gamma_0(N) = \left\{ \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) : N \mid c \right\} \subset \mathrm{SL}_2(\mathbb{Z}).$$

A *cusp form* (of weight 2) and level N is a holomorphic function $f : \mathcal{H} \to \mathbb{C}$ such that

$$f(\gamma z) = f\left(\frac{az+b}{cz+d}\right) = (cz+d)^2 f(z)$$

for all $\gamma \in \Gamma_0(N)$ and such that f vanishes at the cusps. The finite-dimensional $\mathbb{C}$-vector space of cusp forms of level N is denoted $S_2(N)$.

In this talk, we consider the situation where $\mathbb{Q}$ is replaced by a totally real field.

Hilbert modular forms

Let F be a totally real field with $[F:\mathbb{Q}] = n$ and let $\mathbb{Z}_F$ be its ring of integers. Let $v_1, \ldots, v_n : F \to \mathbb{R}$ be the real places of F, and write $v_i(x) = x_i$. For $\gamma \in \mathrm{M}_2(F)$ we write $\gamma_i = v_i(\gamma) \in \mathrm{M}_2(\mathbb{R})$.

The group $\mathrm{GL}_2^+(F) = \{\gamma \in \mathrm{GL}_2(F) : \det \gamma_i > 0 \text{ for } i = 1, \ldots, n\}$ acts on $\mathcal{H}^n$ by coordinatewise linear fractional transformations $\gamma z = (\gamma_i z_i)_i$. For a nonzero ideal $\mathfrak{N} \subset \mathbb{Z}_F$, let

$$\Gamma_0(\mathfrak{N}) = \left\{ \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2^+(\mathbb{Z}_F) : c \in \mathfrak{N} \right\} \subset \mathrm{GL}_2^+(\mathbb{Z}_F).$$

A *Hilbert cusp form* (of parallel weight 2) and level $\mathfrak{N}$ is a holomorphic function $f : \mathcal{H}^n \to \mathbb{C}$ such that

$$f(\gamma z) = f\left(\frac{a_1 z_1 + b_1}{c_1 z_1 + d_1}, \ldots, \frac{a_n z_n + b_n}{c_n z_n + d_n} \right) = \prod_{i=1}^{n} \frac{(c_i z_i + d_i)^2}{\det \gamma_i} \cdot f(z)$$

for all $\gamma \in \Gamma_0(\mathfrak{N})$ and such that f vanishes at the cusps.

Let $S_2(\mathfrak{N})$ denote the space of Hilbert cusp forms of level $\mathfrak{N}$.

Hecke modules

The space $S_2(N)$ is equipped with an action of pairwise commuting *Hecke operators* $T_p \in \mathrm{End}(S_k(N))$ for each prime $p \nmid N$ (arising from correspondences, or an "averaging" operator over lattices of index p, or from a double coset decomposition), so in particular $S_2(N)$ has a basis of *eigenforms*.

Each $f \in S_2(N)$ has a Fourier expansion $f(z) = \sum_n a_n q^n$ where $a_n \in \mathbb{C}$ and $q = \exp(2\pi i z)$. If f is an eigenform, *normalized* so that $a_1 = 1$, then $T_p f = a_p f$ with $a_p \in E \subset \overline{\mathbb{Q}} \subset \mathbb{C}$. (The coefficients a_n are determined by a_p for $p \mid n$.)

In this way, the system of Hecke eigenvalues $(a_p)_p$ for a normalized eigenform f determine the form $f : \mathcal{H} \to \mathbb{C}$. These eigenvalues also determine $L(f, s) = \sum_{n=1}^{\infty} a_n / n^s$ (defined for $\mathrm{Re}\, s > 1$) as well as the ℓ-adic Galois representations $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\overline{\mathbb{Z}}_\ell)$ associated to f.

Hecke modules, continued

In a similar way, the space $S_2(\mathfrak{N})$ of Hilbert cusp forms is equipped with an action of Hecke operators $T_\mathfrak{p}$ for primes $\mathfrak{p} \nmid \mathfrak{N}$. For a normalized eigenform $f \in S_2(\mathfrak{N})$ with Hecke eigenvalues $a_\mathfrak{p}$, we again have a notion of "q-expansion" (more complicated to write down), an L-function

$$L(f, s) = \sum_{\mathfrak{n}} \frac{a_\mathfrak{n}}{N\mathfrak{n}^s}$$

and $\mathfrak{l}$-adic Galois representations $\mathrm{Gal}(\overline{F}/F) \to \mathrm{GL}_2(\overline{\mathbb{Z}}_{F,\mathfrak{l}})$.

For these reasons, the space $S_2(\mathfrak{N})$ is computed as a *Hecke module*: a finite dimensional ($\mathbb{C}$-)vector space equipped with an action of Hecke operators. The Hecke module is thereby determined by the corresponding system of Hecke eigenvalues.

Main algorithm

Theorem (Dembélé-Donnelly, Greenberg-V)

There exists an algorithm which, given a totally real field F and a nonzero ideal $\mathfrak{N} \subset \mathbb{Z}_F$, computes the system of Hecke eigenvalues for the space $S_2(\mathfrak{N})$ of Hilbert cusp forms of level $\mathfrak{N}$ over F.

In other words, there exists an explicit finite procedure which takes as input the field F and the ideal $\mathfrak{N} \subset \mathbb{Z}_F$ encoded in bits (in the usual way), and outputs: a finite set of sequences $(a_{\mathfrak{p}}(f))_{\mathfrak{p}}$ encoding the Hecke eigenvalues for each cusp form constituent f in $S_2(\mathfrak{N})$, where $a_{\mathfrak{p}}(f) \in E_f \subset \overline{\mathbb{Q}}$.

In joint work with Steve Donnelly, we have computed Hecke data for hundreds of thousands of forms over totally real fields up to degree 6.

Example

Let $F = \mathbb{Q}(\sqrt{5})$ and $\mathfrak{N} = (17) \subset \mathbb{Z}_F$. Then $\dim S_2(17) = 5$. There are 2 Hecke irreducible subspaces of dimensions 1 and 4, corresponding to newforms f and g.

$N\mathfrak{p}$	4	5	9	11
$a_{\mathfrak{p}}(f)$	-3	-2	-6	0
$a_{\mathfrak{p}}(g)$	$t^2 - 2$	$t^3 - t^2 - 5t + 1$	$-t^2 + 2t + 5$	$-t^3 + 4t + 1$

Here, the element $t \in \overline{\mathbb{Q}}$ satisfies $t^4 - 6t^2 - 2t + 5 = 0$, and $E = \mathbb{Q}(t)$ is an S_4-field of discriminant 6224. The form f is the base change from $\mathbb{Q}$ to F of the unique form in $S_2(17)$, corresponding to the isogeny class of the modular (elliptic!) curve $X_0(17)$. Using an algorithm of Weinstein, we find that g also arises as the base change of a cusp form in $S_2(425)$.

Geometry

To compute with the space $S_2(N)$, one approach is to use the geometry of the modular curve $X_0(N) = \Gamma_0(N)\backslash\mathcal{H}^*$, where $\mathcal{H}^*$ denotes the completed upper half-plane.

A cusp form $f \in S_2(N)$ corresponds to a holomorphic differential form $f(z)\,dz$ on $X_0(N)$ and so by the theorem of Eichler-Shimura arises naturally in the space $H^1(X_0(N), \mathbb{C}; \text{cusps})^+$ where the $+$ indicates the $+$-space for complex conjugation.

In a similar way, a Hilbert cusp form $f \in S_2(\mathfrak{N})$ gives rise to a holomorphic differential n-form $f(z_1, \ldots, z_n)\,dz_1 \ldots dz_n$ on the *Hilbert modular variety* $X_0(\mathfrak{N}) = \Gamma_0(\mathfrak{N})\backslash\mathcal{H}^{*n}$. But now $X_0(\mathfrak{N})$ has dimension n and f arises in $H^n(X_0(\mathfrak{N}), \mathbb{C}; \text{cusps})$. Yikes! Computing with higher dimensional varieties (and higher degree cohomology groups) is not an easy task.

General strategy

Langlands functoriality predicts that $S_2(\mathfrak{N})$ as a Hecke module occurs in the cohomology of other "modular" varieties. We use a principle called the *Jacquet-Langlands correspondence*, which allows us to work with varieties of complex dimension 0 or 1 by considering twisted forms of GL_2 over F.

Let B be the quaternion algebra over F which is split at all finite places and ramified at all or all but one real place according as $n = [F : \mathbb{Q}]$ is even or odd.

The Jacquet-Langlands correspondence is the isomorphism of Hecke modules

$$S_2(\mathfrak{N}) \xrightarrow{\sim} S_2^B(\mathfrak{N})$$

where $S_2^B(\mathfrak{N})$ denotes the space of quaternionic cusp forms for B (of weight 2) and level $\mathfrak{N}$.

The explicit description of the Hecke module $S_2^B(\mathfrak{N})$ varies accordingly as n is even or odd.

Indefinite method

Suppose first that $n = [F : \mathbb{Q}]$ is odd. Then the quaternion algebra B is split at a unique real place corresponding to $\iota_\infty : B \hookrightarrow \mathrm{M}_2(\mathbb{R})$. We call this the *indefinite method*, since B is indefinite, and it is joint work with Matthew Greenberg.

For expositional simplicity, suppose that F has strict class number 1. Then $\mathbb{Z}_{F,+}^\times = \{x \in \mathbb{Z}_F^\times : x_i > 0 \text{ for all } i\} = \mathbb{Z}_F^{\times 2}$ and hence $\mathrm{GL}_2^+(\mathbb{Z}_F) = \mathbb{Z}_F^\times \mathrm{SL}_2(\mathbb{Z}_F)$. We further assume $B \ncong \mathrm{M}_2(\mathbb{Q})$ for uniformity of presentation.

Indefinite method

Let $\mathcal{O}_0(\mathfrak{N}) \subset B$ be an Eichler order of level $\mathfrak{N}$ let

$$\mathcal{O}_0(\mathfrak{N})_1^\times = \{\gamma \in \mathcal{O}_0(\mathfrak{N}) : \mathrm{nrd}(\gamma) = 1\}$$

and let

$$\Gamma_0(\mathfrak{N}) = \iota_\infty(\mathcal{O}_0(\mathfrak{N})_1^\times) \subset \mathrm{SL}_2(\mathbb{R}).$$

Then $\Gamma_0(\mathfrak{N})$ is a discrete and cocompact subgroup of $\mathrm{SL}_2(\mathbb{R})$; so $X_0^B(\mathfrak{N}) = \Gamma_0(\mathfrak{N})\backslash\mathcal{H}$ is a compact Riemann surface, a *Shimura curve*.

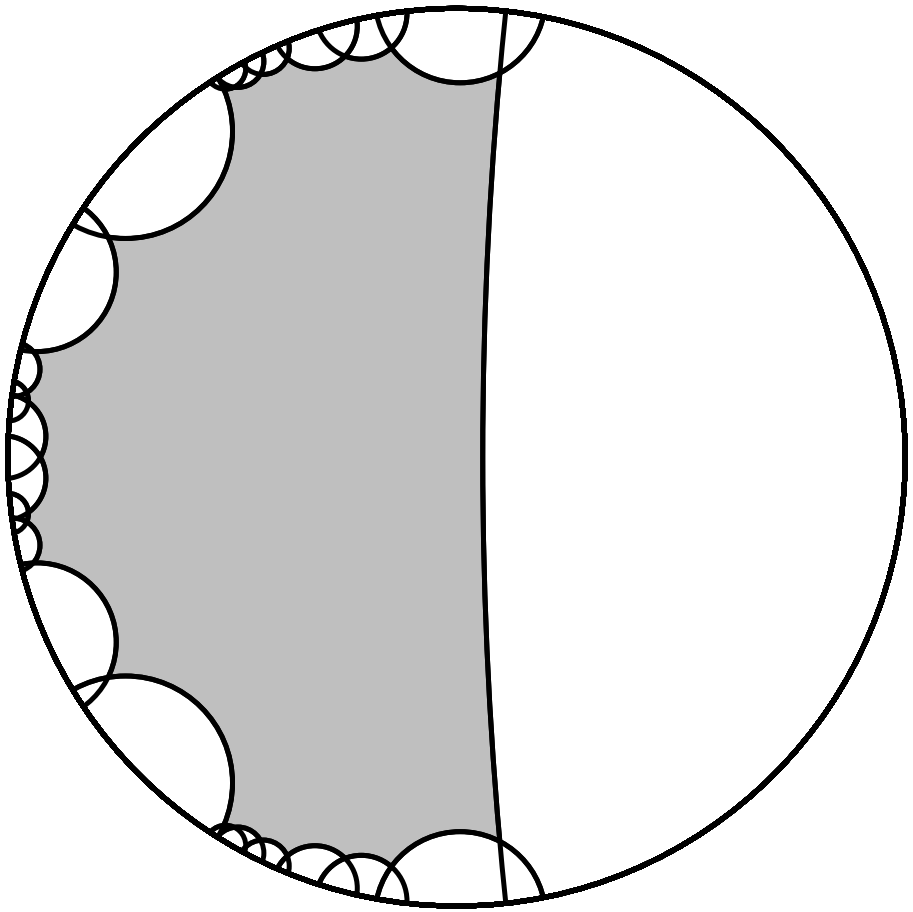
A *quaternionic cusp form* for B is a holomorphic function $f : \mathcal{H} \to \mathbb{C}$ such that $f(\gamma z) = (cz + d)^2 f(z)$ for all $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(\mathfrak{N})$. (No cusps!)

Putting Jacquet-Langlands together with the Eichler-Shimura isomorphism, we have

$$S_2(\mathfrak{N}) \cong S_2^B(\mathfrak{N}) \cong H^1(X_0^B(\mathfrak{N}), \mathbb{C})^+.$$

Example

Let F be the (totally real) cubic field with $d_F = 1101 = 3 \cdot 367$. Then $F = \mathbb{Q}(w)$ with $w^3 - w^2 - 9w + 12 = 0$. The field F has Galois group S_3. Here we work with the Shimura curve $X = X_0^B(1)$ associated to F. The curve X has signature $(1; 2^2, 3^5)$.



Example

We obtain the following Hecke data:

$N\mathfrak{p}$	π	$a(\mathfrak{p})$	$\#J(\mathbb{F}_p)$
2	$w-2$	0	3
3	$w-3$	-3	7
3	$w-1$	-1	5
4	w^2+w-7	-3	8
19	$w+1$	-6	26
23	w^2-2w-1	6	18
31	$2w^2-19$	3	29
31	w^2-5	0	32
31	$3w-5$	4	28

Here, $J = J_0^B(1)$ is the Jacobian of the Shimura curve X.

Example

Using a method of Cremona and Lingham, we find a candidate elliptic curve A to represent the isogeny class of the Jacobian J:

$$A : y^2 + w(w+1)xy + (w+1)y = x^3 + w^2x^2 + a_4x + a_6$$

where a_4 is equal to

$$-13967140935029686 4w^2 - 23568148183993846 8w + 62367237016191282 2$$

and a_6 is equal to

$$110726054056401930182106463w^2 + 186839095087977344668356726w - 494423184252818697135532743.$$

Using the method of Faltings and Serre, we verify that J is indeed isogeneous to A.

Indefinite method: Hecke operators

Recall we compute with the (Hecke module) $H^1(X, \mathbb{C})$, where $X = X_0^B(\mathfrak{N}) = \Gamma\backslash\mathcal{H}$ and $\Gamma = \Gamma_0(\mathfrak{N})$. We have simply

$$H^1(X, \mathbb{C}) = H^1(\Gamma, \mathbb{C}) = \mathrm{Hom}(\Gamma, \mathbb{C});$$

if X has genus g, then $\mathrm{Hom}(\Gamma, \mathbb{C})$ is a vector space of dimension $2g$ with basis given by the characteristic functions of a set of generators for $\Gamma/[\Gamma, \Gamma]$.

The space $\mathrm{Hom}(\Gamma, \mathbb{C})$ is equipped with Hecke operators $T_\mathfrak{p}$ as follows. Let $\mathfrak{p} \subset \mathbb{Z}_F$ be a prime with $\mathfrak{p} \nmid \mathfrak{N}$ and let $k_\mathfrak{p}$ be its residue field. For $f : \Gamma \to \mathbb{C}$, we define

$$(f \mid T_\mathfrak{p})(\gamma) = \sum_{a \in \mathbb{P}^1(k_\mathfrak{p})} f(\delta_a)$$

where α_a for $a \in \mathbb{P}^1(k_\mathfrak{p})$ are generators of the left $\mathcal{O}$-ideals of norm $\mathfrak{p}$ having totally positive reduced norm and

$$\delta_a = \alpha_a \gamma \alpha_{\gamma^* a}^{-1} \in \Gamma.$$

Indefinite method: computational problems

To compute effectively the systems of Hecke eigenvalues in the cohomology of a Shimura curve, we need algorithms to:

1. Compute an explicit finite presentation of Γ;
2. Compute a generator (with totally positive reduced norm) of a left ideal $I \subset \mathcal{O}$; and
3. Given $\delta \in \Gamma$, write δ as an explicit word in the generators for Γ, i.e., solve the word problem in Γ.

Problems 1 and 3 are solved by computing a *Dirichlet domain*, a fundamental domain for Γ equipped with a side pairing. A reduction algorithm is used to solve the word problem. Problem 2 is solved using lattice methods.

Modular symbols

The indefinite method can be viewed as a generalization of the method of modular symbols used in the classical case $\Gamma = \Gamma_0(N)$. There, we have a canonical isomorphism

$$\mathcal{S}_2(N) \cong H_1(X_0(N), \mathbb{Z}) (\cong H^1(X_0(N), \mathbb{Z}))$$

where $\mathcal{S}_2(N)$ is the space of *cuspidal modular symbols*, the space of paths in $\mathcal{H}^*$ whose endpoints are cusps and which are loops in $X_0(N)$. There is an explicit description of the action of the Hecke operators on the space $\mathcal{S}_2(N)$, and the *Manin trick* (the Euclidean algorithm) gives a method for writing a modular symbol as a $\mathbb{Z}$-linear combination of generating symbols $\gamma_i\{0, \infty\}$. This method has been fruitfully pursued by Cremona, Stein, and others.

Dirichlet modular symbols

The Shimura curves $X = X_0^B(\mathfrak{N})$ do not have cusps, and so the method of modular symbols does not generalize directly. However, the side pairing of a Dirichlet domain for Γ gives an explicit characterization of the gluing relations which describe X as a Riemann surface, hence one obtains a complete description for the homology group $H_1(X, \mathbb{Z})$.

Paths are now written $\{v, \gamma v\}$ for v a vertex on a side paired by $\gamma \in G$. The analogue of the Manin trick in our context is played by the solution to the word problem in Γ. And computationally, these points of view are equivalent.

Definite method

Now suppose that $n = [F : \mathbb{Q}]$ is even; the method is due to Dembélé and Dembelé-Donnelly. For expositional simplicity, we again suppose F has strict class number 1.

In this case, the quaternion algebra B is ramified at all real places and so is totally definite. In this case, the Shimura variety associated to B is zero-dimensional: it consists of a finite set of points labelled by the (right) $\mathcal{O}$-ideal classes, where $\mathcal{O} = \mathcal{O}_0(\mathfrak{N})$. We write $X = \mathrm{Cl}\,\mathcal{O}$ for this set and $H = \#X$.

A *quaternionic cusp form* for B of level $\mathfrak{N}$ (and parallel weight 2) is just an element of the space

$$S_2^B(\mathfrak{N}) = \mathrm{Map}(X, \mathbb{C})/\mathbb{C} \cong \mathbb{C}^{H-1}$$

where $\mathbb{C}$ is the space of constant functions.

Definite method: Brandt matrices

The Hecke operators acting on $\bigoplus_i \mathbb{C} I_i = \mathrm{Map}(\mathrm{Cl}\,\mathcal{O}, \mathbb{C})$ are given by Brandt matrices.

This method goes back to Brandt who was working with theta series associated to positive definite quaternary quadratic forms over $\mathbb{Z}$; it was developed further by Eichler, Pizer, and Hijikata-Pizer-Shemanske, Socrates-Whitehouse, and was implemented in Magma by Kohel in the case $F = \mathbb{Q}$.

Definite method: Brandt matrices

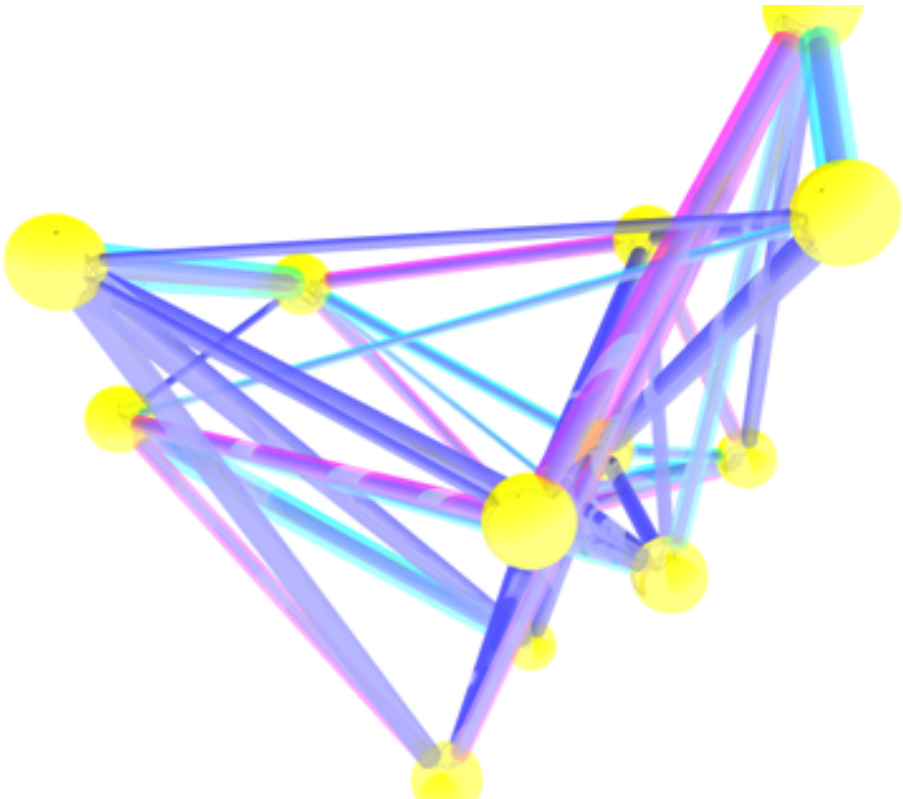
Choose a set of representatives $I_1, \ldots, I_H$ for $\mathrm{Cl}\,\mathcal{O}$, and for simplicity suppose that $\mathrm{nrd}(I_i) = \mathrm{nrd}(I_j) = \mathfrak{q}$ for all i, j. Let $\mathcal{O}_j = \mathcal{O}_L(I_j)$ be the left order of I_j and let $e_j = \#(\mathcal{O}_j)_1^\times$. For a prime $\mathfrak{p} \nmid \mathfrak{q}\mathfrak{N}$, let π be a totally positive generator for $\mathfrak{p}$, and define the $\mathfrak{p}$th-*Brandt matrix* for $\mathcal{O}$ to be the matrix whose (i, j)th entry is equal to

$$\frac{1}{e_j}\#\{x \in I_j^{-1} I_i : \mathrm{nrd}(x) = \pi\}.$$

The Hecke operator $T_\mathfrak{p}$ then acts by this Brandt matrix on $\mathrm{Map}(\mathrm{Cl}\,\mathcal{O}, \mathbb{C})$.

Definite method: Brandt matrices

The Brandt matrix is just a compact way of writing down the adjacency matrix of the graph with vertices $X = \mathrm{Cl}\,\mathcal{O}$ where there is an edge (weighted by units) from I_i to each ideal class which represents an ideal of index $N\mathfrak{p}$ in I_i.



Here, $F = \mathbb{Q}(\sqrt{3})$, $\mathfrak{N} = (11)$, and $\mathfrak{p}$ is an ideal of norm 23.

Definite method: computational problems

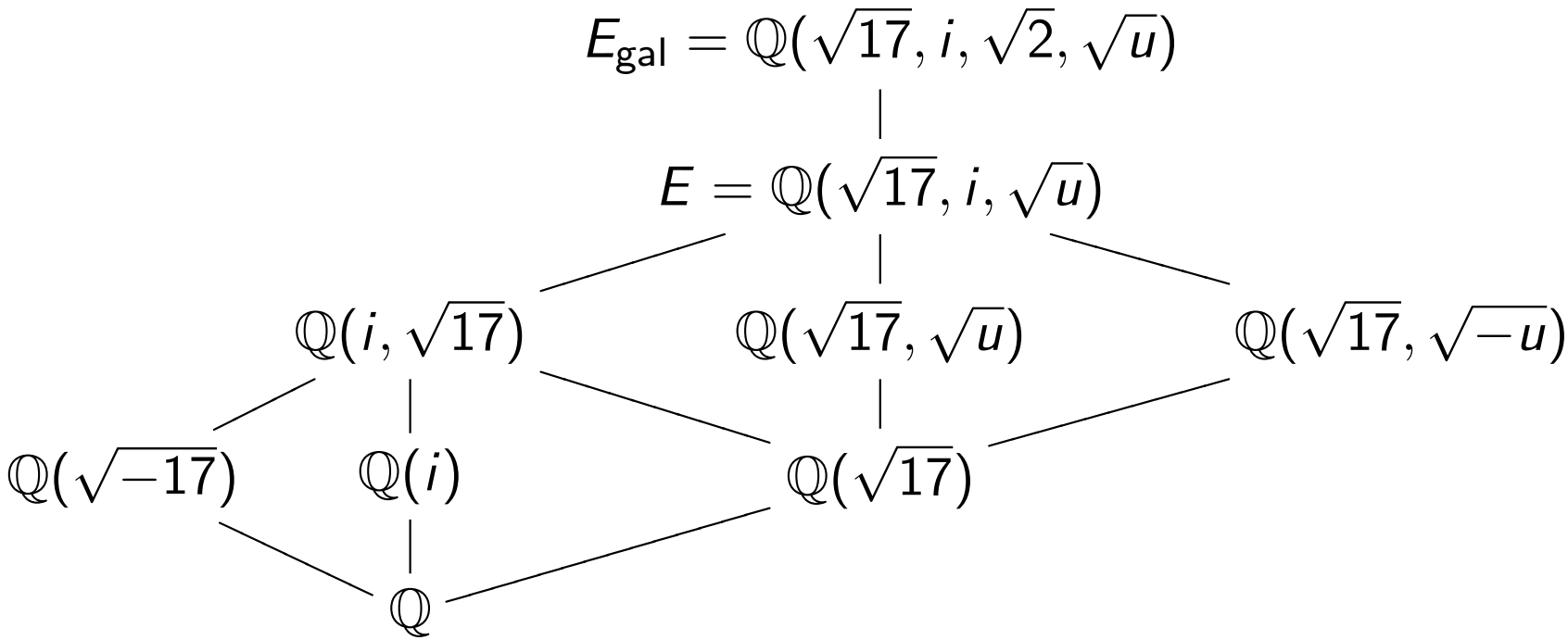
In addition to basic algorithms for working with quaternion orders and ideals, to compute Brandt matrices, we need algorithms to:

1. Compute a set of representatives for $\mathrm{Cl}\,\mathcal{O}$; and
2. Test if two right $\mathcal{O}$-ideals are isomorphic.

Problems 1 and 2 can be solved using direct enumeration and the mass formula, joint work with Markus Kirschmer; we use the fact that $\mathrm{Tr}\,\mathrm{nrd} : \mathcal{O} \to \mathbb{Z}$ gives $\mathcal{O}$ the structure of a lattice in $B \otimes_F \mathbb{R} \cong \mathbb{R}^{4n}$. These lattice methods were made significantly more efficient by an idea of Donnelly.

Example: Inner twists

Let $F = \mathbb{Q}(\sqrt{15})$ and let $\mathfrak{N} = (5, \sqrt{15})$. Then there exists a cusp form of dimension 8 in $S_2(\mathfrak{N})$ such that no single Hecke eigenvalue generates the entire field E of Hecke eigenvalues.



Here, $u = (5 + \sqrt{17})/2$. (There are also examples of this phenomenon over $\mathbb{Q}$, and they are related to inner twists.)

The fine print

We have already mentioned that one can compute in higher weight by modifying the coefficient module appropriately. In fact, it is also much more efficient in practice, in both the definite and indefinite case, to work with an induced module for higher level: in this way, one only ever needs to compute with a maximal order. One can also work with fields F of arbitrary strict class number: in each case, then, the Shimura variety naturally decomposes as a (disjoint) union indexed by the strict class group. One can also obtain eigenvalues for the Atkin-Lehner operators.

The Jacquet-Langlands correspondence implies that the definite and indefinite methods overlap when there is a prime $\mathfrak{p} \parallel \mathfrak{N}$: for then we can consider the quaternion algebra ramified at $\mathfrak{p}$ and all (or all but one) real place. Therefore, in many cases we can use either approach—or both approaches, as a way of verifying the computation.