Modular degrees and congruence numbers for modular abelian varieties over **Q**

Kenneth A. Ribet

UC Berkeley

Southern California Number Theory Day
May 15, 2010

To fix ideas, we focus on $J_0(N)$ over $\mathbf{Q}$. We can work instead with $J_1(N)$ or with the Jacobian of a Shimura curve — as long as we don't need the q-expansion principle.

When f is a newform of weight 2 on $\Gamma_0(N)$, Shimura's construction associates to f an abelian subvariety A_f of $J_0(N)$ and a quotient

$$\iota^\vee : J_0(N) \to A_f^\vee$$

dual to the inclusion ι of A_f into $J_0(N)$.

The kernel

$$B_f := \ker \iota^\vee$$

is an abelian subvariety of $J_0(N)$ that serves as a sort of orthogonal complement to A_f.

We abbreviate, setting

$$J = J_0(N), \qquad A = A_f, \qquad B = B_f.$$

What is essential for us is that A and B are stable under the Hecke operators $T_n : J \longrightarrow J$ for $n \geq 1$. We could alternatively take A to be the sum of a family of different f's as long as we took B to be the complement of A.

We are interested particularly in the finite group

$$A \cap B,$$

considered with its Hecke action and the natural action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on the intersection.

A tremendously important case is that where A is a “strong” modular elliptic curve E over $\mathbf{Q}$. Then naturally $A = E = B$, and the intersection $A \cap B$ is $E[d]$, where $d \geq 1$ is the *modular degree* of E.

Let $\mathbf{T}$ be the subring of $\operatorname{End} J$ generated by all Hecke operators T_n. Then $\mathbf{T}$ is a free $\mathbf{Z}$-module of rank $\dim J$ that is naturally dual to the space of weight-2 cusp forms on $\Gamma_0(N)$ with $\mathbf{Z}$-integral coefficients. It is natural to consider the ideal

$$I_d = \operatorname{Ann}_{\mathbf{T}}(A \cap B);$$

the subscript "d" serves to remind us that this ideal is associated with the modular degree.

Let $\mathbf{T}_A = \mathbf{T}/(\operatorname{Ann}_T B)$ be the image of $\mathbf{T}$ in $\operatorname{End} A$. Because I_d clearly contains the kernel of $T \to \mathbf{T}_A$ (i.e., $\operatorname{Ann}_{\mathbf{T}} B$), I_d is the inverse image in $\mathbf{T}$ of an ideal $\bar{I}_d$ of $\mathbf{T}_A$. When $A = E$ is an elliptic curve, $\mathbf{T}_A$ is just $\mathbf{Z}$, and $\bar{I}_d = d\mathbf{Z}$.

Of course, I_d contains not just $\mathrm{Ann}_{\mathbf{T}}\, B$, but also its analogue with B replaced by A:

$$I_d \supseteq I_c := \mathrm{Ann}_{\mathbf{T}}\, A + \mathrm{Ann}_{\mathbf{T}}\, B.$$

The ideal I_c is the *congruence ideal* associated to A (or its orthogonal complement). Again, I_c is the inverse image of an ideal $\bar{I}_c$ of $\mathbf{T}_A$.

In the special case $A = E$, $\bar{I}_c = c\mathbf{Z} \subseteq \mathbf{Z}$ is the ideal generated by the "congruence number" for the modular form f giving E. This number counts congruences between $\mathbf{Z} \cdot f$ and the space of integral cusp forms in the same space as f that are Petersson orthogonal to f. Because $\bar{I}_d \supseteq \bar{I}_c$, we have $d|c$ in the elliptic curve case; one can say that the modular degree forces congruences.

It is sometimes fruitful to view I_c as the conductor of the ring extension $\mathbf{T} \hookrightarrow \mathbf{T}_A \times \mathbf{T}_B$, where $\mathbf{T}_B$ is the analogue of $\mathbf{T}_A$ with B replacing A.

Theorem

If N is prime, the ideals I_c and I_d coincide.

The proof begins by noting the inclusion that

$$I_d(\mathbf{T}_A \times \mathbf{T}_B) \subseteq (\mathbf{T} \otimes \mathbf{Q}) \cap (\operatorname{End} J),$$

which implies $I_d \subseteq I_c$ whenever $(\mathbf{T} \otimes \mathbf{Q}) \cap (\operatorname{End} J) = \mathbf{T}$.

Mazur's "Eisenstein ideal" article presents the stronger-looking equality $\operatorname{End} J = \mathbf{T}$ in the prime-level case; the theorem then follows.

The essence of Mazur's proof is to show for arbitrary N that $(\mathbf{T} \otimes \mathbf{Q}) \cap (\operatorname{End} J) = \mathbf{T}$ locally at rational primes $p \nmid N$, and also at the particular prime $p = N$ when N is prime.

Mazur's argument proves essentially instantly that $I_d = I_c$ away from the level. With more work, one (Agashe–R–Stein) gets:

Theorem

The ideals I_c and I_d coincide locally away from primes whose squares divide the level.

The proof channels arguments in Wiles's Annals article on Fermat to establish new cases of multiplicity one for the Jacobians $J_0(N)$:

Wiles looks at primes $p|N$ with $p^2 \nmid N$ and proves "multiplicity one" at p even if $J_0(N/p)$ is non-trivial. (In Mazur's article, p was N, and it was somehow important that $J_0(N/p)$ *is* trivial.)

In the elliptic curve case, c and d are integers with $d|c$, and the A–R–S theorem amounts to the inequality

$$2\,\mathrm{ord}_p\left(\frac{c}{d}\right) \le \mathrm{ord}_p N$$

when the right-hand side is ≤ 1. It would be interesting to find an upper bound for $\mathrm{ord}_p(c/d)$ that is valid even when p^2 divides N.

Multiplicity One

The point of “multiplicity one” is to consider the kernels $J_0(N)[\mathfrak{m}]$ for maximal ideals $\mathfrak{m}$ of $\mathbf{T}$. For each $\mathfrak{m}$, there is an associated 2-dimension semisimple representation $\rho_{\mathfrak{m}}$ of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ over $\mathbf{T}/\mathfrak{m}$ whose determinant is the mod ℓ cyclotomic character (if $\mathfrak{m}|\ell$) and whose Frobenius traces are Hecke operators mod $\mathfrak{m}$. For “generic” $\mathfrak{m}$, $\rho_{\mathfrak{m}}$ is irreducible, and one proves that $J_0(N)[\mathfrak{m}]$ is isomorphic to a non-empty direct sum of copies of the representation $\rho_{\mathfrak{m}}$. The number of copies of $\rho_{\mathfrak{m}}$ in the sum is the *multiplicity*, which turns out typically to be equal to 1.

For N prime, Mazur proved that the multiplicity is 1 for all “irreducible” $\mathfrak{m}$ away from 2 and N; he then went on to prove multiplicity 1 both for those $\mathfrak{m}$ dividing N and for those $\mathfrak{m}$ such that $\rho_{\mathfrak{m}}$ is reducible (Eisenstein).

In the early part of this decade, Kilford found examples of multiplicity > 1 in cases when N is prime and $\mathfrak{m}$ has residue characteristic 2. His smallest example occurs for $N = 431$. More recently, Wiese, Emerton and others have explained these examples through a *local* failure of multiplcity 1.

Meanwhile, authors such as Gross, Ribet, Mazur–Ribet and Wiles have established multiplicity 1 in a plethora of cases by variants of Mazur's arguments.

A central component of the argument is to show that the dimension of $H^0(X_0(N)_{/\mathbf{F}_\ell}, \Omega^1)[\mathfrak{m}]$ is ≤ 1 when ℓ is the residue characteristic of $\mathfrak{m}$. If $\ell = p$ is a divisor of N, the statement must be interpreted appropriately (and then proved!).

This same multiplicity-one statement for differentials in positive characteristic is behind the proof that $I_c = I_d$. Namely, one uses this multiplicity-one statement, Serre duality and Nakayama's lemma to show that $H^1(X_0(N), \mathcal{O})$ is free of rank one over $\mathbf{T}$ locally at the prime m. Because $H^1(X_0(N), \mathcal{O})$ can be associated functorially to the Néron model for J, it is naturally a module for $\operatorname{End} J$; we get easily that $\mathbf{T}$ is saturated in $\operatorname{End} J$ locally at m.

Thus there is a link between multiplicity 1 for $J_0(N)[m]$ and the equality $I_c = I_d$ locally at m, but it appears that this link is rather weak because it stems from the fact that both statements arise from the same ingredient.

Observe especially that in the prime-level case we do have $I_c = I_d$ but might not have multiplicity 1 (at least when $\ell = 2$).

I somehow failed to focus until recently on the following:

Proposition

Let $\mathfrak{m}$ be a maximal ideal of $\mathbf{T}$ for which $J_0(N)[\mathfrak{m}]$ has dimension 2. Then $I_c = I_d$ locally at $\mathfrak{m}$.

In the statement, we do not assume that $\rho_{\mathfrak{m}}$ is irreducible and therefore do not know a priori that $\dim J_0(N)[\mathfrak{m}]$ is even. In the reducible (Eisenstein) case, Mazur proved by elaborate arguments that $J_0(N)[\mathfrak{m}]$ is indeed of dimension 2 when N is prime, but little seems to be known in the case where N is composite.

To prove the proposition is to show that if $I_c \subset I_d$ (strict inclusion) locally at m, then $J[\mathfrak{m}]$ has dimension > 2. This is immediate at least in the case where I_d is locally **T**, i.e., where $(A \cap B)[\mathfrak{m}] = 0$. If $I_c \subseteq \mathfrak{m}$, then both $A[\mathfrak{m}]$ and $B[\mathfrak{m}]$ must be non-trivial (and hence of dimension at least 2). Because $A[\mathfrak{m}] \oplus B[\mathfrak{m}]$ injects into J, $J[\mathfrak{m}]$ will have dimension ≥ 4.

It is fun to experiment with examples. For elliptic curves, the first case where the division $d|c$ is strict occurs at conductor 54. For the curve $E = \mathbf{54b}$, we have $d = 2$, $c = 6$. The **T**-module $E[3]$ cuts out an Eisenstein prime m in **T** that contains the particular elements T_3 and $T_2 - 1$. The non-triviality of $B[\mathfrak{m}]$ in this case can be traced back to the old part of $J_0(54)$, which is isogenous to a product of two copies of the elliptic curve $J_0(27)$.

Agashe's new observation is that the multiplicity-one hypothesis $\dim J[\mathfrak{m}] = 2$ implies a structure theorem for the intersection $A \cap B$:

Theorem

If $\dim J[\mathfrak{m}] = 2$*, then (locally at* $\mathfrak{m}$*),* $A \cap B$ *is free of rank 2 over* $\mathbf{T}/I_c$*.*

Since the annihilator of $A \cap B$ is I_d by definition, the theorem implies in particular that $I_d = I_c$ locally at $\mathfrak{m}$.

The proof is quite straightforward: it begins by expressing $A \cap B$ as the cokernel of the map that the surjection $A \times B \to J$ induces on the level of covariant Tate modules. This theorem will appear in an article by Agashe, Mladen Dimitrov and me.