PERSIFLAGE: MATH BLOG POSTS

FRANK CALEGARI

I though your post on swans was your best post ever.

§50

The nilmanifold fibres spread the cohomology around by a Künneth type formula like a Frenchman expectorating over-oaked California Chardonnay into a spittoon

§24

1. Introduction

It does not seem entirely impossible that wordpress might disappear at some point. It would be a shame to lose all my old blog posts. Already, some of my older posts suffer both from link rot and from latex issues related to changes in how wordpress interacts with latex. There are a number of posts which I wrote that contain arguments and remarks (and even Theorems) not proved anywhere else, and (as much for me as anyone else) I thought I would try to preserve them by collating them into a single file. It also makes it easier to search. I've restricted myself to posts with some mathematical content, with the exception of § 50 (OK, perhaps § 96 is not about mathematics either). I've included some of the comments, and I also occasionally added notes which reflect any particularly relevant updates. I've made some latex modifications (such as using theorem and conjecture environments) and I've also included a bibliography and put in citations (inconsistently, sorry) where the blog simply referenced the paper either by name or by link, but otherwise the posts are unmodified. (The conversion to latex may even have introduced new errors.) My rate of posting has slowed over the years. There are two reasons. One, ironically, is that I started collaborating more and doing more mathematics, which meant more time was spent writing papers and less writing blog posts. The second reason is explained later (at least obliquely). Still, I have every intention to continue posting. This is intended to be an organic document which gets updated from time to time as I keep posting, so please free free to point out errors or updates that can be included in future notes.

Contents

The author was supported in part by NSF Grant DMS-2001097.

List of Figures

2. Even Galois Representations mod p

Sun, 7 Oct 2012

Suppose that $\overline{\rho} : G_{\mathbf{Q}} :\to \mathrm{GL}_2(\overline{\mathbf{F}}_p)$ is a continuous irreducible Galois representation. What does the Langlands program say about such $\overline{\rho}$? When $\overline{\rho}$ is odd, the situation is quite satisfactory, the answer being given by Serre's conjecture. For example, having fixed a Serre weight $k \geq 2$ and a Serre level N, one knows that there will only be finitely many such representations and they will all come from classical modular forms for $\mathrm{GL}(2)/\mathbf{Q}$.

When $\overline{\rho}$ is even, however, there is an equally good (if conjectural) description of such representations. First, the dihedral representations are well understood by class field theory, so let us assume we are not in this case. Then, replacing $\overline{\rho}$ by the adjoint representation $\mathrm{ad}^0\overline{\rho}$ and restricting to some (any) imaginary quadratic field, one obtains an irreducible (conjugate) self-dual representation, which, by the generalization of Serre's conjecture [Ser87], should come from an automorphic representation for $U(3)$. It follows that, as in the odd case, there will (conjecturally) only be finitely many such $\overline{\rho}$ for a fixed pair (N, k). However, things are *even better* in the even case. Namely, if one fixes (N, k) but allows p to vary, then there will *still* only be finitely many even representations, in contrast to the odd case where (for $(N, k) = (1, 12)$ for example) such representations occur for infinitely many p. The reason is that all such representations will have to arise from a fixed finite dimensional space of automorphic forms determined by N and k, and thus (by the pigeonhole principle) there will exist an automorphic Π for $U(3)$ whose mod p representation extends to an even representation of $\mathbf{Q}$ for infinitely many p. By multiplicity one, it would follow that $\Pi \simeq \Pi^c \simeq \Pi^\vee$ and hence Π itself must come from the adjoint representation of a form from $\mathrm{GL}(2)$ over $\mathbf{Q}$, which would imply (since we are in regular weight) that the representations are odd. Note that it is important in the definition of Serre weight here that $k \geq 2$; if one allows $k = 1$ then there exist representations in characteristic zero which give rise to mod p representations for all p.

Here's a specific example in which one can prove finiteness. Suppose that we consider representations with $k = 2$ and $N = 1$. Then there are no such even $\overline{\rho}$ for a stupid reason, because the determinant will be cyclotomic (Tate deals with the case $p = 2$.) Now consider the case when $k = 2$ and $N = 4$. In the even case, the determinant must be the cyclotomic character times the unique (odd) character of conductor 4. Let's prove that there are no such representations. Tate like arguments reduce to the case when the representation has image containing $\mathrm{SL}_2(\mathbf{F}_p)$ and $p > 7$. Now take the auxiliary imaginary quadratic field to be $\mathbf{Q}(\sqrt{-1})$. The corresponding adjoint representation now is unramified outside primes above p (the quadratic extension eliminating the ramification at 2) and is Fontaine–Laffaille with weights $[-1, 0, 1]$ at primes dividing p. Using the lifting results of [BLGGT14], we may lift this to a compatible family of self-dual representations of level one

and weight zero which is potentially modular. Because these representations are potentially modular and are not CM, we know that they are all irreducible by Blasius–Rogawski. We now specialize these representations to $p = 5$, and because the Hodge–Tate weights are sufficiently small ($[0, 1, 2]$) and $\mathbf{Q}(\sqrt{-1})$ is also small, we can use results of Fontaine [Fon85] and Abrashkin to deduce that the corresponding 5-adic representation is reducible, which is a contradiction. We thus deduce (using Khare–Wintenberger [KW09a, KW09b] for the odd case) that there do not exist any irreducible finite flat group schemes G of type (p, p) over $\mathrm{Spec}(\mathbf{Z}[\sqrt{-1}])$ whose generic fibre admits descent data to $\mathbf{Q}$. This entire argument is really just a version of the Khare–Wintenberger proof of Serre's conjecture for $U(3)$. Unfortunately, one doesn't quite have enough modularity lifting theorems at this point to deduce Serre's conjecture completely for $U(3)$.

These arguments are quite general. For example, there should only exist finitely many even representations $\overline{\rho} : G_{\mathbf{Q}} :\to \mathrm{GL}_n(\overline{\mathbf{F}}_p)$ (whose image contains $\mathrm{SL}_n(\mathbf{F}_p)$) of fixed Serre weight and level, even when one ranges over all primes p, providing $n \geq 2$.

3. Hilbert modular forms of partial weight one, Part I

Sat, 13 Oct 2012

Let π be an algebraic Hilbert modular cuspform for some totally real field F^+. Then, associated to π, one has a compatible family of Galois representations:

$$r_\lambda(\pi) : G_{F^+} \to \mathrm{GL}_2(\mathcal{O}_\lambda)$$

which are unramified outside finitely many primes (this is the work of many people). The *expectation* is that this representation should satisfy local global compatibility at all primes. This is known if π has *regular* weight, and also if π has parallel weight one. However, this is not known, even for the case $p \neq \ell$ (Here ℓ is the characteristic of $\mathcal{O}/\lambda$). The problem is that these representations are constructed via congruences, not from geometry. Deforming in families does give some control, and indeed one can prove that, for $v|p$ and $p \neq \ell$,

$$\mathrm{WD}(r_\lambda(\pi)|_{G_v})^{\text{F-ss}} \prec \mathrm{rec}(\pi_v)$$

which is a way of saying you get the correct answer *up to the monodromy operator* N, and moreover the monodromy operator on the Galois side can only be *more degenerate* than the automorphic side. In English, if (for example) π_v is Steinberg, then one may deduce (as expected) that the image of inertia on the Galois side is unipotent, but not necessarily that it is non-trivial. In fact, by solvable base change, this is really the *only* problem one has to worry about (so we shall assume we are in this case below).

The usual methods for computing the monodromy N are all geometric (nearby cycles), and, as it seems hopeless to try to construct any (conjectural) motive associated to π, there doesn't seem to be much one can do.

One does, however, have the following strategy, which I learnt from Martin Luu, which should suffice for all but finitely many primes λ for which $r_\lambda(\pi)$ is ordinary. Namely, take the λ-adic Galois representation associated to π, and prove that it is *potentially automorphic* using extensions of the Buzzard–Taylor idea (which has been employed by Sasaki, Kassaei, Pilloni and others in the case of Hilbert modular

forms of *parallel* weight one, but should also apply in this context). The result is that one shows that $r_\lambda(\pi)|G_{E^+}$ for some totally real extension E^+/F^+ is now associated to a cuspidal automorphic form Π *of the right level.* How does this help? Well, now using what we know from local global compatibility (which is ok in the unramified case), we deduce that Π_w for some $w|v$ is associated to the corresponding local Galois representation $r_\lambda(\pi)|_{G_w}$. Now this representation has the property that it looks unipotent on inertia mod λ^n for all n, but, assuming local-global compatibility fails, is actually unramified at p. In particular, the semi-simplification is given by two characters whose ratio is the cyclotomic character, whereas Π_w is an unramified principal series. This implies that the Satake parameters $\{\alpha_w, \beta_w\}$ satisfy $\alpha_w/\beta_w = N(w)$, which contradicts Ramanujan. We are not done yet, because one doesn't have purity in partial weight one. However, one *can* appeal to bounds coming from Rankin-Selberg, and this is enough to obtain a contradiction.

The only obvious *examples* of partial weight one HMF (which are not of parallel weight one) are CM, and since those are potentially unramified, the monodromy operator will always be trivial on the autormophic side (and hence also on the Galois side). So this suggests (but does not beg) the question: do there actually *exist* any partial weight one (but not parallel weight one) Hilbert modular forms which are not CM? Stay tuned for part II!

4. Why it is good to be pure

Mon, 15 Oct 2012

There do not exist any regular pure motives M over $\mathbf{Q}$ which are not essentially self dual. Here is why. M gives rise to a compatible family of Galois representations for each rational prime v such that the characteristic polynomial $R(X)$ of Frobenius is independent of this choice. By purity, the eigenvalues α of $R(X)$ are algebraic integers lying in a CM-field such that $|\iota\alpha|^2 = p^w$ for some integral weight w and any complex embedding ι. In particular, if α is a root of $R(X)$, then $\alpha^c = p^w/\alpha$ is a root of $X^n R(p^w/X)$. Since $R(X)$ has coefficients in $\mathbf{Z}$, it follows that α^c is also a root of $R(X)$, from which one may deduce that $R(X) = X^n R(p^w/X)$ (up to the appropriate constant which makes the RHS monic, this doesn't affect any of the arguments). Yet this implies that $M^\vee(w) \simeq M$, by the Cebotarev density theorem. (Caveat: it really says that the p-adic avatars of M are essentially self-dual. Perhaps deducing the result for M actually requires the standard conjectures.)

This argument no longer applies if one relaxes the conditions slightly; there do exist non-self dual motives of rank three with *coefficients;* Bert van Geemen and Jaap Top [vGT94] found some explicit examples with coefficients in an imaginary quadratic extension of $\mathbf{Q}$. The point where the argument above fails is that it identifies the polynomial $X^n R(p^w/X)$ with the *complex conjugate* polynomial $R^c(X)$, which need not equal $R(X)$ anymore.

Stefan Patrikis and Richard Taylor use a similar argument in their recent paper [PT15] to prove a nice result. Start with a regular pure motive M over $\mathbf{Q}$ (so by the above remarks, it is essentially self dual). Suppose that the corresponding v-adic Galois representation:

$$r_v : G_{\mathbf{Q}} \to \mathrm{GL}_n(\mathbf{Q}_v)$$

is not absolutely irreducible. One may ask: are the irreducible constituents s_v themselves essentially-self dual? They show that the answer is yes. Let $S(X)$ denote the corresponding characteristic polynomials. If $S(X)$ lies in $\mathbf{Z}[X]$, then the same argument above applies to s_v. But it may be the case that the representation r_v only decomposes over an extension of $\mathbf{Q}$. By looking at the eigenvalues, it trivially follows that each of the $S(X)$ may be defined over some CM field F/F^+. More importantly, by a technical argument which I will omit but which is not too difficult, one may find a *fixed* CM field M/M^+ which contains all the polynomials $S(X)$ (one may even do this [in some sense] independently of v, although we won't use that here). Consider the Galois representation $(s_v)^c$, where c is acting on the coefficients. Let α be a root of $S(X)$. Then $\alpha^c = p^w/\alpha$ is now a root of $X^m S(p^w/X)$, and so $S^c(X)$ and $X^m S(p^w/X)$ coincide. Since $X^n R(p^w/X) = R(X)$, we deduce that $(s_v)^c$ is a sub-representation of $(r_v)^c = r_v$. In particular, $(s_v)^c$ and s_v are both sub-representations of r_v. But the Hodge–Tate weights of s_v and $(s_v)^c$ are the same! (Literally, the Hodge–Tate weights of $(s_v)^c$ are the Hodge–Tate weights of ${}^c(s_v)$ where ${}^c(s_v)(g) = s_v(cgc^{-1})$, but since s_v is a representation of $\mathbf{Q}$, conjugation by c is conjugation by a matrix, so there is an isomorphism $s_v \simeq {}^c(s_v)$.) It follows (from the regularity assumption) that $s_v = (s_v)^c$, and then the argument above implies that s_v is self-dual.

One may use this argument as follows. As in [BLGGT14], one may find a prime v such that all of the s_v are residually irreducible, and so (if v is sufficiently large) are also potentially modular (by [BLGGT14] again). In particular, either all of the r_v are reducible or they are irreducible for a set of density one set of primes. Moreover, any regular motive over $\mathbf{Q}$ is potentially modular, which is only three adjectives away from the complete reciprocity conjecture!

Patrikis and Taylor do something slightly more general, instead of pure regular motives over $\mathbf{Q}$, they consider essentially self-conjugate regular compatible systems (with coefficients) of G_F for some CM field F/F^+. For reasons alluded to above, the coefficients live in some CM-field M. This extra generality (mostly) adds some notational complexity to the argument above. (To see the type of complications that arise, consider an elliptic curve E with CM and then restrict to the CM field F. Then any reducible constituent $s_v = \chi_v$ is related not to its complex conjugate χ_v^c acting on M, but the complex conjugate ${}^c\chi_v^c$ of this where complex conjugation is now acting on the coefficients M and on the Galois group F.) As expected, one obtains (using [BLGGT14]) some nice consequences, like potential automorphy of regular polarizable compatible systems, as well as irreducibility (for a density one set of primes) of Galois representations associated to RAESDC automorphic form II.

Comment 4.1 (Persiflage)**.** Regularity is not used anywhere in the first result, so the argument applies to all motives with coefficients in $\mathbf{Q}$. As a sanity check, if χ is a character of a finite group with values in $\mathbf{Q}$, then the dual character is $\overline{\chi} = \chi$, so χ is self-dual. Indeed, this is basically the same argument (in weight $w = 0$.)

5. Remarks on Buzzard–Taylor

Thu, 18 Oct 2012

Let $\rho : G_{\mathbf{Q},S} \to \mathrm{GL}_2(\overline{\mathbf{Q}}_p)$ be continuous and unramified at p. The Fontaine–Mazur conjecture predicts that ρ has finite image and is automorphic. Buzzard and Taylor proved this result under the assumption the natural assumption that ρ is odd, that $\overline{\rho}$ is modular (now uneccessary), but also under the further two assumptions:

(1) $\overline{\rho}$ is irreducible,
(2) $\overline{\rho}$ is p-distinguished.

(caveat: there are mild extra assumptions required when $p = 2$.) The point of this post is to note that is seems possible to remove either of these conditions and to wonder whether both can be removed simultaneously.

Suppose the second condition fails. One may *enlarge* the ordinary Hecke algebra $\mathbf{T}$ to include the operator U_p, call the resulting ring $\widetilde{\mathbf{T}}$. After appropriate localizations, this is different from $\mathbf{T}$ exactly when $\overline{\rho}(\mathrm{Frob}_p)$ is a scalar. On the Galois side, let R_p denote the universal (framed) deformation ring. Then, for any lift of Frobenius, one can define the quadratic extension $\widetilde{R}_p = R_p[\alpha]$ where α is an eigenvalue of Frobenius. Fix a weight $k \geq 2$. If R_p^{ord} denotes the ordinary deformation ring, then there is a corresponding quotient $\widetilde{R}_p^{\mathrm{ord}}$ of $\widetilde{R}_p$ which records ordinary deformations *together* with the action of Frobenius on the unramified quotient. The $\overline{\mathbf{Q}}_p$-points of these local deformation rings are the same (since $k \geq 2$). The usual Taylor–Wiles-Kisin method produces an isomorphism of the type $\widetilde{R}[1/p] = \widetilde{\mathbf{T}}[1/p]$, where $\widetilde{R}$ is the global deformation ring which takes into account the extra data at p. This isomorphism holds for all weights $k \geq 2$, which is enough to get an isomorphism on the corresponding ordinary families.

If $\rho(\mathrm{Frob}_p)$ has distinct eigenvalues, one may now deduce Buzzard–Taylor (by the same argument as BT). If $\rho(\mathrm{Frob}_p)$ is scalar, then one has to make a slight adjustment. To see what to do, note that if $f(\tau)$ is the desired weight one form, then the old space $f(\tau), f(p\tau)$ can no longer be diagonalized with respect to U_p. Instead, it should give rise to a surjective map $\psi : \widetilde{\mathbf{T}} \to \mathcal{O}[\epsilon]/\epsilon^2$ such that the image of $\mathbf{T}$ is $\mathcal{O}$. Conversely, if there is such a map ψ, this produces the ordinary old forms necessary to recover f (both forms have the same Hecke eigenvalues away from p, so one can determine the q-expansions). From the modularity result, it suffices to construct such a map on the Galois side. Yet this exists precisely because ρ comes with two distinct unramified quotients.

David Geraghty and I used these these flavours of deformations rings for a somewhat different purpose (although we required more precise integral information concerning $\widetilde{R}_p^{\mathrm{ord}}$ coming out of a very nice paper of Snowden [Sno18]). On the other hand, as far as the argument above goes, it was apparently known to Richard many years ago (as I learnt by chatting with Toby Gee whilst drinking an $8 can of Boddingtons in Toronto).

Suppose one assumes instead that $\overline{\rho}$ is reducible. Recall that one has maps $R^{\mathrm{ps}} \to R$ and $R^{\mathrm{ps}} \to \mathbf{T}$ for a suitable pseudo-deformation ring R^{ps}. In higher weights, Skinner-Wiles essentially prove that $R[1/p]^{\mathrm{red}} = \mathbf{T}[1/p]$, which should be sufficient to construct the required overconvergent forms f_α and f_β in weight one. While chatting with Patrick Allen over espresso today, it also seems reasonable that (using appropriate framings, as above) one may generalize this to the case where $\overline{\rho}$ is no longer p-distinguished, *as long as* the characteristic zero eigenvalues α and β are distinct. The problem, however, with the $\alpha = \beta$ case is that one needs to promote a

non-reduced quotient $R \to \mathcal{O}[\epsilon]/\epsilon^2$ to a map from $\mathbf{T}$, and the methods of Skinner-Wiles in the reducible case only give information about the reduced quotients of R. Is there any way around this? This seems (pretty close) to the only remaining obstruction for a complete solution to the weight one odd case of Fontaine–Mazur.

6. Jacobi by pure thought

Fri, 26 Oct 2012

Joël Bellaïche asks here whether there is a conceptual proof of Jacobi's formula:

$$\Delta = q \prod_{n=1}^{\infty} (1-q^n)^{24}$$

Here (to me) the best proof is one that requires the least calculation, not necessarily the "easiest." Here is my attempt. We use the following property of Δ, which follows from its moduli theoretic definition: the only zero of Δ is a simple zero at the cusp, moreover, the evaluation of Δ on the Tate curve is normalized so that the leading coefficient is q. Let p be prime. I claim that

$$\Delta(\tau)^{p+1} \prod_{i=0}^{p-1} \zeta^i = \Delta(p\tau) \prod_{i=0}^{p-1} \Delta\left(\frac{\tau+i}{p}\right).$$

Observe that both these expressions are modular forms of level one and weight $(p+1)$ times the weight of Δ. One can prove this "by hand," but also by noting that the RHS is equal to the norm of $\Delta(p\tau)$ on $X_0(p)$ down to $X_0(1)$. On the other hand, the RHS also has a zero of order $p+1$ at $q=0$, from which the result immediately follows, since the ratio will be holomorphic of weight zero. If one defines Hecke operators on q-expansions in the usual way, it also immediately follows that the logarithmic derivative $qd/dq \log(\Delta)$ is (as a q-expansion) an eigenform for T_p of weight two with eigenvalue $p+1$ for all primes p. In fact, the same argument as above (with $X_0(p)$ replaced by $X_0(n)$) implies that this derivative is also an eigenform for T_n with eigenvalue $\sigma_1(n) = \sum_{d|n} d$. This is *almost* enough to determine the q-expansion uniquely: in particular, it implies that

$$q \cdot \frac{d}{dq} \log(\Delta) = 1 + m \cdot \sum_{n=1}^{\infty} \sigma_1(n) q^n$$

for some integer m, from which it follows that

$$\Delta = q \prod_{n=1}^{\infty} (1-q^n)^m.$$

To finish the argument, it suffices to check that $m = 24$, or that $\tau(2) = -24$. One way to do this is to note (by uniqueness) that Δ is a Hecke eigenform, and then use the equation $\tau(2)\tau(3) = \tau(6)$ which implies that $m \in \{0, 1, 2, 3, 24\}$; the cases $m = 1, 2, 3$ are then ruled out by the equations $\tau(2)\tau(7) = \tau(14)$ and $\tau(2)\tau(13) = \tau(26)$, and $m = 0$ is ruled out by the fact that q is not a modular form. Curiously enough, this determines Δ without ever using the fact that it has weight 12. Another (more traditional way) is to show that $1728\Delta = E_4^3 - E_6^2 = q - 24q^2 + \ldots$. Is there a way to do this final step by pure thought?

Comment 6.1 (Emmanuel Kowalski)**.** This argument is very similar to one by Kohnen [Koh05].

7. There are no unramified abelian extensions of $\mathbf{Q}$ (almost)

Tue, 27 Nov 2012

In my class on modularity, I decided to explain what Wiles' argument (in the minimal case) would look like for $\mathrm{GL}(1)/F$. There are two ways one can go with this. On the one hand, one can try to prove (say) Kronecker-Weber using Selmer groups, but avoiding any kind of circularity (by not assuming class field theory). On the other hand, one can allow oneself to be completely circular in an effort to concentrate on the technical details of Wiles' arguments. This post concerns the latter, and we "prove" the following:

Theorem 7.1. *Let F be a number field which does not contain ζ_p. Then the Galois group of the maximal abelian extension of F unramified everywhere is isomorphic to the p-part of the class group.*

To prove this, we will (only) assume the following:

(1) Local class field theory.
(2) For any ray class group of F, there exists a corresponding abelian ray class field whose Galois group is the ray class group (this is half of global class field theory).
(3) The abelian extensions coming from the ray class group are compatible with local class field theory (this is local-global compatibility).
(4) The Wiles-Greenberg Selmer group formula.

The first three assumptions for $F = \mathbf{Q}$ are equivalent to giving oneself the cyclotomic extensions and understanding their ramification properties. The last assumption, of course, contains every part of global class field theory (making the argument circular).

Let $\Gamma_\emptyset$ denote the Galois group of the maximal pro-p abelian unramified extension of F. Let Γ_{Q_N} denote the corresponding group where ramification is allowed at some set of primes Q_N not containing p, where one also insists that the order of inertia at primes in Q_N is at most p^N. Formally, we have the universal deformation rings

$$R_\emptyset = \mathbf{Z}_p[[\Gamma_\emptyset]], \qquad R_{Q_N} = \mathbf{Z}_p[[\Gamma_{Q_N}]],$$

and we also have the universal "modular" deformation rings

$$\mathbf{T}_\emptyset = \mathbf{Z}_p[[(F^\times \backslash \mathbf{A}_F^\times / U)^F]], \qquad \mathbf{T}_{Q_N} = \mathbf{Z}_p[[(F^\times \backslash \mathbf{A}_F^\times / U_Q)^F]].$$

Here M^F denotes the biggest finite quotient of M, U is the obvious maximal open compact, and U_Q is the variant of U such that $U_v = \mathcal{O}_v^\times$ is replaced by $U_{Q_N,v} = \mathcal{O}_v^{\times p^N}$ for $v \in Q_N$. The half of global class field theory we are assuming gives us a compatible diagram of maps $R_{Q_N} \to \mathbf{T}_{Q_N}$ and $R_\emptyset \to \mathbf{T}_\emptyset$. The Wiles–Greenberg formula gives us an equality:

$$\dim |H^1_\emptyset(F, \mathbf{F}_p)| - \dim |H^1_{\emptyset^*}(F, \mathbf{F}_p(1))| = -(r_1 + r_2 - 1),$$

where for this computation we use that $\zeta_p \notin F$. In order to annihilate the dual Selmer group, we need to annihilate classes in $H^1(F, \mu_p)$, which come from extensions $F(\zeta_p, \sqrt[p]{\alpha})$. We can do this in the usual way, but we have to assume

(again) that $\zeta_p \notin F$, since otherwise one cannot annihilate the class defined over $F(\zeta_{p^2}) = F(\zeta_p, \sqrt[p]{\zeta_p})$ using a prime $q \equiv 1 \mod p^2$. We see that we can annihilate the dual Selmer group with $q := |Q_N| = \dim H^1_\emptyset + (r_1 + r_2 - 1)$ primes. What are the auxiliary rings S_N here? As rings, they are

$$S_N = \mathbf{Z}_p[(\mathbf{Z}/p^N\mathbf{Z})^q]$$

the action on R_{Q_N} is via the inertia group at the auxiliary primes. To make this work, one needs local class field theory; this shows that inertia at q is acting via $\mathcal{O}_q^\times/\mathcal{O}_q^{\times p^N}$. The action of S_N on $\mathbf{T}_{Q_N}$ is given by the structure of $\mathbf{T}_{Q_N}$ as a module over $\mathbf{Z}_p[U/U_Q] \simeq S_N$. The compatibility of these actions is given by the compatibility of local and global class field theory. Moreover, if $\mathfrak{a}_N$ is the augmentation ideal of S_N, then $R_{Q_N}/\mathfrak{a}_N = R_\emptyset$ by definition, and $\mathbf{T}_{Q_N}/\mathfrak{a}_N = \mathbf{T}_\emptyset$ by construction. Thus, in the usual way, one ends up with a map $R_\infty \to \mathbf{T}_\infty$ where:

(1) R_∞ is a quotient of a power series ring with $q - (r_1 + r_2 - 1)$ variables.
(2) S_∞ is a power series ring in q variables.

The final thing to understand is the structure of $\mathbf{T}_\infty$ as a module over S_∞. At level Q_N, the annihilator in $S_N \simeq \mathbf{Z}_p[U/U_Q]$ of $\mathbf{T}_{Q_N}$ is given by the image of the global units. By Dirichlet's Theorem, this is generated by at most $r_1 + r_2 - 1$ generators (assuming again that $\zeta_p \notin F$). By patching, it follows (in the limit) that $\mathbf{T}_\infty$ has co-dimension at most $r_1 + r_2 - 1$, and thus (from dimension considerations) that $\mathbf{T}_\infty = R_\infty$, and then (after taking the quotient by $\mathfrak{a}_\infty$) that $R_\emptyset \simeq \mathbf{T}_\emptyset$, which proves that $\Gamma_\emptyset$ is the p-part of the class group. Note that (as expected) when gluing, we need to take into account all the (finitely many) possible $R/\mathfrak{m}^N$, $\mathbf{T}/\mathfrak{m}^N$, the possible maps from the global units to S_N, etc. etc.

In order to see the "circularity" more clearly, one may compute the Selmer groups directly. The group $H^1_\emptyset(F, \mathbf{F}_p)$ is equal to $\Gamma_\emptyset/p$, by definition. On the other hand, the group $H^1_{\emptyset^*}(F, \mu_p)$ by the Kummer sequence is equal to $\mathcal{O}^\times/\mathcal{O}^{\times p} \oplus \mathrm{Pic}^0(\mathrm{Spec}(\mathcal{O}_F))[p]$, and thus the Greenberg-Wiles formula is equivalent to the equality:

$$|\mathrm{Pic}^0(\mathrm{Spec}(\mathcal{O}_F))[p]| = |\Gamma_\emptyset/p\Gamma_\emptyset|$$

or equivalently the claim that the maximal exponent p-quotient of the class group captures all exponent p-unramified extensions. I guess this is very very slightly weaker than $\Gamma_\emptyset \otimes \mathbf{Z}_p \simeq \mathrm{Pic}^0(\mathrm{Spec}(\mathcal{O}_F)) \otimes \mathbf{Z}_p$.

8. Hilbert modular forms of partial weight one, Part II

Sat, 08 Dec 2012

Anyone who spends any time thinking about Hilbert modular forms of partial weight one — see § 3 — should, at some point, wonder whether there actually exist **any examples,** besides the "trivial" examples arising as inductions of Grossencharacters. Fred Diamond asked me this very question at Fontaine's birthday conference in March of 2010. There are various reasons why one should not expect to prove this by pure thought, including the possibility that (for certain levels) there may exist no such forms, and that at any level there may exist only finitely many such forms (more on these heuristics another time). Thus the only way I can really imagine showing that such a beast exists is by explicitly finding an example.

As of today, my students **Richard Moy** and **Joel Specter** have found such a form! Here is (roughly) the strategy they use. As with computing weight one

classical modular forms, one starts by computing a basis of q-expansions in some regular weight, divides by some Eisenstein series, takes the intersection of that space with its Hecke translates, and hopes that the resulting space has bigger dimension than the space of CM forms (which one can compute in advance). There are a few hiccoughs which occur along the way, of course. How does one compute q-expansions of Hilbert modular forms? Since computing uniformizations of surfaces is not realistic, they use the fact that (fortunately!) the q-expansion of a Hilbert modular form can be recovered from its Hecke eigenvalues. On the other hand, by Jacquet–Langlands, a Hilbert modular eigenform over (say) a real quadratic field corresponds to an eigenform on the arithmetic manifold associated to a quaternion algebra which is ramified at all infinite places, which then allows one to pass from the Hilbert modular variety to an adelic quotient which is now a finite set. Lassina Demebele wrote a magma programme which computes the eigenvalues for Hilbert modular eigenforms by this method, although for some reason the programme requires the level to be squarefree, and the character to be trivial. Using Atkin-Lehner theory, one can construct the entire space of forms by this method.

In practice, Richard and Joel worked with $F = \mathbf{Q}(\sqrt{5})$, computed the forms of level $\Gamma_0(N)$ (with N squarefree) and weight $[4, 2]$, then divided by an Eisenstein Series of weight $[1, 1]$ level $\Gamma_1(N)$ and character χ^{-1}, then computed the Hecke operator T_2 on this space and intersected away. Many (many) bugs later, and various annoying steps overcome (to take a random example, magma can compute the L-values of Hecke characters necessary to find constant terms of Eisenstein series [nice] but only as a complex number, not as an algebraic number [not so nice] so "L-value recognition" had to be coded in), the progams finally worked, and after much grinding away (for *all* squarefree N of norm less than 500) they didn't find anything at all (or at least, anything besides CM forms).

So they started working in weight $[6, 2]$, computed away, and *eventually* found a form π of weight $[5, 1]$, level $\Gamma_1(14)$, and character χ, where χ has conductor 7 and is of order 6. The coefficient field of the eigenform is, I believe, $\mathbf{Q}(\sqrt{5}, \sqrt{-3}, \sqrt{-19})$ (note that it must contain the base field as well as the field of the character). Note that this automorphic form π is *Steinberg* at 2! In particular, it is not CM, and one doesn't know whether local-global compatibility holds for the corresponding p-adic Galois representations even restricted to 2.

I should say that *finding* the form actually turned out to be easier than *proving* the form exists rigorously. Theoretically, the proof should be easy: one has found a form F/E for some cuspform F and some Eisenstein series E which looks like it is holomorphic. All one needs to do is square it (so it becomes regular), find a candidate form G of weight $[10, 2]$ such that $GE^2 - F^2 = 0$ (which one can prove since the spaces are finite dimension), and then E/F has no poles and is thus holomorphic. The problem is that the form $[10, 2]$ has non-trivial character, and Lassima's program only works with trivial character. One can take the 6th power and work with a form of weight $[30, 6]$, but this is way beyond what magma can cope with. In the end, Richard and Joel had to come up with a few tricks to do this (which took about three months!), but the final computations are in, and the existence has now been proven.

Comment 8.1 (Akshay Venkatesh)**.** That's a very nice computation indeed. Did they look mod p? Is there really an obstruction to using Hecke action to verify that they have no poles (as in Schaeffer's thesis), or it just seemed annoying to prove?

Comment 8.2 (Persiflage)**.** Dear AV, Some good questions. First, they did not compute anything mod p, and one reason is that it is not clear whether one can compute the *integral* structure of the module of Hilbert modular forms in regular weight (at least not obviously); the only thing one can compute are the eigenvalues of eigenforms, and this only tells you about the rational structure.

As for Schaeffer's approach, it's not obvious how to prove that $TF = \lambda F$ in practice (even for a fixed T) since this runs in to the same computational issues related to the fact that magma only computes spaces without character. There also does seem to be a genuine issue with applying Schaeffer's ideas for operators T dividing the level; we thought about this for a while, but couldn't quite make it work.

Notes 8.3. The corresponding paper is [MS15]. One frustrating aspect of this example was that the base field was strictly bigger than the field generated by F and the value of the character; it would have been nice to find a non-CM form defined over F. However, my student Abhijit Mudigonda pointed out to me that the classical argument that any modular form of odd weight and coefficients in $\mathbf{Q}$ also applies here to show that there are no such forms (more precisely, forms in weight $[1, 2k+1]$ defined over F must be CM).

9. The two cultures of mathematics: a rebuttal

Wed, 12 Dec 2012

Gowers writes thoughtfully about combinatorics here, in an essay which references Snow's famous lectures (or famous amongst mathematicians — I've never met anyone else who has ever heard of them). The trouble, however, starts (as it often does) with the invocation of the word "obvious":

> It is equally obvious that different branches of mathematics require different aptitudes.

I do not think that this claim stands up to scrutiny. By "aptitude," Gowers specifically distinguishes the following two abilities: problem-solving and theory-building. Here algebraic number theory is singled out as area which is firmly tilted towards theory-builders. Yet the vision of algebraic number theory as a rising sea with progress signaled by the application of (to quote Gowers) *deep theorems of great generality* is not, in my opinion, an accurate reflection on reality.

A good lemma is worth a thousand theorems. Gowers describes various principles of combinatorics which (he suggests) play the role of (a direct quotation again) *precisely stated* (general) *theorems*. Yet examples similar to his are readily available in algebraic number theory. Consider, for example, the following Lemma of Ribet (modified from its original formulation):

> If a reducible representation $U \oplus V$ of a group G deforms continuously into an irreducible representation of G, then either there exists a non-trivial extension of U by V, or an extension of V by U.

As a mathematical result, this is not particularly deep. For example, if G is finite, it relates two well known facts: there are no extensions between irreducible representations (Maschke's theorem), and representations of finite groups are defined

over number fields (and so do not deform). Yet this lemma is a crucial ingredient behind many key results (Ribet's construction of unramified extensions, the proof of the main conjecture of Iwasawa Theory by Mazur–Wiles, the non-triviality of the Selmer group of an ordinary Elliptic curve with $L(E,1) = 0$ by Skinner–Urban, and many more). It seems to me (as in the examples Gowers discusses) that the value of this lemma is not in its difficulty, but in the principle it encapsulates: in order to construct extensions of U by V, try to deform $U \oplus V$.

It is a common graduate student error to imagine that mathematics consists merely of judicious applications of highly technical machinery. But I am not accusing Gowers of making this mistake; I would expect him to argue that algebraic number theory is not *exclusively* the domain of theory-builders, but rather only strongly slanted in that direction, and to that end, he might point to Grothendieck. A fascinating essay on Grothendieck may be found here. Grothendieck contrasts his own way of thinking with that of Serre, whom he describes as using the *hammer and chisel* approach, which might loosely be considered synonymous to "problem-solver" (and I would count Serre as someone who has worked in algebraic number theory). Note that, despite the merits of Grothendieck's work, he famously failed to prove the Weil conjectures (by "failing" to prove the standard conjectures) and it required Deligne's use of the tensor power trick (a problem solving technique *par excellence*) to finish the argument. Thus, while Grothendieck's role in modern number theory is significant, it would be an error to imagine that it constitutes the whole subject.

Perhaps Gowers would instead argue that what set combinatorics apart from (say) algebraic number theory is not that it requires problem solvers while the latter field does not, but that (in contrast) it is the *exclusive* domain of problem solvers. There's a hint of this opinion in the following quote:

> One will not get anywhere in graph theory by sitting in an armchair and trying to understand graphs better.

Why is this claim any more convincing than the same statement with the word "graphs" replaced by the word "rings"? I don't see any a priori reasons why there cannot be a Grothendieck of graphs. If the history of mathematics teaches us anything, it is that the nature of a subject can change quite radically over a relatively short period of time (say 30 years). I am not claiming that there is no difference between combinatorics an algebraic number theory. There may well be a difference in the overall structure of the field, the level of background, the need to understand ideas in a broader conjectural framework, etc. And I might also consider agreeing to the claim that these fields, *as they are currently constituted*, may well be better suited to different *personalities*. But it is my opinion that the divide between the type of mathematics required for either subject is not as great as Gowers claims it is.

Gowers *main* point is that a significant part of the mathematical establishment looks down on combinatorics as not being "deep", and that this attitude is both harmful and ignorant. On this point, I think that Gowers criticisms are fair, accurate, and valuable. It's undeniably true that there are many graduate students who fall in love with formalism to the detriment of content, and milder forms of this predujice are pervasive throughout mathematics. To this end, I think Gowers' essay is timely and relevant. However, I can't help but sense *a little* that, perhaps after having spent a career defending combinatorics against ignorant snobs, Gowers

suffers from the *opposite* prejudice, where "theory-builders" are a short distance away from empty formalists, sitting comfortably in their armchairs thinking deep thoughts, studying questions so self referential that they no longer have any application to the original questions which motivated them (this sense also comes from reading some of the remarks on the Langlands programme here).

10. Number theory and 3-manifolds

Sun, 13 Jan 2013

It used to be the case that the Langlands programme could be used to say something interesting about arithmetic 3-manifolds qua hyperbolic manifolds. Now, after the work of Agol, Wise, and others (see [Ago13]) has blown the subject to smithereens, this gravy train appears to be over. It seems to me, however, that the great advance in our knowledge of hyperbolic 3-manifolds has precious little to say about arithmetic 3-manifolds qua lattices in semi-simple groups. As a basic example, suppose that X is a maximal compact arithmetic three orbifold associated to a quaternion algebra Q/F for some field F (with the appropriate behavior at the infinite primes). Then one may ask whether X has positive Betti number after some finite *congruence* cover $\widetilde{X} \to X$. Let's call this the virtual congruence positive Betti number conjecture. (This conjecture should be true — it is a consequence of Langland's conjectural base change for SL(2), which everyone believes but is probably very difficult.) AFAIK, there's not really much one can say about this problem from the geometric group theory/RAAG/LERF/etc perspective, where the arithmetic structure of the tautological SL(2)-representation does not seem to play so much of a role. A related question is the extent to which arithmetic 3-manifolds are intrinsically different from their non-arithmetic hyperbolic brethren. Is the virtual congruence Betti number conjecture (for arithmetic manifolds) something that could plausibly answered using geometric group theory?

Notes 10.1. I think the short answer is that it seems unlikely.

11. NT seminar: a haruspicy

Fri, 18 Jan 2013

Following Jordan Ellenberg's advice, I will blog on something that I know absolutely nothing about. Apologies in advance for mathematical errors!

Simon Marshall gave a number theory seminar this week about the first Betti number of $\Gamma(n)$ — as n varies — for certain lattices in $\mathrm{SU}(2,1)$. In particular, he proved an upper bound of the form:

$$\dim H^1(\Gamma(n), \mathbf{Q}) \ll [\Gamma : \Gamma(n)]^{3/8+\epsilon},$$

which turns out (in certain cases) to be essentially the best possible estimate. As was known to Rogawski, the forms contributing to H^1 all arise via endoscopy. In particular, if Γ is *simple* in the sense of Kottwitz, then the first cohomology vanishes (this also is due to Rogawski). So assume we are not in that case. The argument proceeds mostly as one would expect: Rogawski classifies the endoscopic forms which contribute to cohomology — they come from certain representations $\xi \times \mu$ for $U(2) \times U(1)$. Here I think the choice of Grossencharacter μ is almost

determined by ξ, so I will drop it from the notation below. The possible packets can be described as follows:

(1) Singletons for the split primes.
(2) A set $\{J^+, D^-\}$ for the interesting infinite prime, where J^+ contributes (via $(\mathfrak{g}, K)$ cohomology) to H^1 and another representation D^- which doesn't (although it contributes to H^2, I think).
(3) A set $\{\pi_s, \pi_p\}$ consisting of a supercuspidal representation and another representation at the inert primes.
(4) Something similar to 3. for the ramified primes.

Using Matsushima's formula, in order to count the contribution to cohomology one has to deal with the following:

(1) The global multiplicity: this is either 1 or 0 depending on certain signs related to epsilon factors. As one varies n this should vanish half the time, but one can ignore it as far as an upper bound goes.
(2) Suppose that p divides n, and let K be a hyperspecial maximal compact at p. Then one has to bound the trace of the characteristic function of $K(p^k)$ on the representations π_s and π_p.

Let f be such a characteristic function. One would like to write down a corresponding transfer function f^H on the endoscopic group such that:

$$\mathrm{Tr}(\pi_s, f) + \mathrm{Tr}(\pi_t, f) = \mathrm{Tr}(\xi, f^H)$$

By the Fundamental Lemma, if f is the characteristic function of the hyperspecial K itself, then f^H turns out to be the characteristic function on the maximal compact of $U(2)$. SML shows that (using some of the same computations required for the fundamental lemma for $U(3)$) the *same* identity holds for the corresponding characteristic function for $K(p^n)$, that is, the transfer f^H is the characteristic function of $U(2)(p^n)$. Is this true for any deeper reason? More generally, to what extent do characteristic functions transfer to characteristic functions?

Notes 11.1. Simon's paper is [Mar14]; see [GG23] for more recent developments.

12. Random p-adic matrices

Wed, 23 Jan 2013

Does anyone know if the problem of random matrices over (say) $\mathbf{Z}_p$ have been studied? Here I mean something quite specific. One could do the following, namely, since $\mathbf{Z}_p$ is compact with a natural measure, look at random elements in $M_N(\mathbf{Z}_p)$ and then ask about the distribution of several obvious quantities as N goes to ∞. For example, one can consider the rank of $M \mod p$, which translates into an elementary counting problem over $\mathbf{F}_p$. However, I *don't* mean this, that would just be rubbish for my purposes. What I am looking for is something that models a random *compact* operator, and then I want to understand the behavior of the normalized eigenvectors as the eigenvalue $\lambda \to 0$. To be concrete, let $B = \mathbf{Q}_p\langle T\rangle$ be the Tate algebra corresponding to the open unit ball. Then consider a "random" compact operator U acting on B. What does random mean? This is a good question, to which I do not know the answer. But let me give several properties that it should satisfy. Because the ball B is a disk, it is "dimension 2 as a real manifold", and so

— imagining that our compact operator is a p-adic avatar of $e^{-\nabla}$ for the Laplacian ∇ — the eigenvalues of U should satisfy Weyl's Law:

$$N(T) := \{\#\lambda \parallel -v(\lambda) \leq T\} \sim \frac{\mathrm{Vol}(B)}{4\pi} \cdot T.$$

Here $v(\lambda)$ denotes the valuation of $\lambda \in \overline{\mathbf{Q}}_p$. Ignoring the volume factor, this just means that the Fredholm determinant $\det(1 - UT)$ has a Newton Polygon with certain quadratic growth. I'm not sure exactly what ensembles one can come up with to define such operators, which is one of my questions. Let us also assume, although this may not be necessary, that U is semi-simple and admits nice convergent spectral expansions. We can't quite insist that U is a self-adjoint operator, because one doesn't have p-adic Hilbert spaces. For such an operator, what behavior should one expect of the normalized eigenvalues ϕ_j of U? For example, suppose one knows that the number of zeros of ϕ_j goes to infinity. What limit distribution should the zeros of ϕ_j satisfy when $\lambda \to 0$? (Somewhat troubling here is that the eigenvalues will lie in $\overline{\mathbf{Q}}_p$ in general and $\overline{\mathbf{C}}_p$ has compactness issues...)

As you might guess, this is related to *p-adic arithmetic quantum CHAOS*, a group of subjects which gets *sexier every time an extra adjective is added*, and will form part of my student project at the Arizona Winter School (see here).

13. Small cyclotomic integers

Sat, 26 Jan 2013

Julia Robinson is a famous mathematician responsible for fundamental work in logic and in particular on Hilbert's Tenth problem. Less well known nowadays is that her husband, Raphael Robinson, was a number theorist at Berkeley. One question R.Robinson asked (see [Rob65]) concerned *small* cyclotomic integers. Namely, let α be a cyclotomic integer, and suppose that *every* conjugate of α has absolute value at most R. Then what can one say about α? If $R \leq 1$, then Kronecker's theorem says that α is a root of unity (this statement only requires that α is an algebraic integer). Robinson studied the problem of what happens when $R \leq 2$ and also $R \leq \sqrt{5}$. He made five conjectures concerning these questions, four of which were solved in the 60's by Jones, Cassels, and Schinzel. Five decades later, Frederick Robinson (no relation!) and Michael Wurtz proved the last of these conjectures (while working with me as summer students), and their paper paper has just been accepted by *Acta Arithmetica* (see [RW13]). In particular, they answer the following problem: if α is an algebraic integer the largest of whose absolute values is $R \leq \sqrt{5}$, then what are the possible values of R? Two such families of such numbers are those of the form

$$\zeta + \zeta^{-1}, \qquad i + \zeta + \zeta^{-1}$$

for a root of unity ζ. These give all R of the form

$$2\cos(\pi/N), \qquad \sqrt{1 + 4\cos^2(\pi/N)}.$$

Note that these sets have limit points at $\sqrt{4}$ and $\sqrt{5}$ respectively. It turns out that there exactly two further exceptions, as follows:

$$\frac{\sqrt{3} + \sqrt{7}}{2}, \qquad \sqrt{\frac{5 + \sqrt{13}}{2}}$$

The first element is totally real and cyclotomic, and so manifestly occurs as such an R. The second turns out to be the absolute value of $1+\zeta_{13}+\zeta_{13}^4$. The proof by Robinson and Wurtz actually applies to slightly larger values of R, and after the limit point $\sqrt{5}$ there is another gap, and the next smallest possible R is

$$|1+\zeta_{70}+\zeta_{70}^{10}+\zeta_{70}^{29}| \sim \sqrt{5.017655\ldots}$$

The first two exceptional numbers turn up in relation to subfactors. How about the last example?

Notes 13.1. Kiran Kedlaya has pointed out to me that one of Robinson's five problems remains open, but should also be solvable by the methods of the Robinson–Wurtz paper [RW13]. (It relates to the difference between finding all possible values of $\boxed{\alpha}$ and finding all β with $\boxed{\beta}=\boxed{\alpha}$.)

14. Torsion in the cohomology of co-compact arithmetic lattices

Wed, 06 Feb 2013

Various authors (including Bergeron and Venkatesh) have shown that the cohomology of certain arithmetic groups have a *lot* of torsion. For example, if Γ is a co-compact arithmetic lattice in $\mathrm{SL}_2(\mathbf{C})$, and $\mathcal{L}$ is an acyclic local system, then

$$\log|H^*(\Gamma(N),\mathcal{L})| \gg [\Gamma:\Gamma(N)].$$

The proof relies on the fact that the difference l_0 in ranks of $\mathrm{SL}_2(\mathbf{C})$ and $\mathrm{SU}_2(\mathbf{C})$ is one. As the invariant l_0 grows, one expects there to be less torsion. How much torsion should one expect in general? I'm not sure I have an answer, but the point of this post is that Poincare duality gives a non-trivial bound, at least if one restricts to covers up a p-adic tower. Let $\mathbf{G}$ be a semi-simple group over $\mathbf{Q}$, Let $G=\mathbf{G}(\mathbf{R})$, let K be a maximal compact, let $H^*=\bigoplus H^m$, let Γ be a co-compact lattice, and let $\mathcal{L}$ be an acyclic local system. Suppose that $n=\dim(G)$ and $d=\dim(G/K)$. Then, for a fixed prime p (for which $\mathbf{G}(\mathbf{Q}_p)$ is split) and varying m, I claim that one has the inequality

$$\log|H^*(\Gamma(p^m),\mathcal{L})| \gg [\Gamma:\Gamma(p^m)]^{1-\frac{d}{n}}.$$

An elementary exercise shows that $\mathcal{L}/p\mathcal{L}$ is trivial as a local system for $\Gamma(p^m)$ and large enough m. The inequality above can then be reduced to the following claim: there is an inequality:

$$\dim H_*(\Gamma(p^m),\mathbf{F}_p) \gg p^{m(n-d)}.$$

Assume otherwise. The main point is as follows: taking the inverse limit over all m, we obtain modules $\widetilde{H}_j$ over the Iwasawa algebra Λ. This algebra, by results of Lazard and Venjakob [Laz65, Ven02], is essentially a regular local ring, in particular, it makes sense to talk about the dimension of modules over that ring. If the inequality above does not hold, then these modules will have small dimension, explicitly, co-dimension greater than d. This is so small that Poincare duality will, Ouroboros like – swallow itself completely and collapse into nothingness. However, the only way that could happen is if there was nothing to start with, which is nonsense.

More mathematically, consider the completed homology groups

$$\widetilde{H}_* = \varprojlim H_*(\Gamma(p^m),\mathbf{F}_p)$$

The homology groups may be computed by a complex of free Λ-modules obtain by lifting an initial triangulation on the base. (Here one thinks of group cohomology as the cohomology of the associated arithmetic quotients, of course.) Poincare duality then explains what happens when one takes the dual of this sequence and considers the corresponding homology groups, namely, there is a spectral sequence:

$$\mathrm{Ext}^i(\widetilde{H}_j, \Lambda) \Rightarrow \widetilde{H}_{d-i-j}.$$

This spectral sequence might be more familiar to some readers if one imagines Λ to be a field, in which case the zeroth Ext group is a Hom and the higher Exts vanish, and one obtains the duality isomorphisms between homology and cohomology over a field. Or, if Λ was the integers, then then zeroth Ext group is a Hom, the first Ext group is torsion, the higher Ext groups vanish, and one obtains the usual short exact sequence comparing the dual of homology to cohomology up to a torsion error term.) The dimension assumption we made implies that the limits are small as Λ-modules, in particular that $\mathrm{Ext}^i(\widetilde{H}_j, \Lambda) = 0$ for all $i \leq d$. The key here is a Theorem of Ardakov and Brown relating the size of the cohomology growth under towers to the codimension of the module. Yet putting this assumption into the spectral sequence shows that all terms with $i + j \leq d$ vanish, and hence that $\widetilde{H}_0 = \widetilde{H}_{d-d} = 0$. Yet it is easy to see that

$$\widetilde{H}_0 = \mathbf{F}_p,$$

and thus we have a contradiction.

In fact, this is the same argument that Matthew Emerton and I used to give lower bounds on torsion for p-adic analytic covers of 3-manifolds. There is some slack where the argument can be improved — since one only needs vanishing for a triangular portion of the spectral sequence, you are in good shape if you have extra information about the lower rows. Of course, the *real* answer to the amount of mod p torsion in these towers (which is a different question to the original one of torsion over the integers) should be:

$$\dim H_*(\Gamma(p^m), \mathbf{F}_p) \sim p^{m(n-l_0)},$$

where l_0 was defined above.

In a previous version of this post, I confused the roles of $\dim(K)$ and $d = \dim(G/K)$. For complex groups one has $n = 2d$, and this is asymptotically the correct estimate for simple real groups. In general, one has $n \geq (3/2)d$, with the worse case, ironically, corresponding to (any number of copies of) $\mathrm{SL}_2(\mathbf{R})$. So you get a bound of the form:

$$\log |H^*(\Gamma(N), \mathcal{L})| \gg [\Gamma : \Gamma(N)]^{1/3}.$$

15. Galois representations for non self-dual forms, Part I

Tue, 26 Mar 2013

This is the first of a series of posts discussing the recent work of Harris, Lan, Taylor, and Thorne [HLTT16] on constructing Galois representations associated to regular algebraic automorphic forms for $\mathrm{GL}(n)$ over a CM field F/F^+. I will dispense with any niceties about why one should care, and try simply to decipher the scribbles I made during a talk Richard gave at the Drinfeld seminar. I should

warn the reader of two difficulties: this paper does not exist as a public manuscript, and it also involves technical details which I generally prefer not to avoid thinking about. So caveat emptor.

First, some simplifying assumptions. Let's assume that:

(1) π_∞ has trivial infinitesimal character.
(2) π_p is unramified.
(3) F is an imaginary quadratic field in which p splits.

For examples, I will generally consider the case $n = 1$ and $n = 2$. The goal will be to construct a Galois representation

$$R_p(\pi) = r_p(\pi) \oplus \epsilon^{1-2n} r_p(\pi^{c,\vee})$$

If one can do this for π and for $\pi \otimes \chi$ for enough characters χ, then one can recover $r_p(\pi)$. Naturally enough, $R_p(\pi)$ will be associated to an automorphic form Π for a bigger group. Now $\pi \boxplus \epsilon^{1-2n}\pi^{c,\vee}$ is automorphic for $\mathrm{GL}(2n)/F$; it is, moreover, an essentially conjugate self-dual (RAESD) although no longer cuspidal. It does, however, come from a smaller group, namely, the unitary similitude group G which is ubiquitous in the papers of of Harris and Taylor. Over the complex numbers, G looks like $\mathrm{GL}(2n) \times \mathrm{GL}(1)$, but over the real numbers I think it must look like $\mathrm{GU}(n, n)$. Although it's true that the natural — i.e. occurring in cohomology of $X(G)$ — Galois representations associated to RAESDC forms ϖ for G will actually be nth exterior powers, I don't think that matters so much, since once one has congruences between ϖ and Π one gets Galois representations of the right degree for Π.

OK. Now associated to G and an open compact U of $G(\mathbf{A}^f)$ one has three natural objects: a smooth quasi-projective Shimura variety $Y = Y_U$, a (typically non-smooth) normal minimal compactification $X = X_U$, and a (family of) smooth toroidal compactifications $W = W_U$. The complement of Y in W is SNCD (smooth normal crossing divisor). I'm using somewhat non-standard terminology as far as the letters go because I don't want too many subscripts. If $n = 1$, then Y is an open modular curve, $X = W$ is a smooth compactification, and the complement of Y in W is a finite number of points (cusps). If $n = 2$, then Y has complex dimension 4. More on that example later.

As usual, one has the Hodge bundle $\mathbf{E} = \pi_* \Omega^1_{A/Y}$, from which one may build automorphic bundles ξ_ρ in the usual way for suitable algebraic representations ρ of what I guess amounts to the levi of $G(\mathbf{C})$. In my notes I have written:

$$\xi_{st} = \mathrm{st}_\tau \oplus {\mathrm{st}'}_{\tau'}$$

Here st means the standard n-dimensional representation of GL_n, and st$'$ denotes the complex conjugate representation. One must have has $\mathbf{E} = \xi_{st}$, where the decomposition into a direct sum of two rank n-modules comes from the action of the auxiliary ring on the tangent space to the universal abelian variety (built into the definition of G which I have omitted). I also have written:

$$\mathrm{KS} = \mathrm{st}_\tau \otimes {\mathrm{st}'}_{\tau'}$$

This presumably relates to the Kodaira–Spencer isomorphism. It's certainly consistent with a surjection:

$$\bigwedge^2 \pi_* \Omega^1_{A/Y} \to \Omega^1_{Y/k}$$

Now it turns out that ξ_ρ extends to W in two natural ways, there is the canonical extension ξ_ρ^{can} and the sub-canonical extension ξ_ρ^{sub}; they differ by the divisor corresponding to the boundary. Just as in the case $n = 1$, the bundle ξ^{can} should be though of as having log-poles at the boundary. Last but not least, for the one dimensional representation $\wedge^{2n}(\mathrm{st}_\tau \oplus \mathrm{st}'_{\tau'})$, one has the line bundle ω on Y. Denote the canonical extension of ω to W by ω. Then it turns out that ω is the pull-back of an ample line bundle ω on X. Of course, if $n = 1$, then ω is what you think it is — well, almost, since we are using $GU(1,1)$ Shimura varieties rather than $\mathrm{GL}(2)$. However, for general n, things are a little trickier. For example, ω is ample on X, but not (in general) on W.

If U is maximal at p, then the previous constructions also work over a finite field k of characteristic p and the appropriate smoothness claims are still true. One has the Hasse invariant H, which is a section of ω^{p-1} over X/k. Since ω is ample on X, the complement of the zero divisor of H is affine, it is of course the ordinary locus. In particular, one has Galois representations of the correct flavor associated to forms in the infinite dimensional space

$$H^0(X^{\mathrm{ord}}, \xi_\rho)$$

This follows in the "usual" way; Richard sketched an argument, it goes as expected, although I think the Kocher principle must have slipped in at some point.

So far, I haven't really said anything related to the actual argument, but I think I will stop here for now. The next step is to connect Π in any way to classes in the p-adic modular forms arising in the cohomology group above.

16. Galois representations for non self-dual forms, Part II

Sun, 21 Apr 2013

Let's recap from part I. We have a Shimura variety Y, a minimal projective compactification X, and a (family of) smooth toroidal compactifications W. We also have Galois representations of the correct shape associated to eigenclasses in

$$H^0(X^{\mathrm{ord}}, \xi_\rho).$$

So at this point (well, not only at this point) there is some confusion. In the construction above, I am imagining that we are working with the rigid analytic space corresponding to the ordinary locus. But now there are some remarks in my notes about dagger spaces. Here is what I am imagining is going on. For any sufficiently small radius, we may consider the rigid analytic space $Y[\nu]$ which corresponds (on the moduli level) to the appropriate abelian varieties A (with polarization and level structure and enomorphisms, blah blah) together with a canonical subgroup which (under some measure) is close to being ordinary. Then there is a "dagger space" $Y^\dagger$ which is the limit of all such spaces. The issue (for me) is that I don't really know anything about dagger spaces, but since this is probably not the main point, I will (again) elide the issue here. Of course, the goal is to realize the eigenvalues of the Eisenstein series Π inside this cohomology. Let's assume that Π actually has good reduction at p. Then it is probably going to be true that Π actually has finite slope, and so it lives inside the cohomology of some overconvergent neighbourhood of X^{ord}. So there's some flexibility with exactly what spaces one is working

with. Perhaps working with finite slope eigenforms might help to get local-global compatibility at p.

(It's most natural to work with the dagger spaces (whose cohomology is as described above) since that most naturally corresponds to the rigid cohomology groups occurring below.)

OK, so, we may take the direct limit over all compact subgroups U of the cohomology above, and we want to realize the Eisenstein series Π as a p-adic cusp form inside this space.

To this end, one introduces the following cohomology groups:

$$H^*_{c,\partial}(\overline{X}^{\mathrm{ord}}) := \mathbf{H}^*(W^{\mathrm{ord}}, \Omega^\bullet_{W^{\mathrm{ord}}}(\log\infty)\otimes\mathcal{L})$$

OK. So this is just a definition, it isn't supposed to obviously be functorial: we are taking the special fibre, lifting to characteristic zero, taking a toroidal compactification, then looking at the hypercohomology of the de Rham complex with log poles at the boundary. Well I guess one can do whatever one wants, I suppose.

So what is this? The hycohomology of the de Rham complex of a smooth variety M with log poles along some divisor D with normal crossings should just be the Betti cohomology of the complement of D in M. The factor $\mathcal{L}$ is the difference between the sub-canonical and canonical extensions, not entirely sure why it is there, presumably for some fundamentally important reason. So morally, I think the RHS should be computing something like the Betti cohomology of Y^{ord}, with the proviso that these are dagger spaces, not smooth complex varieties. So one should think of the LHS is some type of algebraic Betti cohomology of the ordinary locus.

Update: the remark about the Betti cohomology of the complement of D is correct, but the presence of the boundary divisor $\mathcal{L}$ is exactly what, in the classical sense, changes the answer from the cohomology of the open variety to the interior cohomology. So the cohomology is somehow compactly supported towards the boundary of W, but not the "other" part of the boundary (that is, the difference between W and W^{ord}. Let's write down a spectral sequence:

$$H^i(W^{\mathrm{ord}}, \Omega^\bullet_{W^{\mathrm{ord}}}(\log\infty)\otimes\mathcal{L}) \Rightarrow H^{i+j}_{c,\partial}(\overline{X}^{\mathrm{ord}}),$$

The existence of this spectral sequence must be a formal consequence of the definition and properties of hypercohomology. Note that the $\Omega^j_{W^{\mathrm{ord}}}(\log\infty)$ are canonical automorphic sheaves of the standard type, so with the boundary piece $\mathcal{L}$ the LHS consists of terms of the form $H^i(W^{\mathrm{ord}}, \xi^{\mathrm{sub}})$. To compute these terms, one can push foward via the map $\pi : W \to X$ from the toroidal compactification to the minimal one. Then one notes that:

(1) The higher direct images $R^i\pi_*\xi^{\mathrm{sub}}$ vanish.
(2) Since X^{ord} is affinoid, its higher cohomology also vanishes.

The second point seems reasonable, I have no idea why the first is true. It is probably a really key point, which I might talk about in part III (note: Richard said nothing about this and there is no pre-print, so I have no idea how to prove this at the moment). Apparently it is important that one uses the subcanonical extension here. This implies that every class which occurs in the RHS in this new cohomology actually occurs in an H^0 term on the LHS. Now one has Galois representations of terms of the form $H^0(W^{\mathrm{ord}}, \xi^{\mathrm{sub}})$, by the first construction - here it must be OK to

pass between W and X using the Kocher principle. So we are reduced to showing that Π contributes to this new cohomology $H^\bullet_{c,\partial}(X^{\mathrm{ord}})$.

Update: here is some more about higher direct images. Let's say a little bit about what the toroidal compactifications look like. Let's even imagine we are working with $\mathcal{A}_2$ and are looking at a cusp where one has purely toric reduction. For the purposes of computing the higher direct images all that matters is the formal completion of W, which at the boundary looks something like Z/Γ for some toric variety Z which is not of finite type. One shows that $H^i(Z, \mathcal{O}_Z) = 0$ using Cech cohomology for $i > 0$, which allows one to think of Z as contractible. Then one would like to say that $H^i(Z/\Gamma, \mathcal{O}_Z)$ is also zero, which comes down to understanding the action of Γ on $H^0(Z, \mathcal{O}_Z)$. Roughly one would like to say that Γ acts with no fixed points and and use Shapiro's Lemma. Back to the specific example, one finds that $H^0(Z, \mathcal{O}_Z)$ corresponds to positive semi-definite 2×2 matrices, and Γ a finite index subgroup of $\mathrm{GL}_2(\mathbf{Z})$. Here one should be reminded of the q-expansions of Siegel modular forms at the cusp — recall that q-expansions are given in terms of such matrices whose coefficients are invariant under $M \mapsto XMX^T$. This action is free as long as $\det(X) \neq 0$; at the level of q-expansions this corresponds exactly to working with cusp forms; this is why working with the sub-canonical extension allows one to restrict the positive definite forms on which the action is indeed free. In the degenerate case when $n = 1$, then Γ is trivial, and so it even acts freely on the non-cusp form 1, which is why it doesn't matter in that case.

Note: the spectral sequences above is, like the Hodge-de Rham spectral sequence, a 1-st page spectral sequence. Thus the vanishing above does *not* imply that it degenerates. Moreover, it certainly won't degenerate, since the RHS will turn out to consist of finite dimensional vector spaces, whereas the terms on the LHS are certainly not (as they are spaces of p-adic or overconvergent forms). (Note to self: compare to work of Coleman.)

The next point is the following. Suppose one now simply replaces X^{ord} by X. Then the cohomology theory $H^\bullet_{c,\partial}$ is probably *literally* computing the Betti cohomology of Y. The Betti cohomology of Y does indeed see the classes coming from the boundary that we would like to find.

Recall that $W \smallsetminus Y$ is a normal crossings divisor. Let ∂_0 denote the variety, ∂_1 the (disjoint) union of the irreducible components of the boundary divisor, ∂_2 the union of the intersection of these components, and so on. One now writes down another 1st page spectral sequence as follows:

$$\mathbf{H}^j(\partial_i, \Omega^\bullet_{W^{\mathrm{ord}}}(\log \infty)) \Rightarrow H^{i+j}_{c,\partial}(X^{\mathrm{ord}}).$$

This is supposed to be an example of the following: in a nice geometric situation (normal crossings divisor) one may compute cohomology with compact supports in terms of the cohomology of the boundary strata. (I'm still a little confused why $H^*_{c,\partial}$ is cohomology with compact supports rather than the cohomology of the interior, but anyway... (**update**: this is explained above: the presence of $\mathcal{L}$ means it has compact supports in the direction of $W \smallsetminus Y$, but not $W \smallsetminus W^{\mathrm{ord}}$)). Moreover, a key point is that the LHS can be interpreted as the rigid cohomology of ∂_i. This allows one to use results of Berthelot and Chiarellotto to deduce that the terms of the LHS are given in terms of the rigid cohomology of (open) varieties. In particular:

(1) They admit a theory of weights,
(2) H^j is mixed of weight at least j.

(3) They are finite dimensional.

We deduce that RHS is also mixed of weight at least $i+j$ and finite dimensional. We want our Π to occur in the RHS, so it certainly suffices to show it actually occurs in H^0. But then by weights it suffices to show that it is coming from the H^0-terms in the LHS. These are simply given by component groups, and so the computation reduces to a problem concerning the combinatorics of the boundary, on which we shall say more in part III.

17. Inverse Galois problems I

Wed, 24 Apr 2013

My favourite group as far as the inverse Galois problem goes is $G = \mathrm{SL}_2(\mathbf{F}_p)$. This is not known to be a Galois group over $\mathbf{Q}$ for any $p \geq 13$, the difficulty of course being that is must correspond to an even Galois representation. A more tractable case is $G = \mathrm{PSL}_2(\mathbf{F}_p)$, and this was recently answered by David Zywina here (see [Zyw15]). Here is a more elementary version of that construction. Suppose that π is a classical modular form of weight three with coefficients in $\mathbf{Z}[\sqrt{-1}]$ and quadratic Nebentypus character χ. Note that there is an isomorphism $\pi^c := \overline{\pi} \simeq \pi^\vee \otimes \| \cdot \|^2 \chi$. For all primes v in $\mathbf{Q}(i)$, one obtains a representation:

$$\varrho = \rho \otimes \epsilon^{-1} : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\mathbf{F}_v).$$

with determinant χ. There are two cases, depending on whether $v|p$ is split or not. If $p \equiv 1 \mod 4$ splits, then, assuming π is not CM, the image of ϱ restricted to the kernel of χ is $\mathrm{SL}_2(\mathbf{F}_p)$ for sufficiently large p which can be explicitly determined in any specific case. Thus the image of ϱ is $\mathrm{SL}_2(\mathbf{F}_p)$ plus the image of complex conjugation:

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Since $p \equiv 1 \mod 4$, there exists an element $\alpha \in \mathbf{F}_p$ of square -1, and hence an element in $\mathrm{SL}_2(\mathbf{F}_p)$ equal to

$$\begin{pmatrix} \alpha & 0 \\ 0 & -\alpha \end{pmatrix}.$$

Hence the image of ϱ contains a scalar element of determinant -1, and thus it has projective image $\mathrm{PSL}_2(\mathbf{F}_p)$.

If $p \equiv -1 \mod 4$, then, from the isomorphism $\pi^c \simeq \pi^\vee \otimes \| \cdot \|^2 \chi$, there is an isomorphism $\varrho^c \simeq \varrho \otimes \chi$, where ϱ^c is the Galois conjugate induced by complex conjugation. It follows that the *projective* image of ϱ lands in $\mathrm{PGL}_2(\mathbf{F}_p)$. The image of ϱ is thus, for sufficiently large p, a subgroup of $\mathbf{F}_{p^2}^\times \mathrm{GL}_2(\mathbf{F}_p)$ with projective image containing $\mathrm{PSL}_2(\mathbf{F}_p)$. We first observe that this implies that ϱ contains $\mathrm{SL}_2(\mathbf{F}_p)$. It suffices to show that it contains all the transvections; yet the lift of any transvection in $\mathrm{PSL}_2(\mathbf{F}_p)$ is a transvection of order p times a scalar of order prime to p, which one can remove by taking an appropriate power. Since the determinant of ϱ is χ, this leaves only the following three possibilities for the image of ϱ:

(1) The subgroup of $\mathrm{GL}_2(\mathbf{F}_p)$ of matrices with determinant ± 1.
(2) The previous subgroup together with the the scalar element I with $I^2 = -1$.
(3) The group $\mathrm{SL}_2(\mathbf{F}_p)$ together with I.

The third group does not have a non-scalar element of order 2 correponding to complex conjugation, and the first has traces which do not generate $\mathbf{F}_{p^2}$. Hence the image must be the second, which has projective image $\mathrm{PSL}_2(\mathbf{F}_p)$.

To conclude the argument, it suffices to show that there exists such a π. Consulting William Stein's tables, one may take

$$f = q + 4i \cdot q^3 + 2 \cdot q^5 - 8i \cdot q^7 + \ldots \in S_3(\Gamma_1(32), \chi),$$

for a quadratic χ where $i^2 = -1$. Since $a_3, a_5, a_7 \neq 0$, this form does not have CM by $\mathbf{Q}(\sqrt{-1})$ or $\mathbf{Q}(\sqrt{-2})$, so $\mathrm{PSL}_2(\mathbf{F}_p)$ is a Galois group for sufficiently large p, which one could compute exactly if one wanted. My impression from the notation in William Stein's tables is that the fixed field of the kernel of χ is $\mathbf{Q}(\sqrt{-1})$, so this is presumably the same family of examples that arises in Zywina. Other examples (in the range of William's tables) are as follows:

$$g = q + 2i \cdot q^2 - 4 \cdot q^4 + (3 - 4i) \cdot q^5 + \ldots \in S_3(\Gamma_1(20)),$$

$$h = q + 3i \cdot q^2 - 5 \cdot q^4 - 3i \cdot q^5 + \ldots \in S_3(\Gamma_1(27))$$

Note that this argument requires slightly more than pure thought; it was key that there existed a non-CM form with coefficient field $\mathbf{Q}(\sqrt{-1})$, and there is no *a priori* reason why there should exist any such form. For example, suppose one wanted to generalize this argument to to $\mathrm{PSp}_4(\mathbf{F}_p)$. Then one would want to look for a non-endoscopic Siegel cusp form of weight (a, b) where (**edit**) $2a + b$ is odd with Hecke eigenvalues in $\mathbf{Q}(\sqrt{-1})$ and quadratic Nebentypus character. Possibly such things exist but perhaps they don't!

Comment 17.1 (David Zywina)**.** Yes, your cusp form of weight 3 and level 32 gives rise to exactly the same representations as in my paper! (I found this post when doing a literature search for a note I am finishing up.) Amusingly, your cusp form of weight 3 and level 27 shows up at the end of Serre's 1987 Duke paper [Ser03]. He shows that the mod 7 representation attached to it produces $\mathrm{PSL}_2(\mathbf{F}_7)$ as a Galois group over $\mathbf{Q}$ (unsurprisingly, the key is that the image contains a scalar matrix with determinant -1). Serre was actually giving an example of his conjecture (he started with the $\mathrm{PSL}_2(\mathbf{F}_7)$-extension and then found the form), so he overlooked that this cusp form also produces $\mathrm{PSL}_2(\mathbf{F}_p)$-extensions for all $p \geq 5$!

Comment 17.2 (Persiflage)**.** Concerning (from some anonymous comment): could you please elaborate on the line "If $p \equiv -1 \mod 4$, then, from the isomorphism $\pi^c \simeq \pi^\vee \otimes \| \cdot \|^2 \chi$, there is an isomorphism $\varrho^c \simeq \varrho \otimes \chi$, where ϱ^c is the Galois conjugate induced by complex conjugation. It follows that the *projective* image of ϱ lands in $\mathrm{PGL}_2(\mathbf{F}_p)$" Could you explain how the existence of an "inner twist" by c implies that the projective image lands in $\mathrm{PGL}_2(\mathbf{F}_p)$? Where does the congruence class of $p \mod 4$ play a role?

The assumption that $p \equiv -1 \mod 4$ means that the residue field of the coefficient ring is $\mathbf{F}_p(i) = \mathbf{F}_{p^2}$ (the case when $p \equiv 1 \mod 4$ is easier and was dealt with previously). Moreover, if $p \equiv 1 \mod 4$, then there are two primes above p in $\mathbf{Z}[\sqrt{-1}]$, and so there is no Galois action on the coefficient field. When $p \equiv -1 \mod 4$, complex conjugation on the coefficients induces the automorphism c of $\mathbf{F}_{p^2}$. Two representations V and W correspond to the same projective representation if and only if $V \simeq W \otimes \chi$ for some character χ. The representation ϱ *a priori* lands in $\mathrm{GL}_2(\mathbf{F}_{p^2})$, and the projective representation lands in $\mathrm{PGL}_2(\mathbf{F}_{p^2})$. The condition that V actually arises from a $\mathrm{GL}_2(\mathbf{F}_p)$ representation is that $V \simeq V^c$. The condition

that the projective representation lands in $\mathrm{PGL}_2(\mathbf{F}_p)$ is that $V \simeq V^c \otimes \chi$ for some character χ.

Notes 17.3. See § 67 for some updates.

18. Galois representations for non-self dual forms, Part III

Sat, 27 Apr 2013

Here are some complements to the previous remarks on [HLTT16], following on from §§ 15–16.

First, in order to deal with non-zero weights, one has to replace the Shimura varieties Y, X, W by Kuga-Satake varieties over these spaces. This "only" adds technical difficulties.

Second, in order to work over the most general bases F, one seems to require good minimal models and compactifications X_U, W_U in characteristic p, for a prime p which may be very ramified in F. This is a genuine problem. The way to avoid this problem is amusing. It turns out that one only needs a good model of X^{ord} and W^{ord}. In other words, one only has to understand integral models and toroidal compactifications at the ordinary cusps. However, the ordinariness is exactly what allows one to give appropriate models at these cusps, without having to deal with the more complicated cusps except in some fairly superficial way (say by taking normalizations over an integral model of a universal moduli space of abelian varieties). This seems quite clever.

Third, I was going to talk in more detail about $n = 2$, but having written down the argument it seems a little pointless now, since it is not going to simplify things very much. The only thing that is (perhaps) easier is to understand why the higher direct images of the pushforward of the subcanonical bundle to the minimal compactification vanishes; yet the example of $\mathcal{A}_2$ in the previous post gives the idea, I think. I was also going to talk about the combinatorics of the boundary and their relationship to the cohomology of $\mathrm{GL}(n)$, but on second thoughts I'm not.

Fourth, how close is $H^*_{c,\partial}(\overline{X}^{\mathrm{ord}})$ to $H^*_{c,\mathrm{Betti}}(X)$, the compactly supported Betti cohomology of the Shimura variety? It's not so clear.

Fifth, the argument really only uses the ordinary locus in a fairly loose way, namely, it is (in the minimal compactification) affinoid, and it is compatible with Hecke correspondences. On the other hand, at finite level, this is pretty much the only possible such choice. However, perhaps at infinite level there may be other possible choices (in a perfect[-oid] world, as it were...).

19. Catalan's constant and periods

Sat, 04 May 2013

There is a 60th birthday conference in honour of Frits Beukers in Utrech in July; I'm hoping to swing by there on the way to Oberwolfach. Thinking about matters Beukers made me reconsider an question that I've had for while.

There is a fairly well known explanation of why $\zeta(3)$ should be irrational (and linearly independent of π^2) in terms of Motives. There is also a fairly good proof that $\zeta(3) \neq 0$ in terms of the non-vanishinjg of Borel's regulator map on $K_5(\mathbf{Z})$. (I

guess there are also more elementary proofs of this fact.) A problem I would love to solve, however, is to show that, for all primes p, the Kubota-Leopoldt p-adic zeta function $\zeta_p(3)$ is non-zero. Indeed, this is equivalent to the injectivity of Soule's regulator map

$$K_5(\mathbf{Z}) \otimes \mathbf{Z}_p \to K_5(\mathbf{Z}_p).$$

(Both these groups have rank one, and the cokernel is (at least for $p \geq 5$) equal to $\mathbf{Z}_p/\zeta_p(3)\mathbf{Z}_p$ by the main conjecture of Iwasawa theory.) It is somewhat of a scandal that we can't prove that $\zeta_p(3)$ is zero or not; it rather makes a mockery out of the idea that the "main conjecture" allows us to "compute" eigenspaces of class groups, since one can't even determine if there exists an unramified non-split extension

$$0 \to \mathbf{Q}_p(3) \to V \to \mathbf{Q}_p \to 0$$

or not. Well, this post is about something related to this but a little different. Namely, it is about the vaguely formed following problem:

Problem 19.1. What is the relationship between a real period and its p-adic analogue?

Since one number is (presumably) in $\mathbf{R} \smallsetminus \mathbf{Q}$ and the other in $\mathbf{Q}_p \smallsetminus \mathbf{Q}$, it's not entirely clear what is meant by this. So let me give an example of what I would like to understand. One could probably do this example with $\zeta(3)$, but I would prefer to consider the "simpler" example of Catalan's constant. Here

$$G = \frac{1}{1} - \frac{1}{3^2} + \frac{1}{5^2} - \frac{1}{7^2} \ldots = L(\chi_4, 2) \in \mathbf{R},$$

is the real Catalan's constant, and

$$G_2 = L_2(\chi_4, 2) \in \mathbf{Q}_2$$

is the 2-adic analogue. (There actual definition of the Kubota-Leopoldt zeta function involves an unnatural twist so that one could conceivably say that $L_2(\chi_4, 2) = 0$ and that the non-zero number is $\zeta_2(2)$, but this is morally wrong, as the examples below will hopefully demonstrate. Morally, of course, they both relate to the motive $\mathbf{Q}(2)(\chi_4)$.)

So what do I mean is the "relation" between G and G_2. Let me give two relations. The first is as follows. Consider the recurrence relation (think Apéry/Beukers):

$$n^2 u_n = (4 - 32(n-1)^2)u_{n-1} - 256(n-2)^2 u_{n-2}.$$

It has two linearly independent solutions with $a_1 = 1$ and $a_2 = -3$, and $b_1 = -2$ and $b_2 = 14$. One fact concerning these solutions is that $b_n \in \mathbf{Z}$, and $a_n \cdot \gcd(1, 2, 3, \ldots, n)^2 \in \mathbf{Z}$. Moreover one has that:

$$\lim_{n \to \infty} \frac{a_n}{b_n} = G_2 \in \mathbf{Q}_2.$$

The convergence is very fast, indeed fast enough to show that $G_2 \notin \mathbf{Q}$ (see [Cal05]). What about convergence in $\mathbf{R}$, does it converge to the real Catalan constant? Well, a numerical test is not very promising; for example, when $n = 40000$ one gets $0.625269\ldots$, which isn't anything like $G = 0.915966\ldots$; for contrast, for this value of n one has $a_n/b_n - G_2 = O(2^{319965})$, which is pretty small. There are, however, two linearly independent solutions over $\mathbf{R}$ given analytically by

$$\frac{(-16)^n}{n^{3/2}}\left(1 + \frac{5}{256}\frac{1}{n^2} - \frac{903}{262144}\frac{1}{n^4} + \frac{136565}{67108864}\frac{1}{n^6} - \frac{665221271}{274877906944}\frac{1}{n^8} + \ldots\right),$$

$$\frac{(-16)^n \cdot \log n}{n^{3/2}} \left(1 + \frac{5}{256}\frac{1}{n^2} - \frac{32261}{7864320}\frac{1}{n^4} + \frac{136565}{67108864}\frac{1}{n^6} - \frac{665221271}{274877906944}\frac{1}{n^8} + \dots\right)$$
$$+\frac{(-16)^n}{n^{3/2}} \left(-\frac{1}{768}\frac{1}{n^2} + \frac{32261}{7864320}\frac{1}{n^4} - \frac{30056525}{8455716864}\frac{1}{n^6} + \frac{1778169492137}{346346162749440}\frac{1}{n^8} + \dots\right),$$

from which one can see that a_n/b_n must converge very slowly, and indeed, one has (caveat: I have some idea on how to prove this but I'm not sure if it works or not):

$$\frac{a_n}{b_n} = G - \frac{1}{(0.2580122754655\dots) \cdot \log n + 0.7059470639\dots}$$

So one has a naturally occurring sequence which converges to G in $\mathbf{R}$ and G_2 in $\mathbf{Q}_2$. So that is some sort of "relationship" alluded to in the original question. Here's another connection. Wadim Zudilin pointed out to me the following equality of Ramanujan:

$$G = \frac{1}{2}\sum_{k=0}^{\infty} \frac{4^k}{(2k+1)^2\binom{2k}{k}} \in \mathbf{R}.$$

This sum also converges 2-adically. So, one can naturally ask whether

$$G_2 \stackrel{?}{=} \frac{1}{2}\sum_{k=0}^{\infty} \frac{4^k}{(2k+1)^2\binom{2k}{k}} \in \mathbf{Q}_2.$$

(It seems to be so to very high precision.) These are not random sums at all. Indeed, they are equal to

$$\frac{1}{2} \cdot F\left(\begin{array}{c} 1,1,1/2 \\ 3/2,3/2 \end{array}; z\right)$$

at $z = 1$. Presumably, both of these connections between G and G_2 must be the same, and must be related to the Picard–Fuchs equation/Gauss-Manin connection for $X_0(4)$. This reminds me of another result of Beukers in which one compares values of hypergeometric functions related to Gauss-Manin connections and elliptic curves, and finds that they converge in $\mathbf{R}$ and $\mathbf{Q}_p$ for various p to algebraic (although sometimes different!) values. Of course, things are a little different here, since the values are (presumably) both transcendental. Yet it would be nice to understand this better, and see to what extent there is a geometric interpretation of (say) the non-vanishing of $L_p(\chi, 2)$ for some odd quadratic character χ. Of course, one always has to be careful not to accidentally prove Leopoldt's conjecture in these circumstances.

Notes 19.2. The claims here are provable but the general question remains vague and mysterious. This post is also related to forthcoming work with Vesselin Dimitrov and Yunqing Tang and will be updated later.

20. Exercise concerning quaternion algebras

Sat, 11 May 2013

Here's a fun problem that came up in a talk by Jacob Tsimerman on Monday concerning some joint work with Andrew Snowden:

Problem: Let $D/\mathbf{Q}(t)$ be a quaternion algebra such that the specialization D_t splits for almost all t. Then show that D itself is split.

As a comparison, if you replace $\mathbf{Q}$ by $\overline{\mathbf{Q}}$, then although the condition that D_t splits becomes empty, the conclusion is still true, by Tsen's theorem.

This definitely *feels* like the type of question which should have a slick solution; can you find one?

21. Equidistribution of Heegner points

Wed, 15 May 2013

I saw a nice talk by Matt Young recently (joint work with Sheng-Chi Liu and Riad Masri, see [LMY13]) on the following problem. For a fundamental discriminant $|D|$ of an imaginary quadratic field F, one has h_D points in $X_0(1)(\mathbf{C})$ with complex multiplication by the ring of integers of F. Choose a prime q which splits in $F = \mathbf{Q}(\sqrt{-|D|})$. One obtains a set of $2h_D$ points in $X_0(q)(\mathbf{C})$, given explicitly as follows:

$$\mathbf{C}/\mathfrak{a} \mapsto \mathbf{C}/\mathfrak{a}\mathfrak{q}^{-1}$$

for $\mathfrak{a}$ in the class group and $\mathfrak{q}$ one of the two primes above q in F. The complex points $X_0(q)(\mathbf{C})$ can be thought of as being tiled by $q+1$ copies of the fundamental domain Ω in the upper half plane.

Problem 21.1. How large does D have to be to guarantee that every one of the $q+1$ copies of Ω contains one of the $2h_K$ CM points by $\mathcal{O}_F$?

This is the question that Young and his collaborators answer. Namely, one gets an upper bound of the shape $|D| \leq O(q^{m+\epsilon})$ (with some explicit m, possibly 20), the point being that this is a polynomial bound. Note that this proof is not effective, since it trivially gives a lower bound on the order of the class group which is a power bound in the discriminant, and no such effective bounds are known.

I idly wondered during the talk about the following mod-p version of this problem. To be concrete, suppose that $p = 2$ (the general case will be similar). We now suppose that D is chosen so that 2 is inert in F. Then all the h_K points in $X_0(1)(\overline{\mathbf{F}}_2)$ are supersingular, which means that they all reduce to the same curve E_0 with j-invariant 1728. Now, as above, choose a prime q which splits in F. The pre-image of $j = 1728$ in $X_0(q)(\overline{\mathbf{F}}_2)$ consists of exactly $q+1$ points.

Problem 21.2. How large does $|D|$ have to be to ensure that these points all come from the reduction of one of the $2h_K$ CM points by $\mathcal{O}_F$ as above?

Since E_0 is supersingular, we know that $\mathrm{Hom}(E_0, E_0)$ is an order in the quaternion algebra ramified at 2 and ∞. In fact, it is equal to the integral Hamilton quaternions $\mathbf{H}$. If E and E' are lifts of E_0, then there is naturally a degree preserving injection:

$$\mathrm{Hom}(E, E') \to \mathrm{Hom}(E_0, E_0) = \mathbf{H}.$$

The degree on the LHS is the degree of an isogeny, and it is the canonical norm on the RHS. In particular, if $E = \mathbf{C}/\mathfrak{a}$ and $E' = \mathbf{C}/\mathfrak{a}\mathfrak{q}^{-1}$, then one obtains a natural map:

$$\psi_{\mathfrak{a}} : \mathfrak{q}^{-1} \simeq \mathrm{Hom}(E, E') \to \mathbf{H}$$

preserving norms. The norm map on $\mathfrak{q}^{-1}$ is $N(x)/N(\mathfrak{q}^{-1})$. The image of the natural q isogeny is simply $\psi_{\mathfrak{a}}(1)$, whose image has norm q. Hence the problem becomes:

Problem 21.3. If one considers all the $2h_K$-maps:

$$\psi_{\mathfrak{a}} : \mathfrak{q}^{-1} \to \mathbf{H}, \qquad \psi_{\mathfrak{a}} : \overline{\mathfrak{q}}^{-1} \to \mathbf{H},$$

do the images of 1 cover the $q+1$ elements of $\mathbf{H}$ of norm q?

Given a field F in which 2 is inert, it wasn't obvious how to explicitly write down the maps $\psi_{\mathfrak{a}}$, but this problem does start to look similar in flavour to the original one. Moreover, to make things even more similar, in the original formulation over $\mathbf{R}$ one can replace modular curves by definite quaternion algebras ramified at (say) 2 and q, and then the Archimidean problem now also becomes a question of a class group surjecting onto a finite set of supersingular points. In fact, this Archimedean analogue may well be *equivalent* to the mod 2 version I just described! Young told me that his collaborators had mentioned working with various quotients coming from quaternion algebras as considered by Gross, which I took to mean the finite quotients coming from definite quaternion algebras as above. Hence, with any luck, they will provide an answer this problem.

Notes 21.4. I wasn't really away of this paper, for example [Mic04, Thm 3]).

22. Finiteness of the global deformation ring over local deformation rings

Sat, 18 May 2013

(This post is the result of a conversation I had with Matt). Suppose that

$$\overline{\rho} : G_F \to \mathrm{GL}_n(\mathbf{F})$$

is a continuous mod-p absolutely irreducible Galois representation. For now, let's assume that F/F^+ is a CM field, and $\overline{\rho}$ is essentially self-dual and odd. Associated to this representation is a global deformation ring R (of essentially self-dual representations) consisting of representations with no local restriction at primes dividing p and the condition of being unramified at primes away from p. One also has a (collection of) local (unrestricted) deformation rings for the set of primes $v|p$, combining to give a ring R^{loc}. Let us also assume that $\overline{\rho}$ has suitably big image (for example, its restriction to $F(\zeta_p)$ is adequate). Then we have:

Proposition 22.1. *The map $R^{\mathrm{loc}} \to R$ is finite.*

(Matt and Vytas prove this in the modular (odd) case when $n = 2$ and $F = \mathbf{Q}$, although I'm not sure whether the paper exists yet [actually, I'm pretty sure it doesn't]. Possibly if I was listening closer to Matt's talk at Fields I might have remembered the argument, since I vaguely think it came up there, although possibly only briefly.)

Here one has to be a little careful defining deformation rings in the local case, of course (for those worried by such issues, simply choose suitable framings). To prove this, it suffices to prove the result after base change, so we may assume that $\overline{\rho}$ is unramified at all primes, and completely trivial at all primes dividing p. By Nakayama's lemma, the problem above reduces to the following:

Proposition 22.2. *Let F^{ur} be the maximal extension of F unramified everywhere. Let Γ be the Galois group of F^{ur} over F. Then Γ does not admit a continuous essentially self-dual representation:*

$$\Gamma \to \mathrm{GL}_n(A)$$

such that A is a complete local Notherian $\mathbf{F}$-algebra of positive dimension.

This is a special case of the generalization of the unramified Fontaine–Mazur conjecture due to Boston. Recall that the group Γ may be infinite (Golod-Shafarevich), but that Fontaine–Mazur predicts that the image of any such representation into any characteristic zero p-adic analytic group has finite image. Boston conjectured that the same finiteness would hold for homomorphisms of Γ into $\mathrm{GL}_n(A)$ for rings like $A = \mathbf{F}[[T]]$. It turns out that even though the Fontaine–Mazur conjecture is hard, when A has characteristic p the conjecture is amenable to modularity lifting theorems by comparison to a new deformation ring in *regular* weight.

The proof is as follows:

Step 1: Using lifting theorems (Theorem 4.3.1 from [BLGGT14]), we may assume, after a finite base change, that $\overline{\rho}$ is potentially ordinarily modular of level one for some regular weight w.

Step 2: Using minimal modularity theorems in the ordinary case (Section 10 from Thorne's Jussieu paper [Tho12], or Theorem 2.2.2 of [BLGGT14], both using work of Geraghty), deduce that the minimal weight w ordinary deformation ring S is finite over $W(\mathbf{F})$, and hence that S/p is finite over $\mathbf{F}$. Strictly speaking, theorems of this kind are required to prove the previous result.

Step 3: Note that the minimal everywhere unramified deformations of $\overline{\rho}$ (i.e., the deformations coming from Γ) of characteristic p are all ordinary of weight w, because everything unramified is ordinary, and in characteristic p any two weights are the same. Hence R/p is a quotient of S/p, from which it follows from the finiteness of S that R is also finite.

While I am using the latest modularity lifting theorems here, weaker versions for $n = 2$ with some local assumptions on $\overline{\rho}$ follow from 90's era technology (say Taylor's Remarks on a conjecture of Fontaine and Mazur paper from 2000, or even earlier if one assumes residual modularity).

Via the usual argument, this result also applies to even Galois representations $\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{F})$ with large image. In particular, the unramified deformation rings in these cases will be finite over $W(\mathbf{F})$, and there will be at most finitely many counter examples to the unramified Fontaine–Mazur conjecture in characteristic zero for a fixed residual representation. One can also apply it to many classes of higher dimensional non-self dual representations by taking irreducible summands of $\rho \otimes \rho^{\vee}$. For example, one can take any representation of $\mathbf{Q}$ whose image contains $\mathrm{SL}_n(\mathbf{F}_p)$ if n is even, since then the associated (n^2-1)-dimensional representation $\mathrm{Ad}^0(\overline{\rho})$ restricted to an auxiliary CM field is irreducible, odd, self-dual, and adequate for large enough p. Similar remarks apply to representations over an arbitrary field F with generic enough image by taking the tensor induction down to $\mathbf{Q}$.

If one starts allowing ramification at auxiliary primes, things become a little harder. One fix is to build the auxiliary primes into the local deformation ring R^{loc}, although this might be considered cheating. The problem is that one cannot deduce

(in general) that more general ordinary deformation rings S are finite in the non-minimal situation. Although perhaps one can get by with the Taylor trick in some contexts. One should be OK with GL_2 by Ihara's Lemma.

Notes 22.3. One can handle the non-minimal case as well, fairly robust arguments are given in my paper with Patrick Allen [AC14].

23. Scholze on torsion 0

Thu, 13 Jun 2013

This will be the zeroth of a series of posts talking about Scholze's recent preprint, available here (now published [Sch15b]). This is mathematics which will, no question, have more impact in number theory than any recent paper I can think of. The basic intent of this post is to commit to future posts in which I will discuss the details. I should remark that Scholze's writing is pretty clear, so these posts will mainly be for my own benefit rather than yours.

Here are some of the specific points that I might cover:

Basics: The Hodge–Tate Period map, Perfectoid spaces, etc. To be honest, I will probably skip the details here to begin with, and discuss them only at points where they become fundamental for understanding.

Theorem IV.3.1: The action of Hecke on the completed cohomology groups $\widetilde{H}^i(\mathbf{Z}/p^n\mathbf{Z})$ for Shimura varieties is detected by the action of Hecke on classical cuspidal automorphic forms. Although it may end up being no easier to consider, this result is already intersting in some quite degenerate cases. For example, this is new even for $X = U(2,1)/\mathbf{Q}$ and $i = 1$ (Gee and Emerton's results, for example, are contingent on the relevant Galois representations being three dimensional — now one knows that they are!). A very similar example is the case of a compact inner form of $U(2,1)$ (so called Rogawski lattices) or, more generally, the simple Shimura variety of Kottwitz-Harris-Taylor type. Can one show in those cases that $\widetilde{H}^i$ vanishes outside degree zero and outside the middle dimension? A weaker question: can one compute the completed cohomology in degree one? Compare with the work of Pascal Boyer.

Local Global Compatibility: Suppose one is in the ordinary case. Then the HLTT approach [HLTT16] (via congruences, discussed previously in § 15, § 16, or § 18, should allow one to establish some cases of local-global compatibility. At ramified primes $\ell \neq p$, the HLTT approach should also work, especially if one is also willing to assume that the residual representation is absolutely irreducible (using base change arguments). What can one do in the torsion case?

The Nilpotent Ideal: Scholze ultimately constructs Galois representations over $\mathbf{T}/I$ for an ideal I such that $I^m = 0$. The necessity of this ideal arises from a spectral sequence argument. (The parameter m only depends on the degree of the field and n.) The Calegari–Geraghty modularity lifting argument (in the minimal case) can still be made to apply even with the presence of this ideal if one is in the *minimal* case, but *not* in the *non-minimal* case which will require $m = 1$ (the Taylor Ihara's avoidance trick requires more precise control than the minimal case). Are there any circumstances (extra assumptions, etc.) in which one can prove that $m = 1$?

Notes 23.1. Of course, § 91 and [ACC+23] is relevant for some questions raised here.

24. Scholze on torsion, Part I

Sun, 16 Jun 2013

This is a sequel to § 23, although as it turns out we *still* won't actually get to anything substantial — or indeed anything beyond an introduction — in this post.

Let me begin with some overview. Suppose that $X = \Gamma\backslash G/K$ is a locally symmetric space, where G is a semi-simple group which does *not* admit discrete series. To be concrete, suppose that $G = \mathrm{SL}_2(\mathbf{C})$, and that Γ is an congruence subgroup of level N in the Bianchi group $\mathrm{SL}(\mathcal{O}_F)$ — recall here that F is an imaginary quadratic field. Since days of yore (Langlands, Clozel, Fontaine–Mazur, etc.), everyone has expected that there should exist a bijection between the following objects:

(1) Regular algebraic cuspidal automorphic representations π for $\mathrm{GL}_2(\mathbf{A}_F)$ of level N and weight zero (= the same infinitesimal character as the trivial representation).
(2) Cuspidal (= non-boundary) cohomology classes $H^1(\Gamma, \mathbf{R})$ which are eigenforms for the ring of Hecke operators $\mathbf{T}$.
(3) Weakly compatible families of two-dimensional Galois representations of F which are irreducible of level N and Hodge–Tate weight $[0, 1]$.
(4) Irreducible semi-stable p-adic Galois representations of weight $[0, 1]$ and level (determined in the usual way) dividing N.
(5) Abelian Varieties of GL(2)-type over F which don't have CM by F.

The equivalence between (1) and (2) follows from Matsushima/Franke. The (conjectural) relationship between (1) and (3) is the problem of *reciprocity* in the Langlands programme. It consists of two directions; *existence* (i.e. constructing Galois representations from automorphic forms) and *modularity* (i.e. showing nice Galois representations are automorphic). Both of these directions are difficult. The passage from (5) $\Rightarrow$ (3) $\Rightarrow$ (4) is easy, whereas (4) $\Rightarrow$ (5) is basically the Fontaine–Mazur conjecture (not so easy). Actually even that last statement is not quite correct: if one knows that V is pure of weight one with non-negative Hodge–Tate weights and arises *up to twist* in the cohomology of some smooth proper algebraic variety, then it *should* actually arise in cohomology without having to twist and hence come from an Abelian variety; proving this, however, is probably hard (as in Standard conjectures hard). Note that (1) $\Rightarrow$ (5) follows, in the case of classical modular forms, from a geometric construction of Shimura, but that idea doesn't work here, and in fact this arrow is completely open and we shall say no more about it. In the particular case of imaginary quadratic fields, Harris, Soudry, and Taylor showed in 1992 that (1) $\Rightarrow$ (3) under certain favourable conditions. This case is slightly exceptional in this regard, since there exist functorial transfers of such π to GSp(4) which *do* contribute to the $(\mathfrak{q}, K)$-cohomology of Shimura varieties and hence can be directly related to the coherent cohomology of Shimura varieties (although not directly to Betti cohomology, because the corresponding weights are not regular.) As readers of this blog know, only very recently, Harris–Lan–Taylor–Thorne [HLTT16] established the same result for GL_n over a CM field. (Small lie: not all the desired

properties of the corresponding Galois representations, — i.e. local-global compatibility — have been established. I think Ila Varma is working this out for her thesis.)

It was observed early on that the cohomology groups of X are not, in general, torsion free. So what then does a torsion class represent? Computations by Grunewald, Helling, and Mennicke in an 1978 paper suggested that torsion classes (specifically, two-torsion with Hecke eigenvalues in $\mathbf{F}_2$) in these groups should be associated to $\mathrm{GL}_2(\mathbf{F}_2) = S_3$ Galois representations of the field F. Apparently there are even some unpublished notes from Grunewald in 1972 doing similar things, although I have only ever heard rumors of their existence (to be fair, I heard those rumours from Grunewald, so they're probably pretty reliable). So very early on there were hints that a further story was going on between torsion classes and Galois representations that wasn't immediately related to the (conjectural) story coming from automorphic forms. The first general and precise conjecture along these lines were formulated by Ash in his 1992 Duke paper, with further refinements by Ash–Sinnott, and Ash–Doud–Pollack. Unlike the previous speculations of Grunewald, these conjectures were precisely formulated and falsifiable, and in the spirit of Serre's original conjecture. Moreover, Herzig did actually come along and falsify them, by finding a more natural prediction for the set of possible Serre weights which turned out to be different from the formulation of Ash *et. al.*, and Herzig's formulation subsequently proved (numerically!) to give the right answer in these cases (**Edit:** see comments, this is not quite correct). At any rate, for quite some time, we have expected that mod-p torsion classes give rise to Galois representations, and following the conjectures of Serre, Ash, and others, one can be quite precise about exactly the local properties the corresponding Galois representations should have at primes of bad reduction. What is perhaps more recent is the idea that, especially for groups G with no discrete series, that torsion is not merely a technical nuisance, but rather is the source of "most" of the interesting Galois representations. In particular,

- The phenomenon whereby Galois representations coming from the countably many classical automorphic forms are *dense* in a suitable universal deformation ring (Böckle, Gouvêa–Mazur, Chenevier) will be totally false when G does not have discrete series. On the other hand, the representations coming from torsion should cut out *all* of the universal deformation ring.
- That in order to answer the most pressing questions concerning reciprocity (even in characteristic zero!) one needs to understand torsion classes.

For one take on this, I might suggest reading Section 1.1, Speculations on p-adic functoriality of [CM09]. For another interesting perspective, you should also read this, as well as the accompanying review.

So let us assume then that studying torsion representations and associating them to Galois representations is an Important Goal. How do we construct them? An observation also going back a long way (I believe to Harder??) is the following. Even though G may not admit discrete series, there may exist a group H containing a parabolic P with G as a Levi. If H *does* admit discrete series, then there will exist a Shimura variety X_H whose Borel-Serre compactification will have at least one boundary component which is a torus bundle over X_G, and as a result one obtains a map (with some mixing of degrees) $H^*(X_G, \mathbf{Z}) \to H^*(\overline{X}_H, \mathbf{Z})$. Now one is theoretically in better shape, because this map should be compatible (in some sense) with Hecke operators, and the latter group has a chance to admit comparisons

to étale cohomology groups which *do* come with Galois representations. There are three immediate problems:

(1) The compactification $\overline{X}_H$ will be singular, except in the case of modular curves.
(2) Given a long exact sequence for (co)homology relative to a boundary divisor, it's not clear whether the cohomology in the boundary ends up in H^{i-1} or H^i.
(3) Just because a class $[c]$ has an interesting Hecke eigensystem doesn't mean that that étale cohomology sees an interesting Galois representation.

The third issue is a genuine problem. If one has a Hecke eigenclass $[c] \in H^*(X)$ in the etale cohomology of a Shimura variety, even in characteristic zero, then all one can deduce is that the corresponding Galois representation is annihilated by the corresponding characteristic polynomial of Frobenius. But this is not always enough to get the *correct* Galois representation! It's probably worthwhile to consider two examples.

First, a somewhat degenerate example. Let $\mathbf{G} = \mathrm{GL}(1)/\mathbf{Q}$, $G = \mathbf{R}^\times$, $\mathbf{H} = \mathrm{SL}(2)/\mathbf{Q}$, and $H = \mathrm{SL}_2(\mathbf{R})$. Now X_G is (for some level structure) a finite set of points, and X_H is a modular curve with cusps. The boundary map realizes X_G as a set of cusps in X_H. The Hecke operators act on cusps by the degree map, i.e. $T_p[c] = (1+p)[c]$. This coincides with the action of Hecke operators on $H^0(X_H)$. So, in the etale cohomology group $H^0(X_H)$ we have a Hecke eigenclass which we imagine (looking at the Hecke operators) to be associated to the Galois representation $1 \oplus \chi$ where χ is the cyclotomic character. Yet $H^0(X_H)$ is one dimensional, and so we only see half of the Galois representation, namely, the trivial character. Now as it turns out, the other piece of the Galois representation *can* be seen in $H^1(X_H)$, which is now mixed because X_H is not projective (so the cuspidal part has motivic weight one, and this other piece of the Eisenstein series has weight two). So even in this trivial case, we see that a Hecke eigenclass in etale cohomology may have a less interesting Galois representation than the Hecke eigenvalues might suggest. From the Eicher–Shimura relation, we *do* get that the (trivial) Galois representation which does occur is annihilated by the characteristic polynomial of Frobenius $(\sigma - 1)(\sigma - \chi(\sigma))$, indeed it is annihilated by the first factor.

Second, let $\mathbf{G} = \mathrm{GL}(2)/F$, $G = \mathbf{GL}_2(\mathbf{C})$, $\mathbf{H} = U(2,2)/\mathbf{Q}$, and $H = U(2,2)$. Here $\mathbf{H}$ is taken to split over F. The cohomology of (a torus bundle over) the Bianchi group maps into the cohomology of $U(2,2)$. The characteristic polynomials of the Hecke operators are, morally, the following. If ρ is the (conjectural) Galois representation associated to an eigenclass on the Bianchi group, then $r = \wedge^2(\rho \oplus \rho^c)$ is a six dimensional (reducible) representation which is a direct sum $r = s \oplus \psi \oplus \psi^c$ for a four dimensional representation $s = \rho \otimes \rho^c$ and a Grossencharacter ψ and its conjugate (which are related to the central character of the original form and its conjugate). Now the characteristic polynomials of Frobenius on this Galois representation are, by Eichler–Shimura, the characteristic polynomials of Hecke on the image of this cohomology class in the cohomology $H^*(X_H)$. Without assuming one has ρ, one can phrase the above purely in terms of Satake parameters, but this way of saying it makes clearer what is going on, even though we don't know yet that ρ actually exists. If one could find the Galois representation r (and in particular s) inside the etale cohomology of X_H one would (almost) be done, but instead, the classes which actually turn up in etale cohomology in these degrees are the *reducible*

terms in r corresponding to the Grossencharacters rather than to the interesting representation s we are looking for. So as above, even in characteristic zero, one has the interesting Hecke eigenclass, but not the Galois representation.

These examples suggest that to understand what is going on we first need to get a better understanding of Shimura varieties. Most of the recent history of understanding Shimura varieties (and the Galois representations associated to automorphic forms) has concentrated on the cohomology arising from *cuspidal* automorphic representations. In this classical setting, the automorphic representations have a classical avatar as global sections of certain *coherent* bundles on X_H. (For example, classical modular forms of weight ≥ 1 are global sections of the line bundle $\omega^{\otimes k}$.) If we want to restrict to cusp forms, we can also take the corresponding extension of these sheaves to minimal (or toroidal, doesn't matter) compactifications which vanish appropriately at the boundary. If we denote these automorphic bundles by $\mathcal{E}_{\mathrm{sub}}$, then another way of saying this is that the action of Hecke operators $\mathbf{T}$ on

$$\bigoplus_{\mathcal{E}} H^0(\overline{X}_H, \mathcal{E}_{\mathrm{sub}})$$

is now understood if X_H is, for example, a Shimura variety of unitary type over a totally real field. Even getting this far is a somewhat monumental task that required, amongst other things, Ngo's work on the Fundamental Lemma, work of Kottwitz, Clozel, some large fraction of Jussieu, the work of Shin, and many more. In fact, as far as local-global compatibility goes, the ink is barely dry on the most recent work. Now we can at least state, in vague terms, the following:

Theorem 24.1. [Sch15b, IV.3.1]*: For (many) Shimura varieties X_H, the action of $\mathbf{T}$ on torsion classes in Betti cohomology factors through the action on coherent cusp forms in characteristic zero.*

Two examples: If X_H is the modular curve, then this says that the action of Hecke operators on $H^1(X_H, \mathbf{Z}/p^n\mathbf{Z})$ can be realized by the action on classical modular cuspidal eigenforms modulo powers of p. Given how we think about modular forms, this is *almost* tautological, because, by Eichler–Shimura, we can pass between cohomology classes and classical modular forms (in this case, we can even do this via the Hodge decomposition of H^1). However, there is a little wrinkle: we do see Eisenstein classes in Betti cohomology, and this theorem says that we can realize these as coming from cusp forms, so this result also implies that there exist cusp forms which are congruent to Eisenstein classes modulo p^n. Since we are ultimately interested in classes coming from the boundary of some compactification, we don't want to ignore this case. Still, it's not so difficult to prove.

If X_H comes from $U(2,1)/\mathbf{Q}$ (so it is a arithmetic complex hyperbolic manifold of real dimension 4, also known as a Picard modular surface), then we can look at the group $H_1(X_H, \mathbf{Z}_p)$. The characteristic zero classes here are known to correspond to endoscopic automorphic representations (and thus to not exist in the co-compact case) and are understood. However, unlike in the modular curve case, we no longer know that this group is torsion free, and in general, it may not be. So, a priori, all we know about the torsion classes and their Hecke operators is that there exists a Galois representation which is annihilated by the characteristic polynomial of T_p, using Eichler–Shimura. These polynomials are all of fixed degree (three in this case), but that doesn't give any lower bound on the dimension of this representation. This is an even more stark example of the well known phenomenon that Eichler–Shimura

is pretty much useless for constructing Galois representations outside the case of dimension two where knowing both the trace and determinant tells you a lot. For example, suppose you have an irreducible representation V of a finite group G in characteristic zero such that all the elements of g have a minimal polynomial of degree at most d: then you can't *a priori* bound the dimension of V! As an example, the extra-special 2-group of order 2^{1+2n} has a representation of dimension 2^n all of whose elements have images satisfying the degree two polynomials $x^2 - 1 = 0$ or $x^2 + 1 = 0$. So, before Scholze, we could not say anything about the dimensions of mod-p Galois representations arising from torsion in the first homology of $U(2,1)$. However, using Scholze, we can now deduce that any such representation comes from a classical cusp form, and hence must (in this case) have dimension three!

Comment 24.2 (Molesworth)**.** In fairness, I'm not sure that Ash etc claimed to have written down a complete list of weights, so it's a little strong to say that FH [Florian Herzig] falsified them.

Comment 24.3 (Persiflage)**.** You are correct: I just checked Ash–Doud–Pollack [ADP02]. They say "Note that the conjecture makes no claim of predicting all possible weights that yield an eigenclass with ρ attached."

25. Scholze on torsion, Part II

Wed, 19 Jun 2013

This post follows on from § § 23–24.

Section V.1: Today we will talk about Chapter V. We will start with Theorem V.1.4. This is basically a summary of the construction of Galois representations in the RACSDC case, which follows, for example, from work of Shin. We know a little bit more than this theorem states (namely, local-global compatibility).

Corollary V.1.7 is just the statement that the cohomology groups $H^0(X, \mathcal{E}_{\mathrm{sub}})$ for the sub-canonical extensions of automorphic sheaves $\mathcal{E}$ are computed by forms π whose transfer to $\mathrm{Res}_{F/\mathbf{Q}}\mathrm{GL}(n)$ are RACSDC. The sub-canonical extension corresponds to imposing a vanishing condition at the cusps. For example, the sub-canonical extension of ω^k on the modular curve is $\omega^k(-\infty)$. There is a nice action of the Hecke algebra $\mathbf{T}$ on this space, which is compatible with the associated Galois representations in all the expected ways (Satake parameters to Frobenius eigenvalues) at the unramified primes. So far, this is all classical (as of 2011).

Determinants: (See [Che14]) We will be using congruences to obtain Galois representations, but the information that gets glued is really the Hecke eigenvalues. So one wants a convenient way to pass from one to the other. The classical approach with modular forms is to remember the "standard" Hecke operators T_x which correspond to the traces of Frobenius. Knowing the trace is enough to determine a two dimensional representation away from characteristic 2, if one has residual irreducibility. This is the theory of pseudo-representations. Naturally enough, for larger dimensional Galois representations, it helps to remember more than the trace, namely, the entire characteristic polynomial. The corresponding theory was worked out by Chenevier. Namely, given an n-dimensional representation ρ of the group G over a commutative ring A, there is a map:

$$D : A[G] \to M_n(A) \to A$$

given by formally extending ρ in the obvious way and then composing with the determinant. For example, if $n = 2$, then

$$T(g) := D([g] + [1]) - D([g]) - D([1]) = \mathrm{Tr}(\rho(g)).$$

Now the map D has to satisfy a bunch of formal properties due to the constraints of coming from an n-dimensional representation. Writing all these down gives the correct notion of Chenevier's generalized "determinant." (Original paper here.) For those who like pseudo-representations, note that when $n = 2$, one can define D using the formula:

$$D(g) = \frac{T(g)^2 - T(g^2)}{2},$$

where T is the trace. So for $n = 2$ in characteristic greater than two, the notions are equivalent. And indeed Chenevier's notion of determinants is the same as a pseudo-representation whenever $n!$ is invertible, but is better behaved in small characteristics. Determinants satisfy the nice properties that pseudo-representations do, and that Galois representations sometimes don't (but do in the residually absolutely irreducible case), namely:

(1) You can glue determinants: $D \to A/I$ and $D \to A/J$ which agree on $A/(I + J)$ to get a determinant $D \to A/(I \cap J)$.
(2) Given a formal variable X, there is a natural determinant map $D : A[X][G] \to A[X]$ such that $D(1 - Xg)$ is the characteristic polynomial of g if the determinant comes from an actual representation.

Here I follow Scholze in using $\mathrm{Det}(I - X \cdot M)$ rather than $\mathrm{Det}(M - I \cdot X)$ as the definition of a characteristic polynomial — this is just a bookkeeping issue (the dreaded arithmetic versus geometric Frobenius). Returning to automorphic forms from coherent cohomology, since H^0 is torsion free, the module $\mathbf{T}$ is flat over $\mathbf{Z}$. Since the characteristic zero forms give rise to Galois representations coming from RACDSC forms, we naturally obtain a determinant map:

$$D : \mathbf{Z}_p[G_F] \to \mathbf{T}$$

such that $D(1 - X \cdot \mathrm{Frob}_x)$ is exactly as one would expect. (This is Corollary V.1.11). Note that the ring $\mathbf{T}_c$, which arises at this point, is just the inverse limit of the corresponding classical $\mathbf{T}$ over all p-power levels; this is defined in Chapter IV which we shall talk about later.

Segue on Completed Cohomology: I have to recall here a few basics about completed cohomology (one reference is [CE12].) I already know about completed cohomology (and so do many of my loyal readers) so I don't really feel obliged to say too much about it, but since most of you have been sent here from Quomodocumque, I will cough up a few pointers. The basic definition (for any congruence arithmetic manifold corresponding to a group $\mathbf{G}$) is as follows:

$$\widetilde{H}^i(X, \mathbf{Z}/p^n\mathbf{Z}) := \lim_{K \to} H^i(X(K), \mathbf{Z}/p^n\mathbf{Z}).$$

Here the limit is over shrinking compact open subgroups K of $\mathbf{G}(\mathbf{Z}_p)$. The tame level is fixed and can be included in the notation somewhere. One can also adorn the cohomology groups in the usual way, namely, by considering compactly supported cohomology. So what's the point of completed cohomology? Apart from having a natural action of $\mathbf{G}(\mathbf{Q}_p)$, which is always the type of group one wants to act on a candidate space for automorphic representations of any kind, a matter of experience and intuition suggested (to Matt and me) that it should be the "correct" space of

automorphic forms modulo p^n when $\mathbf{G}(\mathbf{R})$ does not have discrete series (and even when it does). One way to justify this is via the following four properties, the final one conjectural:

(1) The completed cohomology groups $\widetilde{H}^i(X, \mathbf{Z}/p\mathbf{Z})$ are co-finitely generated over $\Lambda = \mathbf{F}_p[[\mathbf{G}(\mathbf{Z}_p)]]$. This latter ring has nice properties, e.g. after shrinking the group $\mathbf{G}(\mathbf{Z}_p)$ slightly to get a powerful torsion free pro-p group, Λ is a local Noetherian ring which is Auslander regular (see Lazard [Laz65] and also Venjakob, which is [Ven02]).

(2) The Pontryagin dual groups $\widetilde{H}^i(X, \mathbf{Q}_p/\mathbf{Z}_p)^\vee$, which are finitely generated (by part one and Nakayama's Lemma and the usual long exact sequences) are **not** torsion $\Lambda = \mathbf{Z}_p[[\mathbf{G}(\mathbf{Z}_p)]]$-modules if and only if one is in middle degree and the corresponding real group admits discrete series (see [CE09]).

(3) The completed homology groups satisfy a Poincaré duality spectral sequence. The completed cohomology groups are compatible with the Hochschild–Serre sequence from which one can recover classical cohomology groups.

(4) Given a torus bundle, or more generally a nilmanifold, the completed cohomology disappears outside degree zero.

(5) **Conjecturally**: for any reductive algebraic group there will be a dominating term $\widetilde{H}^i(X)$ in degree $i = q_0$ which will have co-dimension l_0 as a Λ-module, where $2q_0 + l_0$ is the real dimension of X_G, and the degrees $[q_0, q_0 + 1, \ldots, q_0 + l_0]$ are exactly the degrees in which tempered automorphic representations contribute to cuspidal cohomology. More directly, l_0 for a semi-simple group is the rank of $\mathbf{G}(\mathbf{R})$ minus the rank of the maximal compact. For example, l_0 is equal to zero if and only if the real group admits discrete series. Hence this bullet point is a conjectural generalization of point (2). As an example, in the case of GL_2 over an imaginary quadratic field, the completed cohomology $\widetilde{H}^1(\mathbf{F}_p)$ should have codimension exactly one.

(For the last three points I'll refer you once again to [CE12].)

Section V.2: The key starting point, as mentioned last time, is that one can relate the cohomology of the group we are interested in — $\mathrm{Res}_{F/\mathbf{Q}}(\mathrm{GL}_n)$ — to the cohomology of Shimura varieties by realizing the first group as the Levi M inside a maximal parabolic P inside a group G corresponding to a Shimura variety. The first step is to compare the cohomology of what we are interested in (coming from the Levi M) to the cohomology of the boundary piece coming from the parabolic P inside G containing M. This is pretty standard: what happens is that the resulting space X_P which actually occurs in the boundary of the Borel–Serre compactification X_G^{BS} of X_G is a torus bundle over X_M. Well, not literally always a torus bundle, but rather a nilmanifold N coming from the unipotent part of P. The nilmanifold fibres spread the cohomology around by a Künneth type formula like a Frenchman expectorating over-oaked California Chardonnay into a spittoon. (Usually this fibration arises as a quotient from a fibration with a contractible fibre, which means that the cohomology really is just the derived product of the cohomology of the base and the cohomology of N, so it's not really so bad.) One way to avoid this mess is by passing to completed cohomology. On the boundary this has the effect of collapsing all the torus like directions in the nilmanold, and obtaining a map from

the completed cohomology of the arithmetic manifold corresponding to the Levi into the completed cohomology of the total space. Compare with equation (1.4) of this survey again.

Hecke Operators from M to G. One thing we have to understand is how to compute the Hecke operators at unramified primes on the completed cohomology of the boundary of X_G in terms of the action of the Hecke operators on the original object of interest X_M. Let us fix an unramified prime x which is prime to everything. To orient you, we are at the top of page 82 of Scholze. I'm going to be more prosaic in my notational choices and write T_G, T_P, and T_M for the local Hecke algebras at the prime x (Scholze does all the unramified primes at once). Yes, I know this is an abuse of notation, because here the groups G, the parabolic P and the Levi M are really the local versions at the prime x. (You will cope.) There are natural maps:

$$T_G \to T_P \to T_M.$$

Let's actually consider what these are in the case when M comes from GL(2) over an imaginary quadratic field F in which x splits, and G comes from $U(2,2)$ which also splits over F. So locally at x, the group G is just GL(4), and M is the Levi GL(2) × GL(2), and P is what it obviously has to be. In this case, we have isomorphisms:

$$T_G \simeq \mathbf{Z}_p[X_1^{\pm}, \dots, X_4^{\pm}]^{S_4}, \qquad T_P \simeq \mathbf{Z}_p[Y_1^{\pm}, Y_2^{\pm}]^{S_2} \times \mathbf{Z}_p[Z_1^{\pm}, Z_2^{\pm}]^{S_2}.$$

Perhaps we are required to adjoin $\sqrt{x}$ to both sides in order to normalize this appropriately. Consider it done. Now the map $T_G \to T_M$ is the one sending $(X_1, X_2, X_3, X_4) \mapsto (xY_1, xY_2, x^{-1}Z_1, x^{-1}Z_2)$. The choice here must be coming from the choice of M (for a fixed torus) corresponding to a choice of subgroup $S_2 \times S_2$ of the Weyl group. One can write down analogous formulas for the inert and ramified primes. The corresponding maps of Satake parameters indicates that the if our original eigenclass has a Galois representation ρ, then the Hecke eigenvalues of the class which has been pulled back is associated to $\rho^\vee \oplus \rho^c$. (**Edit:** In the previous version I omitted the dual. Note that $\rho^\vee \det(\rho) = \rho$ for $n = 2$. **End Edit**) Now this statement seems to be somewhat in conflict with my previous post, where I claimed that the action of the Hecke algebra on the cohomology of $U(2,2)$ corresponded to the Galois representation $\rho \otimes \rho^c$. This is because of a subtlety which I think I can explain. Suppose you start from a classical modular form f and base change it to a Hilbert modular form f_E over a real quadratic extension. Then the corresponding map of Satake parameters is just the obvious one corresponding to the restriction of the Galois representation. In particular, if α, β are the Satake parameters of a local unramified component π_x of f, and if x splits in the quadratic field and y is a prime above x, then π_y of f_E will have the same Satake parameters, and f_E will have the same Hecke eigenvalue for T_y that f has for T_x. However, the actual Galois representation occurring inside the etale cohomology of the Hilbert modular surface is *not* the restriction of the Galois representation to E, but rather the (four dimensional) tensor induction. This also reflects an important point: we will *not* be finding the desired Galois representation inside etale cohomology (which, apparently by an argument of Clozel and Harris, is impossible), but rather we will simply be "following the Hecke eigenclasses." In this context, for example, cuspidal automorphic representations for $U(2,2)$ contain all the information for the associated four-dimensional representations, but the ones occurring in cohomology are (tensor inductions!) of $\wedge^2$. That is why in this post we see the Hecke eigenvalues as looking like the direct

sum $\rho^\vee \oplus \rho^c$, whereas the action on cohomology via Eichler–Shimura looked like $\wedge^2(\rho^\vee \oplus \rho^c)$, which contains $\rho \otimes \rho^c$ up to twist.

The arguments on the lower half of page 82 are just related to the fact that the boundary of the compactification on X_G can have a number of components, and these components can have their own boundary, and so on. If one takes the case where X_G corresponds to $U(2,2)$ over an imaginary quadratic field, then the only boundary components are (torus bundles over) Bianchi manifolds X_M, and the only boundaries that they have are hyperbolic cusps. In particular, in this case, using remark (4) on completed cohomology above, the completed cohomology of the boundary $\widetilde{H}^k(\partial X_G)$ (denoting $X_G^{BS} \smallsetminus X_G$ by ∂X_G), is given by

$$\widetilde{H}^k(\partial X_G, \mathbf{Z}/p^n\mathbf{Z}) = \mathrm{Ind}_{\mathbf{P}(\mathbf{Z}_p)}^{\mathbf{G}(\mathbf{Z}_p)} \left(\widetilde{H}^k(X_M, \mathbf{Z}/p^n\mathbf{Z}) \right).$$

So we are interested in the Hecke action on the right hand side, which we have now transferred to the left hand side. (Of course, the local Hecke algebras combine by taking tensor powers to get the Hecke algebra at all unramified primes, which surjects onto the corresponding global Hecke algebras $\mathbf{T}_G$ and $\mathbf{T}_M$.) There is a natural long exact sequence of completed cohomology associated to a manifold with corners as follows:

$$\ldots \to \widetilde{H}^{k-1}(X_G, \mathbf{Z}/p^n\mathbf{Z}) \to \widetilde{H}^{k-1}(\partial X_G, \mathbf{Z}/p^n\mathbf{Z}) \to \widetilde{H}_c^k(X_G, \mathbf{Z}/p^n\mathbf{Z}) \to \widetilde{H}^k(X_G, \mathbf{Z}/p^n\mathbf{Z}) \to \ldots$$

So to get a Galois representation (or, to begin with, a determinant) on $\widetilde{H}^{k-1}(\partial X_G)$, we can start by finding determinants for the two surrounding terms.

Special Case: Let's continue discussion the special case where X_M is a Bianchi manifold, and X_G comes from $U(2,2)$ which splits over the corresponding imaginary quadratic field. The key term of interest will be (for the Bianchi manifold) $\widetilde{H}^1(X_M)$ or, equivalently, $\widetilde{H}^1(\partial X_G)$. In fact, by Hochschild–Serre, the completed cohomology $\widetilde{H}^1(X_M)$ captures all the interesting Hecke actions coming from torsion in Bianchi groups *as long as* one localizes away from the Eisenstein primes coming from the cusps. The cusps in the Borel–Serre compactification of the Bianchi group are elliptic curves with CM by the underlying imaginary quadratic field. The difference between the classical classes in H^1 and $\widetilde{H}^1$ proved themselves to be a real pain in my book with Akshay, because when one wants a numerical correspondence, one can't ignore Eisenstein terms. Yet blessedly, in this context, we can localize away from them. Hence the key terms are those in the following boundary exact sequence:

$$\widetilde{H}^1(X_G) \to \widetilde{H}^1(\partial X_G) \to \widetilde{H}_c^2(X_G)$$

Let's consider the first term. The group $U(2,2)$ has real rank two. In particular, by super rigidity, any non co-compact lattice in $U(2,2)$ will have the congruence subgroup property. It follows that $\widetilde{H}^1(X_G)$ is trivial! The point is that if all the finite quotients of a lattice in $U(2,2)$ come from congruence quotients, then pulling back over all such quotients kills everything. Actually, this is not strictly correct, because completed cohomology only pulls back over p-power quotients, and there may be cohomology coming from the tame level. However, it is easy to see (by Hochschild–Serre) that any such cohomology will be Eisenstein. In particular, after localizing at a non-Eisenstein (in the appropriate sense) ideal, we get an *injection* from $\widetilde{H}^1(\partial X_G)$ to $\widetilde{H}_c^2(X_G)$, and thus from Theorem IV.3.1, we obtain a determinant to the Hecke algebra of $\widetilde{H}^1(X_M)$ (localized away from Eisenstein ideals) without any need to quotient out by an ideal with fixed zero power as in Corollary V.2.6.

I don't think this trick will really work in any other examples, however, since it's very hard to say anything in general about H^2. (There is recent work on on stable completed co/homology here, but that will never be enough to give something useful in this context.)

General Case: The general case is now quite similar, except know to understand $\widetilde{H}^k(\partial X_G)$ one needs to understand *both* boundary terms. There is also going to be some loss of information coming from the corresponding extension class. If one had determinants on $\widetilde{H}^*(X_G)$ and $\widetilde{H}^*_c(X_G)$, then one would immediately get Corollary V.2.6 with an ideal I with $I^2 = 0$. However, Theorem IV.3.1 (which is being invoked here) only applies to $\widetilde{H}^*_c(X_G)$. Now $\widetilde{H}^*(X_G)$ is related to its compact cousin by a Poincaré duality spectral sequence, but this will once again spread out some terms and necessitate replacing $I^2 = 0$ by some power involving the dimension. At any rate, while there is room for improvement in general, there is still the fundamental problem (mentioned in part zero!) of controlling whether this boundary cohomology is going forwards or backwards in the long exact sequence above (or worse, being mixed). I'm going to give some heuristics next time on what one expects should happen (short answer: after localizing at a nice maximal ideal, it should work out as well as the Bianchi case, but that will be hard to prove.)

Note that Scholze actually works with classical cohomology here, and then relates it back to completed cohomology using Hochschild–Serre on p.86. The point in either argument is that all the terms in the spectral sequence (on every page) are, by Theorem IV.3.1, modules for the Hecke ring $\mathbf{T}_c$ which acts on coherent cohomology. Hence the limit terms have filtrations by a fixed bounded number of such objects.

Next time, I'll say a little more about how one might expect the "simplification" in the Bianchi case above to apply more generally, and I'll talk about the final section **V.3** of chapter **V**, in which we extract the n-dimensional representations from our $2n$-dimensional determinants.

26. Scholze on torsion, Part III

Sat, 22 Jun 2013

This post follows on from § § 23–25. Before I continue along to section **V.3**, I want to discuss an approach to the problem of constructing Galois representations from the pre-Scholze days. Let's continue with the same notation from last time, where X_M is the symmetric space whose cohomology is of interest, and $X = X_G$ is the Shimura variety with Borel–Serre compactification X^{BS} whose boundary contains (simplified assumption: is) a generalized torus bundle X_P over X_M. If we localize at a "non-Eisenstein" ideal, then the completed cohomology groups $\widetilde{H}^n(X_M)$ should vanish outside a single degree q_0. For this discussion, let us define non-Eisenstein classes to be those which do not occur in degrees $\leq q_0$ in $H^*(X_M)$. By Hochschild–Serre, any cohomology class in lowest degree (after localization) always survives in the completed limit, so even if one doesn't assume the expected vanishing in higher degrees, the module $\widetilde{H}^{q_0}(X_M)$ will contain all the information about the classes in H^{q_0} at classical level after localization. Hence, to obtain the desired Galois representations for these classes, one wants to prove:

(1) The vanishing of $\widetilde{H}^{q_0-1}(X^{BS})$ after localization.

(2) There are Galois representations (of the correct form) associated to classes in $\widetilde{H}_c^{q_0}(X^{BS})$.

The hope was that one could try to prove this via the following idealized argument. There is a spectral sequence:

$$\mathrm{Ext}^i(\widetilde{H}_j^{BM}, \Lambda) \Rightarrow \widetilde{H}_{d-i-j},$$

where $d = 2 \cdot \dim(X) = 2n$ is the real dimension of the Shimura variety X. There is an identical sequence with the roles of completed homology and completed Borel-Moore homology reversed. Note that the completed homology groups are (Pontryagin dual) to the cohomology groups, which relates compactly supported cohomology to homology and cohomology to Borel-Moore homology. The non-commutative Ext groups in the spectral sequence vanish for any value of i that is less than the co-dimension of the corresponding module. Recall from last time that $\widetilde{H}_j^*$ is torsion except for the middle degree $j = n$. Now suppose that one can show that the completed homology groups $\widetilde{H}_j^*$ have *sufficiently large* co-dimension outside the middle degree. Then from these bounds (and from trivial bounds on the cohomology of the boundary) the spectral sequence should degenerate, and one should have isomorphisms of the following form (after localization):

$$\widetilde{H}_{n-i} = \mathrm{Ext}^i(\widetilde{H}_n^{BM}, \Lambda),\ i \leq n, \quad \widetilde{H}_{n+i} = 0,\ i \geq 0, \quad \widetilde{H}_n^{BM} = \mathrm{Hom}(\widetilde{H}_n, \Lambda).$$

(To recall, even though we are localizing at an ideal whose avatar on $H^*(X_M)$ is maximally non-Eisenstein, the corresponding ideal on $H^*(X_G)$ *will* be Eisenstein.) From these equalities, we see that to understand the action of the Hecke operators on completed cohomology, we are reduced to understanding the action on the completed cohomology in *middle* degree, which we know to be a module of positive rank and hence (even after localization) contain many cusp forms which are *known* to have interesting Galois representations. At the very least, this would prove the existence of the residual Galois representations associated to such a non-Eisenstein ideal $\mathfrak{m}$. The approach I am outlining here is the one in the (currently non-existent) paper that Matt and I had planned to write. Let's suppose that one attempts to apply this approach in the Bianchi case. There's no issue in defining Eisenstein classes here, since the classes that occur in $H^0(X_M)$ are easy to understand, and $q_0 = 1$. So the first step in the above program is to show that $\widetilde{H}^1(X^{BS})$ vanishes, at least if we pass to finite tame level. As we noted last time, this follows from the congruence subgroup property which is known because $U(2,2)$ has real rank two and the corresponding lattice in this group is (obviously) not co-compact. Here the Shimura variety has complex dimension four. So one *only* has to show that $\widetilde{H}_j$ is small for $j = 2$ and $j = 3$. In particular, one wants, explicitly, that:

$$\mathrm{codim}(\widetilde{H}_2) \geq 4,\ \mathrm{codim}(\widetilde{H}_3) \geq 3$$

The dimension of $\Lambda = \mathbf{Z}_p[[G]]$ is, for reference, $1 + \dim \mathrm{SL}_4(\mathbf{Z}_p) = 16$. As noted previously, we know that these cohomology groups are torsion and so have co-dimension at least one. The proof of this result ultimately relied on facts concerning the growth of spaces of automorphic forms. *However*, it is impossible to determine anything further about the codimension by naïve automorphic considerations, because already Λ/p has co-dimension one but no characteristic zero points. So, to prove this conjecture, one *really* needs to understand the torsion in the cohomology of Shimura varieties. This was where, basically, we were stuck. Note that even understanding

$\widetilde{H}^1$ in this case took a powerful result. Understanding $\widetilde{H}^2$ is already much harder. As the real rank increases, it *won't* be the case that such completed cohomology groups *completely* disappear, since there will exist not only trivial stable classes in characteristic zero, but also exotic torsion classes which will be related to K-theory and regulators (as can be seen [Cal15]). One implication of our conjectures (as noted above) is that the completed cohomology groups vanish for Shimura varieties above the middle dimension. Scholze proves this! (**IV.2.3**). However, he *doesn't* prove it by showing that the $\widetilde{H}_j$ are small for small j, and instead deduces a (weaker form) of such an estimate in reverse. I think it's an interesting problem to understand $H^2(\Gamma, \mathbf{F}_p)$ for groups where the only characteristic zero classes are invariant under G, in both the stable and non-stable range. The first case I mentioned previously, and there is something in this direction (in the second case) in [CV19, §4.5].

Section V.3 OK, continuing on from last time, we now have a determinant of dimension $2n$ with image in $A_0/I = \mathbf{T}/I$ for some ideal I with $I^m = 0$ for an integer m which only depends on $\dim(X)$. The goal is now to extract an n-dimensional determinant, i.e., to recover ρ from $\rho^\vee \oplus \rho^c$. Of course, the idea is not to do this from simply one class, but rather allowing twisting, so that we also know $r_\psi = \rho^\vee \det(\rho)\psi^{-1} \oplus \rho^c\psi^c$ for some Hecke character ψ. We may as well take ψ to be a collection of characters of $\mathbf{Q}$, so that $\psi^c = \psi$.

Let's first make some simplifying assumptions, namely, that the ideal $I = 0$, that we are in characteristic zero, and that the image of r is through a finite group G, and the image of all the twists factors through the group $\Gamma := G \times \mathbf{Z}$ where ψ is a finite order character of the second factor, and $\psi^2 \neq 1$. We would like to imagine that there are equalities:

$$r = W =^? U \oplus V, \quad r_\psi = W_\psi =^? U\psi \oplus V\psi^{-1}.$$

Because the two factors of Γ commute, it follows that $[\psi \otimes W_\psi] - [W]$ is a virtual character of Γ. Evaluating this character on the pairs $G \sim (g, 1) \subset \Gamma$ defines a class function on G. Normalizing by $\psi^2(1) - 1 \neq 0$, this class function applied to Frob_x is the sum of the Satake parameters at x corresponding to U, and we deduce that $[U]$, and hence also $[V]$, are virtual characters (with rational coefficients) of G. It now suffices to promote $[U]$ to an actual character. The virtual characters $[U]$ and $[V]$ tautologically promote to virtual characters of Γ which decompose under the second factor into trivial representations. It follows that $[U\psi]$ and $[V\psi^{-1}]$ are (rational) sums of irreducible representations which decompose under the second factor as direct sums of the representation ψ or ψ^{-1}. Assuming that $\psi \neq \psi^{-1}$, there can be no cancellation in $[U\psi] + [V\psi^{-1}]$, from which it follows that $[U]$ is already an actual character.

In general one has to modify this argument to work more integrally as well as to be compatible with the ideal J. As I told Toby Gee, "without having looking at this yet, it must essentially be trivial." So, if you are like me, you can just ignore the following which took me a non-trivial number of hours to work out:

(1) We take the characters ψ to be characters of $\mathrm{Gal}(\mathbf{Q}(\zeta_{\ell^\infty})/\mathbf{Q})$ of ℓ-power order, where ℓ is prime to two and p and anything else inconvenient including the ramified primes. This auxiliary prime may vary.
(2) Since we are going to allow ψ to have order some arbitrarily large power of various primes, it is convenient to extend scalars to $A = A_0 \otimes W(\overline{\mathbf{F}}_p)$. Here A_0 is the Hecke algebra acting with coefficients modulo some fixed power

of p. It's useful to work with both rings, however, because whilst A accepts characters of all orders, A_0 is literally a finite ring, which is convenient for finiteness arguments. We would like to show that the twisted determinants corresponding to r_ψ have values in A/I_ψ for *the same* A. This amounts to showing that, at the level of our original locally symmetric quotient X_M, we can twist by a sufficiently nice character ψ and not change the Hecke algebra, except for extending scalars. This is straightforward, and is what is going on at the top of page 88.

(3) If we have two determinants with a pair of corresponding ideals I and I_ψ with $I^m = I_\psi^m = 0$, then clearly $\widetilde{I} = I + I_\psi$ satisfies $\widetilde{I}^{2m-1} = 0$. So, at the cost of increasing the nilpotentcy, for any character ψ, we get two determinants with values in $A/\widetilde{I}$. Note that if I and I_ψ are both trivial, then so is $\widetilde{I}$.

(4) We would *also* like the ideal $\widetilde{I}$ to be independent of ψ. Actually, we don't need this, it will suffice to note that we can take $\widetilde{I} \cap A_0$ to be independent of ψ. Because A_0 is finite, there are only finitely many such ideals, and so we can take one that occurs for infinitely many primes ℓ and infinitely many of the corresponding characters ψ.

(5) For any fixed character ψ, our determinant (which has twice the required dimension) will be defined on a finite quotient of

$$\Gamma := \mathrm{Gal}(L_\infty/F) = \mathrm{Gal}(L/F) \times \mathrm{Gal}(F_\infty/F),$$

where L/F is finite and L_∞, F_∞ are the pro-ℓ cyclotomic covers of L, F respectively. This should hopefully look similar to our simplified problem in characteristic zero. We have two determinants D and D_ψ with the property that the characteristic polynomial of Frobenius Frob_x (which exists for determinants) is:

(a) Of the form $P_x^\vee(X)P_x^c(X) \mod I$ for D.
(b) Of the form $P_x^\vee(X/\chi(g))P_x^c(X\chi(g)) \mod I_\psi$ for D_ψ.

These polynomials $P_x^\vee(X)$ and $P_x^c(X)$ are what they obviously should be, namely, the polynomials with inverse roots given by the appropriate Satake parameters. (Or more accurately, with coefficients given by the appropriate Hecke operators.) Because these are determinants, these products are locally constant on the group Γ because they are coming from honest Galois representations of rank $2n$. We would like to decompose these into products of two determinants of rank n. In the characteristic zero case, we took a character ψ such that $\psi^2(1) - 1 \neq 0$ and used this as a fulcrum on which to tease out the representation U. Here we do something similar. A first step is to show that each of the four polynomials above is locally constant. We choose an element $1 \in \mathrm{Gal}(F_\infty/F)$ and a deep enough character χ so that $\chi^{2m}(1) - 1 \neq 0$ for all $m = 1, \ldots, n$. We now find an open neighbourhood of $(G, 1)$ where D and D_χ are constant. Let $a(x)$ be the linear term of $P_x^\vee(X)$, and let $b(x)$ be the linear term of $P_x^c(X)$. Then we deduce that the following two terms are locally constant:

$$a(x) + b(x), \quad a(x)\psi(x) + b(x)\psi^{-1}(x).$$

So, because $\psi^2(x) - 1 \neq 0$, we deduce that $a(x)$ and $b(x)$ are locally constant, and so $a(x) \mod \widetilde{I} \cap A_0$ is also locally constant. Given this, one proves that

the quadratic terms are also locally constant in the same manner, and by induction one has the result for the entire polynomial. Thickening the open neighbourhood up, one proves the same result for the entire group Γ minus the piece coming from $\mathbf{Z} \sim \mathrm{Gal}(F_\infty/F)$, which gives us Lemma **V.3.4**. Then by choosing a different auxiliary prime ℓ', one patches to get a well defined class function on G in Lemma **V.3.5**.

(6) So now we have a class function D on the Galois group $\mathrm{Gal}(L/F)$ with values in characteristic polynomials (now of the right dimension!) in A_0/I (dropping the tildes), and we want to promote it to a genuine determinant. Of course, over finite rings we can't use the language of virtual characters. What Scholze does next is use the fact that we have such a decomposition for infinitely many different characters in order to glue enough of them together to obtain a determinant map

$$D : A[G \times \mathbf{Z}] \to A[t]/I, \quad D(1 - X\gamma^k g) = P_g^\vee(X/t^k)P_g^c(Xt^k),$$

where γ is a generator of $\mathbf{Z}$ and I has the expected properties of nilpotence. This consists of Lemmas **V.3.6** and **V.3.7**.

(7) Now we are at Lemma **V.3.8**. Bugger it, this is taking a long time, and quite possibly nobody is interested in these specific details. Let me cut some corners and replace determinants by pseudo-representations. We deduce from the above that we are in the following situation: we have a degree $2n$ pseudo-representation:

$$T : G \times \mathbf{Z} \to A[t], \qquad T(g, m) = a(g)t^m + b(g)t^{-m}.$$

We want to deduce that $a(g)$ and $b(g)$ are both pseudo-representations of degree n. We are allowed to use the fact (which is obvious) that $a(g)$ and $b(g)$ are *not* pseudo-representations of degree strictly less than n. (Actually, is it obvious? It's certainly obvious for $n = 2$ that $a(g)$ and $b(g)$ are not a character. So let's assume $n = 2$. Ah, I see by passing to the trivial element we can compute that $a(1) = b(1) = n$, so it is obvious.) Now, if we abstract slightly and drop any knowledge about $a(g)$ and $b(g)$ other than they are class functions, the best we can hope to prove is that $a(g)$ and $b(g)$ are both pseudo-representations of degrees A and B respectively, where $A+B = 2n$. This is what we do. Since T is a pseudo-representation of degree $2n$, we have the following identity:

$$\sum_{S_{2n+1}} (-1)^d T_\sigma(g_i, m_i) = 0.$$

In fact, this identity on class functions characterizes pseudo-representations of degree *at most* $2n$, the only other information coming from evaluating on the identity. Suppose we take the m_i to be sufficiently generic integers so that all the sums $\sum \pm m_i$ are distinct. Now let us partition the m_i into two sets M_A, M_B of cardinality $A + 1$ and B respectively, where $A + B = 2n$.

Consider the coefficient of t^C in the sum above, where we take

$$C := \sum_{M_A} m_i - \sum_{M_B} m_i$$

The corresponding coefficient must vanish. Moreover, because of the way that the m_i were chosen, we know exactly what terms can arise with this coefficient: explicitly, the terms in M_A must come from $a(g)$, and the terms in M_B must come from $b(g)$. Hence we recognize the coefficient to be (up to sign)

$$\left(\sum_{S_{A+1} \curvearrowright M_A} (-1)^d a_\sigma(g_i) \right) \left(\sum_{S_B \curvearrowright M_B} (-1)^d b_\sigma(g_i) \right).$$

We deduce that, for any decomposition $A+B=n$, *either* $a(g)$ is a pseudo-representation of degree at most A, or $b(g)$ is a pseudo-representation of degree at most $B-1$. Taking B to be the smallest integer for which $b(g)$ is a pseudo-representation, we deduce the result (such an integer exists because $b(g)$ is at least a degree $\leq 2n$ pseudo-deformation). We are, mercifully, done. Looking at Scholze, I think this lemma (and even roughly the argument) is quite similar to the proof of Lemmas **V.3.8-V.3.15** but this is much easier, at the cost of assuming that $p \geq n$.

It looks as though one can probably skip step 6 simply by choosing the value of $t \sim \psi(1)$ to generate a sufficiently generic extension of A_0 inside A, although I guess that's how one does step 6 anyway.

Section **V.4** is just a matter of putting things together. Next time: onto Chapter **IV**!

27. Scholze on torsion, Part IV

Sat, 29 Jun 2013

This is a continuation of § § 23–26.

I was planning to start talking about Chapter **IV**, instead, this will be a very soft introduction to a few lines on page 72. At this point, we have reduced the problem of constructing Galois representations for torsion classes on a wide class of locally symmetric spaces to the equivalent problem for Shimura varieties. Naturally enough, the Shimura varieties which arise in this context will not be projective. However, the problem of attaching Galois representations to Hecke actions on $\widetilde{H}^*_c(X)$ is still a very interesting one in the *compact* case. The difficulties that arise in the non-compact case are somewhat orthogonal to the issue of constructing Galois representations, so I don't think much is lost (at this point) in considering the compact case. (MH tells me that one of the main ingredients for dealing with issues concerning the boundary may well be the **Hebbarkeitssatz, II.3**.) A good case to keep in mind are the simple Shimura varieties of Kottwitz-Harris-Taylor type, and even the simple case of ball quotients coming from $U(2,1)$ will be of interest. Honestly, even the case of modular curves will be of interest. Modular curves are not compact, of course, but this is the one non-projective case in which the minimal and toroidal compactifications coincide and are smooth, so the boundary causes (relatively) little difficulty. A related problem is to understand the action of Hecke operators on

torsion in *coherent* cohomology. In some sense, Scholze *reduces* the problem to this case, so we shall begin by considering this problem. Note that already in this case the problem is no longer trivial even for classical modular curves, where one may have torsion in $H^1(X, \omega)$.

Coherent Cohomology: Let $\mathcal{E}$ be an automorphic vector bundle on X. Suppose that X is smooth over $\mathbf{Z}_p$, so that it makes sense to impose some nice integral structure on $\mathcal{E}$, and hence to consider the coherent cohomology groups:

$$H^*(X, \mathcal{E}/p)$$

If X is non-compact, then denote (also by $\mathcal{E}$) the sub-canonical extension to a smooth toroidal compactification. This cohomology group has a natural Hecke action. How does one construct Galois representations associated to the Hecke action this object? Let's consider the first non-trivial case, where X is a modular curve and $\mathcal{E} = \omega$. There's no problem understanding H^0, because (via the Hasse invariant) this will be related to classical spaces of modular forms, so the problem is to understand H^1. The first step is to understand what H^1 is as a vector space. To compute the cohomology of a projective curve, we can take a covering by (two) affines and compute Cech cohomology. To do this, we first need to find two affines. In anticipation of having something sufficiently natural in order to understand the action of Hecke, we let S denote the supersingular locus and $U = X \smallsetminus S$. For now let's let the other affine be V. Then the Cech complex is the following:

$$H^0(U, \omega) \oplus H^0(V, \omega) \to H^0(U \cap V, \omega)$$

Here U is the ordinary locus. The space $H^0(U, \omega)$ is the space of ordinary modular forms, and we may relate the Hecke action on this (infinite dimensional) space to the Hecke action on classical modular forms by noting that:

(1) For any section $c \in H^0(U, \omega)$, there exists a power s^n of the Hasse invariant s such that $s^n \cdot c$ extends to $H^0(X, \omega^m)$ for some integer m.
(2) The ordinary locus U is preserved by Hecke operators, and moreover multiplication by the Hasse invariant s is *Hecke equivariant*.

The problem is that it's hard to find a *second* open affine V which is preserved by Hecke, let alone admits an analogue of the Hasse invariant. In this case, we can instead do the following. Take V to be an *infinitesimal* neighbourhood of S, (that is, the completion of X along S). Then V is stable by Hecke. Imagine for convenience that there is only one supersingular point. The cohomology $H^0(V, \omega)$ of V has a filtration by the order of vanishing at (each) supersingular point, the first piece consisting of simply functions $H^0(S, \omega)$ on the supersingular point. There exists a section B^{p-1} (see [Edi92, Prop 7.2]) which *is Hecke equivariant*. This approach is used Emerton/Reduzzi/Xiao to construct Galois representations for torsion classes in the coherent cohomology of Hilbert modular varieties (Note that one would also want these representations to satisfy certain local properties at the prime p, which is more subtle in general, but has been done at least for modular curves at least in the residually irreducible case in my paper [CG18a] with David.) If one thinks about applying this method in the general case, there are two obvious issues. The first, which is perhaps not impossible to overcome, is that one needs to construct a suitable stratification of the Shimura variety by pieces which one understands and for which one can construct suitable Hasse-invariant type sections which allow one to pass to very ample sheaves whose cohomology vanishes, and hence reduce the problem to degree zero. The second is that, at least in the context of Scholze,

one is working at a level which is very ramified at p. Certainly all of the discussion above was predicated on X having *good integral models* at the prime p. It's easier to find good integral models when the corresponding Shimura variety is smooth! At level $X(p^n)$, there do exist integral models (obviously no longer smooth). It's convenient to assume that the open modular curves $X(p^n)$ are projective, because the issues at the cusps are orthogonal to what is happening here. So what do they look like? Well, they are proper and flat, which is nice. The general problem to the construction is that the torsion subgroup $E[p^n]$ of an elliptic curve E is no longer etale (and so certainly not locally isomorphic in the etale topology to $(\mathbf{Z}/p^n\mathbf{Z})^2$), but it is at least finite flat of rank p^2. So all one needs to do is to impose enough extra structure on the finite flat group scheme in order to recover the correct object on the generic fibre and yet have enough points in the special fibre. Katz–Mazur do this by considering a so-called "Drinfeld basis"

$$\phi : (\mathbf{Z}/p\mathbf{Z})^2 \to E[p^n]$$

where there is a corresponding equality of Cartier divisors (see **3.1.2** of KM). In particular, given a point x_n one gets a level structure $P_n, Q_n \in E[p^n]$ given by the image of the two generators.

So how does one understand the tower of varieties $X(1) \leftarrow X(p) \leftarrow X(p^2) \ldots$, either integrally or even just on the generic fibre? The ordinary locus up the tower is easy to understand. Let's first consider the rigid analytic varieties corresponding to the generic fibre. There are sections $X^{\mathrm{ord}}(1) \to X^{\mathrm{ord}}(p^n)_\infty$ from the ordinary locus to the component of the ordinary locus containing infinity, because, for ordinary elliptic curves, we still have etale locally a canonical isomorphism $E[p^n] = \mathbf{Z}/p^n \oplus \mu_{p^n}$, giving an appropriate trivialization. Moreover, the the action of $\mathrm{GL}_2(\mathbf{Z}_p)$ is transitive on the cusps, and so one sees all of the ordinary locus in this way. Thinking more integrally, we can see more directly from Serre–Tate theory that (for all points) at level one the completed local rings will be smooth. However, because $\mathbf{Z}/p^n \oplus \mu_{p^n}$ does not admit any deformations, the covering maps will be smooth at ordinary points and so the complete local rings at any ordinary point will remain smooth. It follows that the *interesting* geometry will be taking place over the supersingular discs. One can try to understand what is happening by looking at the corresponding completed local rings at supersingular points. Suppose one takes a compatible sequence of supersingular points (in the special fibre) in such a tower. The base point corresponds to a supersingular elliptic curve E_0 over $\mathbf{F}_p$ which has a corresponding formal p-divisible group G_0, now of height two. What Weinstein teaches us is that whilst the completed local rings A_n of x_n on $X(p^n)$ will be hard to understand, there is still hope to understand the completion

$$A = (\widehat{\lim_{\to} A_n})$$

over the ring $\mathcal{O}_K$, which is the completion of $W(\zeta_{p^\infty})$. By universality, the Drinfeld level structure gives rise to two parameters X_n, Y_n in A_n which lie inside the maximal ideal. The Weil pairing (we've added a consistent sequence of roots of p-power roots of unity) gives a relation of the form $\Delta_n(X_n, Y_n) = \zeta_{p^n}$. Jared shows that these are essentially all the relations in the limit ring A, which thus has a very nice description. We will come back to this example, because I suspect that understanding this result will be important.

The Lubin–Tate tower There's also a local analogue of this picture, namely the Lubin–Tate tower. Recall that the Lubin–Tate space M_0 is the universal deformation ring of a commutative height h formal group G_0 over $k = \mathbf{F}_p$, where $h = 2$. It turns out that M_0 is smooth of relative dimension $h-1$ over the Witt vectors $W(k)$. The smoothness is the "same" as the smoothness of the modular curve of level one at a supersingular point. It makes sense to consider level structures in the Lubin–Tate context also, where now the nth layer M_n of the Lubin–Tate tower consists of triples (G, ι, α) with Drinfeld level structure, as in the Katz–Mazur model. Quite explicitly, the K-points are given as follows:

(1) G is a formal group over $\mathcal{O}_K$,
(2) G is a deformation of the height h formal group G_0 over k, and $\iota : G_0 \to G \times k$ is an isomorphism,
(3) $\alpha_n(\mathbf{Z}/p^n\mathbf{Z})^h \to G[p^n]$ is an isomorphism.

If we go up the entire tower, there is a natural action of $\mathrm{GL}_h(\mathbf{Z}_p)$ in the limit. If D is the corresponding division algebra, then there is an action of $\mathcal{O}_D^\times$ on (each) piece of the tower, given by replacing G by a prime-to-p isogeny. In order to have richer actions of $\mathrm{GL}_h(\mathbf{Q}_p)$ and $D^\times$ on this tower (not only on the cohomology) it makes sense to modify it slightly (while enlarging the component group in a way that doesn't change the intrinsic geometry) by considering a trivialization of the *rational* Tate module $\alpha : (\mathbf{Q}_p)^h \to V(G)$. Here we now consider deformations *up to isogeny*, although we remember a quasi-isogeny on a nilpotent divided power thickening of k as well so as not to lose the action of $\mathcal{O}_D^\times$. The combined action of these groups on the compactly supported cohomology of the tower realizes the local Langlands correspondence. The proof (for $h = 2$) is to realize this tower geometrically (or at least the cohomology) as the "supersingular part" of the tower of modular curves, and then use global facts concerning automorphic forms. In fact, this is how Harris-Taylor prove local Langlands in general. The corresponding "space" is not literally a rigid space (but more on perfectoid spaces later), but one can ask for a description of the $\mathbf{C}_p$-points of M. To this end, one may construct so called period maps. I plan to come back to this in some detail, but for now let me simply say that these maps (constructed in this context in differing contexts and level of generality by Fargues, Weinstein, and Scholze) have their roots in Tate's p-divisible groups paper, where by taking $\mathcal{O}_{\mathbf{C}_p}$-points one may split the p-divisible group into a p-adic Hodge filtration, and the corresponding period map records the slope of the corresponding line as an element of $\mathbf{P}^1$ (more generally, one obtains a point in a Grassmannian). Let me mention at this point that I have studiously avoided thinking about this whole chapter in the world of Shimura varieties for many years, and it always had the reputation to me as something done by Very Smart People like Mantovan and Fargues, and I have been rewarded in my laziness simply by waiting for the moment where the correct way to view these objects has started to emerge, and there's someone around like Jared Weinstein who (apart from bringing new ideas) writes and lectures so beautifully well. I certainly recommend reading his papers and lecture notes to understand what is going on (instead of having to sort through the partially digested version I have produced for you here.) Scholze also writes well, thank god.

Page 72: Very roughly, one does the following:

(1) Understand the tower (either the Lubin–Tate tower or the corresponding tower of modular curves) as an actual geometric object $\mathcal{X}$ (perfectoid space).

(2) Construct a period map $\pi : \mathcal{X} \to \mathbf{P}^1$ (or $\mathscr{F}$) using p-adic Hodge theory.
(3) Use the first two steps to construct a formal model $\mathfrak{X}$, which will have sections arising via pull-back from some ample line bundle on $\mathbf{P}^1$.
(4) Note that the construction of these sections only depends on the p-tower, and so are Hecke equivariant with respect to all the other Hecke operators and can thus serve as a replacement for the Hasse invariant, and multiplication by these sections allows one to pass back to characteristic zero forms in H^0, which, by virtue of the control one has over the geometric context, one may identify with classical modular forms.

As Matt explained to me, one can understand the image of the ordinary locus under π to be $\mathbf{P}^1(\mathbf{Q}_p)$, which should correspond to the fact that ordinary Galois representations have splittings *already* before having to pass to $\mathbf{C}_p$. This also fits into the Lubin–Tate story and the period map to the Drinfeld upper half plane (which has $\mathbf{P}^1(\mathbf{Q}_p)$ excised), as occurs in the paper of Fargues linked to above. We also see here that the ordinary locus under the period map factors through the component group π_0, with the natural action of $\mathrm{GL}_2(\mathbf{Q}_p)$ permuting the cusps. In particular, all the ordinary points are mapping in the special fibre to $\mathbf{P}^1(\mathbf{F}_p)$, which doesn't look at all like the usual story at all. This is related to footnote 4 on page 72.

Question for the the audience: is it obvious how one can extract the classical coherent cohomology groups $H^*(X, \mathcal{E})$ at level one from $H^*(\mathcal{X}^*, \mathcal{E})$?

28. Effective motives

Wed, 03 Jul 2013

This is a brief follow up concerning a question asked by Felipe. Suppose we assume the standard conjectures. Let M be a pure motive, and consider the following problems:

(1) **Problem A:** ("effectivity") Suppose that M has non-negative Hodge–Tate weights. Then is M effective?
(2) **Problem B:** ("ordinary primes") Does the Hodge polygon = Newton polygon for infinitely primes p?
(3) **Problem C:** ("Katz") Suppose the characteristic polynomials of Frobenius have coefficients in $\mathbf{Z}$. Then is M effective?

An affirmative answer to Problem C implies an affirmative answer to Problem A. Conversely, a positive answer to Problems A & B implies a positive one for Problem C.

The relevance of Problem A was for deducing that a weight zero regular algebraic cuspidal automorphic form for $\mathrm{GL}(2)/F$ could be associated to an abelian variety of GL_2-type over F. I claimed that this was probably "Standard Conjectures hard." It seems that this is partly right and partly wrong. As mentioned previously, if M has weight zero, then Problem A already follows from Kisin–Wortmann (always assuming the standard conjectures), because then M will be an Artin motive.

As was pointed out to me, the case of weight one follows from the Hodge conjecture. Namely, the Hodge realization gives a polarized Hodge structure of weight one which gives a polarized complex torus. By Riemann, such a torus is actually an abelian variety A, which (using the standard conjectures) one can descend to F.

This argument doesn't obviously extend to the general case, because the image of the period map from (say) pure Motives with Hodge–Tate weights $[0, k]$ to polarized Hodge structures will not be surjective for Griffiths transversality reasons. As an aside, it was also pointed out that the Hodge conjecture is not one of the standard conjectures.

When I asked Deligne about Problem A, he politely told me

(1) There's no evidence for Problem A beyond the fact that it would be nice,
(2) The Hodge conjecture is false, and
(3) Grothendieck already mentioned that his (Grothendieck's) modification of the generalized Hodge conjecture implies that the answer the Problem A is positive.

Here the generalized Hodge conjecture says (roughly) that a sub-Hodge structure of H^k with weights in the range $[k-q, q]$ to $[q, k-q]$ arises via the Gysin map from an algebraic cohomology class on an $\geq q$-codimensional subvariety. In particular, if M has non-negative Hodge–Tate weights and is of weight w, and $M(n)$ is effective inside some smooth proper variety X, then M gives rise to a sub-Hodge structure of $H^{w+2n}(X)$ with weights from $[n, n+w]$ to $[n+w, n]$, and hence come from some algebraic subvariety Y of codimension at least n. However, the Gysin map on etale cohomology involves a Tate twist by $\mathbf{Q}_p(n)$, and so (using the standard conjectures) one recovers M effectively in Y. Grothendieck also points out that, in the case when M has weight one, the generalized Hodge conjecture follows from the usual Hodge conjecture after replacing X by $X \times C$ for proper smooth curves C, essentially by the same argument of the previous paragraph. (I guess one also has to use the easy fact that any abelian variety is a quotient of a Jacobian.)

Talking of Deligne and Grothendieck, Benson Farb sent me the following link to an interview of Deligne by MacPherson:

Deligne interview

which contains the following slightly terrifying exchange about Grothendieck:

MacPherson: I've heard people say that he [Grothendieck] was always very kind to students when they didn't understand, but if someone was older and had pretensions he could be less ...

Deligne: That's quite possible, and I think he was completely willing to explain something once, I don't think he would have be willing to explain it three times, even to students.

(In my original memory of this passage, "three times" was replaced by "twice.")

29. Life on the modular curve

Tue, 24 Sep 2013

Alice and Bob live on the modular curve $X_0(1) = \mathbf{H}/\mathrm{PSL}_2(\mathbf{Z})$. What does the world look like to them, assuming that they view the world in hyperbolic perspective?

To those who are not used to hyperbolic geometry, there may be a few mild surprises. Suppose that Alice is at the point $x = i$ and Bob is at $y = 10i$. Let us also imagine that Alice is looking in the direction of the cusp along the projection of the geodesic given by the y-axis. What does she see? Take a moment to think about it if you like; we will give the answer in the next paragraph.

Lifting Bob to the universal cover, there are infinitely many Bobs spaced equally along the horosphere $(10i + t)$. A naive guess is that all of these Bob's would fill out Alice's field of vision. But this can't be true; since geodesics in $\mathbf{H}$ are given by semi-circles perpendicular to the x-axis, most geodesics through $x = i$ don't cross Bob's horosphere. In fact, Bob only takes up about 10° of Alice's vision, and those Bobs who are at $(10i + n)$ for large integers n appear almost to be directly in front of Alice (although a long way away). Of course, Alice also sees copies of herself receding similarly into the distance directly in front of her.

All this and more can be seen in the 80's inspired video game of my undergraduate summer students Jasmine Powell and Justin Ahn (funded by the NSF!). The basic setup is as follows: you are a cube wondering around on $X_0(1)$ and you need to shoot the monsters, which are in the shape of a pill. Occasionally, some bonus feature will appear (extra shields, freeze, extra life, etc.) which you can collect. Some mathematics that is hiding in the background but is only partially relevant for game play: the monsters travel along closed geodesics, and the goodies appear at CM points. The game was also partly inspired by the video not knot. Here's a link to a video capture from the game:

Link to the video

(The transition to video has made it look a little wonky.) If you notice carefully, you will see that at one point in the video you crash into yourself by passing through the cone point i, losing a life.

The α-release of the game itself can also be downloaded here (sorry, macintosh only). Please play around with it and offer suggestions and improvements! Various possibilities include upgrading to a 3-manifold (probably a Bianchi manifold), and also the ability to pass to congruence covers $X_0(p)$ of $X_0(1)$.

30. Virtual congruence betti numbers

Fri, 27 Sep 2013

Suppose that G is a real semisimple group and that $X = \Gamma\backslash G/K$ is a compact arithmetic locally symmetric space. Let us call a cohomology class *tautological* if it is invariant under the group G. For example, if X is a 3-manifold, then the tautological classes are all multiples of either the trivial class in H^0 or the fundamental class in H^3. We say that X has *positive Betti number* if there exist any non-tautological classes in the cohomology of X. One can pose the following question:

Problem 30.1. Show that there exists a finite congruence cover $\widetilde{X} \to X$ such that $\widetilde{X}$ has positive Betti number.

An automorphic way of phrasing this question is as follows: do there exist *any* automorphic forms besides the trivial representation for the $\mathbf{Q}$-group $\mathbf{G}$ associated to Γ. If G admits discrete series, then the result is obvious for automorphic reasons (from the trace formula, by de George-Wallach). If X has non-zero Euler characteristic, then the result is obvious for topological reasons. In fact, as I leant from Gross one day at tea, these two situations coincide (this certainly follows from Borel-Wallach, even in the stronger form that the contribution from each π via Matsushima's formula has zero Euler characteristic if it is not a discrete series; I'm not sure if there's a slicker argument).

The problem is obviously related to the virtual positive Betti number theorem of Agol, but there are a few important subtle differences. The first is that we insist that the cover $\widetilde{X}$ is *congruence*. Hence, the problem remains open for a general arithmetic 3-manifold. Second, we also allow (as we must) cohomology in any degree. Another example to consider is $G = U(2,1)$. In this case, X is a compact complex hyperbolic manifold. It is an open problem whether such manifolds have virtual positive first Betti number. In contrast, by a theorem of Rogawski, they certainly don't have virtual positive first Betti number in *congruence covers*, although they clearly do have virtual positive Betti number in congruence covers for the two equivalent reasons given above.

What I want to do in this post is discuss a related problem, namely, can one find arbitrarily large congruence covers $\widetilde{X}$ which all *fail* to have positive Betti number? Specific examples of this kind (for a compact arithmetic 3-manifold X) were given in my paper with Dunfield (conditional on local-global compatibility of certain Galois representations, now known), and Boston–Ellenberg shortly thereafter found a different (unconditional) argument using group theory (which applied to the same example). I want to explain how to generalize these results to higher dimension, contingent on computations which might be hard to carry out explicitly.

Choose:

- An imaginary quadratic field F.
- A prime p which splits as $\mathfrak{p}\overline{\mathfrak{p}}$ in F.
- A central simple algebra D/F with local invariants $1/N$ and $-1/N$ at the primes dividing p.

Associated to D is a maximal lattice Γ in $G/K = \mathrm{SL}_N(\mathbf{C})/\mathrm{SU}_N(\mathbf{C})$ whose quotient is a compact finite volume orbifold of real dimension N^2-1. For sufficiently large n, the congruence covers $X(\mathfrak{p}^n)$ are manifolds which are $K(\pi,1)$ spaces with fundamental group $\Gamma(\mathfrak{p}^n)$. When $F = \mathbf{Q}(\sqrt{-2})$, $N = 2$, and $p = 3$, one recovers the manifolds considered in my paper with Nathan.

Let me now make another definition. Let F_S be the maximal pro-p extension of $F(\zeta_p)$ unramified outside the primes dividing p.

Definition 30.2. The prime p is **very regular** in F if the map:

$$\mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p) \to D_v \subset \mathrm{Gal}(F_S/F)$$

is surjective for either $v|p$.

The notion of very regular primes arose in my latest paper on K-theory [Cal15] and completed cohomology in the stable range, but more on that later. One last definition: say that an ideal $\mathfrak{m}$ of a Hecke algebra $\mathbf{T}$ is Eisenstein if the image of any Hecke operator T in $\mathbf{T}/\mathfrak{m}$ coincides with multiplication by the degree $\deg(T)$. This is how $\mathbf{T}$ acts on the trivial representation. We then have the following:

Conditional Theorem 30.3. *Suppose that p is very regular, and that $\mathfrak{m}$ is an Eisenstein maximal ideal. Then for all n there is an equality:*

$$H^*(X(\mathfrak{p}^n), \mathbf{Z}_p)_{\mathfrak{m}} \otimes \mathbf{Q}_p = H^*(\mathrm{SU}(N), \mathbf{Q}_p)$$

In particular, if the only *maximal ideals of $\mathbf{T}$ on $H^*(X(\mathfrak{p}), \mathbf{Z}_p)$ are Eisenstein, then all the $X(\mathfrak{p}^n)$ are rational $\mathrm{SU}(N)$-homology spaces.*

Example 30.4. The prime $p = 3$ is strongly regular for $F = \mathbf{Q}(\sqrt{-2})$, and – by a computation – the only maximal ideals of $\mathbf{T}$ on $H^*(X(\mathfrak{p}), \mathbf{Z}_p)$ for $N = 2$ are Eisenstein. Of course, a rational SU(2)-homology space is a homology 3-sphere.

Proof. Suppose that there is exists a non-trivial class in the cohomology of $X(\mathfrak{p}^n)$. It will give rise to an automorphic representation π which is tempered, because X are Shimura manifolds for which we can show (reference?) have no endoscopic forms. Hence, by HLTT or Scholze [HLTT16, Sch15b], there exists a corresponding Galois representation

$$r(\pi) : G_F \to \mathrm{GL}_n(\mathbf{Q}_p)$$

that is unramified away from p. We now assume (this may be proved soon, but this is the reason for the "conditional" in the statement) that we know enough about local-global compatibility to deduce that this representation is also *ordinary* at the prime $\mathfrak{p}'$. Note that the reason it *should* be ordinary is that the level is prime to $\mathfrak{p}'$, and since the quaternion algebra is ramified at this prime we know that $\pi_{\mathfrak{p}'}$ is Steinberg. We deduce that $r(\pi)$ is completely reducible after restriction to D_v for $v = \mathfrak{p}'$. The Eisenstein assumption and the ramification assumption imply that $\overline{r(\pi)}$ and hence $r(\pi)$ factor through $\mathrm{Gal}(F_S/F)$. Hence, using the fact that p is very regular, we immediately deduce that $r(\pi)$ itself is reducible and ordinary. It follows that, after semisimplification, $r(\pi)$ is a direct sum of characters, which leads to an easy contradiction. □

Experts will recognize this argument as a generalized and more streamlined version of what appears in my paper with Nathan. One may naturally ask whether there is a generalization of the Boston–Ellenberg argument as well. Emerton and I already explained that the correct way to view that argument was as follows. What one really wants to prove is that the partially completed cohomology groups:

$$\widetilde{H}^*(\mathfrak{p}) = \lim_{\to} H^*(X(\mathfrak{p}^n), \mathbf{F}_p)$$

all vanish identically outside degree zero. For 3-manifolds, it suffices to prove this for $\widetilde{H}^1$. For what X might one be able to prove such vanishing? As Matt and I explained in our paper on 3-manifolds, for all these groups to vanish there has to be a delicate balancing act between the dimension of the group acting on completed cohomology and the dimension of the manifold. For example, it is crucial that there is an equality

$$\dim(G/K) = \dim(\prod_S G(F_v))$$

where one partially completes at primes S above p. (Otherwise one obtains an immediate contradiction by Hochschild–Serre.) In the case at hand, this inequality is satisfied, since:

$$\dim(G/K) = \dim(\mathrm{SL}_N(\mathbf{C})) - \dim(\mathrm{SU}_N(\mathbf{C})) = N^2 - 1 = \dim(\mathrm{SL}_N(\mathbf{Z}_p))$$

Hence, it is really possible that all the completed cohomology groups may vanish in this case. In fact, if one instead considers the *split* group $\mathrm{GL}(N)/F$, then the partially completed cohomology groups **do** vanish in the stable range exactly for very regular primes. (This is where the definition of very regular primes comes from.) By Nakayama's Lemma, one can explicitly compute at some finite level to

determine whether the $\widetilde{H}^*(\mathfrak{p})$ vanish or not. In fact, it suffices to compute that the maps:

$$H^*(G(p), \mathbf{F}_p) \to H^*(X(\mathfrak{p}), \mathbf{F}_p)$$

are isomorphisms, where $G(p)$ is the congruence subgroup of $\mathrm{SL}_n(\mathbf{Z}_p)$. If one wanted to find an explicit example where these theorems applied for $N \geq 3$, the first place to look would probably be to take $F = \mathbf{Q}(\sqrt{-2})$, $p = 3$, and $N = 3$. One would then have to compute the cohomology of a certain 8-dimensional manifold! (The resulting manifolds would potentially all be rational SU(3)-homology space = rational $S^5 \times S^3$-homology space). This computation is within the realms of plausibility. To rule out characteristic zero representations, we can pass by functoriality to the split side. So, if there is a characteristic zero class which is not Eisenstein mod-p, that residual representation also has to occur at low(ish) level inside the cohomology of $\mathrm{GL}_3(\mathbf{Z}[\sqrt{-2}])$. This is the sort of cohomology that people like Gunnells might almost be able to compute!

Notes 30.5. Certainly the required local–global compatibility results are now known, e.g [ACC+23], so Conditional Theorem 30.3 is unconditional.

31. Abelian varieties

Wed, 30 Oct 2013

Jerry Wang gave a nice talk this week on his generalization of Manjul's work on pointless hyperelliptic curves to hyperelliptic curves with no points over any field of odd degree (equivalently, Pic^1 is pointless). This work (link here, also [BGW17]) is joint with Manjul and Dick, so the exposition is predictably of high quality. But I wanted to mention a result that arose during the talk which I found quite intriguing. Namely, given the intersection X of two quadrics P and Q in projective $(2n+1)$-space, the variety of projective n-spaces passing through X turns out (over the complex numbers) to be an abelian variety. For $n = 1$ this is pretty familiar, but, for general n, I hadn't seen any construction like this before. It gives, for example, explicit constructions of equations for abelian varieties in surprisingly low degree. It brought me back to a lecture I once went to by Beauville as a graduate student when he talked about intermediate Jacobians (wait — perhaps this construction also has to be isomorphic to an intermediate Jacobian ...). Is it possible (in some weak sense) to classify all varieties whose variety of maximal linear subspaces is an abelian variety of suitably high dimension? Are there varieties in which this construction gives rise to abelian varieties which are *not* isogenous to Jacobians? The geometric result is due (independently) to several authors, but, in a solo paper here, Jerry showed that the result is true *arithmetically*, and, even better, the construction can more precisely be described as giving an explicit torsor for the corresponding Jacobian. This very nicely generalizes the classical picture between pairs of quadrics and 2- and 4-descent.

Comment 31.1 (Jack Thorne)**.** Dear Persiflage, I cannot resist mentioning my favorite example of this kind of construction, which relates to smooth hyperplane sections H of $G(4, 8)$. H has dimension 15; it has primitive cohomology only in the middle degree, which gives a Hodge structure of dimension 6 and level 1. Thus the intermediate Jacobian is a PPAV.

Over the complex numbers, the PPAV which arise this way are exactly the Jacobians of the non-hyperelliptic curves X of genus 3. What about over a general field K of characteristic zero? Then the Jacobian of X arises from a K-rational hyperplane section exactly when the curve X has a K-rational flex in the canonical embedding.

Comment 31.2 (Wholesome Breakfast)**.** The variety of $\mathbf{P}^{n-1}$s in X is indeed an intermediate Jacobian, see [Don80].

32. Local representations occurring in cohomology

Tue, 05 Nov 2013

Michael Harris was in town for a few days, and we chatted about the relationship between my conjectures on completed cohomology groups with Emerton and the recent work of Scholze. The brief summary is that Scholze's results are not naively strong enough to prove our conjectures in full, even for PEL Shimura varieties. Motivated by this discussion, I want to give two quite explicit challenges concerning the mod-p cohomology of arithmetic locally symmetric spaces. The first I imagine will be very hard — it should already imply a certain vanishing conjecture of Geraghty and myself which has strong consequences. However, the formulation is somewhat different and so might be helpful.

Fix an arithmetic locally symmetric space X corresponding to a reductive group G over $\mathbf{Q}$. Let ℓ and p be distinct prime numbers. Consider the completed cohomology groups

$$\widehat{H}^d(\overline{\mathbf{F}}_\ell) = \varinjlim H^d(X(K), \overline{\mathbf{F}}_\ell), \qquad \widehat{H}^d(\mathbf{C}) = \varinjlim H^d(X(K), \mathbf{C}),$$

where we take the completion over all compact open subgroups. The limit has an action of $G(\mathbf{A})$ for the finite adeles $\mathbf{A}$, and so, in particular, has an action of $G(\mathbf{Q}_p)$. What irreducible $G(\mathbf{Q}_p)$ representations can occur in $\widehat{H}^d(\overline{\mathbf{F}}_\ell)$? Here is a guess:

Conjecture 32.1. *If the smooth admissible representation π of $G(\mathbf{Q}_p)$ occurs as an irreducible sub-representation of $\widehat{H}^i(\overline{\mathbf{F}}_\ell)$, then there exists an irreducible representation Π of $G(\mathbf{Q}_p)$ in characteristic zero such that:*

1. *The Gelfand–Kirillov dimension of Π is at least that of π. Equivalently,*
$$\dim\, \Pi^{K(p^n)} \gg \dim\, \pi^{K(p^n)}.$$
2. *Let $\mathrm{rec}(\Pi)$ and $\mathrm{rec}(\pi)$ be the Weil-Deligne representations associated to Π and π respectively by the classical local Langlands conjecture and the mod-ℓ local Langlands conjecture of Vigneras. Then*
$$(\overline{\mathrm{rec}(\Pi)})^{\mathrm{ss}} \simeq (\mathrm{rec}(\pi))^{\mathrm{ss}}.$$
3. *The representation Π occurs in $\widehat{H}^j(\mathbf{C})$ for some $j \leq i$.*

Roughly speaking, this conjecture says that the irreducible representations occurring in characteristic p are no more complicated than those which occur in characteristic zero. One naive way to try prove this conjecture would be to show that any torsion class lifts to characteristic zero, at least virtually. This conjecture is too strong, however, as can be seen by considering K-theoretic torsion classes in stable cohomology — the mod 3 torsion class in $H^3(\mathrm{GL}_N(\mathbf{Z}), \mathbf{F}_3)$ can never lift to

characteristic zero for sufficiently large N because the cohomology over $\mathbf{Q}$ is zero for all congruence subgroups by a theorem of Borel. The conjecture as stated seems very hard.

In a different direction, here is the following challenge to those trying to understand completed cohomology through perfectoid spaces. (I expect one can prove this by other means, but I would like to see a proof using algebraic geometry.)

Problem 32.2. Fix an integer d, and let X_g be the Shimura variety corresponding to the moduli space of polarized abelian varieties of genus g. Prove that, for g sufficiently large, the completed cohomology group $\widetilde{H}^d(X_g, \mathbf{F}_p)$ is finite over $\mathbf{F}_p$.

An equivalent formulation of this problem is to show that the only smooth admissible $\mathrm{GSp}_{2g}(\mathbf{Q}_p)$-representations π which occur inside $\widetilde{H}^d(X_g, \mathbf{F}_p)$ are one dimensional.

Notes 32.3. This is still completely open and really interesting. It is of course related to the congruence subgroup problem.

33. Daleks

Thu, 05 Dec 2013

I've wanted to write a post about the new Doctor Who series for a while, but this is not that post. Instead, this post is about a Macintosh game called Daleks, which I first played on a Mac 512 (running OS 3) in the mid-'80s. Research indicates that this game was based on a Unix game called robots, and that some wag came up with the idea of rebranding it under the name of the classic Doctor Who monster. The first version I played had a very peculiar high score table: all the high scores were attributed to a fellow named "fingers," and the high scores were not in any sort of numerical order. Moreover, no matter what one scored, it was impossible to permanently make it onto the high score list. My second encounter with the game was during a summer research program with Alf van der Poorten in 94/95, where I was impressed to find that he had broken 10000. Later, I had a copy on an ancient laptop given to me by my brother, and still later, I played classic Daleks in classic mode under OS X. I am not ashamed to say that I am proud of my high score, 15670, a feat which is probably meaningless to almost everyone. Anyway, today's post is about some mathematical problems related to this game. If you have a mac computer, I recommend playing around with some current incarnations of the game, for example super daleks (presumably robot is available on Gnome games as well):

Consider the following game: the Doctor is positioned on the lattice $\mathbf{Z}^2$ at the origin $(0, 0)$, and daleks are distributed on the rest of the lattice with uniform density $\rho \in (0, 1)$. It turns out that it is more convenient to work with the parameter $q = 1 - \rho$, although all the graphs below are drawn with respect to ρ. On each move, the dalek at point P moves to the unique neighbouring square (out of 8) which is closest to the origin in the taxicab metric. In particular, daleks always move diagonally towards the origin unless they lie on one of the axes. If two or more daleks occupy the same square, then they crash and are destroyed, leaving a pile of debris which remains at that square forever. Moreover, any other dalek which later moves on to the same square now occupied by the debris is also destroyed. If a dalek reaches the origin unscathed, the Doctor is exterminated. However, if the

debris resulting from dalek collisions prevents all other daleks from reaching the origin, then the Doctor survives. What is the probability that the Doctor survives? (In the computer game the Doctor can also move about, but not in our simplified version.)

If a dalek starts on either the diagonal or the anti-diagonal then it will never crash with another dalek (in general, daleks can only crash on the axes). Hence, we modify the game by forbidding daleks from either of these diagonals. This effectively separates the playing area into four quadrants which do not interact, and so we may as well confine ourselves to a single quadrant, and assume that all daleks lie in the quadrant $(1,0)+Q$ where $Q=(x,y)$ with $x \geq |y|$. A sample game is as follows, with the positions at time $t=0$, 1, and 2. The Doctor will win this game, because the debris at $(1,0)$ will prevent all other daleks in the quadrant from reaching the origin (they will crash into the debris and be destroyed):

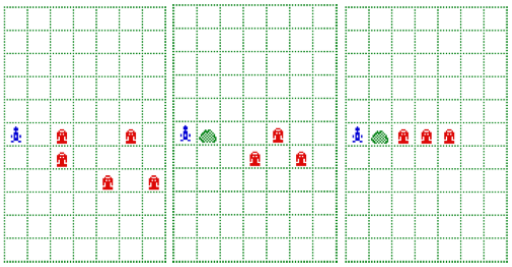

FIGURE 1. Three consecutive frames: the daleks are all destroyed

Definition 33.1. Let Q_d be the truncated quadrant consisting of (x,y) with $|y| \leq x < d$. Let w_d denote the probability of surviving the game where daleks only exist with density ρ in the quadrant $(1,0)+Q_d$.

For example $w_0 = 1$, and $w_1 = q = 1-\rho$. It is clear that

$$1 - w_d = \sum_{P \in (1,0)+Q_d} E(P),$$

where $E(P)$ is the expectation of being exterminated by a dalek which originates at point P. (If there is a dalek which kills the Doctor, it is unique.) Note that $E(P)$ is independent of d, providing that $P \in (1,0)+Q_d$.

Definition 33.2. The **occludation** $O(P)$ of P consists of the squares R different from P where a dalek at square R will reach the origin before or at the same time as P, and which reach the x-axis at least as near to the origin as P reaches the x-axis.

For example, here is a point $P=(1,0)+(4,3)$ together with its occludation shaded in blue (the green region to be described later):

The daleks $R \in O(P)$ are those for which P is in the shadow of R, namely, those R which eventually occlude P from the origin (thus the name). It is not a great name, but I couldn't think of anything better. Explicitly, if $P=(x,y)$, then

$$O(P) \cup P = \{(a,b) \in (1,0)+Q \text{ such that } a < x, \text{ and } a - |b| \leq x - |y|\}$$

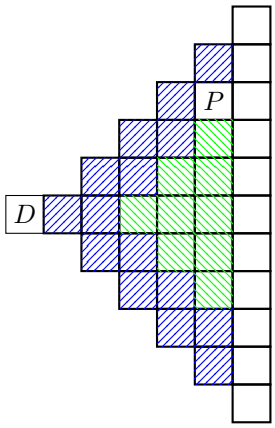

FIGURE 2. The occludation of $P = (5, 3)$

The Doctor can only be killed by dalek at point P if the occludation $O(P)$ is empty of daleks. The reason is that any dalek R in the occludation can only crash at points on the x-axis where P must eventually travel, and R will reach this point either at or before P does. We have

$$|O(P)| = |O((x, y))| = x^2 - y^2 - 1.$$

On the other hand, conditional on the assumption that the occludation contains no daleks, then the probability that P exterminates the Doctor only depends on y; namely, it is equal to the probability of surviving the game with Q_d and $d = |y|$. This the quadrant not in the occludation of P. For example, with $P = (5, 3)$ as in Figure 2 above, the dalek at P (if the occludation is empty) will reach the doctor only if the daleks in the green region, corresponding to Q_3, do not survive. It follows that

$$E(P) = q^{|O(P)|}(1 - q)w_{|y|} = q^{x^2-y^2-1}(1 - q)w_{|y|},$$

and hence

$$1 - w_d = \sum_{n=1}^{d} \sum_{|m|\le n} E((n, m)) = \sum_{n=1}^{d} \sum_{|m|\le n} q^{n^2-m^2-1}(1 - q)w_{|m|}.$$

We may simplify this slightly by writing

$$w_{d-1} - w_d = (1 - w_d) - (1 - w_{d-1}) = \sum_{|m|\le d} q^{d^2-m^2-1}(1 - q)w_{|m|},$$

This simplifies even further to

$$\begin{aligned}
&(w_d - w_{d+1}) - q^{2d+1}(w_{d-1} - w_d) \\
&= \sum_{|m|\le d+1} q^{(d+1)^2-m^2-1}(1 - q)w_{|m|} - q^{2d+1} \sum_{|m|\le d} q^{d^2-m^2-1}(1 - q)w_{|m|} \\
&= 2q^{2d}(1 - q)w_d,
\end{aligned}$$

and hence, subject to $w_0 = 1$ and $w_1 = q$,

$$w_{d+1} = (3q^{2d+1} - 2q^{2d} + 1)w_d - q^{2d+1}w_{d-1}.$$

The resulting recurrence relation for w_d gives a decreasing convergent sequence (for each fixed q and also in $\mathbf{Z}[[q]]$) with limit

$$w_\infty = q - 3q^3 + 3q^4 - 2q^5 + 2q^6 + 4q^7 - 13q^8 + 13q^9 + \dots$$

Here is a graph of this function in Figure 3 (with respect to ρ, remember that $\rho = 1 - q$):

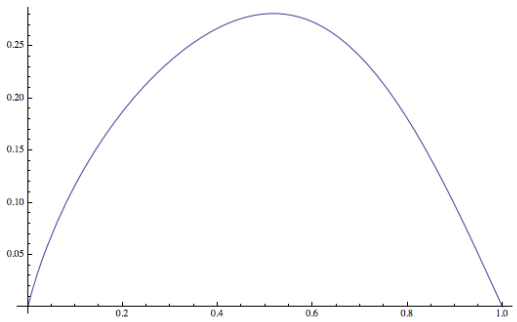

FIGURE 3. The chance of survival w_∞ for dalek density ρ

Although it appears from the graph that the maximum occurs at $\rho = q = 1/2$, closer inspection reveals that the optimal density is $\rho = 0.517208\dots$ The maximum value is approximately $\sim 0.28116\dots$, which means that, on an entire plane with all four quadrants, the largest possible chance of winning (with daleks on the diagonal and anti-diagonal removed) is approximately 1 in 160. Note that as $\rho \to 1$, we certainly have $w_\infty \to 0$. As $\rho \to 0$, it is also clear that one should expect the first dalek on a line to survive, which means that w_∞ should tend to 0 as $q \to 1$ (as see in Figure 3), that is not yet apparent from the formula above.

33.3. **Reverse The Polarity.** Here is a different way to estimate w_∞, this time from below. In order for the Doctor to survive, two or three daleks must eventually coincide at $(1, 0)$. Call such daleks **savior** daleks. All savior daleks must be in the same row, and at least one such dalek must lie on the edge of the quadrant. Let us now consider the probability s_n that one will be "saved" by a dalek in the nth row. If $P = (n, n-1)$ is a savior dalek, then the dalek P creates the first crash at the point $(0, 1)$, and no dalek exterminates the Doctor before this point. It follows that no daleks may occlude P, and hence $O(P)$ must be free of daleks, with the possible exception of $-P$. Note that $|O(P)| = n^2 - (n-1)^2 - 1 = 2n$. Suppose that $-P$ is not occupied. Then (assuming that $O(P)$ is empty) P will be a savior dalek if and only if the remaining restricted quadrant of size $n-1$ would otherwise result on the Doctor being exterminated at the final term, equivalently, the probability that, from a quadrant of size $n-1$, the Doctor would be exterminated by a dalek in the last row. Yet the probability of this is

$$(1 - w_{n-1}) - (1 - w_{n-2}) = w_{n-2} - w_{n-1},$$

and hence the contribution to s_n is

$$2q^{2n-2}(1-q)(w_{n-2} - w_{n-1}).$$

On the other hand, if both P and $-P$ are to be savior daleks, then one simply requires that, in addition to the rest of occlusion $O(P)$ being empty, that in the remaining quadrant of size $n-2$ (removing the final row and the occlusion) no Doctor is exterminated, and this has probability w_{n-2}. Hence

$$s_n = 2q^{2n-2}(1-q)(w_{n-2} - w_{n-1}) + q^{2n-3}(1-q)^2 w_{n-2}.$$

Let t_n be the probability that there exists a savior dalek at a row at most n. Then clearly

$$t_n = \sum_{m=1}^{n} s_m.$$

Moreover, we naturally have inequalities $w_n \geq t_n$, and

$$w_\infty = \lim_{n\to\infty} w_n = \lim_{n\to\infty} t_n,$$

where the limit is pointwise and $q \neq 1$. However, the behavior of w_n and t_n is quite different in the regime $\rho \to 0$ or $q \to 1$, as (Figure 4) of $w_5 \geq t_5$ shows. This is not so surprising, as $\rho \to 0$ one expects that $w_\infty = t_\infty = 0$, but the dalek which destroys the doctor will be expected to become further and further away.

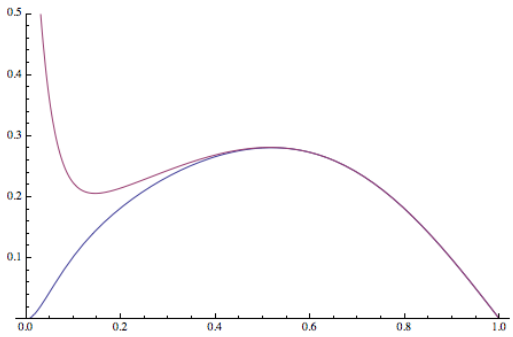

FIGURE 4. Upper and lower bounds: $t_5 \geq t_\infty = w_\infty \geq w_5$

33.4. **Behavior as $\rho \to 0$.** In order to estimate the behavior of w_∞ as $q \to 1$, we consider the following problem: What is the probability that the first row with any dalek contains exactly two daleks, and that at least one of these daleks lies at the edge of the quadrant? In such a situation, the daleks necessarily annihilate one another at $(1,0)$, and the Doctor is saved. Call the resulting function $A(q)$, so $w_\infty \geq A(q)$. Since there are $(4n-1)$ pairs of elements in $1,\ldots,2n+1$ which contain

at least one of the end points, we have

$$\begin{aligned} A(q) &= \sum_{n=1}^{\infty}(4n-1)q^{n^2}(q^{2n-1}(1-q)^2) = (q-1)^2q^{-2}\sum_{n=2}^{\infty}(4n-5)q^{n^2} \\ &= (q-1)^2q^{-2}\left(5+q+\sum_{n=0}^{\infty}4nq^{n^2}-5\sum_{n=0}^{\infty}q^{n^2}\right) \\ &= (q-1)^2q^{-2}\left(\frac{5}{2}+q+2\sum_{n=-\infty}^{\infty}|n|q^{n^2}-\frac{5}{2}\sum_{n=-\infty}^{\infty}q^{n^2}\right) \\ &= (q-1)^2\left(2\sum_{n=-\infty}^{\infty}|n|q^{n^2}-\frac{5}{2}\sum_{n=-\infty}^{\infty}q^{n^2}\right)(1+O(1))+O((q-1)^3). \end{aligned}$$

Let $q = e^{-\tau}$. As $q \to 1$, we have $\tau \to 0$, and so $1-q \sim \tau$. On the other hand, by Poisson summation, we have

$$\sum_{n=-\infty}^{\infty} e^{-n^2\tau} \sim \sqrt{\frac{\pi}{\tau}} + O(\tau^N),$$

$$\sum_{n=0}^{\infty}|n|e^{-n^2/\tau} \sim \frac{1}{\tau}-\frac{1}{6}-\frac{\tau}{60}-\frac{\tau^2}{252}+O(\tau^3),$$

from which it follows easily that $A(q) \sim 2\tau \sim 2(1-q)$, and thus

$$\limsup_{q\to 1} w_\infty \geq 2(1-q).$$

In fact, we can actually prove that

$$w_\infty(e^{-\tau}) = \frac{2}{\tau} + O\left(\frac{1}{\sqrt{\tau}}\right).$$

In other words, the simple model above is very accurate in the limit $q \to 1$. However, the combinatorics required to prove this are actually somewhat involved and annoying, and this is a blog, so I will omit it here. (The arguments are somewhat timey–wimey.)

33.5. **A conditional game.** Consider the game which is pre-conditioned on the first square $(1,0)$ being empty. Since that square containing a dalek is not consistent with survival, the new game results in a win with probability:

$$c_\infty = \frac{w_\infty}{1-\rho} = \frac{w_\infty}{q}.$$

Apropos of nothing, here's Davros enjoying a cuppa in Figure 5.

33.6. **The TARDIS.** Suppose that the Doctor has a TARDIS. This allows him, at any point, to dematerialize and the materialize somewhere else. In the context of the classic daleks game, the player appears at a random point in the plane with uniform distribution. Although this doesn't quite make sense on an infinite plane, we can take it to mean that we have moved sufficiently far away from the axes that it is as if the game has started again. Hence this will be the context in which we shall consider rematerialization, namely, as if the game has started again. The catch with using the TARDIS is that the Doctor may materialize next to a dalek, in which case he is immediately exterminated. The optimal strategy is to continue to continue rematerializing until one has a winning game. The chance of surviving

FIGURE 5. Davros enjoying a cuppa

a rematerialization is $(1-\rho)$; the resulting game is the same, but now conditional on not being annihilated by the initial dalek, hence is equivalent to the conditional game described above. It follows that the chances of survival are:

$$d_\infty := w_\infty + (1-w_\infty)(1-\rho)(c_\infty + (1-c_\infty)(1-\rho)(c_\infty + \ldots = \frac{(2-q)w_\infty}{1-q+w_\infty}.$$

The asymptotic behavior of this function as $\rho \to 1$ (or $q \to 1$) requires the correct asymptotic $w_\infty \simeq 2(1-q)$, and from this we can deduce that

$$d_\infty \to 2/3 \text{ as } \rho \to 0.$$

In this case, we see that the optimal probability is that the density ρ tends to zero. A graph of d_∞ is in Figure 6.

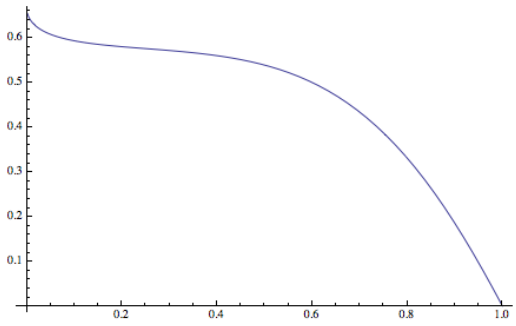

FIGURE 6. The chance d_∞ of surviving with a TARDIS

33.7. **The Sonic Screwdriver.** Like John Nathan-Turner, I find the sonic screwdriver to be somewhat ridiculous. Although it does exist in some versions of the game, I will only mention a minor modification here. The "sonic" in the game allows the Doctor to survive for one round when he would otherwise be exterminated; it has only one use. We shall additionally assume that the sonic can only be used on the very first round. This essentially changes the game (at the beginning) into the conditional game described above. If one is allowed to use the TARDIS as above, the resulting probability of winning is

$$(c_\infty + (1-c_\infty)(1-\rho)(c_\infty + (1-c_\infty)(1-\rho)(c_\infty + \ldots) = \frac{w_\infty}{q(1-q+w_\infty)}.$$

As $\rho \to 1$, this function tends to 1, and as $\rho \to 0$, it tends to $2/3$. The behavior of this function in a neighbourhood of 0 appears to be of the form

$$2/3 - A\rho^{1/2} + \dots$$

for some constant A, possibly around 0.3. Note that this function is not monotone (see Figure 7); the most dangerous density of daleks is approximately $\rho = 0.127$, where the resulting probability of surviving dips below $3/5$.

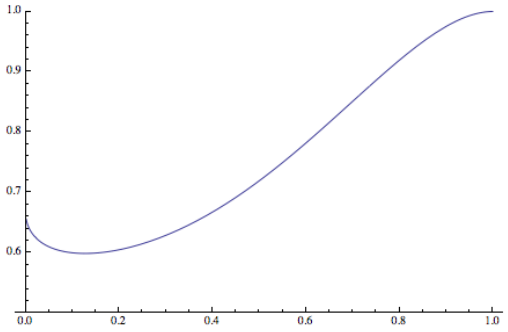

FIGURE 7. The change of winning with the sonic screwdriver

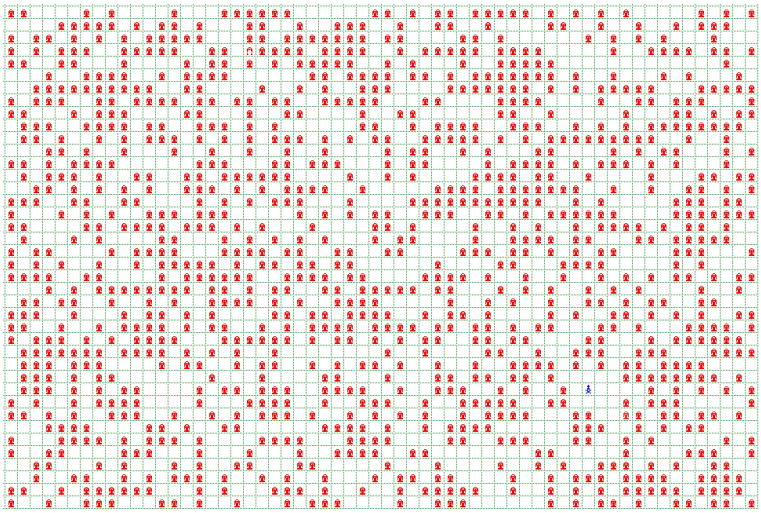

FIGURE 8. The vanilla game at the optimal value $\rho \sim 0.517$ — the Doctor lives!

Notes 33.8. I'm not sure if anyone ever read this post. I made some small clarifications and added another diagram.

34. The mystery of the primes

Sat, 04 Jan 2014

No, this is not the sequel to Marcus du Sautoy's book, but rather a curious observation regarding George Schaeffer's tables of "ethereal" weight one Katz modular eigenforms (which you can find starting on p.64 of his thesis, ultimately downloadable from Proquest but available more directly from the unstable link here). Let N be a positive integer, let χ be an odd quadratic character of conductor dividing N, and let p be an a prime not dividing N. Recall that the reduction map between spaces of Katz modular forms:

$$M_1(\Gamma_1(N), \chi, \mathbf{Z}_p) \to M_1(\Gamma_1(N), \chi, \mathbf{F}_p)$$

is not surjective in general, although it will be surjective for all but finitely many p. For what pairs (N, p) is the map not surjective? As originally observed by Mestre (and predicted by Serre), such pairs do exist. One way to think of the primes which arise in this way are as the primes dividing the torsion subgroup of $H^1(X_H(N), \omega)$, where $H \subset (\mathbf{Z}/N\mathbf{Z})^\times$ is the subgroup of squares, and $X_H(N)$ is the corresponding modular curve (as a stack, if necessary) over $\mathbf{Z}[1/N]$. The reduction mod-p map is Hecke equivariant; let $\mathfrak{m}$ denote a maximal ideal of $\mathbf{T}$ in the support of the cokernel. Associated to $\mathfrak{m}$ is a Galois representation:

$$\overline{\rho} = \overline{\rho}_{\mathfrak{m}} : G_{\mathbf{Q}} \to \mathrm{GL}_2(\overline{\mathbf{F}}_p)$$

which is unramified at primes not dividing N (including p). It is not necessarily the case that $\overline{\rho}$ does not lift to characteristic zero, although this is typically the case for the examples arising in the tables (and is always the case if the image of $\overline{\rho}$ contains $\mathrm{SL}_2(\mathbf{F}_q)$ for some $q \geq 5$). Not surprisingly, it turns out there are no such forms for small N. The reason is that the fixed field K of the kernel of $\overline{\rho}$ would be a high degree field with a root discriminant which (for very small N) would violate the GRH discriminant bounds of Odlyzko, and for smallish N would still give fields of unusually low root discriminant.

Of course, as N increases, there do exist many such forms, sometimes in quite large characteristic. However, something peculiar happens in the range of the tables, namely, there is not a single example with N prime. This leads to the (incredibly) vague question: can this be predicted in advance? If there is going to exist a $\mathrm{PGL}_2(\mathbf{F}_{199})$ representation unramified outside N for small N, is it more likely that $N = 82$ (see here) rather than $N = 83$? One can try to use heuristics predicting the number of fields with certain ramification behavior, but these heuristics are much better behaved for fixed Galois groups $G = \mathrm{PGL}_2(\mathbf{F}_p)$ or $G = \mathrm{PSL}_2(\mathbf{F}_p)$ and increasing discriminant, not in the regime of fixed root discriminant and Galois group G as above for varying p. Is there any conspiracy ruling out certain kinds of number fields with small root discriminant ramified at a single prime? For example, if you fix some arbitrary constant, say $M = 1000$, do there exist infinitely many primes p such that there is a number field K different from $\mathbf{Q}$ which is unramified away from p and has root discriminant less that M?

These questions are hard to pin down, because they are really questions concerning the law of small numbers. Namely, they ask about the behavior/distribution of various quantities in the range before asymptotic behavior begins. Since the asymptotic behavior is (in these contexts) already mostly conjectural, it's probably hard to say anything intelligent about these even more delicate questions. (Idle question:

are there similar problems for which one *does* understand what happens before the asymptotic regime begins, even heuristically?)

Here's one reason to consider these questions. Suppose one wants to compute "ethereal" Siegel modular forms. At what level does one first expect to find such forms? The numerics above suggest that it might be easier to find such forms at small composite levels rather than prime levels. Is that a reasonable inference?

Comment 34.1 (Dick Gross)**.** I thought about this question when writing my paper on companion forms [Gro90], and basically gave up. It's predicting when a line bundle of small degree on a curve has a larger space of sections (mod p). That's why Serre's criterion is so subtle — this jump occurs precisely when you have an odd 2-dimensional modular representation which is unramified at p which does not come from a 2-dimensional representation over $\mathbf{C}$. Serre's conjecture in weight 1 was what attracted me to the subject of companion forms in the first place, and I found it amusing that it was precisely the weight 1 situation that I couldn't resolve completely. Fortunately, Robert Coleman understood what I was doing much better than I did, and finished it off [CV92]. I should say that Mestre's computations for $p = 2$ were more convincing than any of the proofs!

Comment 34.2 (Persiflage)**.** In response, I noted: even classical (odd) Artin representations can give rise to torsion classes, exactly when their mod-p reductions admit "extra" unramified p-adic deformations. For example, consider a modular representation:

$$\rho : G_{\mathbf{Q}} \to S_3 \hookrightarrow \mathrm{GL}_2(\mathbf{C}).$$

Let $K/\mathbf{Q}$ be (any of the) corresponding imaginary cubic fields inside the fixed field of the kernel of ρ. If $p \geq 3$ is prime, then $\overline{\rho}$ admits a non-trivial unramified deformation to $\mathbf{F}_p[\epsilon]/\epsilon^2$ exactly when p divides the class number of K. This deformation will be (by [CG18a]) Katz modular but does not come from characteristic zero, so it will give rise to torsion in $H^1(X, \omega)$, or equivalently mod-p classes which don't lift to characteristic zero. The smallest example (of the exact flavour above) occurs for the field

$$K = \mathbf{Q}(\theta)/(\theta^3 - \theta^2 + 7\theta - 6)$$

of discriminant $-3 \cdot 521$, with class number $h_K = 5$. As Dick then noted: the cubic field K has a unit group of rank 1, so its class number will rarely be divisible by p — by the Cohen–Lenstra heuristic — and the existence of unramified deformations is still a sporadic phenomenon.

Notes 34.3. See the remark of George Boxer in §112

35. Gross Fugue

Sat, 11 Jan 2014

Here are some variations on the theme of the last post § 34, which is also related to a problem of Dick Gross.

In this post, I want to discuss weight one modular forms where the level varies in the "vertical" aspect (that is, N is a growing power of a fixed prime, rather than simply an increasing integer). First of all, consider the spaces

$$S_1(\Gamma(M \cdot \ell^n), \mathbf{C})$$

for fixed M and growing n. For example, if $M = 1$, the corresponding Galois representations are associated to number fields unramified outside a single prime ℓ. Given a cusp form f, the twists $f \otimes \chi$ by a finite order character of ℓ-power order will also be modular (possibly with larger n), so all the finiteness statements below should be interpreted "up to twist."

The first observation is that there exist only finitely many exceptional cusp forms (with projective image A_4, S_4, A_5) because, by a theorem of Hermite, there are only finitely many fields with a fixed Galois group unramified outside a fixed set of primes. (This echos a very general conjecture which says [very loosely] that if one fixes an infinitesimal character and varies the level in a ℓ-adic tower, one should only see finitely many automorphic forms which do not arise via functoriality from constructions using discrete series.)

The second observation is that all the other cusp forms are easy to describe: they are induced from finite order characters of a fixed number of easily determined quadratic fields K.

So far so good. But what happens if one replaces $\mathbf{C}$ by $\mathbf{F}_p$, or more generally $\mathbf{Q}_p/\mathbf{Z}_p$? Here is the following optimistic guess:

Question 35.1. For a fixed prime $p \neq \ell$, are there only a *finite* number of non-liftable forms in the p-power tower?

Here we have to take the usual caveats — not only do we have to take into account twisting, but also the $\mathrm{GL}_2(\mathbf{Q}_\ell)$-action (old forms).

This question is supposed to be a GL(2)-analogue of Washington's famous theorem on the p-part of the class group in the ℓ-adic cyclotomic tower. We shall see that it is more than an analogy. What will the source of torsion classes be?

(1) One source are Galois representations: $\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{F}_q)$ with **big image** that are unramified outside ℓ (with q a power of p). Of course there are only finitely many such representations for any fixed q, but some heuristics I learnt from Akshay convince me that there should only be finitely many even if one varies q over all powers of p (taking into account twisting, of course).

(2) Another source of torsion comes from **deformations** of big image Galois representations $\overline{\rho}$ as above, or from representations with projective image one of the exceptional groups. Since each unramified deformation ring will be finite, each $\overline{\rho}$ should only give rise to finitely many extra torsion classes.

(3) A third source of torsion classes comes from **reducible indecomposable representations**. The residual representations $\overline{\rho}$ which arise in this way occur when $L(0, \chi)$ is divisible by p for an odd character χ of finite order. In particular, there are only finitely many such representations which occur exactly if all but finitely many L-values $L(0, \chi)$ are prime to p, where χ is an odd character of conductor M times a power of ℓ. But this *exactly* the content from Washington's Theorem (the oddness assumption is not, however, necessary).

(4) The final class come from deformations of **dihedral representations**. If $\overline{\rho}$ is the induction of a character ψ of $K/\mathbf{Q}$, then the tangent space to the unramified deformation ring of $\overline{\rho}$ gives rise to torsion classes when there are no everywhere unramified classes in $H^1(\mathbf{Q}, \mathrm{Ind}(\psi/\psi^c))$ — the unramified dihedral representations in $H^1(\mathbf{Q}, \eta_K)$ are seen globally. By inflation-restriction, this is equal to a certain invariant part of the class groups of the

anti-cyclotomic tower. There are non-vanishing results concerning L-values of Hida that are relevant here, although I haven't checked to see if they imply the finiteness statement or not.

The only way to start thinking about answering this question is to think in terms of the torsion in the cohomology of modular curves. But, I confess, I do not really have any ideas on how to prove it. (To be honest, I still find Washington's proof very mysterious.)

On related matters, it would be nice if one could prove — say by analytic means — that $H^1(X_H(N), \omega)$ has torsion (prime to N) for all sufficiently large N. Taking N to be a power of a prime, this would give a different construction of non-solvable Galois representations unramified outside a single prime (for all ℓ) from the one suggested by Dick and carried out in for $\ell \in \{2, 3, 5, 7\}$ by Dembélé and others. Moreover, although (as in those examples) it would involve the group $\mathrm{PSL}_2(\mathbf{F})$ as a simple factor, the residue characteristic would be different from ℓ rather than equal to ℓ in the previous constructions. (George Schaeffer told me he tried computing torsion coming from $X_H(343)$ but didn't find any.) It might also (for suitable H) give a lower bound for $\pi_1(\mathcal{O}_K)$ where $K = \mathbf{Q}(\sqrt{-D})$ which is better (at least for some primes) than one gets from class numbers.

36. Local crystalline deformation rings

Sat, 08 Feb 2014

I just returned from a very pleasant conference in Puerto Rico courtesy of the Simons Foundation (general advice: if you live in Chicago, always accept invitations to conferences in January). One thing I learnt from Toby Gee was the following nice observation. Suppose that

$$\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{F}_p)$$

is a modular Galois representation, which for convenience we shall assume is unramified outside p. Consider deformations of this representation which are crystalline with fixed Hodge–Tate weights $[0, k-1]$ where k is even. According to Kisin, the global minimal crystalline deformation ring contains a point on every component of the corresponding local crystalline deformation ring. (All discussions of components refer to the generic fibres.) One natural question is how many components the local deformation rings actually have (when the weight is very small, it's usually the case that there is only one such component and it is smooth — this was crucial in the original Taylor–Wiles method before Kisin). For higher weight, one can distinguish between components which are "ordinary" and "not ordinary", but it is not clear what else there is. (Indeed, Kisin seemed to think some years ago that this would be it, using the meta-argument that amongst any finite set one should be able to distinguish different points by some naturally available property.)

Now suppose we also now assume that $\overline{\rho}$ is locally reducible. According to Buzzard's conjectures, all the slopes of the global crystalline lifts of $\overline{\rho}$ will be integral. Suppose one wants to prove this by local methods. Then one is ultimately led to conjecturing that each component of the local crystalline deformation ring has a fixed integral slope (recall we are in the locally reducible case, this is certainly false for locally irreducible representations in general). As a first consequence, one sees that in very high weights there will be many different components. Moreover, if one

takes a different global representation $\overline{\varrho}$ which is the same locally as $\overline{\rho}$, then the set of slopes arising from lifts of $\overline{\varrho}$ will be the same as for $\overline{\rho}$. These ideas do not quite give a complete conjectural explanation of why Buzzard's slope conjectures are true, but it is a good start.

Something that is a little disturbing in this picture, however, is the case when $\overline{\rho}$ is reducible. It becomes clear that, in high weight, there will be many crystalline representations with reducible residual representations, but the set of components of local crystalline deformation space which have a global point will be a proper subset of the set of components (assuming that components can be distinguished by slope). For example, all the slopes at level one when $p = 2$ are (besides the Eisenstein series) ≥ 3, but there certainly exist modular forms of higher tame level with the same local residual representation of slope one. So is there any way to predict when a reducible representation will have a global lift on any component of local deformation space?

In fact, the failure of lifts in the reducible case is an old problem. In the most naive sense, one can find reducible representations at levels where there are no cusp forms, but to play the game honestly we should also allow (globally) reducible lifts. Perhaps the first genuine example corresponds to extensions:

$$1 \to \mathbf{Z}/p\mathbf{Z} \to V \to \mu_p \to 1$$

where the extension is completely split at p but ramified at an auxiliary prime N. These representations are locally split and so certainly admit local lifts (namely, $\mathbf{Z}_p \oplus \mathbf{Z}_p(1)$). If $p \geq 3$, then such extensions exist whenever $N \equiv \pm 1 \mod p$, but (by Mazur) one knows that there exist weight two level $\Gamma_0(N)$ lifts only when $N \equiv +1 \mod p$ (in fact, one can prove the analogous claim that there only exist global crystalline lifts with the appropriate conductors under the same congruence condition). This is related to the general problem of understanding when certain reducible representations can be lifted to cusp forms, which seems to be a tricky problem (Ken Ribet's student Hwajong Yoo has thought about this, see [Yoo19]).

This also reminds me of a fact I learnt from Kevin Buzzard. Take the representation

$$\rho : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{Q}_2)$$

associated to the cusp form Δ. Then there exist lattices for this representation such that the corresponding residual representation is any one of the four (three non-split) extensions of $\mathbf{Z}/2\mathbf{Z}$ by itself which are unramified outside 2. (Question: does this immediately imply the same is true for all 2-adic representations coming from level one modular forms?)

37. The thick diagonal

Fri, 14 Mar 2014

Suppose that F is an imaginary quadratic field. Suppose that π is a cuspidal automorphic form for $\mathrm{GL}(2)/F$ of cohomological type, and let us suppose that it contributes to the cohomology group $H^1(\Gamma, \mathbf{C})$ for some congruence subgroup Γ of $\mathrm{GL}_2(\mathcal{O}_F)$. Choose a prime p which splits in F so that π is ordinary at $v|p$. Hida proves that the corresponding cohomology class lives in a Hida family $\mathcal{H}$ over the appropriate weight space, which in this case is (up to connected components) just $\Lambda = \mathbf{Z}_p[[X, Y]]$. However, unlike the classical situation, this Hida family will not

be flat, because the specialization to any local system which is not invariant under complex conjugation is necessarily finite. Thus the support D of $\mathcal{H}$ has co-dimension at least one over Λ. Hida proves that it does indeed have co-dimension one.

What does the support D of $\mathcal{H}$ look like? Let us suppose that we are normalizing Λ so that the point $X = Y = 0$ corresponds to π. One can imagine two possibilities:

(1) D contains the diagonal $\Delta : X = Y$.
(2) the components of D passing through $X = Y = 0$ only contains finitely many classical points.

It seems as though these are the only possibilities. Certainly, by a Zariski closure argument, D either contains the diagonal Δ or intersects it in finitely many points. Hence, it is true that if the first condition does not hold, then the components passing through $[0, 0]$ contain only finitely many *crystalline* automorphic forms. However, there could be more classical points on D, namely, those of parallel weight but non-parallel finite order nebentypus character. To be concrete, the possible points of $\mathrm{Spec}(\Lambda)$ which may give rise to automorphic forms have (with some normalization) the following shape:

$$1 + X \mapsto (1 + p)^k \zeta, \qquad 1 + Y \mapsto (1 + p)^k \xi,$$

where ζ and ξ are p-power roots of unity, and k is a non-negative integer. So one is really considering not simply the intersection of D with the diagonal Δ, but with the thick diagonal $\mathbb{\Delta}$, which is the union of the infinitely many translates of Δ by p-power roots of unity. In particular, the Zariski closure of $\mathbb{\Delta}$ is all of weight space.

I wrote a paper with Barry Mazur [CM09] where, as an illustrative example, we found an explicit Hida family which did not satisfy the first condition and claimed that it therefore satisfied the second, whereas we should only have made the weaker claim that D (which was irreducible in this particular case) contains only finitely many crystalline points. (The main point of the paper was, by studying infinitesimal deformations of Artin representations, to give evidence that D should only ever contain the diagonal when π is either a base change form or CM.) The error was pointed out to me by David Loeffler [Loe11, §5.1].

I am pleased to say, however, that my student Vlad Serban has overcome this error [Ser22] (see also [Ser18])! Namely, suppose one has a non-trivial power series $\Phi(X, Y) \in \mathbf{Z}_p[[X, Y]]$, and suppose that

$$\Phi((1 + p)^k \zeta - 1, (1 + p)^k \xi - 1) = 0$$

for infinitely many triples (k, ζ, ξ) with k a non-negative integer, and ζ, ξ, p-power roots of unity. Let D be a component of the zero set $\Phi(X, Y) = 0$ passing through $(0, 0)$. Then, after possibly replacing the roles of X and Y, Vlad proves the following. Either:

(1) D contains the diagonal Δ,
(2) $\Phi(\zeta - 1, \zeta^N - 1) = 0$ for *all* p-power roots of unity ζ, for a *fixed* $N \in \mathbf{Z}_p$.

Certainly the latter is possible, because one could have $\Phi(X, Y) = (1 + X)^N - (1 + Y)$. In fact, he proves a more general theorem than this for all the components (not necessarily passing through $(0, 0)$. After translation, this amounts to working over ramified extensions of $\mathbf{Z}_p$.

This theorem allows one to prove (with finite computation) that any particular D only contains finitely many points (when that is true). It also shows, without any computation at all, that D either contains Δ, or it only contains finitely many

classical points of weight different from π. A nice way to think about this theorem is that it is of the flavour as the multiplicative Manin–Mumford conjecture. That is, one is intersecting a sub-variety with a particular arithmetically defined discrete set (inside $\mathbb{A}$), and one wants to deduce that this can only happen for a well defined geometric reason. In fact, if one replaced $\Phi(X,Y)$ by a polynomial with coefficients over $\mathbf{C}$ and specialized to the case when k is always zero, then this would *exactly* be the Multiplicative Manin-Mumford conjecture in two dimensions.

As a special case, letting $k = 0$, one ends up with the following pretty result. Suppose that $\Phi(X,Y) \in \mathbf{Z}_p[[X,Y]]$ is a power series, and suppose that

$$\Phi(\zeta_1 - 1, \zeta_2 - 1) = 0$$

for infinitely many pairs of p-power roots of unity. Then the zero set of Φ contains a translate of $\mathbf{G}_m$. This exactly answers the puzzle asked by Jordan here. Explicitly, it says that the only quotients of $\mathbf{Z}_p[[\mathbf{Z}_p^2]]$ of co-dimension one which have lots of "arithmetic" points really do come from a one-dimensional subgroup!

I think that this special case (with $k = 0$) is probably easier than the general case, because one has other methods available. The argument was, however, inspired by a result of Hida which came up during his last number theory seminar at Northwestern. Translated into the language of this post, Hida's rigidity lemma corresponds to the puzzle of Jordan above in the case when $\Phi(X,Y) = Y - F(X)$ for some function $F(X) \in \mathbf{Z}_p[[X]]$.

38. The congruence subgroup property for thin groups.

Sun, 09 Mar 2014

I finally had a chance to visit Yale, which (by various orderings) is the fanciest US university at which I had never given a talk (nor even visited). The town itself struck me, at first, as a cross between Oxford and New Jersey. That aside, my coffee research led me to Blue State Coffee, which was more than up to the task of preparing a decent 8 ounce latte. (As a comparison, it is significantly better than Small World Coffee in Princeton. Small World has all the correct hipster attitude without enough of the corresponding aptitude.) Mathematically, I had a great chat with Hee Oh and Gregg Zuckerman over several hours. At one point, I raised the following idle question about thin groups.

Problem 38.1. Let $G = \mathrm{SL}_N(\mathbf{R})$ where $N \geq 2$. Let Γ be an arithmetic lattice in G. Suppose that $\Phi \subset \Gamma$ is a subgroup such that the following two conditions are satisfied:

(1) The Zariski closure of Φ in G is G.
(2) The induced map of profinite completions: $\widehat{\Phi} \to \widehat{\Gamma}$ is injective.

Then is Φ necessarily of finite index in Γ?

If $\Gamma = \mathrm{SL}_N(\mathbf{Z})$, then the first condition implies that the image of the induced map of profinite completions has finite index; I presume this is true more generally. Hence the question asked can be phrased as follows: "can congruence subgroups be determined by their pro-finite completions?" Alternatively, in the opposite direction, one can ask: "are there thin groups which satisfy the congruence subgroup property?" I have no particular reason to believe that the answer to the question above

is positive, and I might even guess that one could write down a counter-example, but I don't know how to write one down myself.

On the other hand, suppose that the answer to the question *is* positive. Then it might prove useful for determining whether, given a finitely presented group $H := \langle G \mid R \rangle$ and an explicit homomorphism:

$$\phi : H \to \Gamma$$

whether its image has finite index or (even more strongly) whether ϕ is an isomorphism onto a finite index subgroup. Namely, if the image of ϕ does not have finite index, then a positive answer to the question above would imply that H must have a finite quotient which does not come from Γ, and (since finite quotients of H may be enumerated) this leads to an algorithm which terminates if ϕ has infinite index. On the other hand, if H does have such a quotient, then certainly ϕ will not be an isomorphism onto a finite index subgroup.

This problem explicitly came up in some work of Curt McMullen (see question 5.6 of this paper), who produced explicit maps of various finitely presented groups into lattices (not quite in $\mathrm{SL}_N(\mathbf{R})$, but one can of course ask the more general question for lattices in semi-simple groups of rank at least two) and asked whether these maps were isomorphisms onto finite index subgroups. So the hope is that (in the contexts in which one expected the answer to be negative) this could always be answered by considering the pro-finite completion of the finitely presented group in question. Alas, I believe that I explicitly tried to find non-congruence quotients of the associated explicitly presented groups (in contexts where one expected ϕ to have infinite index) and didn't find any (not that I carried out this computation in anything approaching a sophisticated manner, of course).

Comment 38.2 (Anon)**.** For some related (mainly negative) results see Bridson–Grunewald's paper [BG04b].

39. Robert Coleman

Tue, 25 Mar 2014

I was very sad to learn that, after a long illness with multiple sclerosis, Robert Coleman has just died.

Robert's influence on mathematics is certainly obvious to all of us in the field. Most of my personal interaction with him was during my last two years as a graduate student at Berkeley. We would chat in his office, and sometimes have lunch at Nefeli caffe. Kevin and I had recently made some modest progress on Kevin's crazy slope conjectures, and much of that time with Robert was spent with me presenting crazy ideas and predictions on the white board in Evans Hall while Robert looked on with his classic look of amused skepticism. There would also be the occasional wine and cheese in his office, especially if an old visitor was in town.

I certainly didn't know him as well as many others did, but I felt very honored that he asked me to accompany him (as a grad student assistant) to China for his ICM address. As it happened, the relevant hotels in China would not allow him to bring Bishop (his guide dog) along with him, so he didn't end up going.

Mathematically, Robert was very original. I have no plans to attempt to summarize his research, but I just want to discuss one problem which he had thought

about in recent years, namely, what the eigencurve looked like at the boundary of weight space — especially in light of the description given by Kevin and Lloyd Kilford when $N = 1$ and $p = 2$. Suppose one is given a Fredholm determinant

$$\det(1 - UT) = P(T) = 1 + \sum_{n=1}^{\infty} a_n T^n$$

where $a_n \in \Lambda = \mathbf{Z}_p[[X]]$, and one wants to understand the spectrum of U at the "boundary" of weight space, that is, when the valuation of X goes to zero. For example, an interesting collection of points near the boundary are the classical points with highly ramified nebentypus character. If a_n is not divisible by p, then the valuation of a_n at a specialization of X close to one will coincide with the valuation of the reduction mod-p of a_n as an element of the discrete valuation ring $\mathbf{F}_p[[T]]$, that is, it will be determined by the smallest non-zero coefficient of a_n modulo p. Robert's idea was to study the "halo" of the eigencurve, which intuitively speaking, should be an object cut out by a compact operator U_χ in characteristic p with characteristic power series $P(X) \mod p$. If the valuations of the elements $a_n(X) \mod p$ define a Newton Polygon N, then the Newton Polygon at some point on the eigencurve which is sufficiently close to the boundary should be a simple multiple of N. This is one of my favourite problems! I know Robert has some ideas on how to approach this problem, but unfortunately I don't know exactly what they were or how much progress he had made. One natural question is whether this structure will ultimately be purely explainable in terms of p-adic local Langlands. One even more basic question is what happens numerically on components of the eigencurve corresponding to a representation $\overline{\rho}$ which is absolutely irreducible after restriction to a decomposition group at p; I presume one sees the same behavior, but has anyone checked this? Perhaps the easiest example to check would be to compute the slopes of forms on $S_2(\Gamma_1(11 \cdot 2^n), \chi)$, where χ has conductor 2^n.

Matt Baker has some further recollections of Robert here, and he also invites his readers to share there memories there.

Comment 39.1 (Toby Gee)**.** It won't surprise you to hear that I wanted to attack this problem with p-adic local Langlands and $R = \mathbf{T}$, but we never got anywhere. I think Kevin Buzzard and I did think about this a little in 2006 — of course these representations are still trianguline, so you can look in Colmez and see a concrete description of p-adic Local–Langlands, and then try to compute reductions mod p. Given that the expected answer is so simple, you might hope that there was some nice structure that you'd see that would explain it, but we didn't spot anything. Then again, I think we were sufficiently disillusioned with the whole approach to these kinds of questions that we didn't even explicitly bash out a single example, which is presumably possible.

40. Are Galois deformation rings Cohen–Macaulay?

Wed, 02 Apr 2014

Hyman Bass once wrote a paper on the ubiquity of Gorenstein rings [Bas63]. The first time they arose in the context of Hecke algebras, however, was Barry's Eisenstein ideal paper, where he proves (at prime level) that the completions $\mathbf{T}_\mathfrak{m}$ are Gorenstein for all non-Eisenstein maximal ideals $\mathfrak{m}$ of $\mathbf{T}$ except possibly those

which are ordinary of residual characteristic two. He also shows that the completions at Eisenstein primes are also Gorenstein, although this is trickier and makes fundamental use of the assumption that the level is prime. The Gorenstein property of various Hecke at non-Eisenstein maximal ideals was crucially used by Wiles to deduce non-minimal modularity lifting theorems. In the late 90's, including around the time I started graduate school, it seemed as though all Hecke algebras in weight two were going to be Gorenstein (localized at non-Eisenstein ideals). One case remained, however, namely when $\mathrm{char}(k) = 2$, and

$$\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_2(k)$$

has the property that $\overline{\rho}$ is unramified at 2 and, moreover, the image of Frobenius at 2 is a scalar. (The other cases having been dealt with by results of Mazur, Wiles, Ribet, and Buzzard.) But then it turned out, amazingly, that $\mathbf{T}$ was *not* always Gorenstein. Lloyd Kilford found a counter-example at level $N = 431$. The natural place to look, of course, is at $\mathrm{GL}_2(\mathbf{F}_2) = S_3$-representations. They have to come from a quadratic field K with class number divisible by three and such that 2 splits completely in the corresponding unramified degree three extension of K. It also makes sense to work at prime level, because this will make computing the integral Hecke ring easier. The condition that 2 splits in K forces Δ_K to be congruent to 1 mod 8, which certainly means the class number is odd. The condition that 2 split in the corresponding cubic field is more subtle; if the class number of the field was 3, then this would be equivalent to the primes in K above 2 splitting principally in K, but this can't happen for norm reasons. So one has to start with a quadratic field K with $\Delta_K \equiv 1 \mod 8$ and class number $h = 3h'$ for some $h' \geq 1$, and such that the class given by $[\mathfrak{p}]$ for the prime above 2 does not generate the 3-Sylow subgroup. The smallest prime number with this property is $\ldots N = 431$. So it fails at the first opportunity!

Nowadays we know, at least in the analogous context when p is odd and we are in weight p, that the appropriate Hecke algebras are Cohen–Macaulay. But we understand that the reason that these global Hecke algebras have these properties is because the *local* Hecke algebras have nice properties. The idea of deducing facts about the global Hecke algebra in the process of proving modularity lifting theorems started with Diamond, who found the first improvement to the Taylor–Wiles method. Essentially, given an $R = \mathbf{T}$ theorem, one has a presentation of $\mathbf{T}$ as a quotient of a (power series over a) local deformation ring by a sequence of parameters. If the local deformation rings are nice (Complete Intersections, Gorenstein, Cohen–Macaulay, etc.) then so is the global Hecke ring. Now this is only true in the contexts where $\ell_0 = 0$; otherwise one is taking the quotient by "too many" relations (that is, not a sequence of parameters), and so there's no longer any reason to expect that $\mathbf{T}$ has those nice properties unless $\ell_0 = 1$ and $\mathbf{T}$ is finite.

So now we come to the question: are all *local* deformation rings Cohen–Macaulay? Well, perhaps there is not really any reason to suppose that they are. Perhaps even worse, there is a paper [San14] by Fabian Sander, a student of Vytas, proving that a certain deformation ring is *not* Cohen–Macaulay. But I am not deterred. My issue is that one has to take the *correct* deformation ring. And the correct deformation ring is the one that should *include* the extra data corresponding to the local Hecke operators which may not come (at an integral level) from the Galois representation.

To take a well known example, consider ordinary p-adic representations of weight p. From a characteristic zero ordinary representation, one can always recover the

(unique) eigenvalue of Frobenius on the unramified quotient. But this is not possible at the integral level, because (for example) $\overline{\rho}$ could be locally trivial. This exactly corresponds to the fact that in weight p, the Hecke operator T_p does not have to lie in the algebra generated by the other Hecke operators (the "anemic" Hecke algebra — was that term coined by Ken Ribet?). In order to prove modularity theorems, it usually suffices to work with the anemic Hecke algebra, but when one does include data which captures T_p (or U_p) the local deformation ring *is* (in this case) Cohen–Macaulay, as was shown by Snowden. So, for example, I would conjecture that the ordinary deformation ring (in any dimension) which *includes* the local Galois information corresponding to *all* the Hecke operators is Cohen–Macaulay.

Is there any real evidence for this guess besides the fact that it would be useful? Well, not really. But it would provide a systematic local Galois explanation for why deformation rings are torsion free, which is consistent with the guess that, appropriately defined, one should latex $R = \mathbf{T}$ theorems on the nose, not just after looking at (say) MaxSpec. Of course, all of this is in the residually globally irreducible setting. Note that one reason to care about integral modularity statements is that most of the time, one would expect both R and $\mathbf{T}$ to be torsion anyway.

Notes 40.1. Some conjectural progress on these questions has been made by my student Chengyang Bao, see § 152.

41. A Preview of Barbados/Bellairs

Mon, 21 Apr 2014

This post is probably not so interesting unless you plan to travel to the Caribbean in a few weeks. The website for the conference is offline, so I thought I might update attendees on what might be happening, at least those who read my blog.

There are two hours of talks in the morning by me and two hours of talks in the evening. **Warning**: the paragraphs below are not necessarily in one-to-one correspondence with talks.

Part I: I will give an overview of the Taylor–Wiles method in something approaching its original formulation (so without Kisin's modifications). I may give the circular proof of modularity for GL(1) as an example. I will then start talking about modular forms of weight one. I will give the details of local-global compatibility as proved in my paper with David, first in the irreducible case, and then via a modification of this method in the general case (using results which will be in Joel Specter's thesis).

Background I: Jared and Peter will give a background talk on the geometry of Shimura varieties, with an emphasis on the case of modular curves, and (possibly) also that of Siegel 3-folds.

Part II: I will introduce the general strategy developed by myself and David in [CG18a] to prove modularity lifting in the $\ell_0 = 1$ and $\ell_0 = \ell_0$ situations, in particular, the details of our patching lemma. I will outline how the method naturally breaks up into several different constituent problems (constructing Galois

representations, proving local-global compatibility, proving vanishing of cohomology outside certain ranges, representation theoretic problems arising from Taylor–Wiles primes). I will then apply these strategies to prove minimal modularity lifting theorems for weight one modular forms in the residually irreducible setting.

Background II: David will talk about Kisin's modification of the Taylor–Wiles method. Toby Gee will talk about how to prove local-global compatibility and what that means in a (somewhat) general setting.

Part III: I will discuss the geometry of local deformation rings for GL(2). Topics to be covered here include classical questions of multiplicity one and two, as well as non-minimal modularity lifting theorems in weight one.

Background III: Sug Woo will discuss the relation between cohomology and automorphic forms and how the Eichler–Shimura isomorphism generalizes to higher dimensions. Jack Thorne will discuss Taylor–Wiles primes for GL(n).

Part IV: I will talk about my work with David concerning minimal modularity lifting theorems for low weight Siegel modular forms. This will consist of generalizing some of the ingredients from GL(2), such as local-global compatibility results, and vanishing results of Lan-Suh. I also discuss an approach to Taylor–Wiles primes in the torsion setting for GL(n).

Part V: I will talk about completed cohomology in low degree. I shall explain my results with Emerton on the stability of completed cohomology, and the computation of these groups using K-theory.

Related Research: I have asked a number of people to talk about their recent work on topics related to this conference. This includes George Boxer who has agreed to talk about coherent cohomology and generalized Hasse invariants, and Ila Varma who will talk about local-global compatibility for non-self-dual representations at $\ell \neq p$.

42. A postview of Bellairs/Barbados

Wed, 14 May 2014

I am just recovering from my trip to Barbados for the McGill sponsored conference at the Bellairs institute (§ 41). I thought it was a wonderfully enjoyable conference, for many reasons. The first is that I got to give 14 hours or so of talks, and I like the sound of my own voice. What was unique, however, was the really high level of the audience, not just in terms of technical strength, but in terms of their knowledge of the particular topics which were being discussed. Usually when you have a chance to talk to a specialized audience, you only have 50 minutes to speak, and for at least for the first 20 minutes or so you should not assume that your audience is *au fait* with all the latest technical developments in the subject. On the other hand, the contexts in which one has multiple hours to give details (such as a mini-course or graduate class) it's often the case that the target audience is graduate students first encountering the material. At this conference, practically half the audience had written papers proving modularity lifting theorems! I surveyed some participants beforehand on how long I should spend reviewing the basic theory of Galois deformations, and the answers typically ranged from 1 to 5 minutes. In reality, I gave a 150 minute "background" talk on the first morning, although

by background here I really mean Wiles' proof of minimal modularity lifting for irreducible modular Galois representations of $G_{\mathbf{Q}}$.

I broke the mold of previous Bellairs conferences by scheduling an additional talk in the afternoon, so typically we had some 6-7 hours of lectures per day. This sounds a lot, but when it is divided up into only three speakers and spread out from early morning to late evening, it didn't seem so much at all. (We still had plenty of time every day to snorkel at the reef, and even one free afternoon to go on a boat tour and swim with the turtles. Even Sug Woo's 200+ minute talk just flew by, although it was accompanied by rum drinks.) In addition to the background talks I mentioned previously, , there were also research talks by Peter Scholze, Jack Thorne, George Boxer, Ila Varma, and David Geraghty (I may blog about some of these talks later). I think this was the first conference in which I learned something from every single talk. Of course, I did get to suggest many of the participants, so in a way this conference was designed for me.

One outcome of the conference is that I feel confident that we will have unconditional modularity lifting theorems for $\mathrm{GL}(n)/\mathbf{Q}$ in the next five years. Of course, it's always dangerous to make predictions.

Finally, apropos of nothing, I hope to have more posts in the future whose keywords include both "Richard Taylor" and "Turtles."

Notes 42.1. Since [ACC+23] was written within five years of this conference, the prediction was correct.

43. Thurston, Selberg, and random polynomials, Part I.

Wed, 21 May 2014

Apart from everything else, you could always count on Bill Thurston to ask interesting questions. This is the first of a small number of posts which were motivated in part by figure two from this paper, and this accompanying MO question. I liked this problem enough to give it as a thesis problem to my student Zili Huang, and much of what I discuss below arose from this project (see [CH17]).

Say that an algebraic integer α is Perron if $|\alpha| \geq |\sigma\alpha|$ for every conjugate $\sigma\alpha$ of α. One immediately observes that α must be real. Say that a monic polynomial is Perron if it is irreducible and has a Perron integer as a root. Thurston's question is (roughly) to describe the distribution of Perron algebraic integers, especially those chosen in some (small) fixed interval in $\mathbf{R}$. This question has several interpretations, but one experiment Thurston does is to take 20,000 monic polynomials of degree 21 with integer coefficients in $[-5, 5]$, and plots the quantities $\sigma\alpha/\alpha \in B(1)$ for all the conjugates of the 5,932 resulting Perron polynomials such that the corresponding Perron integer was in the interval $[1, 2]$. The result is this:

The first observation is that this graph has (apart from some noise coming from real roots) rotational symmetry. The next observation is that the roots tend to be concentrated in a ring of some radius, which (from experiment) becomes more concentrated the more one restricts the range in $\mathbf{R}$ of the Perron integers one is considering. The first question is: can one explain this graph, and does it reflect reality (that is, the *actual* distribution of Perron integers)?

The answers to these questions turn out to be: yes, and no. The first problem is that it is hard (a priori) to "randomly" generate Perron algebraic integers of large

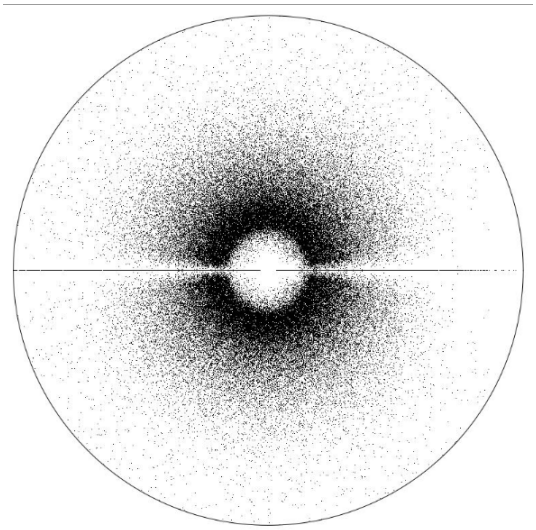

FIGURE 9. $\sigma\alpha/\alpha \in B(1)$ for Perron $\alpha \in [1,5]$ which are roots of random monic polynomials with coefficients in $[-5,5] \cap \mathbf{Z}$

degree in $[1,2]$. Knowing a bound on the roots places a bound on the coefficients, but a randomly chosen polynomial with coefficients satisfying the required bounds will almost always have a root larger than 2. Thus Thurston "cheats" with his algorithm, making the coefficients of his polynomials very small in order to increase the probability that the largest root will also be small. (Full disclosure, Thurston makes no claims that his algorithm reflects reality, and explicitly asks whether it does so or not.) The issue is then whether this will skew the distribution of the roots. It turns out that it does! To explain why this might not be surprising, let's talk about the size of the spaces over which Thurston is sampling. Let Ω_{21}^P be the set of monic polynomials of degree 21 with real coefficients and with a unique largest real root $\lambda \leq 2$. Thurston is sampling over a space with 11^{20} lattice points and volume 10^{20}. On the other hand, it turns out that the volume of Ω_{21}^P is equal to

$$\frac{2^{399}}{3^{24}5^{12}7^{10}11^{11}13^{9}17^{5}19^{3}} \sim 2.249 \times 10^{60}.$$

So Thurston was only really sampling a 10^{-40}th of the entire space! Thurston's picture can be explained as follows: polynomials with (suitably) small coefficients (contingent on the initial and final coefficients not being too small) tend to have all their roots clustering uniformly around the disc of radius one. This follows in the radial direction by a famous theorem of Erdös and Turán, and for the absolute values it follows (in a related way using Jensen's formula) from a paper of Hughes and Nikeghbali here. So the apparent "radius" in Thurston's picture is just representing $1/R$, where R is the approximate size of the Perron integers being considered. It turns out that, in reality, most of the conjugates of Perron integers have size comparable to the Perron integer itself. That is, the correct version of Thurston's picture should show the roots clustering (roughly) uniformly around the boundary.

OK, now a pause when I look at Thurston's graph and see that the radius is not something like a half as I claimed above, but something much smaller. So I just

repeated Thurston's experiment, and out of 20,000 monic polynomials with coefficients randomly chosen in [-5,5], only 1011 were Perron polynomials with largest root less than 2, and the resulting picture came out like this:

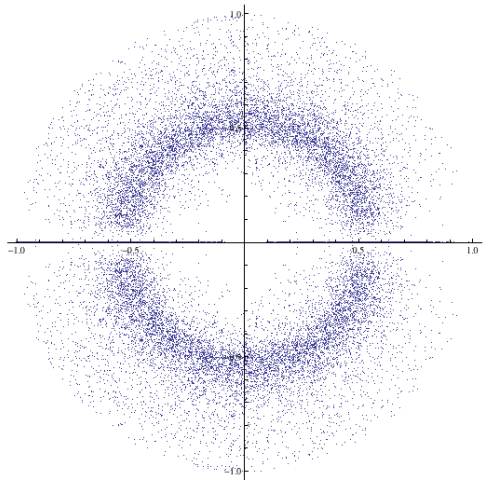

FIGURE 10. $\sigma\alpha/\alpha \in B(1)$ for Perron $\alpha \in [1,2]$ which are roots of random monic polynomials with coefficients in $[-5,5] \cap \mathbf{Z}$

Here one really sees the (misleading) accumulation around the radius 1/2. I'm guessing that Thurston actually kept all polynomials whose largest root was in $[1,5]$, which would account for the larger success rate for choosing Perron polynomials as well as the smaller radius. This is also consistent with how Thurston describes the corresponding graphs in the MO question rather than in Figure 2 of his preprint.

So how does one study Perron integers? Let us re-wind slightly and discuss a more elementary problem. How does one count algebraic integers? The most natural way to count algebraic integers is to order them by *height*. However, Thurston's problem clearly suggests a different measure, namely, to count by the size of the largest conjugate. This has a profound effect on some of the statistical properties under consideration. Roughly, algebraic integers ordered by height are much more likely to have a small number of "outliers" with large absolute value, whereas when one orders by the size of the largest conjugate, most of the other conjugates accumulate around the circle with radius the size of the largest root as the degree goes to infinity.

The problem of understanding algebraic integers of bounded size (where by bounded we mean a bound on the largest conjugate) amounts to understanding the lattice points in a certain region of $\mathbf{R}^N$. Now as long as one fixes the degree and increases the bound, such counting problems (including this one) typically reduce to a volume problem. (One also uses the fact that almost all polynomials are irreducible, and that the regions are "nice" in some explicit way, i.e. not Cantor sets.) Moreover, the corresponding regions are essentially (up to a simple stretching) independent of the bound. Hence the key region to understand is the region $\Omega_N \subset \mathbf{R}^N$ of monic degree N polynomials all of whose roots have absolute value at most one, and the region $\Omega_N^P \subset \Omega_N$ consisting of such polynomials whose largest

root is real. Of course, one is not only interested in the volumes of these regions, but also the integrals of various quantities. As an example, one can consider the integral

$$C_N(T,\alpha) = \int_{\Omega_N} P(T)|a_N|^{\alpha-1}dV$$

where $P \in \Omega_N$ represents the monic polynomial at any point, and a_N is the constant term. Evaluating this integral at $\alpha = 1$ and taking the leading term (in T) recovers the volume. On the other hand, there are some other relations. A fairly simple computation shows that

$$\mathrm{Vol}(\Omega_N^P) = \frac{4}{N(N+1)} C_{N-1}(1,1),$$

which is how one can compute the left hand side exactly for any N. In order to evaluate these integrals, it makes more sense to integrate not over the "coefficient space" of polynomials, but rather the "configuration space" of roots. The coefficient space is naturally stratified by the number of real and complex roots. For that reason, it makes sense to decompose Ω_N as

$$\coprod_{R+2S=N} \Omega_{R,S}$$

where $\Omega_{R,S}$ corresponds to polynomials whose roots all have absolute value at most one and have signature (R,S) (since we are interested only in integrals, we elide issues concerning whether one wants these spaces to be open or closed or somewhere in between). As a special case, let's think about the integral $C_{N,0}(T,\alpha)$ where one restricts the integrand to $\Omega_{N,0}$. The configuration space is simply $[-1,1]^N$. On the other hand, the map from configuration space to coefficient space is just given in terms of the symmetric polynomials, and the corresponding Jacobian matrix is the Vandermonde determinant. Hence, taking into account the action of S_N on the fibres, one finds that

$$C_{N,0}(T,\alpha) = \frac{1}{N!}\int_{[-1,1]^N} \left|\prod x_i\right|^{\alpha-1} \prod (T-x_i) \prod |x_i - x_j| dx_1 \dots dx_N.$$

This is now very reminiscent of the classical Selberg Integral. There is some beautiful mathematics related to the Selberg integral; let me direct you here for a nice survey (see [FW08]). The integrals arising here are, however, not *quite* Selberg integrals except for some very degenerate cases.

Once you start writing these integrals down, and computing some of them (by hook or crook), there are a number of problems which naturally come to mind. For example,

Question 43.1. What is the probability that a random polynomial all of whose roots have absolute value at most one is Perron?

Answer 43.2. By explicitly computing the ratio of the volume of Ω_N^P to Ω_N, you find that the answer is $1/N$ if N is odd and $1/(N-1)$ if N is even (this checks out for $N = 1, 2$).

Question 43.3. Given a polynomial all of whose roots have absolute value at most one, what the expected number of *real* roots?

Question 43.4. What the probability is (at least in even degree) that the polynomial has no real roots at all?

Having asked these questions, it is then sensible to ask the same questions for other ways of choosing random polynomials. The classical way to choose a real random polynomial is to write

$$f(x) = a_N x^N + \ldots + a_0$$

where the a_i are independent normal variables with mean zero (this is the Kac ensemble). To what extent do the statistics of random polynomials with this measure mirror the constrained problem consisting of polynomials all of whose roots have absolute value at most one? Obviously, it depends on the type of problem one considers. The most classical problem for real polynomials concerns counting the expected number of real roots. A famous theorem of Kac says that, under the ensemble above, the expected number of real roots is approximately $2/\pi \cdot \log(N)$. I recommend reading this paper for an introduction to the subject; I learnt these things from chatting with Peter Sarnak at the IAS.) The methods of Kac also show that the real roots concentrate for large N around -1 and $+1$. In fact, the complex roots also concentrate along the unit circle as well. How does this compare to our constrained model? First of all, the real roots in Kac model either lie in $[-1,1]$ or in $[-\infty,1] \cup [1,\infty]$. Certainly our polynomials have no roots in the larger region. If one restricts the Kac polynomials to $[-1,1]$, then the expected number of real roots decreases to $1/\pi \cdot \log(N)$. This is in some sense easy to see from the previous formula, because the map on coefficients $a_k \to a_{N-k}$ is measure preserving and inverts the roots. In fact, a stronger result follows from Kac. If one takes an interval $[a,b]$ strictly contained inside $[-1,1]$, then the expected number of real roots in the polynomial for sufficiently large N converges to

$$\frac{1}{\pi} \int_a^b \frac{1}{1-T^2}.$$

This gives another strong indication of how the roots are concentrating at the points $+1$ and -1. OK, so now let us return to our constrained model consisting of monic polynomials all of whose roots have absolute value at most one. How many real roots does one expect such a polynomial to have? There's a natural map

$$\Omega_{N-1} \times [-1,1] \to \Omega_N$$

which sends $P(x)$ to $P(x)(x-T)$. The Jacobian of this matrix turns out to be equal to $|P(T)|$. On the other hand, the map is not one to one, rather, the image of $\Omega_{R,S}$ has multiplicity R. Hence, if $Z(P)$ denotes the number of real roots of the polynomial P, then

$$\int_{\Omega_N} Z(P) = \int_0^1 \int_{\Omega_{N-1}} |P(T)| dV$$

The left hand side (after dividing by the volume) gives the expected number of real roots. So one is again reduced to a Selberg type integral. In this case, one apparently has (based on some Zagier-like integral guessing mojo, but unfortunately not yet Zagier-like integral proving mojo) for $N = 2m$,

$$\frac{1}{D_N} \int_{\Omega_N} |P(T)| = \frac{1}{2^{2m}\binom{2m}{m}} \left(\sum_{k=0}^m \frac{2m-2k+1}{2m+1} \binom{2m-2k}{m-k} \binom{2k}{k} T^{2k} \right) \left(\sum_{k=0}^m \binom{2m-2k}{m-k} \binom{2k}{k} T^{2k} \right),$$

and there is a similar formula for $N = 2m+1$. After some analysis to estimate the resulting integral of the RHS from $T = -1$ to 1, it turns out that, for large N, the

expected number of real roots is approximately

$$\frac{1}{\pi}\log N,$$

which is *exactly* in accordance with the Kac model! (See [CH17] for details). Indeed, if one restricts to real roots in an interval $[a,b]$ strictly in $[-1,1]$, then one also obtains the same integral formula as in the Kac ensemble. So, somewhat surprisingly to me, the number of real roots in $[-1,1]$ behaves in a very similar way whether one considers Kac polynomials or monic polynomials all of whose roots have absolute value at most one.

What then of the other problems? Given a polynomial in the Kac model of even degree $2N$, what is the probability that is has no roots in the interval $[-1,1]$? This problem was explicitly addressed in [DPSZ02] by Dembo, Poonen, Shao, and Zeitouni here, where they show (in a wide class of models) that this occurs with probability $O(N^{-b/2+o(1)})$ for some universal constant $b/2$ which they do not determine, although they estimate based on numerical evidence that $b/2 = 0.38 \pm 0.015$. What happens in our constrained model? Once more it comes down to a Selberg-like integral, this time computing the ratio of volumes:

$$\frac{\displaystyle\int_{\Omega_{0,N}} dV}{\displaystyle\int_{\Omega_{2N}} dV}$$

It turns out that one can compute this explicitly as a product of factorials. Moreover, one can compute the exact asymptotic in this case as $N \to \infty$, and the resulting probability is

$$\frac{2C}{\sqrt{2\pi}(2N)^{3/8}}, \text{ where } C = 2^{-1/24}e^{-3/2\cdot\zeta'(-1)} = 1.24512\ldots$$

(It may be hard to read in the exponent, but that is the derivative of the Riemann zeta function $\zeta'(-1)$ at -1. That may seem strange, but in fact this is a fairly typical constant that comes up in asymptotics of the Barnes-G function, which is exactly the type of expression (a product of factorials) which turns up in the evaluation of the relevant integrals.) Now the result of [DPSZ02] does not apply in our case (where the coefficients are a long way from being independent), but given the similarity in the distribution of real roots between our polynomials and the Kac model, we naturally make the following conjecture:

Conjecture 43.5. *The constant $b/2$ is equal to $3/8$.*

Optimistically, one might even try to prove this conjecture by showing that the statistics of our collection of polynomials mirror those of the Kac polynomials for sufficiently large N.

Next time: we discuss a more concrete relationship between random polynomials and our models in terms of limits of gap probabilities. But let me also leave you with the following teaser question: What is the probability that the largest root of a polynomial of degree N is real?

Notes 43.6. Conjecture 43.5 was proved in this paper [PS18] by Mihail Poplavskyi and Grégory Scheh, (not indexed by MathSciNet)

44. Thurston, Selberg, and random polynomials, Part II.

Sat, 24 May 2014

Problem 44.1. What is the probability that the largest root of a polynomial is real?

Naturally enough, this depends on how one models a random polynomial. If we take polynomials of degree N which are constrained to have all of their roots to be of absolute value at most one (with respect to the normalized Lebesgue measure on $\mathbf{R}^N$), then, as mentioned last time, the probability that the largest root is real is either $1/N$ in odd degree N and $1/(N-1)$ in even degree. A priori, this seems surprisingly small. However, the roots of such polynomials are accumulating on the unit circle, and it's easier for complex roots to be near the unit circle than real roots. So let's instead consider the Kac model of polynomials $f(x)$, where the coefficients are chosen to be independent normals with mean zero. If you ask for the probability that the root whose absolute value is closest to one is real, then I suspect that the answer will be approximately $1/N$. However, what about the largest root? The first observation is that the expected number of real roots is $2/\pi \log N$, so a the most naïve guess is that the probability that the largest root is real is approximately $(2/\pi N)\log N$. If you like, you can pause here and guess whether you think this is too high, too low, or about right.

A useful observation is that, instead of considering the largest root, we can consider the smallest root. This is because the map $a_k \to a_{N-k}$ is measure preserving and inverts the roots. On the other hand, the behavior of random Kac polynomials in large degree inside the unit circle starts to approximate the behavior of random *power series*

$$f(x) = a_0 + a_1 x + a_2 x^2 + \dots$$

where the a_i are all normally distributed with mean zero and standard deviation one. It's easy to see that $f(x)$ will have radius of convergence 1 with probability one. So we might instead consider what the probability is that the smallest root of a random power series is real. However, in this case, it is quite elementary to see that this probability P_∞ is strictly between zero and one. Quite explicitly, consider the subspace of power series such that the following inequality holds:

$$|a_0| + \frac{1}{2}|a_1 - 2| + \frac{1}{2^2}|a_2| + \frac{1}{2^3}|a_3| + \dots \leq 1.$$

This region has positive measure (easy exercise). On the other hand, for all such power series, one can apply Rouché's theorem for the contour $|2x| = 1$ to see that $f(x)$ and $2x$ have the same number of zeroes inside this disc, and hence $f(x)$ has exactly one root of absolute value less than $1/2$. By the reflection principle, this root is real. It follows that the probability that the smallest root of $f(x)$ is real is positive. Equally, one can consider the region:

$$|a_0 - 1| + \frac{1}{2}|a_1| + \frac{1}{2^2}|a_2 - 8| + \frac{1}{2^3}|a_3| + \dots \leq 1,$$

and by applying Rouché along $|2x| = 1$ and comparing with $1 + 8x^2$, the corresponding $f(x)$ will have exactly two roots inside this ball, and from the inequalities above it follows that neither of them will be real, and hence $P_\infty \leq 1$.

The same argument shows that if P_N is the probability that the smallest (or largest) root of a Kac polynomial is real, then there are uniform (independent of

N) estimates $0 \leq a \leq P_N \leq b \leq 1$ for all N. Naturally enough, one should expect that P_N converges to P_∞. This is true, and the rough idea is to show that, with probability approaching one (as $N \to \infty$), one can apply Rouché's theorem to deduce that the smallest root of $f(x)$ is real if and only if the smallest roots of its truncation $f_N(x)$ is real. The key idea here is that, for the truncation $f_N(x)$, most of the roots of $f_N(x)$ will be uniformly distributed along the unit circle, and so the contribution of the relevant factor $\prod |x - \alpha|$ to $f_N(x)$ will not be too small. Hence one can usually apply Rouché along the contour $|x| = \beta$ as long as there are no roots of $f_N(x)$ of absolute value too close to β.

The computations above also allow one to give effective gaps between P_∞ and either zero or one (by estimating the measure of the corresponding regions as translates of $|a_i| \leq 1/2^{i+1}$), although these estimates are not so sharp. Namely, the probability that the smallest root of a random power series is real is at least 0.256% and at most 99.9999999999999917%. Some numerical data suggests, however, that the probability that the largest root of a random Kac polynomial (of large degree) will be real is approximately 52%. I have some undergraduates working with me this summer, and one of their projects will be to see if they can prove that the probability is really strictly larger than 50%, or at least to find a good an estimate as they can.

One may ask what happens for other ensembles of polynomials. One natural class to consider is the so-called binomial polynomials, where the a_i are now normal with mean zero and variance $n!/i!(n-i)!$. Here the previous argument doesn't (a priori) work. On the other hand, as Boris Hanin (a Steve Zelditch student from Northwestern who is leaving for a postdoc at MIT next year) pointed out to me, it actually does: to fix it, one should *scale* all the roots of the relevant polynomials by $\sqrt{N}$, and then there really is a limit distribution as $N \to \infty$, given by power series

$$f(x) = \frac{a_0}{0!} + \frac{a_1 x}{\sqrt{1!}} + \frac{a_2 x^2}{\sqrt{2!}} + \dots$$

where the normalized a_i are normals with standard deviation one. Note that these power series have an infinite radius of convergence with probability one. The probability that the smallest root is real will once again be strictly between 0 and 1. In order to prove convergence of P_N (by applying Rouché's theorem), one needs to know that the relevant 2-point correlation functions behave reasonably enough; I'm hoping to get Boris to work out and write down the details here. Numerically, the limit probability in this case is somewhere around 62%.

44.2. **Gap Probabilities.** I speculated last time on some conjectural relationship between the space of real monic polynomials Ω_N all of whose roots are at most one, and the space of random Kac polynomials of degree N as N goes to infinity. But now I wanted to point out a more direct an elementary relationship between ensembles of random real polynomials and our space Ω_N. A gap probability is the probability that the eigenvalues/roots of some ensemble avoid some region of the corresponding parameter space. Let's compute this for a very large gap. That is, let's compute the probability that a random polynomial has all of its roots less than T as $T \to 0$.

Let's consider the Kac model of random polynomials

$$f(x) = a_0 x^N + a_1 x^{N-1} + \dots + a_N$$

where the a_i are chosen independently from a normal distribution with mean zero and standard deviation one. Hence we are asking: what is the probability that all the roots of $f(x)$ have absolute value at most T? This is simply the integral

$$\left(\frac{1}{\sqrt{2\pi}}\right)^{N+1} \int_{P\Omega_N(T)} e^{-|x|^2/2} dx$$

where $P\Omega_N(T)$ is the space of polynomials (not necessarily monic) all of whose roots are at most T. There is a map

$$\Omega_N(T) \times [-\infty, \infty] \to P\Omega_N$$

given by $(\lambda, P) \mapsto \lambda P$ with Jacobian $|\lambda|^{N+1}$. Hence we can write our quantity as an integral over $\Omega_N(T)$, which turns out to be

$$\left(\frac{1}{\sqrt{2\pi}}\right)^{N+1} \int_{-\infty}^{\infty} \int_{\Omega_N(T)} |\lambda|^{N+1} e^{-(\lambda^2 - a_1^2\lambda^2 - \ldots - a_N^2\lambda^2)/2} dx.$$

We can now compute the integral over λ directly, and then scaling $\Omega_N(T)$ to Ω_N in the usual way, we find that the probability that all the roots have absolute value T is

$$T^{N(N+1)/2} \left(\frac{2}{\pi}\right)^{(N+1)/2} \Gamma(N/2+1) \int_{\Omega_N} \frac{dV}{(1 + T^2 a_1^2 + T^4 a_2^2 + \ldots + T^{2N} a_N^2)^{N/2+1}}$$

Now suppose that $T \to 0$. Then the integral over Ω_N converges to the volume of Ω_N, and we obtain an exact asymptotic that all the roots are (highly) concentrated at zero. In fact, one can do this computation with any probability measure μ which decays sufficiently at infinity.

Curiously enough, we can also ask (in the setting of random polynomials subject to some reasonable measure μ for each i) what the probability is that a random polynomial has R real roots *contingent* on all the roots of that polynomial being less than T. It turns out that, as $T \to 0$, the answer in this case is simply the ratio of the volume of $\Omega_{R,S}$ to Ω_N (with $R + 2S = N$). This answer does not depend at all on μ. The explanation for this is that, having subjected the polynomials to the constraint that all the roots have absolute value at most T for small T, one is restricting to some tiny region where the measure is constant, and so it is converging to a scaled version of Lebesgue measure.

Notes 44.3. The summer students did not succeed, I would still like to see this question answered. Problem 44.1 was recently asked here (this time in another model where the a_i are uniformly chosen from $[-1, 1]$). Unfortunately, the question is raised without any of the accompanying insight or explanation given here, and additionally comes with a speculative conjecture (and clearly intuitively wrong) which at least is debunked in one of the answers. A certain combination of lack of insight with a speculative nonsense conjecture which takes some effort to disprove is mother's milk to that website.

45. I don't know how to prove Serre's conjecture.

Fri, 30 May 2014

I find it slightly annoying that I don't know how to prove Serre's conjecture for imaginary quadratic fields. In particular, I don't even see any particularly good strategy for showing that a surjective Galois representation — say finite flat with cyclotomic determinant for $v|p$ —

$$\overline{\rho} : G_F \to \mathrm{GL}_2(\mathbf{F}_3)$$

is modular of the right level. The first problem is that the strategy used by Wiles does not work. The results of Langlands-Tunnell imply the existence of an automorphic form π for $\mathrm{GL}(2)/F$ which has an associated finite image Galois representation into $\mathrm{GL}_2(\mathbf{Z}[\sqrt{-2}])$ with projective image A_4 that is "congruent" to $\overline{\rho}$ modulo a prime above 3, but there is no way to realize this congruence in cohomology. An analogous example over $\mathbf{Q}$ would be that the (known) modularity of a surjective even Galois representation:

$$\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{SL}_2(\mathbf{F}_4) = A_5$$

has no implications for the modularity of the corresponding even complex representation with projective image A_5 (which is "congruent" modulo 2), because there is no way to relate them via Betti or coherent cohomology.

One context in which we have a fairly satisfactory answer to Serre's conjecture over imaginary quadratic fields is for representations $\overline{\rho}$ which are the restriction of an odd representation of $G_{\mathbf{Q}}$. (I guess one also has modularity in some CM cases, that is, inductions from CM extensions H/F.) So, if we give ourselves modularity lifting results (surely a requirement to get anywhere), one could imagine trying to play some sort of game using the 3-5 switch to construct a chain between a representation which comes from $\mathbf{Q}$ and the target representation. Or, perhaps, one can play the 3-3 game using abelian surfaces with real multiplication by $\mathbf{Z}[\sqrt{7}]$. However, there's a big hole in this strategy: the 3-5 game *presupposes* that once you know that $\overline{\rho}$ is modular of *some* level, you know it at minimal level. So now one runs into the problem of level lowering. Alternatively, if you want to play the 3-5 game Khare–Wintenberger style, you really have to construct *minimal* lifts. But such lifts will not (in general) exist over imaginary quadratic fields.

This seems to be a serious problem. The only general strategies I can imagine involve being able to push the torsion classes around to different groups using some (as yet unknown) functoriality for torsion classes. (For example, find minimal lifts over some large CM extension F'/F, prove modularity over F', and then invoke non-abelian base change for torsion classes to recover modularity of the original representation.) The other argument would be to examine the corresponding Eisenstein classes for $U(2,2)/F$. This seems a little fishy, however; one would really want to see these representations inside (etale) cohomology in order to invoke some kind of Mazur principle, but as we have noted previously, the Galois representations of interest don't actually live inside the etale cohomology groups that one might want them to. Ultimately, the basic problem is that the classical (Mazur–Ribet) style arguments make strong use the geometry of modular curves (which is certainly missing here) and the more modern approaches (starting with Skinner–Wiles) rely on base change.

Notes 45.1. The optimal torsion version of Serre's conjecture in this case is certainly open, but there certainly have been a number of developments with practical consequences for the modularity of elliptic curves over imaginary quadratic fields, see [AKT23] and more recently [CN23].

46. There are non-liftable weight one forms modulo p for any p

Tue, 10 Jun 2014

In this post, we show:

Theorem 46.1. *Let p be any prime. There exists an integer N prime to p such that $H^1(X_1(N), \omega_{\mathbf{Z}})$ has a torsion class of order p.*

Almost equivalently, there exists a Katz modular form of level N and weight one over $\mathbf{F}_p$ which does not lift to characteristic zero. We shall give two different arguments. The first argument will have the virtue that the torsion class is non-trivial after localization at a maximal ideal $\mathfrak{m}$ which is new of level N. The second argument, in contrast, will produce torsion classes at fairly explicit levels. Neither proof, unfortunately, implies the existence of interesting Galois representations unramified at p with image containing $\mathrm{SL}_2(\mathbf{F}_p)$. Rather, the classes will come from deformations of characteristic zero classes. (This post is an elaboration of my comment in § 34.)

46.2. **A first proof of Theorem 46.1.** Let $K/\mathbf{Q}$ be an imaginary cubic extension unramified outside p with Galois closure $L/\mathbf{Q}$ with Galois group S_3. There is a corresponding Galois representation:

$$\rho : G_{\mathbf{Q}} \to \mathrm{Gal}(L/\mathbf{Q}) = S_3 \to \mathrm{GL}_2(\mathbf{Q}_p).$$

This representation is modular. Suppose for convenience that $p \geq 3$. Associated to ρ is an absolutely irreducible residual representation $\overline{\rho}$. Let R denote the corresponding universal unramified deformation. The only characteristic zero deformations are dihedral. Let R^{dh} denote the corresponding universal unramified dihedral deformation ring. It's easy to identify this ring explicitly; it is

$$R^{\mathrm{dh}} = \mathbf{Z}_p[C_E \otimes \mathbf{Z}_p],$$

where C_E is the class group of the imaginary quadratic subfield E of L. The ring R will fail to be $\mathbf{Z}_p$-flat exactly when $R \neq R^{\mathrm{dh}}$. Fortunately, this can be determined purely from the reduced tangent space of R. Note that

$$\mathrm{ad}^0(\rho) \simeq \rho \oplus \eta,$$

where η is the quadratic character of $E/\mathbf{Q}$. The reduced tangent space of R^{dh} is the Bloch–Kato Selmer group $H^1_f(\mathbf{Q}, \overline{\eta})$, where H^1_f denotes the subring of cohomology classes which are unramified everywhere. So it all comes down to finding K so that $H^1_f(\mathbf{Q}, \overline{\rho})$ is non-zero. However, an elementary argument using inflation-restriction shows that this is equivalent to showing that the class number h_K of K is divisible by p. So we are done provided that we can find a suitable K with class number divisible by p. (I should mention, of course, that we are using the theorem that $R = \mathbf{T}_{\mathfrak{m}}$ which was proved by me and David [CG18a].) The last step follows from the lemma below; the argument is essentially taken from this paper of Bilu–Luca (see [BL05]).

Lemma 46.3. *Fix a prime $p \geq 3$. There exists an imaginary cubic field $K/\mathbf{Q}$ of discriminant prime to p and class number divisible by p.*

Proof. Consider the field $K = \mathbf{Q}(\theta)$, where

$$(\theta^2 + 1)(\theta - t^p + 1) - 1 = 0,$$

and t is an element of $\mathbf{Q}$ to be chosen later. Note that $\theta^2 + 1$ is manifestly a unit in K. We may compute that

$$(\theta^2 + \theta + 1)\theta = (1 + \theta^2)t^p.$$

Since $(\theta, \theta^2 + \theta + 1) = (1)$ is trivial, it follows that $(\theta) = \mathfrak{a}^p$ for some ideal $\mathfrak{a}$. We shall show that, for a suitably chosen t, the element $\mathfrak{a}$ is non-trivial in the class group. If $\mathfrak{a}$ is trivial, then, up to a unit, θ is a pth power. On the other hand, the rank of the unit group of K is one, and $\theta^2 + 1$ is a unit. Hence, it suffices to choose a t such that:

(1) $\theta^2 + 1$ generates a subgroup of $\mathcal{O}_K^\times$ of index prime to p. Equivalently, $\theta^2 + 1$ is not a perfect pth power in K.
(2) None of the elements $\theta(\theta^2 + 1)^i$ for $i = 0, \dots, p - 1$ is a perfect pth power in K.
(3) The polynomial defining K is irreducible.
(4) The discriminant of K is not a square.
(5) The discriminant of K is prime to p.

By working over the function field $\mathbf{Q}(t)$ instead of $\mathbf{Q}$, one finds that the first four conditions hold for all $t \in \mathbf{Q}$ outside a thin set. (The discriminant Δ is always negative, so the signature of the field is always $(1, 1)$.) On the other hand, the discriminant of the defining polynomial is $-3 \mod t$, so if one (for example) takes t to be an integer divisible by p then the discriminant will be prime to p. Note that the set of integers divisible by p will contain elements not in any thin set, because the number of integral points of height at most H in a thin set is $o(H)$. □

46.4. **A second proof of Theorem 46.1.** Let $E = \mathbf{Q}(\sqrt{-23})$, and let $L = E[\theta]/(\theta^3 - \theta + 1)$ be the Hilbert class field of E. There is a weight one modular form of level $\Gamma_1(23)$ and quadratic character corresponding to the Galois representation:

$$\rho : \mathrm{Gal}(L/\mathbf{Q}) \to \mathrm{GL}_2(\mathbf{Q}_p).$$

Lemma 46.5. *Let $p \geq 3$. Let $q = x^2 + 23y^2$ be a prime such that $q \equiv 1 \mod p$. Equivalently, let q be a prime which splits completely in $L(\zeta_p)$. Then*

$$\#H^1(X(\Gamma_1(23) \cap \Gamma_0(q)), \omega)^{\mathrm{tors}}$$

is divisible by p. More generally, for any prime q, the quantity above is divisible by any prime divisor of $a_q^2 - (1+q)^2$, and $a_\ell \in \{2, 0, -1\}$ for a prime ℓ is the coefficient of q^ℓ in $q\prod(1 - q^n)(1 - q^{23n})$.

Proof. This follows from "level–raising" in characteristic p for weight one forms. Under the hypothesis that $a_q^2 - (1 + q)^2$, we find that there is more cohomology (over $\mathbf{Z}_p$) in level $\Gamma_1(23) \cap \Gamma_0(q)$ than is accounted for by oldforms. Assuming that there is no torsion, this is inconsistent with the fact that there are no newforms in characteristic zero, because weight one forms cannot be Steinberg at any place. (The easiest way to see this is that the eigenvalue of U_q would have to be non-integral — it also follows on the Galois side from local-global compatibility, but this is overkill.) Note that level–raising in this context does not follow from classical level–raising — for the details I refer to you my fifth lecture in Barbados on non-minimal modularity lifting theorems in weight one. □

In a weak sense, the second argument is the same as the first, except one replaces class groups with *ray* class groups, and every field has a ray class group of order divisible by p if one is allowed to choose the conductor.

Comment 46.6 (Akshay Venkatesh)**.** I still don't understand why there is the experimental "suppression" of torsion classes here by characteristic zero classes. Hopefully the Arakelov version of analytic torsion will clarify the situation eventually — at first glance, the regulator seems negligible in this case, but something else must be going on.

Notes 46.7. An echo of Akshay's comments are make by Dick in Comments 34.1. Since the Barbados lectures are not available in any form, one can also refer to [Cal18].

47. An obvious claim

Sun, 06 Jul 2014

It's been a while since I saw Serre's "how to write mathematics badly" lecture, but I'm pretty sure there would have been something about the dangers of using the word "obvious." After all, if something really is obvious, then it shouldn't be too difficult to explain why. It is especially embarrassing when someone asks you to clarify a remark/claim in one of your papers which you claim is "obvious" and you find yourself having no idea what the implicit argument was supposed to be. Such a thing happened recently to me, when Toby asked me to explain why the following was true:

Lemma 47.1. *Let $N \equiv 3 \mod 4$ be prime, and let ϵ be the fundamental unit of $K = \mathbf{Q}(\sqrt{N})$. Then $\epsilon = a + b\sqrt{N}$ where a is even and b is odd.*

Proof. Between Toby, Kevin, and myself, we managed to come up with the argument below, following a suggestion of Rebecca Bellovin: It's easy enough to see (obvious) that a and b are integers and $N(\epsilon) = 1$. Hence, it suffices to rule out the case that b even and a odd. Write $a^2 - Nb^2 = 1$. It follows that $a^2 \equiv 1 \mod N$, and since N is prime, that $a \equiv \pm 1 \mod N$. Assuming that a is odd, write $a = 2NA \pm 1$, and $b = 2B$. Then the equation above becomes

$$A(NA \pm 1) = B^2.$$

Without loss of generality, assume that A is positive. Then this equation implies that A and $NA \pm 1$ are squares, say $A = d^2$ and $NA \pm 1 = c^2$. But then

$$c^2 - Nd^2 = (NA \pm 1) - NA = \pm 1,$$

and hence $\eta = c + N\sqrt{d}$ is a (smaller) unit (in fact, $\eta^2 = \pm\epsilon$), contradicting the assumption that ϵ was a fundamental unit. □

This argument is really a 2-descent on the unit group. As Kevin remarked: "So this is a descent argument in a completely elementary situation which I don't think I'd ever seen before and which proves something that I don't think I knew . . . What's ridiculous is that if the equation had been a cubic and we were after rational solutions then I would have instantly leapt on descent as one of my main tools for attacking it :-/ We live and learn!"

So what was I thinking when I wrote the paper? The actual claim in the paper is this: "If H' is the (2 part of the) strict ray class group of K of conductor (2), then $H = H'$, where H is the (2 part of the) class group. The "argument" is as follows:

The proof of [the above] is even more straightforward: it follows immediately from a consideration of the units in $\mathcal{O}_K^\times$ and the exact sequence

$$\mathcal{O}_K^\times \to (\mathcal{O}_K/2\mathcal{O}_K)^\times \to H' \to H \to 0.$$

Well, at least the word *obvious* was only implicit here. I could try to place the blame on my co-author Matt here, but honestly the phrasing of the claim does sound a little like something I would write.

Comment 47.2 (Florent Jouve)**.** Using the descent argument in your proof one can also show e.g. that the negative Pell equation $x^2 - py^2 = -1$ is soluble in integers x, y if p is a prime congruent to 1 mod 4. In a beautiful series of paper "Higher descent on Pell conics I, II and III" (available on the arXiv) Lemmermeyer gives historical background on these questions and seems to claim that the descent argument you use goes back to (at least) Legendre. Also (if you don't mind me self advertising a bit) together with E. Fouvry we recently made crucial use of the descent argument to study the size of the solutions to Pell's equation and the (related question of the) size of regulators of real quadratic fields, e.g. here (see [FJ13]).

Comment 47.3 (Persiflage)**.** The claim about the existence of a norm -1 unit, on the other hand, is more directly obvious on the Galois side, since otherwise $\mathbf{Q}(\sqrt{p})$ would admit a quadratic extension unramified at all finite primes, which it does not. I'm sure there will be a similar argument for the claim above, except now one has to rule out the existence of a degree eight extension $E/\mathbf{Q}$ containing $\mathbf{Q}(\sqrt{N}, \sqrt{-1})$ with limited ramification properties at two. I'm sure this is not so hard, but, perhaps, not "obvious."

48. Report from Luminy

Tue, 08 Jul 2014

For how long has Luminy been infested with bloodthirsty mosquitoes? The combination of mosquitoes in my room with the fact that my bed was 6 foot long with a completely unnecessary headboard (which meant that I had to sleep on an angle with my ankles exposed) did not end well.

As for the math, there were plenty of interesting talks, most of which I will not discuss here. Jan Nekovar gave a nice talk (on [Nek18]) explaining how one could prove that the cohomology of compact Shimura varieties of GL(2)-type were semi-simple. For concreteness, imagine that X is a Hilbert modular surface associated to a real quadratic field F. Suppose that ρ is the Galois representation associated to a cuspidal Hilbert modular form of parallel weight two. Then the Langlands-Kottwitz method shows that the semi-simplification of $\rho^{\otimes 2}$ should occur inside H^2. On the other hand, this argument only ever deals with the trace of Hecke operators and so cannot say anything about semi-simplifications. Nekovar's argument is to use the Eichler–Shimura relation applied to partial Hecke operators for primes which split completely in both F and the corresponding reflex field. The point is that these operators satisfy a quadratic relation (with distinct eigenvalues for generic elements of the Galois group), and so act semi-simply on H^2 (imagine everything is compact

here). Then, by pure group theory, if the image of ρ is large enough, the sheer number of such elements is enough to force semi-simplicity. It is perhaps useful to note that if V is a representation such that $V = (\rho^{\otimes 2})$ and V is geometric and pure, then V should automatically be semi-simple. This follows from any number of combinations of bits of the standard conjectures, but one way to see it is that if W is geometric and of weight zero, then (by Bloch-Kato) one should have $H^1_f(F, W) = 0$. The relevant W in the example above is $\mathrm{ad}^0(\rho)$. So in fact one can give an alternate proof of the theorem using the full power of modularity lifting theorems, providing one is willing to omit finitely many primes p. This is really an explanation of why Jan's result is nice! For example, as soon as one replaces $\rho^{\otimes 2}$ by $\rho^{\otimes n}$, one has to start dealing with $H^1_f(F, \mathrm{Sym}^{2n}(\rho)(-n))$, which gets a little tricky.

Ana Caraiani talked about a very nice result (now [CLH16]) concerning the sign of Galois representations associated to torsion classes for $\mathrm{GL}(n)/F^+$ for totally real fields F^+ (this was joint work with Bao Le Hung). Namely, the trace of any complex conjugation lies in $\{-1, 0, 1\}$ (in fact, the result identified the exact characteristic polynomial, which is more general in small characteristic). The basic strategy is to follow Scholze's construction and "reduce" the problem to the case of essentially self-dual forms, where one has previous results by Taylor, Bellaïche-Chenevier, and Taïbi. However, there is a problem, which is that the regular self-dual automorphic forms one finds congruences with need not be globally irreducible, and perhaps not even cuspidal. Suppose one can show that they decompose into an isobaric sum $\pi = \boxplus \pi_i$ where the π_i are self-dual. One runs into problems if too many of the π_i are of dimension n_i with n_i odd. However, by considering the weights, only one of the π_i can be odd, because otherwise the Hodge–Tate weight zero would occur with multiplicity which would violate the fact that π is regular. There is still something to check for the even n_i also, because previous results required some sign assumption on the character η such that $\pi = \pi^\vee \eta$. I believe that even getting to this point required a further assumption on the torsion class not coming from the boundary. In the boundary case, there was also a reduction/induction case which also required careful handling of "odd" dimensional pieces, and some computation of a restriction of Hecke operators from the relevant Parabolic/Levi which required a sign to come out correctly. One clever technical step was working with the cohomology of adelic quotients $G(F)\backslash G(\mathbf{A})/UK$ where K is a maximal compact of $G(\mathbf{R})$ rather than the connected component K^0. The advantage of this is that, in the odd dimensional case, this pins down the trace of complex conjugation to be $+1$ rather than ± 1. This is clear when $n = 1$, and that one should expect it to be true follows for n odd by taking determinants.

Peter Scholze gave a talk on his new functor. The basic elements in the construction of this functor are as follows. The Gross-Hopkins period map allows one to view the (infinite level) Lubin–Tate tower as a $\mathrm{GL}_n(F)$-torsor over the ($D^\times$ Severi–Brauer variety) $\mathbf{P}^{n-1}_{\mathbf{C}_p}$. So, given an admissible representation π, one can form the "local system" $\mathscr{F}_\pi$ on the base, and then take its cohomology. The key technical point of this construction is to show that the result is admissible for $D^\times$, which amounts to proving finiteness of K-invariants for suitable compact open K of $D^\times$. The first step is to pull back to the (lowest level) part of the Lubin–Tate tower, which one can do because the GH map splits. Now the map from infinite level to the base of the Lubin–Tate tower is really a $\mathrm{GL}_n(\mathcal{O}_F)$-torsor, so one only has to consider the restriction of π to $\mathrm{GL}_n(\mathcal{O}_F)$. But then using the admissibility of π, one can look

instead at the regular representation of $\mathrm{GL}_n(\mathcal{O}_F)$. Now, by some sort of Shapiro's Lemma, one can pull everything back up to infinite level. At infinite level, however, one can replace the Lubin–Tate space by the corresponding Drinfeld tower. Now taking K-invariants is something that is "easy" to do, because there is an action of K on the space, and the quotient by K is some sufficiently nice object for which one has (again by Peter) some nice finiteness theorems for cohomology. I should probably have mentioned that at some point we are working with coefficients in $\mathcal{O}^+/p$, i.e. in the almost world. The main application in the talk was to show that when one patches completed cohomology (a la [CEG+16]), then one can recover the Galois representation from the result. This essentially amounts to showing that when one patches together suitable admissible π_i, one can also patch the functor. This requires more than admissibility of the functor, but some sort of "uniform" admissibility (which is always required for patching). I think the key point here is that if π_i is something patched with a group of diamond operators Δ, then π_i has a filtration by $|\Delta|$ copies of the original π, and so $\mathscr{F}_{\pi_i}$ has a corresponding uniformly bounded filtration by $\mathscr{F}_\pi$, and so $H^{n-1}(\mathbf{P}^{n-1}_{\mathbf{C}_p}, \mathscr{F}_\pi)^K$ has length at most $|\Delta|$ times the corresponding length for the (fixed for all time) version for π. On the other hand, Peter instead pulled out a new piece of kit by patching using ultra-filters. My own feeling about logic is that it is never really necessary to prove anything, and I think PS agreed that it wasn't strictly required for this particular application. Now I understand that my prejudice may not be justified (for example, it is probably hard to prove various identities concerning orbital integrals in small characteristic directly), but I think it applies in this case. Plus, as a purely expositional remark, if you are going to whip out ultrafilters during a number theory talk then everyone is just going to talk about ultrafilters rather than the beautiful construction!

49. A public service announcement concerning Fontaine–Mazur for GL(1)

Sat, 12 Jul 2014

There's a rumour going around that results from transcendence theory are required to prove the Fontaine–Mazur conjecture for GL(1). This is not correct. In Serre's book on ℓ-adic representations, he defines a p-adic representation V of a global Galois group G_F to be *rational* if it is unramified outside finitely many primes and if the characteristic polynomials of Frob_λ actually all lie in some fixed number field E rather than over $\mathbf{Q}_p$. Certainly being rational is a consequence of occurring inside the etale cohomology of a smooth proper scheme X, and one might be motivated to make a conjecture in the converse direction assuming that V is absolutely irreducible. But being "rational" is just a rubbish definition (sorry Serre), a mere proxy for the correct notion of being potentially semistable at all primes dividing p ("geometric," given the other assumptions on V). And the implication

$$\text{A character } \chi : G_F \to \overline{\mathbf{Q}}_p \text{ is Hodge–Tate} \Rightarrow \chi \text{ is automorphic}$$

doesn't require any transcendence results at all. One can't really blame Serre for not coming up with the Fontaine–Mazur conjecture in 1968. The reason for this confusion seems to be the proof of Theorem stated on III-20 of Serre's book on abelian ℓ-adic representations (with the modifications noted in the updated version of Serre's book), namely:

Theorem 49.1 (Serre–Waldschmidt)**.** *If V is an abelian representation of G_F which is rational, then V is locally algebraic.*

This argument (even for the case when F is a composite of quadratic fields, the case considered by Serre) requires some transcendence theory. But the implication V is abelian and Hodge–Tate $\Rightarrow$ V is locally algebraic (also proved in Serre) only uses Tate era p-adic Hodge theory. The other ingredients for Fontaine–Mazur are as follows: First, there is the classification of algebraic Hecke characters (due to Weil, I think). A key point here is that the algebraicity forces the unit group to be annihilated by some element in the integral group ring. However, the representation V occurs in $\mathcal{O}_F^\times \otimes \mathbf{C}$ with dimension $\dim(V|c = 1)$ if V is non-trivial, so this forces the existence of representations V of G on which $c = -1$, corresponding to CM subfields. The final step is the theory of CM abelian varieties. So although the result is non-trivial, you can be rest assured, gentle reader, that you are not secretly invoking subtle transcendence results every time you twist an automorphic Galois representation by a Hodge–Tate character and claim that the result is still automorphic.

50. 100 Posts

Wed, 20 Aug 2014

Meaningless numerical milestones are a good a reason as any for an indulgent post. Today, I will discuss some facts from this blog which you might not otherwise know about. It will be in the form of an (mercifully short) interview with myself.

Question 50.1. When did you start this blog?

I originally started it when I went to the IAS for a special year in 2010–2011, but I never ended up making the blog public at that time. The irreverence has been toned down for the current version. **Sample post from the IAS**: "Who wears short shorts? Deligne wears short shorts!"

Question 50.2. What topics would you like to blog about in the future?

No promises, but here are some thoughts:

- How does an NSF panel work?
- What are letters of recommendation really like?
- Who wore it better: piano v. orchestral arrangements.
- Langlands versus the world.
- Book reviews: Frenkel, Ellenberg, Harris.
- India's greatest mathematician: Harish–Chandra.
- The top 1%: class and privilege in academia.

Question 50.3. Does that mean you are planning to have less math in the future?

No, the math posts are not really planned in advance, they are just what I happen to be thinking about at the time. The math posts are really the main (if not exclusive) focus of this blog. Although, as one of my graduate students once remarked: "*I though your post on swans was your best post ever.*" Yeah, thanks for that, I'm working hard to bring you occasional insights into the vast edifice of algebraic number theory, and you like the guy who can wobble around to Saint–Saëns.

Question 50.4. What can your readers do for you?

More audience participation! A lot of what I write is speculative, so please don't refrain from giving your partially formed thoughts in the comments. As the readership of this blog went up, the number of comments has gone down. I think I understand this phenomenon, especially when it comes to math posts. The worst thing that can happen, however, is that you say something completely ridiculous in front of a bunch of senior number theorists. But, if you are not occasionally saying stupid things in front of smart people, then you are doing it wrong.

Question 50.5. Does the audience have any questions?

I'm going to take audience questions in the form of random search terms which led to this blog. Perhaps those who came here were disappointed with their search results at the time, but perhaps if they search again this post will provide some answers.

- 아리조나 윈터스쿨
  I recommend going here.
- *review my paper*
  No thanks!
- *how can i tell how many pages my paper is*
  Form a bijection with one of the sets defined in Part II of this volume.
- *paskunas conference 2013*
  Sounds good! Alas, I was not invited.
- *maximal unramified abelian extension of a local field*
  It's procyclic, and is generated by roots of unity of order prime to the residue characteristic. Assuming, of course, your local field is of mixed characteristic, which all the interesting ones are.
- *bush is the messiah*
  This seems to me to be an unverifiable claim.
- *galois representations matrix*
  Unfortunately, no one can be told what the Matrix is. You have to see it for yourself.
- *honorarium + editor + elsevier*
  $60 for any processed paper. It is taxable income, however.
- *galoisrepresentations+blog+who?*
  It's me!
- *bach mit pedal schiff*
  Bach *without* pedal, surely?
- *how do i find out how my paper is being reviewed*
  Oh, I can tell you that. If you are lucky, the reviewer has completely forgotten about it. Otherwise, the reviewer is currently cursing you for generally ruining his or her life.

- *compute the average rate of change from $x = -10$ to $x = 10$. enter your answer as a fraction in simplest terms using a slash (/). do not include spaces in your answer*
  $(f(10) - f(-10))/20$.
- *local even galois representations*
  I presume you are asking about representations of a local Galois group which are even. For this to make sense, you should probably talk about Galois representations at the infinite prime, that is, representations of $\mathrm{Gal}(\mathbf{C}/\mathbf{R})$ on which complex conjugation acts trivially. Let me classify those for you: they are all trivial!
- *peter scholze gowers*
  I don't think they wrote any joint papers.
- *ila varma grothendieck*
  Same answer as above.
- *ila varma galois*
  Same answer as above.
- *danny calegari brilliant; jacob lurie genius*
  self-googling, I imagine.
- *joel specters math thesis*
  I shall link to it on this blog after it has been written.
- *representation galois change of characteristic*
  I assume you are asking about the p-q-switch? There are plenty of good expositions available online.
- *affirmative motives*
  I guess these are motives which just need a little support. For only a $5 donation to this blog, I will help turn a poor motive into a bold and effective one, simply by twisting.
- *kevin buzzard chess*
  I'd be surprised if he had the time. I fancy my chances.
- *xxx agol in school*
  Personally, I rate the way Agol schooled 3-manifolds as **T**18+, suitable for topologists of ages 18 and above.
- *the math behind a waffle*
  Aah, sorry about that. I'm more an expert in the waffle behind the math. On the other hand, you can learn about the chemistry of waffles here.

Notes 50.6. Note that this is only the 50th section, which gives some indication that math blog posts account for approximately 50% of all posts. (That ratio persists, with 305 or so posts in total, of which a touch over 150 are "math".) I wrote about NSF panels in § 59. A review of Michael Harris's Book is here, a review of Jordan Ellenberg's book is here. Joel Specter's thesis is here, though you can also

find his thesis-adjacent papers [MS15, CS19, Spe18]. Some consider my post on swans still my best post ever.

51. The distribution of Hecke eigenvalues, Part I

Tue, 29 Jul 2014

Here is a question I raised at the Puerto-Rico conference during one of the "problem sessions." Toby Gee seems to remember that I had some half-baked heuristics that predicted both **A** and **B** below, but perhaps one of my readers has a more sophisticated suggestion, or even a similarly wild guess (or even a similarly contradictory collection of guesses).

Fix a pair of distinct odd primes p and l. Now consider a random normalized Hecke eigenform $f = \sum a_n q^n \in \overline{\mathbf{Z}}$ of weight two and level $\Gamma_0(N)$, where N is squarefree and prime to both p and l. Now take the Hecke eigenvalue a_l and reduce it modulo a random prime $\mathfrak{p}$ above p.

Question: As one ranges over all newforms of conductor $\leq X$, what is the resulting distribution — if it even exists — of $a_l \in \overline{\mathbf{F}}_p$?

Let me not be too precise about what random means — for example, there's a question about whether one wants to normalize in some way for Galois conjugates of eigenforms, but none of this will really matter for the very weak questions I have in mind. For example, consider the following two possibilities:

(1) **A**: The element a_l lies in $\mathbf{F}_p$ at least 100% of the time.
(2) **B**: The element a_l lies in $\overline{\mathbf{F}}_p \smallsetminus \mathbf{F}_p$ at least 100% of the time.

Here by 100% I mean as a proportion of all forms as $X \to \infty$, although I confess that I can't even rule out the extreme version of **A** where 100% really means every single form.

The specifications on the level are designed to rule out some "trivial" examples. At level $\Gamma_0(N)$ with N squarefree there will be no CM-forms, which is one cheap way to generate large coefficient fields. The level also prevents twisting by characters (an even cheaper trick). Finally, in general, it is possible to generate large coefficient fields for Galois representations by imposing a local condition at some auxiliary prime q. For example, one can impose some supercuspidal condition so that the *local* residual representation at q does not land in $\mathrm{GL}_2(\mathbf{F}_{p^m})$ for any m not divisible by an arbitrary fixed integer chosen in advance. However, this too is not possible if π_q is forced to be either unramified or special (up to unramified quadratic twist).

Note that there do exist infinitely many semi-stable modular elliptic curves, so a_l will lie in $\mathbf{F}_p$ at least infinitely often. This disproves the "extreme" version of **B**, but doesn't go very far towards disproving the asymptotic version of **B**. As for **A**, every single time you write down a normalized eigenform with coefficients in some field $E \neq \mathbf{Q}$, you disprove the extreme version of **A** for a positive density of pairs (p, l). But no finite collection of such forms can disprove **A** even for a single l and varying p, because there will always be (many) primes which split completely in any finite collection of number fields.

Here are three questions:

(1) Can you disprove the extreme version of **A** for all p and l?
(2) Can you disprove the super-extreme version of **A**, namely, show that for all primes p, there exists a newform of squarefree level N prime to p such that the residual representation is not defined over $\mathbf{F}_p$? (equivalently, replace a_l by the collection of all a_l with l prime to Np.)
(3) Can you give any heuristic that suggests that either **A** or **B** (in the weaker form) is either strong or true?
(4) Do you have any guesses as to the distribution of the a_l?

Right now, as you read this, Kevin's computer is churning away in sage generating some data, which will be the topic of Part II. But until then, I would like to hear your opinions/guesses. For me, I think that **A** is probably false, but I honestly have no feeling for **B**.

Comment 51.1 (Toby Gee)**.** I'm not sure why you're saying "at least 100%". My emails do suggest that you originally conjectured (during the problem session?) that most of the time they were in $\mathbf{F}_p$, and possibly 100% of the time asymptotically, but that you then flipflopped on that. I think the second conversation took place in the sea, though, so unfortunately I have no notes to back this claim up.

Comment 51.2 (Akshay Venkatesh)**.** Wow, what heuristic supports (**A**)? I would have thought that, most of the time the eigenvalues live in bigger and bigger extensions, so to speak, and therefore no limiting distribution? I'm sure you've thought of both of these, but both thinking about charpoly(T_l) like a random polynomial, and thinking on the Galois side seems to point against (A). On the Galois side it is a bit unclear, perhaps, how much a fixed residual representation deforms, but without thinking carefully I'd imagine that Cohen–Lenstra predicts "not too much".

Comment 51.3 (Persiflage)**.** To be fair, my flirtation with (**A**) was relatively short. (In other words, I was for it before I was against it, or something like that.) Suppose you just count $\overline{\rho}$. Then is the expected number of level $\Gamma_0(N)$ forms with image containing $\mathrm{SL}_2(\mathbf{F}_q)$ something like a constant C_q that decreases rapidly with q? I remember you told me the heuristics here, but I can never quite remember the numbers. If you are correct, though, then surely it's embarrassing that one can't disprove (**A**)?

Comment 51.4 (Akshay Venkatesh)**.** I think actually C_q doesn't decrease with q. You are right, it is embarrassing. One thing we could try is this: if the strong form of (**A**) holds, then the trace of $(T_l^p - T_l)T'$ will be zero $\mod p$ for any other Hecke operator T', and we can try to show this doesn't happen (at least for many N) via trace formula. At the least, the class numbers that show up here don't depend on N, and thus we could at the least hope to show this way that a fixed (l, p) doesn't satisfy extreme-**A** for most N with a finite amount of computation. There are also terms in the trace formula like the genus of $X_0(N)$, which we could arrange to be indivisible by p even if we know nothing about class numbers, so one might optimistically hope to get more this way.

52. The Abelian house is not closed

Sun, 03 Aug 2014

Today I will talk about $\dfrac{97+26\sqrt{13}}{27} = 7.064604\ldots$

For an algebraic integer α, the house $\boxed{\alpha}$ is the absolute value of the largest conjugate of α. Kronecker proved the following:

(1) If $\boxed{\alpha} \leq 1$, then either $\alpha = 0$ or a root of unity.
(2) If $\boxed{\alpha} \leq 2$ and α lives in a CM field, then $\boxed{\alpha} = 2\cos \pi/N$.

The first claim is well known. The second claim follows from the first: the CM condition implies that the conjugates of the squares of the absolute value are the squares of the absolute values of the conjugates. Hence, if $\zeta^2 + \zeta^{-2} = \boxed{\alpha}^2 - 2$, then ζ must be a root of unity by part one. On the other hand, beyond these two results, the respective values of $\boxed{\alpha}$ are dense in $[1, \infty)$ (general case) and $[2, \infty)$ (CM case). There are a number of ways to modify this problem. One way is to replace the largest conjugate $\boxed{\alpha}$ by the d-power mean of the absolute values:

$$M_n(\alpha) = \left(\frac{1}{[\mathbf{Q}(\alpha) : \mathbf{Q}]} \sum |\sigma\alpha|^n \right)^{1/n}.$$

For such a construction, it makes the most sense to assume either that α is totally real or lives in a CM field, so that $|\sigma\alpha| = \sigma|\alpha|$. For example, if one lets

$$\mathfrak{M}_n = \{x \in (1, \infty) | x = M_n(\alpha),\ \sigma c\alpha = c\sigma\alpha\},$$

then Chris Smyth shows [Smy84] that, for all $n \geq 0$, the smallest elements of $\mathfrak{M}_n$ are isolated, whereas $\mathfrak{M}_n$ is dense for sufficiently large x. In this post, we shall be interested in what happens when one restricts to the class of cyclotomic integers. Namely, let

$$\mathfrak{M}_n^{\mathrm{ab}} = \{x \in (1, \infty) | x = M_n(\alpha)\ \alpha \in \mathbf{Q}^{\mathrm{ab}}\}.$$

In particular, when $n = \infty$, we obtain the set $\mathfrak{M}_\infty^{\mathrm{ab}}$ consisting of the values $\boxed{\alpha}$ for cyclotomic integers α. We call $\mathfrak{M}_\infty^{\mathrm{ab}}$ the *Abelian House*. As already noted, the values of $\mathfrak{M}_\infty^{\mathrm{ab}} \cap [1, 2]$ consist of elements of the form $2\cos(\pi/N)$, which includes 2 as a limit point. However, the spectrum of $\mathfrak{M}_\infty^{\mathrm{ab}}$ for a short while beyond 2 is once again discrete. For example, the main theorem of [RW13] (previously discussed here) completely computes $\mathfrak{M}_\infty^{\mathrm{ab}}$ in the interval $[0, (5.04)^{1/2}]$ — it has a second limit point at $\sqrt{5} = 2.2360679\ldots$ and is once again discrete beyond this point. The case $n = 2$ was studied in Cassels [Cas69] and in [CMS11]. In particular, [CMS11, Theorem 9.1.1] is equivalent to:

Proposition 52.1. *The set $\mathfrak{M}_2^{\mathrm{ab}} = \overline{\mathfrak{M}_2^{\mathrm{ab}}} \subset \mathbf{R}$ is closed.*

Note that $M_2(\alpha)^2 =: \mathcal{M}(\alpha) \in \mathbf{Q}$ (the notation $\mathcal{M}$ being used in *ibid*, so this closed subset is countable and is thus very far from being dense. Moreover $M_{2n}(\alpha)^n = M_2(\alpha^n)$, so the theorem above implies that the closure $\overline{\mathfrak{M}_{2n}^{\mathrm{ab}}}$ is also countable and lives inside $\mathbf{Q}^{1/n} \cap \mathbf{R} \subset \overline{\mathbf{Q}} \cap \mathbf{R}$. The main goal of the current post is to generalize this result to the abelian house.

Theorem 52.2. *The closure of the abelian house $\overline{\mathfrak{M}_\infty^{\mathrm{ab}}}$ is a subset of $\overline{\mathbf{Q}} \cap \mathbf{R}$. If $S \subset \mathfrak{M}_\infty^{\mathrm{ab}}$ is bounded, then $\liminf S \in \mathfrak{M}_\infty^{\mathrm{ab}}$. However, $\limsup S$ is not necessarily in $\mathfrak{M}_\infty^{\mathrm{ab}}$, that is, the abelian house itself is not closed.*

One application of this is to the possible index of subfactors (see here and here for an overview of the problem):

Corollary 52.3. *Let $\alpha \in \mathbf{R} \setminus \mathbf{R} \cap \overline{\mathbf{Q}}$ be a real transcendental number. Then there does not exist a finite depth subfactor $A \leq B$ of index in the range $(\alpha - \epsilon, \alpha + \epsilon)$ for some $\epsilon \geq 0$.*

Corollary 52.4. *Let $\alpha \in \mathbf{R} \cap \overline{\mathbf{Q}}$ be an algebraic number. Then there does not exist a finite depth subfactor $A \leq B$ of index in the range $(\alpha, \alpha + \epsilon)$ for some $\epsilon \geq 0$.*

Corollary 52.5. *The set of indices of finite depth subfactors is a well-ordered subset of $\mathbf{R}$ of ordinal type ω^ω.*

Remark 52.6. Just like volumes of 3-manifolds according to Thurston and Jørgensen.) I'm assuming here that it is easy enough to construct subfactors of index

$$\prod_{i=1}^{n} 4\cos^2(\pi/p_i)$$

for distinct odd prime numbers p_i.

Since the main context here is that such indices arise as the spectral eigenvalue of graphs, it might be helpful (for contrast) to note that this latter spectrum is dense in $[\sqrt{2+\sqrt{5}}, \infty)$ (see [She89]).

This theorem came from my bag of thesis problems. I actually expected it to be the case that $\mathfrak{M}_\infty^{\mathrm{ab}}$ was closed, but this turns out to be completely false. On the other hand, the argument I had in mind to prove this theorem was roughly correct. On the third hand, it turns out that the solution to this problem was almost entirely included in a paper of Antonia Jones from the '70s [Jon75] (using the method I roughly had in mind).

52.7. **Rational Linear subspace of $(\mathbf{R}/\mathbf{Z})^k$.** Consider the standard torus $\mathbf{T} := (\mathbf{R}/\mathbf{Z})^k$ with coordinates $(x_1, \ldots, x_k)$. We define a rational linear subspace V of $\mathbf{T}$ to be the subspace cut out by any number of equations of the form:

$$\sum a_{i,j} x_i = c_j$$

for integers $a_{i,j}$ and elements $c_j \in \mathbf{Q}/\mathbf{Z}$. Topologically, V is finite disjoint union of tori. Any connected component of V is also a rational linear subspace. If all the $c_j = 0$, then we call V a rational linear subgroup. Call a point $\underline{x} \in V$ rational if $\underline{x} = (x_1, \ldots, x_k)$ where $x_i \in \mathbf{Q}/\mathbf{Z}$. Given V, the rational numbers c_j have a common denominator; let M denote *some* integer divisible by this common denominator. The map $[m] : \mathbf{T} \to \mathbf{T}$ given by multiplication by m preserves V whenever $m \equiv 1 \mod M$.

Definition 52.8. For any rational point $\underline{x}$ on V and an admissible integer M, let $L(\underline{x}) = L_M(\underline{x})$ denote the (finite) set of rational points of the form $[m]\underline{x} \in V$ for all m satisfying the following two conditions:

(1) $m \equiv 1 \mod M$,
(2) m is prime to N, where $\underline{x} = (x_1, \ldots, x_k)$ are elements of $\mathbf{Z}[1/N]/\mathbf{Z}$.

Of course, this definition comes from looking at the exponent of the conjugates of root of unity which fix an Mth root of unity. We call $L(\underline{x})$ the line through $\underline{x}$. The notion of line depends on a choice of integer M, although replacing M by a multiple only (at worst) decreases the size of $L(\underline{x})$. Our main technical lemma is the following:

Lemma 52.9. *Let $S \subset V$ be any set of rational points, and let M be admissible for V. Then the closure of $W = \bigcup L(\underline{x})$ of all lines $L(\underline{x}) = L_M(\underline{x})$ for $\underline{x} \in S$ is a union of connected rational subspaces W^0 of V.*

We shall apply this theorem to $V = \mathbf{T}$ with $M = 1$. However, in order to prove the result (by induction), it is easier to prove this more general statement.

Example 52.10. Suppose that $n = 2$, $V = \mathbf{T}$, and $M = 2$. If $\underline{x} = (1/2, 1/q)$, then the line $L(\underline{x})$ consists of points of the form $(1/2, p/q)$ with p odd and prime to q. The closure of all such points is the rational subspace $x_1 = 1/2$.

Proof. We proceed by induction on the dimension of V. We may first claim that we can assume $V = V^0$ is connected. The connected components of a rational subspace are obtained by replacing the linear equations by their saturation. However, this requires introducing numerators into the constants c_j, and so for this step (as well as several others) we must allow the auxiliary integer M to increase. If V is connected, it suffices to show that if the closure is not dense, then the points all lie on a (finite union of) co-dimension ≥ 1 rational subspaces W of V, and then apply the inductive hypothesis.

Choose a rational base point $\underline{v} \in V$. After increasing M again if necessary, we may assume that $M\underline{v} = 0$. Under this assumption, translation by $\underline{v}$ preserves lines and sends V to a connected rational linear subgroup of $\mathbf{T}$. After an integral change of basis, any connected rational linear subgroup is linearly equivalent to one of the form $a_i x_i = 0$ for a_i either zero or one, and thus, again without loss of generality, we may assume that $V = \mathbf{T}$. Now suppose that $\underline{v} \in V$ is a point which is not in the closure of the set of lines. Because the complement of the closure is open, we may assume that $\underline{v}$ is rational. Hence, once more translating by $\underline{v}$ and increasing M if necessary, it suffices to show that either 0 is in the closure of the set of lines, or the points are all contained in a subvariety defined by a linear equation. Let $\underline{x} = (x_1, \ldots, x_k) \in S$, where one may think of the x_i as being lifted to $\mathbf{Q}$. The problem is to construct an integer n with $n \equiv 1 \mod M$ and $(n, N) = 1$ such that if $\|x\|$ denotes the nearest integer to x, then $\|nx_i\| \leq \epsilon$ for all i, or to show that all the x_i satisfy some linear relation in $\mathbf{Q}/\mathbf{Z}$. Without the congruence condition on M, this is exactly a lemma proved by Davenport and Schinzel in [DS67]. Their proof does not obviously extend to this case, however. I had an idea to replace this analytic argument by using an idea of Cassels using the geometry of numbers. Write $x_i = a_i/N$ where N is the smallest common denominator (so the greatest common divisor of the a_i is one). Let $\Lambda \subset \mathbf{Z}^k$ denote the lattice

$$\Lambda := \{\lambda \equiv m(a_1, a_2, \ldots, a_k) \mod N, \quad m \in \mathbf{Z}\}.$$

The basic idea is to break up the problem into two steps: first, find an element of Λ of small length. If this element reduces under the natural map to $\mathbf{Z}/N\mathbf{Z}$ to a multiple m of $(a_1, \ldots, a_k)$ which is prime to N and $1 \mod M$, then one wins. If not, deform the element both other small vectors, and use the fact that (in an arithmetic progression) one doesn't have to go very far to find elements prime to N (by Iwaniec, $\log(N)$ or so will suffice). In the end, it turns out that this improved version is essentially proved by Jones in [Jon75]. (For me, it is easiest to modify her proof of Theorem 1 than read the notation in some of the latter theorems, but all of the required content is here.) In fact, the application Jones had in mind was almost identical to the topic of this post, namely, to study the higher derived sets

of $\mathfrak{M}_n^{\mathrm{ab}}$. For some reason, however, she did not seem to notice (or mention) the implication that $\overline{\mathfrak{M}_n^{\mathrm{ab}}}$ was a subset of $\overline{\mathbf{Q}}$, possibly because her formalism was less algebraic than what we consider below (or she wasn't interested!) □

Consider an infinite set S of roots of k-tuples of roots of unity $(x_1, \ldots, x_k)$ which is closed under the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, and view it as a subset of $\mathbf{G}_m^k$. Say that a set of k-tuples of units are constrained by a k-tuple of integers $h = (h_1, \ldots, h_k)$ if, for all such tuples,

$$(x_1)^{h_1}(x_2)^{h_2} \ldots (x_k)^{h_k} \in \zeta^{\mathbf{Z}},$$

for some fixed root of unity ζ. Since this property is preserved under taking dth roots for any fixed $d \in \mathbf{Z}$, we also insist that each constraining k-tuple consists of co-prime integers. A constraint cuts out a subvariety Z_h of $\mathbf{G}_m^k$, which in general is not geometrically connected. The intersection of any finite number of subvarieties Z_{h_i} is determined by the saturation of the subgroup of $\mathbf{Z}^k$ generated by the h_i. In particular, there exists a maximal finite set of h_i such that $Z := \cap Z_{h_i}$.

Theorem 52.11. *The supremum of the elements*

$$|y_1 + y_2 + \ldots + y_k|^2$$

in S is equal to the supremum of the quantity

$$|z_1 + z_2 + \ldots + z_k|^2,$$

over $(z_1, \ldots, z_k) \in (S^1)^k \cap Z^0$, where $Z^0 \subset Z$ is some geometrically connected component of Z^0, and Z is a variety cut out by constraints for finitely many k-tuples. The infimum of the houses

$$\overline{|y_1 + y_2 + \ldots + y_k|}$$

is realized by an element of S.

Proof. Pulling back under the isomorphism $\exp : \mathbf{T} \to (S^1)^k \subset \mathbf{G}_m^k$, the pre-image of any geometrically connected component Z^0 is a connected rational linear subspace of $\mathbf{T}$, and conversely any connected rational linear subspace gives rise to such a Z^0. Write the pre-image of $\underline{y} \in S$ as $\underline{x} = (x_1, \ldots, x_k)$. Suppose that $\underline{y} = (y_1, \ldots, y_k)$ where each y_i is a roots of unity in $\mathbf{Q}(\zeta_N)$ (with N divisible by M), so that the denominators of the x_i divide N. The action of $G := \mathrm{Gal}(\mathbf{Q}(\zeta_N)/\mathbf{Q}(\zeta_M))$ on $\underline{x}$ via $\exp^*$ sends $\underline{x}$ to mx for some m with $(n, m) = 1$. In particular, the conjugates on V precisely cut out the line $L_M(\underline{x})$ of $\mathbf{T}$ (with $M = 1$). It follows from Lemma 52.9 that the closure of $\exp^*(S)$ consists of a finite union of connected rational subspaces $W = \coprod W^0$, and hence the closure of S is the finite union of the sets $(S^1)^k \cap Z^0$ for a finite number of geometrically connected Z^0. This proves the claim concerning the supremum. For the infimum, we argue as follows. There exists a component Z^0 such that the infimum of the largest conjugate of $y_1 + \ldots y_k$ *on* Z^0 is equal to the infimum of the houses of elements of S. Let the supremum of $|z_1 + z_2 + \ldots + z_k|^2$ on this space be β. If the desired infimum is equal to β, then all elements must the same house, and the result follows immediately. If not, there exists a subset of S whose largest conjugates are bounded by $\beta - \epsilon$. But such a set can no longer be dense in $(S^1)^k \cap Z^0$. Hence, replacing S by this smallest set, we may reduce the dimension of Z^0. Continuing this process, we reduce to the case when either Z^0 is a point or all the houses of elements are the same, and in either case the result follows. Note that the supremum of an algebraic function

on $(S^1)^k \cap Z$ will automatically be algebraic — essentially by a rigidity argument. Alternatively, one can write down the equations required for a point to be a local minima, and observe that they are algebraic. To finish the proof of the main claim (except for the claim that $\mathfrak{M}_\infty^{\mathrm{ab}}$ is not itself closed) it suffices to note, following a result of Loxton, that any cyclotomic integer of absolute value at most B can be written as a sum of (at most) $L(B)$ roots of unity, so, when dealing with the closure of $\mathfrak{M}_\infty^{\mathrm{ab}}$, it suffices to consider sums of k roots of unity for a fixed k. □

Returning to what Jones does, her main result is to consider the sets $\mathfrak{M}_\infty^{\mathrm{ab}}(k)$ of cyclotomic integers which are the sum of k roots of unity, and then prove that the $k-1$st derived set consists precisely of the element $\{k\}$. In our context, it may seem as though the n-th derived set should consist precisely of the maxima of the natural function on the sets $(S^1)^k \cap Z$ where Z has codimension n. However, there is an extra degeneracy coming from the fact that multiplication by a root of unity doesn't change the house — so we may insist from the start that 1 is always one of the roots of unity of S, imposing the condition $x_1 = 1$.

52.12. **The abelian house is not closed.** We now prove that $\overline{\mathfrak{M}_\infty^{\mathrm{ab}}} \neq \mathfrak{M}_\infty^{\mathrm{ab}}$ by constructing an explicit element of $\overline{\mathfrak{M}_\infty^{\mathrm{ab}}}$ not in $\mathfrak{M}_\infty^{\mathrm{ab}}$. Indeed, the corresponding element will neither be cyclotomic nor an algebraic integer (although it will be algebraic). Consider the set of cyclotomic algebraic integers:

$$\beta = \zeta^2 + \zeta - \zeta^{-1}$$

$$\gamma = \zeta^2 + \zeta + \omega\zeta^{-1}$$

where ω is a cube root of unity, and ζ is (say) and pth root of unity for prime p. For large p, the Galois conjugates of ζ become dense in the unit circle. It follows that the supremum of $\overline{|\beta|}^2$ is the square of the maximum of the quantity

$$|X^2 + X - X^{-1}|$$

over $|X| = 1$, and similarly the supremum of $\overline{|\gamma|}^2$ is the maximum of the two quantities

$$|X^2 + X + \omega X^{-1}|, \quad |X^2 + X - \omega^{-1}X^{-1}|,$$

over the same region. One can compute this maximum, and it turns out, perhaps surprisingly, that it is equal to the value

$$\frac{97 + 26\sqrt{13}}{27} = 7.064604\ldots$$

in the first case, which is not an algebraic integer and so not in $\mathfrak{M}_\infty^{\mathrm{ab}}$, and is equal in the second case to

$$\frac{1}{27} \cdot \theta = 8.096242\ldots$$

where

$$\theta^5 - 446\theta^4 + 62377\theta^3 - 3023244\theta^2 + 57168180\theta - 351065988 = 0,$$

and $K = \mathbf{Q}(\theta)$ has discriminant $2^2 \cdot 3^5 \cdot 15619$ and Galois closure S_5. These are, perhaps, surprisingly ugly numbers for fairly simple looking maximization problems. It is clear, of course, that neither of these numbers lies in $\mathfrak{M}_\infty^{\mathrm{ab}}$, so this proves $\overline{\mathfrak{M}_\infty^{\mathrm{ab}}} \neq \mathfrak{M}_\infty^{\mathrm{ab}}$. Moreover, I think it quite likely (and quite provable, perhaps with

a certain amount of computational effort) that $\dfrac{97+26\sqrt{13}}{27} = 7.064604\ldots$ is the *smallest* number in $\overline{\mathfrak{M}_\infty^{\mathrm{ab}}} \smallsetminus \mathfrak{M}_\infty^{\mathrm{ab}}$.

Notes 52.13. I think it's still open that $\dfrac{97+26\sqrt{13}}{27} = 7.064604\ldots$ is the *smallest* number in $\overline{\mathfrak{M}_\infty^{\mathrm{ab}}} \smallsetminus \mathfrak{M}_\infty^{\mathrm{ab}}$. More generally, I think some of the ideas here would be worth writing down in more detail.

53. The distribution of Hecke eigenvalues, part II

Sat, 02 Aug 2014

Here are some numbers from Kevin promised in 51. "For the first 61595 newforms of squarefree level coprime to 15 here's the field extension of $\mathbf{F}_3$ generated by the a_5 field extensions:" Actually, I have suppressed most of this table — but it is available in source code downloadable from the arXiv.

$[\mathbf{F}_3(a_5) : \mathbf{F}_3]$	Total Number	$G_{\mathbf{Q}}$-orbits	Density of forms	Density of orbits
Totals:	61595	10740	1	1
1	4623	4623	0.07505	0.4304
2	2492	1246	0.04046	0.1160
3	2397	799	0.03892	0.07439
4	2476	619	0.04020	0.05764
5	2600	520	0.04221	0.04842
6	2142	357	0.03478	0.03324
7	2289	327	0.03716	0.03045
8	2008	251	0.03260	0.02337
9	1962	218	0.03185	0.02030
10	1530	153	0.02484	0.01425
11	1837	167	0.02982	0.01555
12	1656	138	0.02689	0.01285
13	1612	124	0.02617	0.01155
14	1638	117	0.02659	0.01089
15	1455	97	0.02362	0.009032
⋱				
102	0	0	0	0
103	0	0	0	0
104	104	1	0.001688	0.00009311

I've presented the numbers Kevin send me in various ways. The first column simply counts the field generated by a_5. The second column normalizes by the order of the field. This is a little like counting two representations which differ by an automorphism of the coefficient field as being "the same." The final two comments are then the proportion of the first two columns overall.

I'm really not quite sure what to make of this data. It does suggest that **A** is false, which is perhaps not surprising. It's not terribly overwhelming evidence for **B**, but then, law of smaller numbers and all.

Akshay's suggestion in the comments that the constants C_q could be independent of q must refer to the constants in the second last column, I believe. Of course,

it might be the case that $\mathbf{F}_3(a_5)$ is smaller than $\mathbf{F}_3(a_2, a_5, a_7, a_{11}, \ldots)$, so these numbers aren't exactly the same as the fields generated by the mod-p reductions of the eigenforms. If you squint, the numbers in this column do look somewhat constant for $n \leq 10$ or so. One can even argue that $n = 1$ might be artificially inflated exactly because the phenomenon of "slipping into a subfield" mentioned above. So I'm giving the points to AV. (Yes, that's right, there were points available and *you missed out because you didn't play the game.*)

54. Horizontal vanishing conjectures.

Fri, 08 Aug 2014

Let F be a number field, and let $\mathbf{G}$ be a reductive group over F, and let Γ be a congruence subgroup of $\mathbf{G}(\mathcal{O}_F)$. I can hear Brian objecting that this doesn't make sense without extra choices; if you have such an objection, please make such choices. Matt and I have made various conjectures concerning the vanishing of the completed cohomology groups $\widetilde{H}^n$ in the range $n \geq q_0$, where q_0 has been defined for all time by Borel and Wallach. (And what is q, you ask? Well, having just consulted [BW00] by downloading a pirated djvu copy, I can tell you that $2q = \dim(G/K)$ [BW00, §4.3, p.67]. What's that, you say — q isn't even always an integer? Nope!) Several cases of this conjecture were proved by Peter (in particular, in the Shimura variety context), but the general conjecture seems quite hard (not that the Shimura variety case was a cakewalk!). For example, when $G = \mathrm{GL}(1)$, then $q_0 = 0$ and the conjecture is equivalent to Leopoldt's conjecture. To remind you, one way of stating Leopoldt's conjecture is that the profinite topology on $\mathcal{O}_F^\times \times \mathbf{Z}_p$ coincides with the topology coming from the p-adic topology — that is, units are close if they are close modulo powers of p. In contrast, one can ask for the weaker statement that that the profinite topology on $\mathcal{O}_F^\times \times \mathbf{Z}_p$ coincides with the congruence topology, namely, the topology coming from looking at units modulo N for any ideal N. This turns out to be unconditionally true and not too difficult, although it is not quite as obvious as it may seem (the same can be said of LC). It motivates, however, the following conjecture:

Conjecture 54.1. *Horizontal Vanishing Let $n \geq q_0$. Then the following direct limit vanishes*

$$\lim_K H^n(X(K), \mathbf{F}_p) = 0$$

as K ranges over all compact open subgroups of $\mathbf{G}(\mathbf{A}_F^f)$.

There is an equivalent formulation of this conjecture in terms of group cohomology for arithmetic lattices. Because the conjecture is known for $\mathrm{GL}(1)$, one can also pass easily enough between SL and GL. For example, for $\mathrm{SL}_N(\mathbf{Z})$ and $n \geq 2$ it has the following formulation: Any cohomology class in $H^n(\mathrm{SL}_N(\mathbf{Z}), M)$ for a finite discrete module M capitulates in some congruence subgroup, providing that

$$n \geq \left\lfloor \frac{N^2}{4} \right\rfloor.$$

This vanishing is related to the concept of virtual cohomological dimension. The virtual cohomological dimension of a group G is the smallest integer m such that there exists a finite index subgroup $H \subset G$ such that every cohomology class in degree $\geq m$ capitulates in H. The notion being considered here is what one gets by

reversing the quantifiers — one only insists that the classes capitulate in smaller and smaller groups (in addition, we insist that H is a congruence subgroup, although that is not too restrictive when the rank is ≥ 2). There is a trivial bound $m \geq q_0$, but this bound is not at all sharp. Since this seems an a priori interesting notion, let's define it:

Definition 54.2. pro-virtual cohomological dimension: Let G be a group, and let p be a prime. Say that $\mathrm{pvcd}_p(G) \leq m$ if, for every discrete G-module M annihilated by p, and every cohomology class $[c] \in H^n(G, M)$ for some $n \geq m$, there exists a finite subgroup H so that the restriction of $[c]$ to $H^n(H, M)$ vanishes. Say that G has pro-virtual cohomological dimension $\leq m$ if $\mathrm{pvcd}_p(G) \leq m$ for all p.

I have nothing profound to say about whether this concept is relevant beyond the example at hand. As you can see for $\mathrm{SL}_N(\mathbf{Z})$, the virtual cohomological dimension and pro-virtual cohomological dimension are conjecturally quite different, the latter being given conjecturally by the formula above (at least when $N \geq 2$), and the former by $\binom{N}{2}$.

I wanted to remark in this post that the Horizontal Vanishing conjecture is, at least after localization at a non-Eisenstein maximal ideal $\mathfrak{m}$, a consequence of modularity lifting results (in the spirit of [CG18a]). Namely, the entire point of that method is that the patched complex has cohomology concentrated in a single degree (q_0), which amounts to saying that cohomology classes in $H^{q_0+i}(X(K), \mathbf{F}_p)$ can be annihilated after passing to some auxiliary level coming from some choice of Taylor–Wiles primes. Now many aspects of this argument are still conditional (note that to annihilate classes of deep p-power level, one would need corresponding local-global compatibility relating Galois representations associated to torsion classes to quotients of Kisin deformation rings, at the very least), but perhaps it is a less hopeless task than trying to prove Leopoldt's conjecture.

It's instructive to consider what is possibly the simplest case of this conjecture beyond Shimura varieties, namely, GL(2) over an imaginary quadratic field (here $q_0 = 1$, so the claim is that one can kill classes in H^2). Here at least one doesn't have to worry about vanishing of cohomology outside the indicated range. Local-global compatibility is still a problem, but one possibly way to get around this is to work at all p-power levels at once, namely, to patch the completed cohomology groups. (Matt, Toby, and I chatted over roast duck at Sun Wah what patching completed cohomology for general groups should look like.) Since one certainly has Galois representations, one gets "for free" the fact that the patched modules are modules over the appropriate power series ring of the local deformation ring. On the other hand, as Matt cautioned at the 'Pig, it is no longer so easy to do naive arguments with codimensions, because the patched objects are not finitely generated over the ring of diamond operators, but only over a non-commutative group algebra, which leads into questions relating to the size of the corresponding p-adic representations, which leads back to questions concerning local-global compatibility in p-adic local Langlands.

I wonder, however, if there are any softer arguments in any special cases.

Comment 54.3 (Summary)**.** There was some clarification that this is quite different from the notion defined by Serre about "good" groups which it's somewhat

orthogonal to what is being considered here. The "interesting" cohomology of arithmetic groups is precisely the cohomology that is *not* coming from the "local" (in the number theory sense) cohomology of the congruence completion, so the general concept of goodness is in some sense opposite to what one wants in this case.

55. Is Serre's conjecture still open?

Sun, 10 Aug 2014

The conjecture in [Ser87] has indeed been proven. But that isn't the entire story. Serre was fully aware of Katz modular forms of weight one. However, Serre ~~was too timid~~ was prudently conservative and made his conjecture only for weights $k(\rho) \geq 2$.

Well, perhaps I am overstating the case; we may as well quote Serre himself here:

> Au lieu de définir les formes paraboliques à coefficients dans $\mathbf{F}_p$ par réduction à partir de la caractéristique 0, comme nous l'avons fait, nous aurions pu utiliser la définition de Katz [Kat73], qui conduit à un espace a priori plus grand ... Il serait également intéressant d'étudier de ce point de vue le cas $k = 1$, que nous avons exclu jusqu'ici ; peut-être la définition de Katz donne-t-elle alors beaucoup plus de représentations ρ_f?
>
> Instead of defining the cusp forms with coefficients in $\mathbf{F}_p$ by reduction from characteristic 0, as we did, we could have used the definition of Katz [Kat73], which leads, a priori, to a larger space ... It would also be interesting to study from this viewpoint the case $k = 1$ we have ruled out so far; Perhaps Katz's definition gives more representations ρ_f?

In his Inventiones paper [Edi92] on the weight in Serre's conjecture, Edixhoven does give the correct formulation where one allows $k(\rho) = 1$ and correspondingly also Katz modular forms. The bridge between the two conjectures essentially consists of two further conjectures: first, that Galois representations associated to residual weight one forms are unramified, and second, unramified modular representations come from weight one.

The first progress on this problem was actually pre-Edixhoven, namely, Gross' companion form paper [Gro90] in Duke. Gross deals with both directions in the case when $\rho(\mathrm{Frob}_p)$ has distinct eigenvalues (I guess the assumption in the direction weight one $\Rightarrow$ unramified is that the eigenvalues of $X^2 - a_p X + \chi(p)$ are distinct). Of course, there was the famous matter of the "unchecked compatibilities," (I'm not one for checking compatibilities myself, to be honest) which have certainly been resolved at this point (does Bryden Cais do this in his thesis? I think he does) The next step was the work of Coleman–Voloch [CV92], who deal with the remaining case under the additional assumption that p is odd. So this leaves the case $p = 2$. Somewhat more recently, Gabor Wiese showed (see [Wie14]) that weight one Katz modular forms *do* give rise to unramified representations without any assumptions. So this leaves:

Conjecture 55.1 (Serre's Conjecture as formulated by Edixhoven)**.** *Let*

$$\rho : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{F}_q)$$

be an absolutely irreducible modular representation of characteristic 2. Assume that ρ is unramified at 2 and that the semi-simplification of $\rho(\mathrm{Frob}_2)$ is scalar. Then ρ is modular of weight one.

Wiese also explicitly dealt with the case when ρ was (projectively) dihedral, so we can assume that ρ is absolutely irreducible with non-dihedral image. Suppose that the Serre level is N. Let $\mathfrak{m}$ denote the maximal ideal of the weight two Hecke algebra which does not include the Hecke operator T_2. Let's imagine we are working with Hecke algebras over some sufficiently large extension $\mathcal{O}_E$ of $\mathbf{Z}_2$ with residue field k so to include enough Frobenius eigenvalues. It suffices to prove that

$$\dim_{\mathbf{T}/\mathfrak{m}} H^0(X_1(N)/k, \omega^{\otimes 2})[\mathfrak{m}] \geq 2,$$

because then we will have found two modular forms f and g which are Hecke eigenvalues for all Hecke operators away from p, and by the q-expansion principle, some linear combination of f and g will have to be the square of the desired weight one form.

Let R_{loc} denote the Kisin deformation ring at two for $\rho|D_2$ for the decomposition group D_2 at 2, (this is just the ordinary deformation ring, in the sense of Geraghty). Let $R^\dagger_{\mathrm{loc}}$ denote the augmented deformation which also includes the crystalline Frobenius eigenvalue T_2 (or, to put it differently, the eigenvalue of Frobenius on the "unramified quotient" U_2, where the former is meant in a sense that can and does make sense integrally. By Hensel's lemma, both pieces of added data are equivalent.) Now one uses the modularity machine, which is OK by Khare–Wintenberger for $p = 2$ because we are in the non-dihedral setting. Let's patch the Betti cohomology of modular curves following KW, except now working with the modified global Kisin deformation ring $R^\dagger$ which remembers crystalline Frobenius, and the full Hecke algebra $\mathbf{T}^\dagger$ which includes T_2. Now $R^\dagger_{\mathrm{loc}}$ is a domain with formally smooth generic fibre (this is proved in Snowden's paper [Sno18] — the ring in question is denoted $\widetilde{R}_3$ in *ibid.*). Hence, by Kisin-Khare–Wintenberger method, we obtain an isomorphism $R^\dagger[1/\varpi] = \mathbf{T}^\dagger[1/\varpi]$. However, because $R^\dagger_{\mathrm{loc}}$ is in addition Cohen–Macaulay, this can be upgraded to an $R^\dagger = \mathbf{T}^\dagger$ theorem. (It might be cleaner to instead patch coherent cohomology — multiplicity one [which always holds with T_2 included] implies that the patched module is free of rank one, which makes it easy to deduce the integral $R^\dagger = \mathbf{T}^\dagger$ theorem.) By considering the action of $\mathbf{T}^\dagger$ on coherent cohomology, however, our multiplicity one assumption allows us to deduce by Nakayama that $\mathbf{T} = \mathbf{T}^\dagger$ (more trivially: the space of modular forms with coefficients in $E/\mathcal{O}_E$ with $\mathcal{O}_E/\varpi = k$ is co-free of rank one over both of these rings) and so $R \to R^\dagger = \mathbf{T}^\dagger = \mathbf{T}$ is surjective. However, there cannot be a surjection $R \to R^\dagger$, because there is a map $R^\dagger \to k[\epsilon]/\epsilon^2$ which is trivial as a Galois deformation but is non-trivial for (the Galois avatar of) T_2. For example, in the trivial case, this just amounts to saying that the trivial representation of $G_{\mathbf{Q}_2}$ to $\mathrm{GL}_2(k[\epsilon]/\epsilon^2)$ can be thought of as "ordinary with eigenvalue $1 + \epsilon$." It follows that multiplicity one without T_2 cannot hold. Thus Serre's conjecture is true!

56. $K_2(\mathcal{O}_F)$ for number fields F

Fri, 05 Sep 2014

Belabas and Gangl have a nice paper [BG04a] where they compute $K_2(\mathcal{O}_E)$ for a large number of quadratic fields E. There main result is a method for proving upper bounds for $K_2(\mathcal{O}_E)$ in a rigorous and computationally efficient way. Tate had previously computed these groups for small imaginary quadratic fields by hand; — the problem is finding an efficient way to do this in general. (Brownkin and Gangl had previously found a non-rigorous way of computing these groups using $K_3(\mathcal{O}_E)$ and regulator maps, but more on that later.) A good analogy to keep in mind is the problem of computing the class groups of imaginary quadratic fields. In the latter case, however, there are rigorous ways to determine whether an element in the class group is non-trivial, and this is missing from the computation of $K_2(\mathcal{O}_E)$. To produce lower bounds, [BG04a] use theorems of Tate and Keune to relate the p-primary part of $K_2(\mathcal{O}_E)$ to class groups of $E(\zeta_p)$, which they can then compute in some cases. One nice example they give is

$$K_2\left(\mathbf{Z}\left[\frac{1+\sqrt{-491}}{2}\right]\right) = \mathbf{Z}/13\mathbf{Z}.$$

Akshay and I used this as one of the examples in our paper; in our context, it implies that the order of the group

$$H_1(\Gamma_0(\mathfrak{p}), \mathbf{Z})$$

is always divisible by 13 where $\Gamma_0(\mathfrak{p})$ is the congruence subgroup of $\mathrm{PGL}_2(\mathcal{O}_E)$ for $E = \mathbf{Q}(\sqrt{-491})$ and $\mathfrak{p}$ is any prime — even though the group $H_1(\Gamma, \mathbf{Z}) = (\mathbf{Z}/2\mathbf{Z})^{26}$ is *not* so divisible. (Because we are talking about PGL rather than PSL, the cusps are quotients of tori by involutions, so only contribute 2-torsion to H_1. This group is occasionally infinite; we use the convention that ∞ is divisible by 13.) It's always nice to see a theoretical argument come to life in an actual computation — fortunately, Aurel Page was kind enough to compute a presentation for Γ in order for us to do this. Now that I think about it, this and many other interesting things didn't make it into the submitted version of the paper; you'll have to read the "directors cut" to learn about it.

Alexander Rahm pointed out to me that the computation of $K_2(\mathcal{O}_E)$ we used was annotated with an asterisk in [BG04a], meaning that what was proved was only an upper bound. The issue is as follows. Let $p = 13$, and let $F = E(\zeta_p)$, let $G = \mathrm{Gal}(F/E) = (\mathbf{Z}/p\mathbf{Z})^\times$, and let $\mathrm{Cl}(F)$ denote the class group of F. What is required is to show, in light of Tate's work on K_2, is that

$$(\mathrm{Cl}(F)[p])^{G=\chi^{-1}} \neq 0,$$

where $\chi : G \to \mathbf{F}_p^\times$ is the cyclotomic character. The problem is that F has degree 24, and it is difficult to compute class groups explicitly in such cases. Let $H = \mathrm{Gal}(F/\mathbf{Q})$, so there is a canonical decomposition $H = G \times \mathbf{Z}/2\mathbf{Z}$. There are two extensions of χ to H, given (with some abuse of notation) by χ and $\chi\eta$, where η is the non-trivial character of $\mathrm{Gal}(F/\mathbf{Q})$. The main conjecture of Iwasawa Theory (Mazur–Wiles) allows one to easily compute minus parts of class groups in terms of L-values without actually computing with explicit number fields. However, we should not expect this to help us here. Namely, it's not hard to show that there is an isomorphism $(\mathrm{Cl}(\mathbf{Q}(\zeta_p))[p])^{G=\chi^{-1}} \simeq (\mathrm{Cl}(F)[p])^{H=\chi^{-1}}$. However, the former is trivial by Herbrand's theorem, because $B_2 = 1/6$ is not divisible by 13. That leaves us with the problem of proving that $(\mathrm{Cl}(F)[p])^{H=\chi^{-1}\eta} \neq 0$, which is a statement about the class group of a totally real cyclotomic extension. Since $\chi\eta^{-1}$ is an even

character, we get some savings by working in the totally real subfield F^+ of degree 12. Now pari happily tells me via `bnfinit` and `bnfclgp` that the class group of this field is $\mathbf{Z}/13\mathbf{Z}$, so it looks like we are in good shape. However, pari has the habit when computing class groups of assuming not only GRH but something stronger. What information does bnfinit actually contain? It certainly gives, *inter alia*:

(1) The Galois automorphisms of F^+, using `nfisisom(nf,nf)`.
(2) A finite index subgroup V of the unit group $U := \mathcal{O}_{F^+}^\times$, using bnfinit[8][5]

Let me show how, just with this data, one can prove that the relevant part of the class group we are interested in is non-zero. BTW, if you tell pari can you confirm this answer is really correct? (using `bnfcertify`) it complains, and says the following:

> *** bnfcertify: Warning: large Minkowski bound: certification will be VERY long. *** bnfcertify: not enough precomputed primes, need primelimit 59644617.

A rough guess (in part) as to what it might be doing: to compute all the invariants necessary for class field theory, one needs to know the full unit group. To do this, one can take the units V found so far and saturate them in the entire unit group U. For each prime q, one can do this by taking representatives in V/qV and determining whether or not they are perfect qth powers. By taking enough primes, on either rules out the existence of such an element, or finds a candidate $v \in V$ and then checks whether it is a qth power. On the other hand, from V, one can compute a pseudo-regulator R_V, which is related to the actual regulator R_U by the unknown index. So to make this computation finite, it suffices to have some a priori bound on the regulator (to give an upper bound on the index), which will ultimately come down to some a priori bound on an L-value at one, which GRH probably tells you something useful about.

One can identify the automorphisms of F^+ computed by pari with the elements of the Galois group given by the corresponding quotient of $H = G \times \mathbf{Z}/2\mathbf{Z}$ by $(-1, 1)$. This group is generated by the image of $\sigma = (2, 1) = \mathrm{Frob}_2$, so it is enough to find the automorphism σ such that $\sigma\theta - \theta^2 \equiv 0 \mod 2$. View $\chi\eta$, a character of degree 12, as being valued in $\mathbf{F}_{13}^\times$. Now choose a random unit, say `bnf[8][5][6]` (Warning! I have a feeling that `bnfinit` does something different each time you run it, which means you might have to tweak the choice of index 6 above if you are doing this at home. And by "you," I really mean "me" in six months time. I guess I should also tell myself that the relevant pari file is

`\~fcale/Zagier/BG491`

We may write down a second unit as follows:

$$\epsilon = \prod_{i=0}^{12} (\sigma^i(u))^{\chi\eta(\sigma^i)} \in (\mathcal{O}_F^\times/\mathcal{O}_F^{\times p})^{H=\chi^{-1}\eta}$$

What we have done is apply the appropriate projector in the group ring $\mathbf{F}_{13}[H]$ to u. Naturally enough, we can lift ϵ to an actual unit in F^+.

Now choose an auxiliary prime q which splits completely in F, say $q = 38299$. I chose this because it actually splits completely in $\mathbf{Q}(\zeta_{13\cdot 491})$, which will make a computation below slightly easier. We reduce ϵ modulo a prime $\mathfrak{q}$ above q in $\mathcal{O}_{F^+}$ and we find that

$$\epsilon^{(q-1)/13} \not\equiv 1 \mod \mathfrak{q}.$$

What this last computation proves is that ϵ actually generates

$$(\mathcal{O}_F^\times/\mathcal{O}_F^{\times p})^{H=\chi^{-1}\eta},$$

which has dimension one by Dirichlet's theorem. Note also that the inequality above does not depend on the choice of $\mathfrak{q}$ — any other choice is conjugate to $\mathfrak{q}$ which replaces ϵ by $\sigma\epsilon$ and the latter is a non-zero scalar multiple of the former modulo 13th powers by construction.

On the other hand, let ζ be a primitive $13 \cdot 491$th root of unity. Then we may consider the projection of $1-\zeta$ modulo 13th powers to the $\chi^{-1}\eta$ eigenspace (the latter is naturally also a character on $F(\zeta_{491})$). Remember this eigenspace is generated by ϵ. Take $q = 38299$ again, so $q - 1 = 13 \cdot 491 \cdot 6$. Then 2^6 is a primitive $13 \cdot 491$th root of unity modulo q. On the other hand,

$$\left(\prod_{(\mathbf{Z}/13\cdot 491\mathbf{Z})^\times} (1-2^{6n})^{n\left(\frac{n}{491}\right)} \right)^{(q-1)/13} \equiv 1 \mod q$$

(The exponent of $(1-2^{6n})$ is just the value of $\chi\eta(n)$ — remember that the character gets inverted in the projection formula — and that $\eta^{-1} = \eta$.) This implies that the projection of $(1-\zeta)$ to the $\chi^{-1}\eta$-eigenspace of units modulo $p = 13$ is trivial, because the image of $\epsilon^{(q-1)/13}$ computed above was not $1 \mod q$. The same is trivially true for the units in $\mathbf{Q}(\zeta_{491})$ and $\mathbf{Q}(\zeta_{13})$, because the projection of any unit in a subfield of F can only be an eigenvalue for a character of the corresponding quotient of the Galois group. In particular, if C denotes the group of *circular* units, we have shown that the map

$$(C/13C)^{\chi^{-1}\eta} \to (\mathcal{O}_F^\times/\mathcal{O}_F^{\times 13})^{\chi^{-1}\eta}$$

is the zero map. This proves that the index of the circular units in the entire units is divisible by 13. This is enough to prove that 13 divides h_F^+, but even better, by the Gras conjecture (also proved by Mazur–Wiles, following Greenberg) it follows that the $\chi^{-1}\eta$-part of the class group is non-zero, and hence, given the previous upper bound, this gives a proof that

$$K_2(\mathcal{O}_E) = \mathbf{Z}/13\mathbf{Z}.$$

56.1. **Further Examples.** Let's now look at other examples in [BG04a]. Consider the following example:

$$K_2\left(\mathbf{Z}\left[\frac{1+\sqrt{-755}}{2}\right]\right) =^? \mathbf{Z}/41\mathbf{Z} \oplus \mathbf{Z}/2\mathbf{Z}.$$

Let $F = E(\zeta_{41})$, and let F^+ be the totally real subfield of F of degree 40. Well we certainly won't be able to say so much about the class group of F^+. On the other hand, we can do the latter part of the computation, namely, testing that the $\chi^{-1}\eta$-eigenspace in the circular units looks like it has index divisible by p in the entire units. For example, if $q = 123821 = 1 + 41 \cdot 755 \cdot 4$, we can compute that

$$\left(\prod_{(\mathbf{Z}/41\cdot 755\mathbf{Z})^\times} (1-2^{n(q-1)/(41\cdot 755)})^{n\left(\frac{n}{755}\right)} \right)^{(q-1)/13} \equiv 1 \mod q$$

For good measure, the same congruence holds for the next seven primes which split completely in $F(\zeta_{755})$. (One also has to check that the multiplicative order of 2 for all these primes is co-prime to $41 \cdot 755$.) But, although this is compelling

numerically, it doesn't prove anything. If ϵ is a generator of $(\mathcal{O}_F^\times/\mathcal{O}_F^{\times p})^{\chi^{-1}\eta}$, it might be the case that $\epsilon^{(q-1)/p} \equiv 1 \mod \mathfrak{q}$ for $\mathfrak{q}$ above the first thousand primes of norm $q \equiv 1 \mod p$. This would simply correspond to a certain ray class group being divisible by p. By Cebotarev, we know that we can find *some* prime q for which this congruence does not hold, but explicit Cebotarev bounds tend to be rubbish in practice.

If we re-think our original computation, what we really want is a "generic" unit of F^+ in order to project. Since F is abelian, we actually know how to compute a finite index subgroup of the unit group, namely, by projecting (via the norm map) the group of circular units from some cyclotomic overfield. Of course, this exactly won't be good enough to find a candidate unit ϵ. One approach is to take our lattice $V \subseteq U = \mathcal{O}_{F^+}^\times$ and saturate it. Now we only have to saturate it at $p = 41$. In fact, we only need to saturate the $\chi^{-1}\eta$ eigenspace, which is one dimensional. That is, it suffices to show that

$$e_{\chi^{-1}\eta} N_{F(\zeta_{41})/F^+}(1-\zeta)$$

is a pth power in F^+. (Before taking the norm, the element is already in F^+ up to pth powers, and $[F(\zeta_{41}) : F^+]$ has order prime to 41.) But if I ask pari to compute the following:

$$\left(\prod_{(\mathbf{Z}/41\cdot 755\mathbf{Z})^\times} (1-\zeta^n)^{n\left(\frac{n}{755}\right)}\right)^{(q-1)/13} \mod \Phi_{41\cdot 755}(\zeta)$$

it complains and conks out. Well, probably René Schoof could do this computation, but let's think about these things a little differently.

56.2. **Higher Regulators.** So far, we've been relying on the fact that the fields E we are considering are abelian, in order to be able to explicitly write down some finite index subgroup of the full unit group using circular units. But what if we want to compute $K_2(\mathcal{O}_E)$ for non-abelian fields E? For this, I want to talk about an earlier paper of Gangl with Brownkin [BG99]. Their approach is through the study of higher regulators. Borel constructs a higher regulator map for odd K-groups (the even ones are trivial after tensoring with $\mathbf{Q}$). For imaginary quadratic fields and K_3, this amounts to a map

$$K_3(\mathcal{O}_E) \to (2\pi i)^2 \cdot \mathbf{R},$$

where the co-volume of the image is a rational multiple of $\zeta_E(2)$. The Quillen–Lichtenbaum conjecture predicts that the covolume differs exactly from $\zeta_E(2)$ by a factor coming from the torsion in $K_3(\mathcal{O}_E)$, which has order dividing 24, some slightly mysterious powers of 2 which I will ignore, and — the most relevant term for us — the order of $K_2(\mathcal{O}_E)$. Now the Quillen–Lichtenbaum conjecture is true. So how does this help to compute anything? Well, first one has to ask how to compute $K_3(\mathcal{O}_E)$. As an abelian group, it is easy to compute, but this is not enough to compute the regulator map. One could give explicit classes in $\pi_3(\mathrm{BGL}(\mathcal{O}_E)^+)$, of course, but that may not be the most practical approach. It turns out that the group K_3 is computable in a natural way because of its relation to the Bloch group $B(E)$, due to theorems of Bloch and Suslin. (That is, via the Hurewicz map we get classes in $H_3(\mathrm{GL}_N(\mathcal{O}_E), \mathbf{Z})$ which turn out to be seen by GL_2.) To recall, the Bloch

group is defined as the quotient of the pre-Bloch group:

$$\sum n_i[x_i], x_i \in E^\times, \text{ such that } \sum n_i(x_i \wedge (1-x_i)) = 0 \in \bigwedge^2 E^\times$$

by the 5-term relation

$$[x] - [y] + \left[\frac{y}{x}\right] - \left[\frac{1-y}{1-x}\right] + \left[\frac{1-y^{-1}}{1-x^{-1}}\right] = 0, x, y \in E^\times \smallsetminus 1.$$

Now the Bloch group admits a very natural regulator map

$$B(E) \to \mathbf{R}^{r_2}$$

(where E has signature (r_1, r_2)) given by (under the various complex embeddings) the Bloch–Wigner dilogarithm

$$D(z) = \mathrm{Im}(\mathrm{Li}_2(z)) + \arg(1-z)\log|z| \in \mathbf{R}.$$

Now all of this is (almost) very computable. Namely, one can replace $E^\times$ by the S-units of $\mathcal{O}_E$ for some (as large as you can) set S, compute the pre-Bloch group, then do linear algebra to find the quotient. Since (roughly) $K_3(\mathcal{O}_E) = \mathbf{Z}^{r_2} \oplus T$ for an easy to understand finite group T which has order dividing 24, as soon as one has a enough independent elements in the Bloch group (which can be detected by computing $D(z)$) you can compute a group $B_S(E)$ which has finite index in $B(E)$. Moreover, the dilogarithm is also easy to compute numerically, and so one can compute a regulator $D_S(E)$ coming from the Bloch group. Now this regulator map is known to be rationally the same as Bloch's regulator map (by Suslin and Bloch). Assuming this is also true integrally, we expect there to be a formula:

$$\frac{3|d_E|^{3/2}}{\pi^2 D(E)} \cdot \zeta_E(2) =^? K_2(\mathcal{O}_F),$$

at least for primes $p \geq 3$. (The 3 is coming from the torsion of K_3, and this formula is probably only true up to powers of 2 — this formulation above comes from Brownkin–Gangl [BG99].) For S big enough, $D_S(E)$ should stabilize to $D(E)$, which gives a method of computing the order of $K_2(\mathcal{O}_E)$. This is what Brownkin and Gangl do. There are two issues which naturally one has to worry about. The first is that it's not known that the regulator map coming from dilogarithms is the same on the nose as Bloch's map. However, even granting this (and it should be true), this algorithm will not certifiably end, because one can never be sure that $D_S(E) = D(E)$. If you compare this to the computation that pari is doing with the class group, the problem is that there is no *a priori* bounds on the size of the corresponding regulators. Well, I guess this algorithm can *sometimes* end, namely, when one can be sure if the indicated upper bound for $K_2(\mathcal{O}_E)$ matches with a known lower bound. However, we are exactly in a situation in which we are trying to *prove* a lower bound. For example, when $E = \mathbf{Q}(\sqrt{-755})$, Brownkin and Gangl predict that $|K_2(\mathcal{O}_E)| = 2 \cdot 41$ because, for a set of larger and larger primes S, the index formula above stabilizes. So, beyond the issue of relating two different higher regulator maps, we have the problem of determining whether a class in $K_3(\mathcal{O}_E)$ is *divisible* by a prime p or not. This seems harder than our previous problem of determining whether a unit was divisible or not! (To be fair, however, it seems impossible to find units in $E(\zeta_p)$ once E is non-abelian and p is in any sense large.)

56.3. **Chern Class Maps.** We want to understand whether a class in the Bloch group $B(E)$ or in $K_3(\mathcal{O}_E)$ is divisible by p or not. Instead of working over $\mathbf{R}$, another approach is to work modulo a prime q. (It may seem a little strange to work modulo q to detect divisibility by p, but bear with me.) Soulé constructed certain Chern class maps, which include a map:

$$c_2 : K_3(\mathcal{O}_E) = K_3(E) \to H^1(E, \mathbf{Z}_p(2)).$$

These maps are the boundary map in the Atiyah–Hirzebruch spectral sequence for étale K-theory. Now compose this maps with the reduction modulo p map. Then, after restricting to $F = E(\zeta_p)$, we may identify $\mathbf{Z}_p(2)/p$ with μ_p, and so, by Kummer and Hilbert 90, we get a map:

$$c_2 : K_3(\mathcal{O}_E)/p \to H^1(F, \mathbf{Z}_p(2)/p) \simeq H^1(F, \mathbf{Z}_p(1)/p) = F^\times/F^{\times p}.$$

Keeping track of the various identifications, the image lands in the χ^{-1} invariant subspace, where χ is the cyclotomic character of $G = \mathrm{Gal}(F/E)$.

Lemma 56.4. *Let $p \geq 3$ be a prime which is totally ramified in $E(\zeta_p)/E$ and suppose that p does not divide the order of $K_2(\mathcal{O}_E)$. Then the Chern class map induces an isomorphism*

$$(\mathbf{Z}/p\mathbf{Z})^{r_2} = K_3(\mathcal{O}_E)/pK_3(\mathcal{O}_E) \to (\mathcal{O}_F^\times/\mathcal{O}_F^{\times p})^{\chi^{-1}}.$$

That is, the image of c_2 in $F^\times/F^{\times p}$ may be taken to land in the unit group, and the ranks of all the groups are the same and equal to r_2, the number of complex places of the field E.

This lemma follows from Quillen–Lichtenbaum, but it can also be proved directly from the surjectivity of the Chern class map as proved by Soulé, the known rank of $K_3 \otimes \mathbf{Q}$ by Borel, and some knowledge of the torsion of K_3 proved by Merkuriev and Suslin. It turns out that the hypothesis on $K_2(\mathcal{O}_E)$ is necessary not only for the proof but for the lemma to be true.

To detect whether a class in K_3 is divisible by p, it suffices to "compute" the Chern class map above and see whether it is zero. If one ever wants to compute anything, it makes sense to work with the Bloch group $B(E)$ instead. On the other hand, it seems hopeless to give a "concrete" map:

$$B(E) \to F^\times/F^{\times p}.$$

Even though one can write down elements in the first group somewhat explicitly, it's hard to imagine a recipe that would produce explicit elements in $F^\times$ with the correct Galois action.

Instead, what we do is reduce modulo q for some prime $q \equiv -1 \mod p$. That is, we pass from the Bloch group over E (which will be generated by S units for some S) to the Bloch group of the field $\mathbf{F}_q$. The construction over $\mathbf{F}_q$ is just the same. By a theorem Hutchinson, this group will have order $q + 1$. The numerology here is intimately related to Quillen's result that $K_3(\mathbf{F}_q) = \mathbf{Z}/(q^2 - 1)\mathbf{Z}$. Now there are some commutative diagrams one has to check commute here; I think the key point to keep in mind is that Quillen's computation of $K_3(\mathbf{F}_q)$ can already be realized in the cohomology group $H^3(\mathrm{SL}_2(\mathbf{F}_q), \mathbf{Z})$, and so the map of Bloch groups will be the same as the map on K-groups via comparison with the Hurewicz map.

Let's choose a prime $q \equiv -1 \mod p$ which splits completely in $E(\zeta_p + \zeta_p^{-1}) \subset F = E(\zeta_p)$. So we have a map $B(E) \to B(\mathcal{O}_E/\mathfrak{q}) \otimes \mathbf{F}_p = B(\mathbf{F}_q) \otimes \mathbf{F}_p = \mathbf{F}_p$. The Bloch group can be thought of in terms of (a quotient of a subgroup of) the free

abelian group of elements of $\mathbf{P}^1(E)$, so there's no issue about this reduction map. Moreover, given an element of the Bloch group, we can explicitly compute its image in the latter group. If this image is non-zero, that gives a certificate that the original element is not divisible by p. This will be enough to compute $K_2(\mathcal{O}_E)$ as long as the Bloch regulator map agrees with the dilogarithm map.

This argument is still yoked to real regular maps. Let's try to work entirely with c_2 and finite auxiliary primes $q \equiv -1 \mod p$. Another manifestation of the map $B(E) \to B(\mathbf{F}_q) \otimes \mathbf{F}_p$ is the map:

$$c_2 : K_3(\mathcal{O}_E) \to \mathcal{O}_F^\times / \mathcal{O}_F^{\times p} \to (\mathcal{O}_F/\mathfrak{Q})^\times \otimes \mathbf{F}_p = \mathbf{F}_p,$$

where $\mathfrak{Q}$ is a prime above $\mathfrak{q}$ in F. Let's go back to considering the case when E is an imaginary quadratic field. The image of a generator of $K_3(\mathcal{O}_E)$ will map exactly to a non-zero multiple of the non-trivial element unit $\epsilon \in F^\times/F^{\times p}$. If $K_2(\mathcal{O}_E)$ is prime to p, it will even land in $(\mathcal{O}_F^\times/\mathcal{O}_F^{\times p})^{\chi^{-1}}$. The latter map is exactly computing (up to a non-zero scalar) $\epsilon^{(q-1)/p} \mod \mathfrak{q}$, and so, purely using the Bloch group, we can check whether this is trivial or not. In particular, given an element of the Bloch group $B(E)$ which (we think) is a generator, or at least not divisible by p, we can find a prime $q \equiv -1 \mod p$ such that the reduction to $B(\mathbf{F}_q) \otimes \mathbf{F}_p$ is non-zero, which will imply that the image of c_2 is non-zero, which will imply that

$$\epsilon^{(q-1)/p} \not\equiv 1 \mod p.$$

This gives an explicit value of q for which this is true without ever having to compute ϵ. For such a prime q, we can then check that the circular units project to the identity in this space, which will prove unconditionally that $K_2(\mathcal{O}_E)$ is divisible by p. (Part of this computation assumed that p did not divide $K_2(\mathcal{O}_E)$, but that's OK, because to prove that p does divide this group we are allowed make that assumption anyway). Back to our example. We now want a prime $q \equiv -1 \mod 41$, which is also a square modulo 755. We take $q = 163$. Now this is not the most attractive computation in the world, because the root of unity ζ of order $37 \cdot 755$ cuts out the extension $\mathbf{F}_{q^{300}}$, as we can see by computing the multiplicative order of $q = 163$ modulo $41 \cdot 5 \cdot 151$. Let's do it in baby steps. By choosing a suitable prime $\mathfrak{Q}$ in $E(\zeta_{755})$, we can ensure that

$\zeta^{755} + \zeta^{-755} = \zeta_{41} + \zeta_{41}^{-1} \equiv 4 \mod \mathfrak{Q}$.

We write

$$\zeta^{1510} - 4\zeta^{755} + 1 = F(\zeta)G(\zeta) \mod 163,$$

where $F(\zeta)$ is any of the four factors of degree 300 (there are also two factors of degree 150, and factors of degrees 2, 4, and 4.) Now we want to compute, with $p = 41$, $q = 163$, and $r = 755$,

$$\eta := \left(\prod_{(\mathbf{Z}/pr\mathbf{Z})^\times} (1-\zeta^n)^{n\left(\frac{n}{r}\right)} \right)^{(q^{300}-1)/p} \mod \mathfrak{Q} = (163, F(\zeta))$$

Of course, one should first reduce the exponents $\chi^{-1}\eta(n) = n(n/r)$ modulo $p = 41$ before taking the powers. (Actually, it's probably kind of stupid to take a product over $\varphi(pr) = 24000$ different terms, and one can surely set this up much more efficiently, but whatever.) We find (drum roll) that:

$$\eta \equiv 1 \mod 163.$$

To finish, we have to take an element in the Bloch group $B(\mathcal{O}_E)$ and show that it doesn't vanish in $B(\mathcal{O}_E/\mathfrak{q}) \otimes \mathbf{F}_p = B(\mathbf{F}_{163}) \otimes \mathbf{F}_{41}$. At this point, I email Herbert (Gangl), and he sends me an email with the following beautiful element of $B(E)$, where $\alpha^2 = -755$:

$$-8\left[\frac{3-\alpha}{10}\right] - 10\left[\frac{7-\alpha}{10}\right] - 8\left[\frac{3-\alpha}{100}\right] + \ldots + 6\left[\frac{7\alpha+221}{972}\right].$$

(There are 114 terms in all! This should be a generator of $B(E)$.) Into my magma programme it goes, which cheerily reports that the image of this element is non-zero in $B(\mathbf{F}_{163}) \otimes \mathbf{F}_{41}$! So K_2 is really divisible by 41. (You might question the veracity of my programme's output, but more on that below.)

56.5. **Stark and Beyond.** Here are some more general remarks. Let's still suppose that E is imaginary quadratic. Take the image of a generator $[M]$ of $B(E)$, which is defined up to torsion and up to sign. The image of the Chern class map for some $p \geq 3$ and p not dividing $K_2(\mathcal{O}_E)$ gives a *canonical* unit in $\mathcal{O}_F^\times/\mathcal{O}_F^{\times p}$, where $F = E(\zeta_p)$. Let me a be a bit more careful here: by writing F as $F = E(\zeta_p)$, we are choosing a root of unity (this unit depends on this choice). There's also an automorphism of $\mathrm{Gal}(E/\mathbf{Q})$ which acts, but this changes the sign of $[M]$, so that is the same ambiguity we had before. What is this canonical unit? It is not just a circular unit, but a canonical one (modulo pth powers). What is it? More generally, when $r_2 = 1$, both K_3 and $(\mathcal{O}_F^\times/\mathcal{O}_F^{\times p})^{\chi^{-1}}$ have rank $r_2 = 1$, so if p is prime to $K_2(\mathcal{O}_E)$ we are generating canonical units. It's tempting here to conjecture some relation to Stark units here, and in particular to the special value of $L(1, E, \chi^{-1})$, but let me say no more about this. When $r_2 \geq 1$, one is no longer in the Stark world, but there is still a canonical map from the Bloch group to the unit group (the group $\mathbf{Z}^{r_2}$ has no canonical generator when $r_2 \geq 1$ — but in the manifestation of this group as a Bloch group, one does have explicit elements.)

Actually, I haven't even explained how to *compute* c_2. So far, I have only explained how to compute whether it is zero or not modulo p. To evaluate it exactly requires a further threading of the needle through the previous maps (on the Bloch group), and ultimately uses a test element coming from torsion in $B(\mathbf{Q}(\zeta_p + \zeta_p^{-1}))$. Although this is somewhat delicate, and I have not yet proved all of the appropriate diagrams commute (blech), one can work with it in practice and it gives many consistency checks on all the computations. (So, for example, once one has the image of c_2, one can compute the reduction of the corresponding element in the Bloch group in $B(\mathbf{F}_q) \otimes \mathbf{F}_p$ for one prime $q \equiv -1 \mod p$ knowing its image in the corresponding group for another such prime. Generating the same element of $\mathbf{F}_p$ for $p = 13$ and twenty different primes q is pretty convincing.)

In fact, computing this map exactly is exactly the problem that I was thinking about in the first place. I did compute it explicitly for $K = \mathbf{Q}(\sqrt{-491})$ and $p = 13$ (and also $K = \mathbf{Q}(\sqrt{-571})$ for $p = 5$), and the image of a generator of the Bloch group is *not* a unit. Instead, it gives a generator of $\mathfrak{a}^{13}$ for a non-trivial ideal in the class group $\mathrm{Cl}(F)$ of $F = E(\zeta_p)$, indeed, an element of $\mathrm{Cl}(F)[p]^{\chi^{-1}\eta}$. (In particular, it gives, having fixed a root of unity, a canonical element of this class group, which is also somewhat mysterious.) Let me also mention the Coates–Sinnott conjecture, higher Stickelberger elements, and work of Banaszak and Popescu which are closely related to the topics in this post (in particular, using Chern class maps to construct Euler systems generalizing the circular unit Euler system, although not so much

question of identifying these elements in some explicit way — especially because much less is known about higher analogues of the Bloch group). But perhaps this is enough for now.

Notes 56.6. The relevant `pari/gp` file mentioned above can also be found here. This post is clearly closely related to some of the material in [CGZ23].

57. The Artin conjecture is rubbish

Thu, 11 Sep 2014

Let $\rho : G_{\mathbf{Q}} \to \mathrm{GL}_N(\mathbf{C})$ be a continuous irreducible representation. Artin conjectured that the L-function $L(\rho, s)$ is analytically continues to an entire function on $\mathbf{C}$ (except for the trivial representation where the is a simple pole at one) and satisfies a functional equation of a precise shape. Langlands later had the profound insight to link this conjecture to functoriality in the Langlands program, which would additionally imply that ρ is *automorphic* which implies, *inter alia*, that $L(\rho, s) = L(\pi, s)$ for a cuspidal automorphic representation π for $\mathrm{GL}(N)(\mathbf{Q})$.

This is a beautiful and fundamental conjecture. However, it does appear to be completely useless for any actual applications. The most natural application of Artin's conjecture is to prove ... the Cebotarev density theorem. This is why Cebotarev's density theorem is so amazing! True, one can upgrade the error estimates if one knows Artin, but to do this one *also* has to know GRH. And if you know GRH, you are not too far away from Artin anyway, because then $L(\rho, s)$ at worst has poles on the critical strip, and so you can (essentially) get close to optimal bounds for Cebotarev anyway.

I thought a little bit about applications of Artin's conjecture when I wrote a paper about it, but I came up empty. Then recently, I had occasion to look at my paper again, and found to my chagrin that when Springer made the final edit they lopped off a sentence in the statement of one of the main theorems. I guess that's why the good people at Springer get paid the big bucks. (My best ever copy editing job, by the way, was for a paper in an AMS journal.) In a different direction, I guess it also reflects the deep study of this paper by people in the field that nobody has asked me about it. However, I did notice a statement in the paper that *could* be improved upon, which I will mention now.

To set the context, let $K^{\mathrm{gal}}/\mathbf{Q}$ be a Galois extension with Galois group S_5, and suppose that complex conjugation in this group is equal to (12)(34). Now suppose that ρ is a representation of $\mathrm{Gal}(K^{\mathrm{gal}}/\mathbf{Q})$. We already know that $L(s, \rho)$ is meromorphic, as proved by Brauer and Artin. One thing that can be proven is that, in the particular case above, $L(s, \rho)$ is holomorphic in a strip $\mathrm{Re}(s) \geq 1 - c$ for some constant $c \geq 0$ which I described as "ineffective." But looking at it again, I realized that it is not ineffective at all, due to a result of Stark. What one actually shows is that if $L(s, \rho)$ has a pole in the strip $\mathrm{Re}(s) \geq 1 - c$, then there must also be another L-function for the same field which has a zero *on the real line in this interval*. Note that, again from by Artin, it is trivially the case that a pole of one L-function must come from the zero of another L-function, since the product of all such L-functions is the Dedekind zeta function. So the content here is that the offending pole has to be on the real line. One consequence is that, in any particular case, one can rigorously check that the L-function in question has no such zeros,

and hence (combined with other results in this paper) that ρ is automorphic. With help from Andrew Booker, I was able to compute one such example (Jo Dwyer has since gone on to compute a number of other examples.) On the other hand, back to the general case, we do have effective results for zeros on the real line! The result in the paper is stated in terms of the existence of a zero of $\zeta_H(s)$ for a certain subfield H of K^{gal} of degree twelve. (The definition of H was exactly what was swallowed up by Springer, so it's not actually *defined* in the paper. To define it, note that S_5 has a faithful representation on six points. There is a degree six extension E which is the fixed field of the stabilizer of a point; then H is the compositum of E and the quadratic extension inside K^{gal}.) However, the actual argument produces a zero in an Artin L-factor of $\zeta_H(s)$ which is not divisible by the Dirichlet L-function for the quadratic character of S_5. Stark shows (*Some Effective Cases of the Brauer–Siegel Theorem*) (see [Sta74]) that such an L-function does not have Siegel zeros, and also gives an explicit estimate for the largest zero on the real line. In particular, for the $L(\rho, s)$ of interest, one deduces that they are analytic on the strip $\mathrm{Re}(s) \geq 1 - c$ where one can take

$$1 - c = 1 - \frac{1}{4 \log |\Delta_H|}.$$

The result of Stark, BTW, is why one could effectively solve the *class number at most X* problem for totally complex CM fields which were *not* imaginary quadratic fields *before* Goldfeld–Gross–Zagier.

Comment 57.1 (Dick Gross)**.** Frank, your title reminds me of a great Dan Aykroyd skit on an early SNL, where he played a late night talk radio host, and kept proposing more and more outrageous topics in a (futile) effort to get someone to call in. But I'll bite. What "applications" do you have in mind for the Riemann Hypothesis, or the conjecture of Birch and Swinnerton-Dyer, or the Hodge conjecture?

Comment 57.2 (Persiflage)**.** The *GRH* undisputedly has many interesting consequences (the other Artin's conjecture, effective Brauer–Siegel, and many more). Even the classical RH has consequences concerning the approximation of $\pi(x)$ by $\mathrm{Li}(x)$. BSD? It does what it says on the bottle: take an elliptic curve and you can determine whether it has infinitely many points or not. The Hodge conjecture? Throw in the standard conjectures and one gets a robust theory of motives. If you wanted to give famous conjectures with absolutely zero interesting consequences, then surely additive number theory supplies the ne plus ultra of such problems: the Goldbach and twin prime conjectures.

Of course I agree with your implicit thesis that the purpose of proving theorems is to advance understanding rather than merely as a means to proving ... more theorems. (It can safely be said that we learnt quite a lot from the proof of Fermat.) But doesn't it seem a little sad that one can't deduce anything from Artin's conjecture? Unlike (say) in the case of elliptic curves, the results of Cebotarev and Artin–Brauer allow one to extract all of the relevant juices out of $L(\rho, s)$ without knowing that it is holomorphic. In part, this is because the critical values for Artin motives are on the edge of the strip, and one can happily talk about the special value $L(\rho, 1)$ without having to know modularity.

I guess one consequence of a particular case of Artin's conjecture is ... more cases of Artin's conjecture! More precisely, if one knows that ρ and ϱ satisfy Artin's conjecture in the strong Langlands sense (is automorphic), then $\rho \otimes \varrho$ satisfies Artin's conjecture in the weak sense, by Rankin–Selberg.

Comment 57.3 (Matthew Emerton)**.** When I think of applications of modularity of elliptic curves, I not only think of all the theoretical applications (Gross–Zagier, Kato, Skinner–Urban, ...) but also of the fact that the tables of modular elliptic curves are complete tables of elliptic curves, ordered by conductor.

Similar, I think of modularity for odd two-dim'l Artin reps. of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ as meaning that the table of weight one forms gives a complete list of certain number fields. Maybe this doesn't count as an application, but it means something. It doesn't seem that much less effective as a GL_2 class field theory than classical CFT does for (say) totally real fields, when one has no idea how big/small the various ray class groups are.

Comment 57.4 (Persiflage)**.** Threading the needle slightly, I both agree and disagree with you here. You are talking about a a very special case when you select odd 2-dimensional Artin representations. In almost every other context, the automorphic forms corresponding to Artin representations can't actually be computed at all in an exact way. Using class field theory, for example, you can really compute whether any particular field has an extension with any particular solvable Galois group G unramified outside a given finite set of primes. And yet we still don't know Artin for solvable extensions of $\mathbf{Q}$. So while the *existence* (conjectural or otherwise) of a bijection between finite representations of Galois groups of number fields and certain automorphic forms is a beautiful one, it's a stretch to say that it is "computable."

Comment 57.5 (Minhyong Kim)**.** Regarding BSD, I've always felt that the most important portion is the finiteness of Sha, which implies that the standard algorithm used to compute the Mordell–Weil group actually terminates. From this point of view, the importance of the L-function can be questioned in the elliptic curve context as well. However, people far wiser than I seem to believe that trying to get at Sha forces the L-function on you, one way or another.

Comment 57.6 (Persiflage)**.** Dear MK, I both completely agree and disagree with what you say. First, I think that the finiteness of Sha is a really fundamental problem. On the other hand, I think your resulting corollary (that the Mordell–Weil group of E can be computed algorithmically) is not particularly interesting at all! I mean, as a practical matter, one can compute $E(\mathbf{Q})$ anyway, and knowing that Sha is finite doesn't help. It seems a little like saying that the finiteness of the class group is a fundamental problem because it allows one to algorithmically perform explicit computations in number fields (rather than just saying that it is a fundamental structural result). I'll also push back slightly on other aspects of BSD — proving BSD modulo finiteness of Sha would be a brilliant result.

On a different matter, I also think I don't necessarily agree with your experts. It seems possible that one could prove Sha is finite without automorphy — and I even have my own wise expert to back me up. To discuss a related result, I don't think that the Leopoldt conjecture will necessarily be proved using automorphic methods either.

58. The nearly ordinary deformation ring is (usually) torsion over weight space

Sat, 20 Sep 2014

Let $F/\mathbf{Q}$ be an arbitrary number field. Let p be a prime which splits completely in F, and consider an absolutely irreducible representation:

$$\rho : G_F \to \mathrm{GL}_2(\overline{\mathbf{Q}}_p)$$

which is unramified outside finitely many primes. If one assumes that ρ is *geometric*, then the Fontaine–Mazur conjecture predicts that ρ should be *motivic*, and the Langlands reciprocity conjecture predicts that ρ should be *automorphic*. This is probably difficult, so let's make our lives easier by adding some hypotheses. For example, let us assume that:

- **A:** For all $v|p$, the representation $\rho|G_v$ is crystalline and nearly ordinary,
- **B:** The residual representation $\overline{\rho}$ has suitably big image (Taylor–Wiles type condition.)

Proving the modularity of ρ under these hypotheses is still too ambitious — it still includes even icosahedral representations and Elliptic curves over arbitrary number fields. Natural further hypotheses to make include conditions on the Hodge–Tate weights and conditions on complex conjugation.

We prove the following:

Theorem 58.1. *Assume, in addition to conditions* **A** *and* **B**+*, that*

- **C:** *The Hodge–Tate weights* $[a_v, b_v]$ *at each* $v|p$ *are sufficiently generic,*
- **D:** *If* F *is totally real, then there exists at least one infinite place such that* ρ *is even.*

Then ρ *does not exist.*

The condition **B**+ (which will be defined during the proof) is more restrictive than the usual Taylor–Wiles condition — we shall see from the proof exactly what it entails. Condition **C** will also be explained — but let us note that, for any suitable method of counting, almost all choices of integers are generic, even after imposing some condition on the determinant (say $a_v + b_v$ is constant) to rule out stupidities.

One should think of this theorem as follows. If F is totally real, then condition **D** should be sufficient to rule out the existence of any automorphic ρ in regular weight, because (for motivic reasons) such representations should be totally odd. On the other hand, if F is not totally real, then the weights of any motive (with coefficients) should satisfy a certain non-trivial symmetry property with respect to the action of complex conjugation. So, for example, if F has signature $(1, 2)$, then *either* condition **C** or **D** should be sufficient, but we will require both. In fact, even condition **C** is stronger than what *should* be necessary. In addition to assuming regularity at all primes, it amounts to (on the representation theoretic side) insisting that *none* of the $\mathrm{GL}_2(\mathbf{C})$ weights are fixed by any conjugate of complex conjugation, whereas a single such example should be enough for a contradiction.

Perhaps a useful way to think about Theorem I is to make the following comparison. Hida proves the following theorem:

Theorem 58.2 (Hida)**.** *The nearly ordinary Hida family for* $\mathrm{SL}(2)/F$ *is finite over weight space and has positive rank if and only if* F *is totally real and the corresponding* $\overline{\rho}$ *is odd at all infinite places.*

On the other hand, a consequence of Theorem 58.1 is:

Theorem 58.3. *The fixed determinant nearly ordinary deformation ring of a residual representation* $\overline{\rho}$ *satisfying condition* **B**+ *is finite over weight space and has*

positive rank if and only if F is totally real and the corresponding $\overline{\rho}$ is odd at all infinite places.

In both cases, I am only considering the deformation rings up to twist — the deformation ring of the character is torsion over the corresponding weight space whenever $\mathcal{O}_F$ has infinitely many units. Also in both cases, it is of interest to determine the exact co-dimension of the ordinary family — this is a difficult problem, because strong enough results would allow you do deduce Leopoldt by considering induced representations.

OK, so what is the argument? If you have read some of my papers, you can probably guess.

Proof. Assume that ρ exists. Let U be the representation corresponding to ρ. Now replace U by $V = \mathrm{Sym}^2(U)$. Now replace V by the tensor induction:

$$W = \bigotimes_{G_F/G_{\mathbf{Q}}} V$$

of dimension $3^{[F:\mathbf{Q}]}$. We now let **C** be the condition that W has distinct Hodge–Tate weights. To see that this is generic, it really suffices to show that there is at least *one* choice of weights for which this is true. But one can let the weights of U up to translation consist of the 2-uples $[0,1]$, $[0,3]$, $[0,9]$, etc. and then the weights of W are, again up to translation, $[0,1,2,\ldots,3^{[F:\mathbf{Q}]}-1]$. We now let **B**+ be the condition that the residual representation is absolutely irreducible, and that the prime $p \geq 2 \cdot 3^{[F:\mathbf{Q}]} + 1$. This is *generically* true, and amounts to saying that the conjugates of $\overline{\rho}$ under $G_{\mathbf{Q}}$ are sufficiently distinct. Since the dimension of W is odd, and because it is essentially self-dual (exercise), orthogonal (obvious), nearly ordinary (by assumption), has distinct Hodge–Tate weights (by construction), satisfies the required sign condition (automatic in odd dimension), we deduce that it is potentially modular by [BLGGT14]. In order to win, it suffices to show, by a theorem I made Richard prove, that the action of complex conjugation on W has trace ± 1. However, this is equivalent to condition **D** (see below). □

One can relax condition **B**+ slightly by only inducing down to the largest totally real subfield of F. On the other hand, there are plenty of examples of representations $\overline{\rho}$ to which Theorem 58.3 applies. I think one can take any elliptic curve E/F without CM and such that $j_E \in F$ does not lie in any subfield of F, and then take p to be any sufficiently large ordinary prime which splits completely in F (*caveat emptor*, I didn't check this). Of course, the condition that p splits isn't really necessary either, I guess.

The proof of Theorem |refthm:two follows along the exact same lines — the conditions are strong enough to ensure, using results of Thorne [Tho12], that the nearly ordinary deformation ring of (the now residual) representation W is finite over weight space, which translates back into finiteness of deformations of U over weight space. The result is obvious if F is totally real and $\overline{\rho}$ is odd. Otherwise, we choose a sufficiently generic point in weight space (in the sense of **C**), and then, by Theorem 58.1, we see that the specialization of the nearly ordinary deformation ring at that point must be torsion.

It remains to compute the sign of W. This is an exercise in finite group theory, we only recall enough of the details for our purposes. Let V be a representation of

H of dimension d. Consider the tensor induction:

$$\bigotimes_{G/H} \sigma V.$$

Let T denote a set of representatives of right cosets of H in G. Let $tg \in T$ denote the corresponding choice for the coset Htg. For $g \in G$, let $n(t)$ denote the size of the $\langle g \rangle$-orbit which contains T. If $g = c$ has order 2, then either $n(t) = 1$ or $n(t) = 2$ Certainly

$$tc^{n(t)}t^{-1} \in H, \quad t \in T.$$

Let T_0 be a set of representatives for the $\langle g \rangle$ orbits on T_0. Then (proof omitted)

$$\phi^{\otimes G}(c) = \prod_{t \in T_0} \phi(tc^{n(t)}t^{-1}).$$

We observe that:

(1) If $n(t) = 2$, then $\phi(tc^{n(t)}t^{-1}) = \phi(tt^{-1}) = \phi(1) = d$.
(2) If $n(t) = 1$, then $tct^{-1} \in H$ and $\phi(tct^{-1})$ is what it is. For example, it is $0, \pm 1$ if and only if V is GL-odd with respect to tct^{-1}.

Now suppose that $G = G_{\mathbf{Q}}$ and $H = G_F$. The elements tct^{-1} are exactly the different complex conjugations of the representations of the conjugates of H. We deduce:

(1) If $\dim(V)$ is even, then W is GL-odd if and only if there exists at least one real place of F such that V is GL-odd.
(2) If $\dim(V)$ is odd, then W is GL-odd if and only if F is totally real and V is GL-odd at every real place.

Equivalently, a product of even integers can equal zero only if at least one of them is zero, and a product of odd integers can equal ± 1 if and only if all of them are ± 1.

59. Applying for an NSF grant

Fri, 26 Sep 2014

It's not easy to write a good grant proposal. But it can be even harder to write one for the first time, *especially* if you're not quite sure who will be reading your proposal. So today, I want to say a little bit about how an NSF mathematics panel is run, and give you some idea of who your target audience should be.

Before I start, I want to include a pseudo-legal disclaimer. For fairly obvious reasons, you are not supposed to reveal that you served on any particular panel. But I *am* allowed to say that I have served on *some* panels, and there is enough uniformity in the process to make me confident that what I say should resemble your reality if you decide to apply. (Let me also mention that I had some help on this post from a friend (whom I shall refer to as **the Hawk**) who is much more of an NSF pro than I am. He made various corrections and suggestions on a first draft of this blog, and I even included a few of his remarks verbatim in the text.)

The NSF administers many different types of grants. I'm not just talking about graduate fellowships, postgraduate fellowships and research grants here. There are FRG grants, RTG grants, CAREER NSF grants, REUs, conference grants, and so on. However, for the purpose of this email, I want to concentrate on research grants.

59.1. **The Mechanics.** The panel is comprised of approximately 10 or so mathematicians, who consider approximately 40-50 or so proposals. About six weeks before the panel takes place, each panelist is given the list of proposals and asked to rank the proposals 1,2,3,C based on the following criteria:

1 = I feel comfortable reviewing this proposal
2 = I could review this proposal if necessary
3 = It would be very difficult for me to review this proposal
C = I have a conflict of interest with this proposal

Here "conflict of interest" is defined in a fairly precise way. It includes some obvious things (recent co-authors, people at your institution, family members) and some non-obvious ones (people with whom you serve with on an editorial board, people at institutions that have paid you an honorarium for giving a recent talk). You are also free to declare a conflict of interest which is not on that list. About a month before the panel meets, each panelist is given 12 or so files to read (all the files are online, of course). It is not unusual for a panelist to be given (in their suite of proposals) one or two grants they graded as a "3" above — it depends on how parsimonious they were in their initial grading. For each of these files, the panelist is asked to grade the proposal on both intellectual merit and broader impact. Many panelists also unofficially rank the proposals that they read. In addition to grading the proposals, the panelist writes a brief summary indicating what they feel are the strengths and merits of each proposal. A panelist can, if they wish, also read other proposals.

The next step is that the panel meets at the NSF headquarters in Virginia, sometime between November and March. A typical panel may last 2.5 days. The panel is chaired by the relevant program officer and three or so other NSF employees (usually professional mathematicians who have taken a leave of absence for a two year position at the NSF), so there will typically be 14-15 people in a conference room, each with either their laptop or a supplied computer. The first 1.5 days of the panel consist of going through the files one by one. For each file, the three (or so) panelists who were assigned the proposal read out their review of the proposal. During this time, other panelists (especially those with some expertise) will also offer their opinions. During this period, anyone who is conflicted with the proposal has to leave the room. At the end of each discussion (which takes about 10 minutes), a yellow sticky sheet with the PI's name has to be placed on a white board with three columns. The columns are officially designated as "strongly recommended for funding," "recommended for funding if possible," and "not recommended for funding." The desired outcome is to have 10% of proposals in the first column, 30% in the second, and 60% in the third. Within the first two columns the names are ordered, although, during the process, certain proposals can float up or down as they are re-evaluated in light of other proposals. During each discussion, a panelist who was not assigned to read the proposal is assigned to be a scribe and record the highlights of each discussion. Each panelist is a scribe on 3-4 proposals.

The final step is for each panelist to write up a summary of the panel discussions for which they were a scribe, highlighting what the panel thought were the strengths and weaknesses of the proposal, indicating "which column" the panel placed the name, and reflecting the extent to which there was uniform agreement or not. Everyone then goes over these summaries to confirm that the summary does reflect the panel discussion. If you ever apply for a grant, you will be able to read this

summary, together with the evaluation of the three members who read your proposal in depth. (The panelists assigned to your proposal have an opportunity to modify their evaluations during the meeting if they change their minds in light of the discussion.)

Then the panel ends; the panel has given the program officer a (roughly) ordered set of names, and it is up to the NSF to decide whom to fund. I'm not sure the extent to which the recommendations of the panel exactly mirror the actual results, although I suspect that it is quite close. I can imagine, however, that a programme officer feels that a certain proposal suffered because nobody on the particular panel was an expert in that area, and they may decide to send that proposal off for further review.

The Hawk says: The actual results can deviate significantly from the advice of the panel. I think it's safe to say that the 'highly recommended' proposals always get funded. After that, there are various other objectives that the program officers are trying to achieve – gender diversity, racial diversity, support for young PIs, support for worthy PIs at undergraduate-only institutions. The panel list is typically the default in cases where none of those other objectives apply, though you can imagine reasons to deviate from it (e.g. you might not let the same person suffer the bad luck of being the first person after the cutoff two years running, you might support a proposal in a sub-discipline that has otherwise been shut out, etc). So in the 'recommended' zone there are certainly some inversions. It's also not unheard of for a 'not recommended' proposal to end up being funded. One way this can happen is for the proposal to be looked at by a second (perhaps more appropriate) panel that likes the proposal much better. But also, the program officers can simply decide that the panel's conclusions about a proposal were unjust for some reason, and raise the proposal up in the rankings.

59.2. **How narrow is the focus of each panel.** As I mentioned, there are approximately 40-50 proposals for each panel, of which maybe 15 are funded. So take the 80 or so people who are research active and applying for grants who are closest to you mathematically, and that gives you a rough idea. If you study Galois representations and modular forms, or Iwasawa theory, or the arithmetic of Shimura varieties, or arithmetic geometry of some kind, your proposal may well end up in the same panel as mine was (it can happen — as it did to me last year — that your proposal ends up being evaluated by *two* panels — this is possibly done in order to normalize the orderings in some way. Because I wasn't there, I can't quite tell what the difference was between the two panels).

The Hawk says: This is the first time I've heard a suggestion that normalization is the reason that some proposals are looked at by two panels. I think this happens either because the program officers feel that the proposal straddles two panels to such an extent that they feel both opinions could be useful; or because the proposal has two very different parts that genuinely fit in separate panels; or because the assigned panel decided that there were parts of a proposal that they didn't have the expertise to comment on, and so they suggest getting the input of another panel.

On the other hand, I'm pretty sure that my proposal would not be on the same panel as someone like Ken Ono or Soundararajan. Could my proposal be on the same panel as Akshay's? I'm not sure. I probably would have said no if my proposal didn't end up on two panels last time. And Akshay is a collaborator of mine! So

it's pretty focused. On the other hand, there are certainly areas in each field which are smaller than others, and if you work in such a sub-field, then it's more likely that the panelists will not be experts in your area.

59.3. **Who serves on the panel.** There are no formal NSF requirements for the constitution of any panel. Who is a typical member of the panel? Well, of course, one goal of the program officer is to make the panel is not *too* uniform. But, for example, I would expect that there would always be at least one person on the committee who knows as much (say) about modular forms and Galois representations as I do. So if that is what you do, then you can be pretty sure that whomever that person is will be reading your file. But you can also be sure that someone who is *not* an expert will also be reading your file, perhaps someone in Iwasawa theory, say. And this already should give you a pretty good idea of your target audience. In other words, you have to do two things:

- You have to explain to Iwasawa theory person why the modularity theorems *you* are going to prove are interesting. When is math interesting? Well, there are plenty of ways it can be interesting. You may have an idea of how to apply previous machinery in a novel way. You may have an interesting application in mind. You may have a completely new approach to an old theorem. You may have a completely new idea on how to solve an open problem. This is what you want to get across when you are talking to Iwasawa theory person — to give a sense of why the general problem you are studying is interesting, and how you are going to make a contribution to that field.
- You have an easier job convincing *me* (or equivalent) why your modularity theorems are broadly interesting, but you still have to convince me that you *particular* proposal is interesting. More importantly, you have to convince me that *you can carry out your proposal successfully, or at least to the point of producing interesting mathematics.*

  The Hawk says: I think it would be worth mentioning here the fine line between saying enough about how you intend to carry out your plans that the panel is convinced you can do it, and saying so much that they think you've done it already. I think new proposers often struggle on this point.

Of course, if you do something other than what I do, then replace "Iwasawa theory person" above by me or equivalent and "me" by someone with expertise in your field.

59.4. **What should I take away from this?** First up, I think that an NSF grant proposal is probably the most technical audience you will write for in a context that is not one of your research papers. So you don't need (beyond a cursory mention) to say how modular forms played a role in the proof of Fermat's Last Theorem which you might do (say) in a job application. Nor do you need to define the class group of a number field, or explain what a modular curve is. But, at the same time, and this is very important, it can still be incredibly useful to place your work in a broader context. For example, on my last NSF proposal, I started out by reminding the reader briefly how there are very general conjectures linking Galois representations coming from geometry to automorphic L-functions. I reminded the reader that special degenerate cases of this conjecture correspond to very classical objects like

the Riemann zeta function. I then mention how the work of Wiles addresses the case when the representation comes from the cohomology of an elliptic curve over **Q**. Then I explain how all the generalizations of Wiles' theorem share a common assumption, namely, that the Galois representations over **Q** that one can study by this method have the property that they are, up to a twist, self-dual. So already, in perhaps not much more than a half a page, I have given the context to explain how proving that a *non-self-dual* Galois representation is modular is "interesting." Of course, then I have to go on and explain *how* I am going to say anything interesting about non-self-dual representations.

59.5. **Do fat cats just get their grants without trying?** Every proposal is evaluated on its merits, but of course "prior success" is taken into account when judging future chances of success, and so it should be. But if Peter Scholze (say, to take someone who is not in the US so I can use his name) sends in an application consisting solely of "I am working on several projects that I decline to disclose but that I expect to be of the same importance as my prior results," he would not be funded. More realistically, I have heard that it has been the case that fields medalists have been turned down for grants, but because all grants that are turned down are never officially acknowledged, this is just hearsay. My feeling is that, on the whole, the panels do a pretty good job, and (apart from the occasional controversial case) there is more of a uniform agreement than you might guess. **The Hawk** brings up the key point that this opinion only concerns number theory panels. It may be the case (and I occasionally hear rumours to this effect) that other areas are not run as well. I would also say that the fat cats (on the whole) seem to put as much effort into writing their NSF proposals as everyone else.

59.6. **How can I compete with the fat cats given I'm only just starting out?** This is taken into account. If you are at most 6 years from your PhD, your proposal is evaluated in that context; an effort is made to fund promising young people, and also people who have never received prior NSF support. That said, it's not easy to get a grant the first time you apply coming straight out of a postdoctoral position.

59.7. **What about broader impact?** This is hard for younger people. But everyone on the panel realizes this and so the expectations are lower. You probably don't have any grad students yet, so what can you say? Perhaps you have given expository talks at a workshop? Perhaps you have written up detailed notes on otherwise hard to access topics? Perhaps you have gone into the public schools in some hardscrabble inner suburban neighbourhood and taught calculus? (Not the last one? Then don't suggest that you might if there's no reason to suspect that you have any previous inclination to do so.)

59.8. **Don't Imagine.** that you are going to be held account for what you say you are going to prove in future proposals. Future proposals will be evaluated on their own merits (as well as prior research), and nobody is going to know or remember what you said in your previous NSF grants. It's expected that some of problems you are working on might not work out, and that you will have new ideas while working on the proposal.

Two further suggestions from **the Hawk**:

59.9. **When will I hear back?** Answer: who the hell knows. Usually within six months from the deadline, but not always, especially these days when the federal government is funded from continuing resolution to continuing resolution. If you hear in January, either you are Peter Scholze or it's bad news. By May, no news is good news: you probably weren't in the 'not recommended' pile, and they're waiting to see how far they can stretch the money in the 'recommended' pile.

59.10. **If I get the grant, how much money will I get?** Answer: probably less than what you asked for in your budget, and if not, you probably didn't budget enough. Less glib answer: the program officers do adjust the award sizes in order to hit their target funding rates. You shouldn't fret that if you ask for too much and the person who's next on the list asks for a lower number, that could hurt your chances. The natural followup: "If program officers have that kind of discretion, wouldn't it be better if they gave smaller awards to more people?" You can certainly argue that in theory that might be better, but in practice the answer is emphatically no. DMS's (DMS = division of mathematical sciences at the NSF) funding rate is already much higher than that of other divisions, as high as can politically be sustained within NSF. If the funding rate went up, DMS's budget would be cut, and the rate would go back down again.

59.11. **Do you have any other thoughts?** The fact that approximately 30% percent of proposals get accepted is a fairly immutable law of nature. It is no doubt depressing to be continually rejected by the NSF, and good people simply stop applying, in some sense making it then harder for everyone else. If, for some reason, the number of applications suddenly doubled, it wouldn't be the case that the success rate would halve, but more proposals would be awarded. So, there is a real sense in which the more people who apply the more grants are awarded.

60. In brief

Wed, 22 Oct 2014

The start of the academic year has a habit of bringing forth distractions, not least of all to someone as disorganized as me. So here are a few remarks in brief.

60.1. **The class number of $\mathbf{Q}(\zeta_{151})^+$ is one.** John Miller, a student of Iwaniec at Rutgers, wrote the following nice paper (see [Mil15b]), which improves upon a previous result of Schoof. (Related: here, [Mil15a]) One technique that is useful in computing the class numbers of fields with small discriminant is to make use of the Odlyzko bounds. Here's a typical example. If $K = \mathbf{Q}(\zeta_{37})^+$, then the root discriminant of K is 30 or so. However, by consulting Odlyzko, one sees that any totally real field with this root discriminant has degree at most 40. Hence the class number of K is either one or two, and it is easy to rule out the second possibility by using genus theory. More generally, whenever one has an *a priori* bound on h^+, one can compute h^+ by relating h^+ to the index of the circular units (Schoof did this in a previous paper.) This trick only works if the root discriminant of the totally real field is at most 60 (or so), which seems to prevent one from applying this to real cyclotomic fields for $p \geq 67$. (There's always a bound on the class group by Minkowski, but that is a terrible bound.) The idea behind this paper is that Odlyzko's bound can be improved if one in addition knows that certain primes of

small norm are principal. And since one has explicit fields, it is possible to show that the relevant ideals are principal by exhibiting explicit elements with the appropriate norm. I can't quite tell how lucky the author was to find such elements (he searches for cyclotomic elements expressible as a small number of roots of unity), but it works! Perhaps, *a postiori*, it is useful that these fields do actually turn out to have class number one.

60.2. **Stickelberger's Theorem.** I proved Stickelberger's theorem in class the other day — well, with one caveat. I proved that all the ideals $\mathfrak{q}$ of prime norm are annihilated by the Stickelberger ideal. This certainly implies the result, because the class group is generated by such ideals. This follows, for example, by the Cebotarev density theorem applied to the Hilbert class field (which was my argument in class). But then I worried that this was an anachronistic argument, and indeed Stickelberger's theorem was a solidly 19th century result. So what did Stickelberger do?

Comment 60.3 (Danny)**.** Speaking of fields with class number one, have you seen this preprint by Darren Long and Morwen Thistlethwaite: here? (see [LT16]).

Comment 60.4 (John Miller)**.** Long and Thistlethwaite's paper is quite interesting. They mention that to have even a chance of producing a prime clique big enough to show class number 1, the root discriminant must be less than $16\pi e$. This raises the question: Although we believe that fields with class number 1 are quite plentiful, do we have any reason to expect there are infinitely many fields of class number 1 with bounded root discriminant? Towers of fields with class number 1 should have increasing root discriminant; and on the other hand most towers with bounded root discriminant are built out of Hilbert class fields of fields with nontrivial class groups.

Comment 60.5 (Persiflage)**.** The only clearly convincing heuristics (to me) about class numbers concern families of fields of fixed degree (so the discriminant necessarily tends to infinity). The problem of understanding fields with bounded root discriminant (where the bound is bigger than the limit of the Odlyzko bound) on the other hand seems very quite difficult. It's *possible* that one could have a tower of fields all with class number one and the same root discriminant (if the corresponding Galois groups are perfect), although it doesn't seem particularly plausible on some sketchy heuristic grounds.

61. Mysterious formulae

Sat, 15 Nov 2014

I'm not one of those mathematicians who is in love with abstraction for its own sake (not that there's anything wrong with that). I can still be seduced by an explicit example, or even — quell horreur — a definite integral. When I was younger, however, those tendencies were certainly more pronounced than they are now. Still, who can fail to appreciate an identity like the following:

$$e^{-2\pi}\prod_{n=1}^{\infty}(1-e^{-2\pi n})^{24}=\frac{\Gamma(1/4)^{24}}{2^{24}\pi^{18}}.$$

But man cannot live on identities alone, and ultimately one's efforts turn in other directions. So it's always nice when the old and new words coincide, and an identity is revealed to have a deeper meaning. The formula above is a special case of the Chowla–Selberg formula [SC67], which is, possible typos in transcription aside,

$$\sum_{CM(K)} \log\left(y^6|\Delta(\tau)|\right) + 6h\log(4\pi\sqrt{\Delta_K}) = 3w_K \sum \chi(r)\log\Gamma(r/\Delta_K).$$

Here the notation is as you might guess — y is the imaginary part of τ, which is ranging over the equivalence classes of CM points for a fixed ring of integers in an imaginary quadratic field (there is presumably a version for orders as well). The existence of this identity (and a vague sense that it was related to the Kronecker limit formula) was basically all that I new about this identity, but Tonghai Yang gave a beautiful number theory seminar this week explaining the geometric ideas behind this formula, and some generalizations (the latter being the new work). So, just as in the Gross-Zagier paper [GZ85] on the special values of j at CM points, one now has *two* proofs of this result which complement each other, one analytic, and one geometric. (I apologize in advance for not being able to attribute all [or really any] of the ideas, Tonghai certainly mentioned many names but I never take notes and this was 5 days ago.) The first remark is that the RHS is essentially the logarithmic derivative of the corresponding Artin L-function. On the other hand, it turns out (non-obviously) that the left hand side can be related to the Faltings height(s) of the corresponding Elliptic curves with CM by $\mathcal{O}_K$. I think this relation was discovered by Colmez in his '93 Annals paper [Col93]. The Faltings height has always been a slippery concept to me, and in fact the theory of heights in general has always struck me as being connected to the dark arts. In particular, various definitions depend on certain choices of height function, although they actually don't depend on that choice in the end. So when actually doing a calculation, it's always nice if you can magically produce some choice which makes calculation possible. And of course, when making a choice of function on some (tensor power of) ω over the modular curve, what better choice is there (if one wants to control the zeros and poles) than Δ. (Tonghai mentioned another version of the formula where one instead used certain forms which are Borcherds products — of which Δ is a highly degenerate example. I had the sense that this formulation was more generalizable to other Shimura varieties, but I never understood Borcherds products so I shall say no more.) Key difficulties in understanding generalizations of these formulas involve ruling out certain vertical components in certain arithmetic divisors on Shimura varieties, which I guess must ultimately be related to understanding the mod-p reduction of these varieties in recalcitrant characteristics (blech).

Colmez also formulated a conjectural generalization of the CS-formula, which is what Tonghai was talking about, and on which he (and now he together with his co-authors) have made some progress. The viewpoint in the talk was to re-interpret these identities in terms of arithmetic intersection numbers of arithmetic divisors on Shimura varieties. Of course, this is intimately related to the ideas of Gross-Zagier and its subsequent developments, especially in the work of Kudla, Rapoport, Brunier, Ben Howard, and Tonghai himself (and surely others... see caveat above). In light of this, one can start to see how special values of L-functions and their derivatives might appear. I can't possibly begin to do this topic justice in a blog post, but I will at least strongly recommend watching Ben Howard talk about this at MSRI in a few weeks (Harris-fest, Tuesday Dec 2 at 11:00). I'll be there to watch

in person, but for those of you playing at home, the video will certainly be posted online. Ben is talking about exactly this problem. Since he is an excellent lecturer, I can safely promise this will be a great talk.

Added: Dick Gross emailed me the following (which also gives me the chance to say that Tonghai did indeed mention Greg Anderson during his talk):

> ... if you want to read a nice analytic treatment of the Chowla–Selberg formula, using Kronecker's first limit formula, you can find it in the last chapter of Weil's book "Eisenstein and Kronecker".
>
> I found an algebraic proof of [SC67] when I was a graduate student, using the moduli of abelian varieties with multiplication by an imaginary quadratic field (what we would now call unitary Shimura varieties). Deligne figured out what I was actually doing, and generalized it to prove his wonderful theorem that Hodge cycles on abelian varieties are absolutely Hodge.
>
> Greg Anderson formulated a generalization of [SC67] for the periods of abelian varieties with complex multiplication. This was refined by Colmez, and we know how to prove all the refinements when the CM field is abelian over $\mathbf{Q}$. Tonghai and Ben have been making progress in some non-abelian cases.

62. Harris 60

Mon, 15 Dec 2014

I've just returned from the excellent MSRI workshop which honored Michael Harris' 60th birthday, and here is a brief summary of some of the gossip [excised from this version] and mathematics I picked up when I was there.

Akshay gave a very intriguing talk on integral structures in cohomology. It reminded me of a question that we discussed a long time ago near Washington Square Park in NYC. Recall that, for tempered automorphic forms contributing to the cohomology of $\mathrm{GL}(n)/\mathbf{Q}$, a computation with $(\mathfrak{g}, K)$ cohomology shows that each such form occurs in cohomology in degrees $q_0, \dots, q_0 + \ell_0$ and contributes (a multiple) of

$$\binom{\ell_0}{k}$$

dimensions in degree $q_0 + k$. Is there any analogue of this for torsion classes or in characteristic p? Assume here that the residual representation also occurs only inside the relevant ranges of cohomology, which should be the case as long as the corresponding residual representation is not Eisenstein. If $\ell_0 = 0$, there is nothing to say. If $\ell_0 = 1$, then the result follows from an Euler characteristic argument; that is, the cohomology in each of the two non-zero degrees over $\mathbf{F}_p$ will have the same dimension. If $\ell_0 = 2$, then Poincaré duality shows that the cohomology groups in the two outer degrees will have the same dimension (again we are working with non-Eisenstein classes and trivial coefficients, so from this point of view things look compact), and then an Euler characteristic argument shows that the space appears in some multiplicities of dimensions $(1, 2, 1)$, as in characteristic zero. Now suppose that ℓ_0 is arbitrary. I will assume we are in a multiplicity one situation and that all the local deformation rings are smooth. The CG-method (under suitable

hypotheses) produces a resolution $P^\bullet$ of R_∞ consisting of finite free S_∞-modules, where R_∞ is smooth of relative dimension q and S_∞ is free of relative dimension $n := q + \ell_0$. To recover the cohomology over a finite field, one takes the quotient of this resolution by the maximal ideal of S_∞ and then takes cohomology. In other words, what we are really computing is

$$\mathrm{Ext}^*_{S_\infty}(R_\infty, k).$$

This answer depends only on the rings involved and not on the resolution. For example, suppose that $q = 0$, which is the same as saying that there are no non-trivial local infinitesimal deformations. Then $R_\infty = \mathbf{Z}_p$, and one can compute the cohomology of this ring using the Koszul complex, and one gets the expected dimensions. Note that if $R = \mathbf{Z}_p$, then this recovers the automorphic computation, but this is already slightly interesting if $R = \mathbf{F}_p$ or even $\mathbf{Z}/p^k\mathbf{Z}$. However, it seems a little optimistic to expect this pattern to hold in general. I spent some time trying to prove it using commutative algebra, but one problem is that it is not true in that generality. For example, suppose that $\ell_0 = 3$, and that

$$R_\infty = \mathbf{Z}_p[[y_1, y_2, y_3]], \quad S_\infty = \mathbf{Z}_p[[x_1, x_2, x_3, x_4, x_5, x_6]],$$

where the map from S_∞ sends the generators to the six possible monomials of degree two. Then the appropriate dimensions of the ext groups are $4, 14, 14, 4$. (Thanks to Daniel Erman for this example.) Now this example actually can't occur globally, because the same computation implies that one the Betti numbers over $\mathbf{Q}_p$ are also given by these numbers, which violates the previously referenced computation with $(\mathfrak{g}, K)$-cohomology. However, one should easily be able to deform it very slightly to kill off any cohomology in characteristic zero, for example replacing y_i by $y_i - \epsilon_i$ for small constants ϵ_i. Of course this doesn't disprove anything, but it does strongly suggest that the dimensions over $\mathbf{F}_p$ could be all over the place, subject to the Poincaré and Euler characteristic conditions. Akshay has also pointed out that the case $\ell_0 = 3$ is interesting from a different but related perspective: the analytic torsion will vanish in this case, which implies that, at least morally (since we have localized at a maximal ideal), that the alternating product of the the orders of the cohomology groups over $\mathbf{Z}_p$ should equal one. Is this a consequence of the Taylor–Wiles method? I just thought of this question right now and it may have an obvious answer which Akshay knows, I'll ask him today and report back. A second obvious question is what happens if one looks only at $\mathfrak{m}$-torsion rather than $\mathbf{F}_p$-torsion; perhaps that is the more sensible generalization of the characteristic zero question anyway.

63. Derived Langlands

Mon, 22 Dec 2014

Although it has been in the air for some time, it seems as though ideas from derived algebraic geometry have begun to inform developments in the Langlands program. (A necessary qualifier: I am talking about reciprocity in the classical arithmetic Langlands program here.)

I want to describe a very simple instance of this which came up in Akshay's MSRI talk which I linked to in the post above. Start by fixing a global residual

(GL-)odd Galois representation:

$$\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_N(\mathbf{F}_p)$$

Let us suppose that $\overline{\rho}$ is surjective. Associated to this representation is a fixed determinant unrestricted global deformation ring R, and a fixed determinant unrestricted local deformation ring which I will call S. (Apologies for the notation, but wordpress is not great with lots of subscripts.) I assume that the reader can make the appropriate adjustments to these definitions if the local representation is not irreducible by adding framings. One knows (by [AC14]), at least if $p \geq 2N+1$, that the map

$$\mathrm{Spec}(R) \to \mathrm{Spec}(S)$$

is finite. Let us now suppose that $\overline{\rho}$ restricted to $G_{\mathbf{Q}_p}$ admits a crystalline lift of some regular weight; associated to this weight is a Kisin local deformation ring which I shall call T. If you like, you can even imagine that we are working in small weight so that T has nice properties; perhaps it is even smooth. Barry Mazur has made a conjecture for what the (relative) dimension of R should be over $\mathbf{Z}_p$. Namely, it should be given by the Euler characteristic of the adjoint representation, which is equal (see Def.2.1 and the subsequent comment) to

$$\dim B - \ell_0,$$

where B is a Borel of $\mathrm{SL}_N(\mathbf{R})$, and ℓ_0 is the difference between the rank of $\mathrm{SL}_N(\mathbf{R})$ and $\mathrm{SO}_N(\mathbf{R})$. Of course, these quantities can easily be calculated explicitly in this (or any) case; for $\mathrm{SL}(N)/\mathbf{Q}$, ℓ_0 is the integer part of $(N-1)/2$. On the other hand, we may also compute the (relative) dimensions of the rings S and T, and we find that

$$\dim(S/\mathbf{Z}_p) = N^2 - 1, \quad \dim(T/\mathbf{Z}_p) = N(N-1)/2.$$

(The notation here means the relative dimension.) The Fontaine–Mazur sanity check is to see that, on the associated rigid analytic spaces, the assumption that R and T meet transversally inside S should imply that their intersection only has finitely many points. Indeed, we can compute that the expected dimension of the intersection is exactly:

$$N(N-1)/2 + \dim(B) - \ell_0 - (N^2-1) = -\ell_0.$$

When $N = 2$, we have $\ell_0 = 0$, and everything is as expected. However, as soon as $N \geq 2$, we have $\ell_0 \geq 0$, and so the expected dimension is negative. This says that regular algebraic automorphic forms for such N are much rarer beasts than their counterparts for $N = 2$, where modular forms are abundant. For example, it is not known if there exists a regular algebraic cusp form for $\mathrm{GL}(N)/\mathbf{Q}$ giving rise to a $\overline{\rho}$ as above for any $N \geq 5$. (Note that forms from smaller groups coming via functorial lifts will fail to give rise to representations with such large image.) Now all of this is a philosophy that has been known and exploited for some time. But suppose we actually try to interpret this heuristic a little more literally. For a start, we do expect that forms of characteristic zero do exist. This means that, in general, there *are* unlikely intersections of $\mathrm{Spf}(R)$ and $\mathrm{Spf}(T)$ inside $\mathrm{Spf}(S)$. That is, R and T will *not*, in general, be transverse! This is exactly a context in where, to understand the intersection, it makes sense to introduce the derived world (see, for example, the introduction to DAG-V).

In the classical picture, to recover the usual minimal deformation ring, one considers the intersection $X := R \otimes_S T$. However, science now tells us it is more natural to consider the derived tensor product

$$Y = R \otimes_S^{\mathbf{L}} T.$$

If $\ell_0 = 0$, then the cohomology of Y should exactly recover X in degree zero and be zero otherwise — that is, the classical context should be related to a completely transverse intersection, and we are still in the usual word of schemes (or even complete local Noetherian rings). However, this will (essentially) never happen when $\ell_0 \geq 0$. Classically, the ring X may be identified with the ring of endomorphisms generated by Hecke operators on a single extremal degree of cohomology. More generally, the cohomology of Y should be identified with the ring of Hecke operators acting on the cohomology now in degrees $q_0, \dots q_0 + \ell_0$, where the notation is as in Borel-Wallach and is fixed for all time. In particular, the only context in which one should expect the intersection to be transverse (beyond $\ell_0 = 0$) is the case when $\ell_0 = 1$ and X is a finite ring (which can happen). Indeed, in such contexts, the cohomology over $\mathbf{Z}_p$ also occurs exactly in one degree. It might be worth noting here that the ring S is not in general regular, and so Y, *a priori*, is not even bounded.

On the other hand, science also tells us that the complex Y has more information than its cohomology; and so one should really think of Y as the correct object. Unfortunately, I don't have anything clever to say about derived arguments, but let me use the Fontaine–Mazur heuristic to extract some tiny amount of juice. Since I am not so DAGgy, I will only use algebra that goes back 30 years or more.

Instead of looking at the intersection of R and T inside the formal spectrum of S, let us look at their intersection over $\mathbf{Z}_p$. In this case, all the dimensions have been shifted by one, so, when $\ell_0 = 0$, their intersection should be infinite (that is, the length over $\mathbf{Z}_p$). This is obviously the case, because X (which by assumption exists) is flat over $\mathbf{Z}_p$ and so automatically infinite. Well, obvious modulo Serre's conjecture and Fontaine–Mazur for $\mathrm{GL}(2)/\mathbf{Q}$, at least. On the other hand, when $\ell_0 \geq 1$, then the dimensions don't add up and the intersection multiplicity should be zero. Thus, assuming the dimensions of the rings are correct, we should have:

(1) If $\ell_0 = 1$, then the intersection multiplicity is finite and non-zero;
(2) If $\ell_0 \geq 1$, then the intersection multiplicity is zero.

In the latter case, we are invoking the conjecture of Serre (proved by Roberts and Gillet–Soule that the intersection multiplicity is zero when the dimensions are too small. (Heuristically, one can "move around" classes whose codimension is sufficiently large so they don't intersect at all, although one cannot literally do this in a local ring!) Actually, even this is a lie, because S is not necessarily regular, as mentioned above, but pretend for this paragraph that it is. Serre's formula tells us that the intersection multiplicity is given by the Euler characteristic of the complex. Let me now suppose that Y has no characteristic zero cohomology. For example, we could be working in a weight corresponding to a (strongly) acyclic local system (as long as $\ell_0 \neq 0$). What we want to compute is the alternating product of the cohomology groups, which should be equal to the alternating product of the integral cohomology groups (everything localized at the appropriate maximal ideal) of the corresponding congruence subgroup of $\mathrm{GL}_N(\mathbf{Z})$. Yet this product (without localizing at a maximal ideal) is basically equal to the Reidemeister torsion, which

is equal to the analytic torsion, which always vanishes when $\ell_0 \geq 1$! Under our finiteness conditions, when $\ell_0 = 1$, all the cohomology occurs in just a single degree, and so the multiplicity is just the length of X. But when $\ell_0 \geq 1$, this gives (the expected) refinement of the vanishing of analytic torsion after localizing at a non-Eisenstein maximal ideal, which was one of the questions implicitly raised in the last blog post. To be more precise about our assumptions and conclusions, we have:

Proposition 63.1. *Let $\mathfrak{m}$ be a non-Eisenstein maximal ideal of the cohomology of a congruence subgroup of $\mathrm{SL}_N(\mathcal{O}_F)$ for a CM number field F, and let $\overline{\rho}$ be the corresponding Galois representation. Assume that:*

(1) The required assumptions in [CG18a] hold — vanishing outside degrees $q_0, \dots, q_0 + \ell_0$ after localization at $\mathfrak{m}$, the representation $\overline{\rho}$ has big image, local-global compatibility, etc.;
(2) The cohomology localized at $\mathfrak{m}$ vanishes in characteristic zero.

Then the alternating product of the orders of the cohomology groups localized at $\mathfrak{m}$ is non-zero if and only if $\ell_0 = 1$.

One issue with proving this directly using the above argument is that we don't actually know the dimension of R in general. So, instead of working with Y, we work instead with the output of the Taylor–Wiles method as in C-G, namely:

$$Y_\infty = R_\infty \otimes^{\mathbf{L}}_{S_\infty} S_\infty/\mathfrak{a}$$

Here S_∞ is an Iwasawa algebra of diamond operators of dimension $q + \ell_0$, the ideal $\mathfrak{a}$ is the augmentation ideal with $S_\infty/\mathfrak{a} = \mathbf{Z}_p$, and R_∞ is a patched minimal deformation ring of dimension q. These are relative dimensions, so the transverse case over $\mathbf{Z}_p$ corresponds exactly to the case when $\ell_0 = 1$. We note here that the ring S_∞ is regular, so we are in the appropriate context of Serre's multiplicity formula, and the result follows. (**Exercise:** we are only using a very special case of the vanishing claim for intersection multiplicities when one component is $\mathbf{Z}_p$ inside S_∞; the vanishing should be easy to prove directly in this case.) This is good, because it gives a purely Galois theoretic proof (well, really only a heuristic because of all the conjectures one needs to assume) of a result (vanishing of analytic torsion) which is not at all obvious. (Well, not quite; the result is localized at a maximal ideal — which one can't do analytically — but it only applies to non-Eisenstein maximal ideals.) At any rate, thinking through this example after Akshay's talk has convinced me that this derived perspective is a very good one.

64. Stable completed homology without Quillen–Lichtenbaum

Wed, 21 Jan 2015

Having just made (hopefully) the final revisions on my paper on stable completed cohomology groups [Cal15], I wanted to record here a few remarks which didn't otherwise make it into the paper.

The first is that, in addition to the result that $\widetilde{H}_2(\mathrm{SL}, \mathbf{Z}_p) = \mathbf{Z}_p$ for $p \geq 2$, one may also compute $\widetilde{H}_3(\mathrm{SL}, \mathbf{Z}_p)$ for $p \geq 3$. Namely:

$$\widetilde{H}_3(\mathrm{SL}, \mathbf{Z}_p) = 0.$$

This result is proved in the paper up to a finite group, so the point here is the integral refinement. The computation of $\widetilde{H}_2$ comes from the Hurewicz isomorphism

$$\pi_2(SK(\mathbf{Z},\mathbf{Z}_p);\mathbf{Z}_p) \simeq \widetilde{H}_2^{\mathrm{cont}}(\mathbf{Z}_p).$$

However, the Hurewicz theorem also gives an epimorphism

$$\pi_3(SK(\mathbf{Z},\mathbf{Z}_p);\mathbf{Z}_p) \to \widetilde{H}_3^{\mathrm{cont}}(\mathbf{Z}_p),$$

and one finds that the first group lives in an exact sequence

$$H^2(\mathbf{Q}_p,\mathbf{Z}_p(2)) \to \pi_3(SK(\mathbf{Z},\mathbf{Z}_p);\mathbf{Z}_p) \to K_3(\mathbf{Z})\otimes\mathbf{Z}_p$$

Since both flanking groups vanish for $p \geq 3$, the middle group is zero, and the claim follows.

The second remark is that, throughout the paper, I assume the Quillen–Lichtenbaum conjecture, which is now a theorem due do Voevodsky and others. However, I must confess, I do not have the fine details of the argument at my fingertips. How much can one say without it? The answer is quite a lot. Due to work of Borel, Soulé, and Quillen (see [Bor74, Sou79, Qui72], all of which is much more familiar to me, at least relatively speaking), we know that the K-groups of number fields are finitely generated abelian groups, we know their ranks, and we know that the Chern class maps to the appropriate Galois cohomology groups are surjective. Moreover, we understand $K_2(\mathscr{O}_F)$ completely in terms of Galois cohomology by work of Tate [Tat76]. (In this game, I am also giving up the results of Hesselholt and Masden on the K-theory of local fields, and instead using the results of Wagoner [Wag76], which similarly give everything in very small degree and up to a finite group in higher degrees.) In particular:

(1) The computation of $H_2(\Gamma_N(p),\mathbf{F}_p)$ for large N, where $\Gamma_N(p)$ is the principal congruence subgroup of $\mathrm{GL}_N(\mathbf{Z})$, is unaffected. This also uses the computation of $K_3(\mathbf{Z})$ by Lee and Szczarba [LS76].
(2) The identification of the completed K-groups with Galois cohomology groups still holds up to a finite group.
(3) The computation of the rational stable completed homology groups

$$\widetilde{H}_*(\mathrm{SL},\mathbf{Z}_p)\otimes\mathbf{Q} = \mathbf{Q}[x_2,x_6,x_{10},x_{14},\ldots]$$

under the assumption that either p is regular or $\zeta_p(3),\zeta_p(5),\zeta_p(7)$ etc. are all non-vanishing still holds.

Something that does require Quillen–Lichtenbaum is the vanishing of the partially completed K-group for very regular primes.

Regarding the computation of the rational stable completed homology groups, the referee made a very interesting point (I will come back in a later post to the refereeing of this paper and some other of my recent papers in a post on "what a great referee report should be"). I prove that the rational stable completed homology groups are the continuous homology of the homotopy fibre

$$SK(\mathbf{Z},\mathbf{Z}_p) \to SK(\mathbf{Z}) \to SK(\mathbf{Z}_p)$$

(The definition of $SK(\mathbf{Z},\mathbf{Z}_p)$ is just homotopy fibre of this map.) Now $SK(\mathbf{Z},\mathbf{Z}_p)$ is an infinite loop space, which under the assumption that p is regular or on the non-vanishing of the p-adic zeta function at integral arguments, has the property that the homotopy groups with coefficients $\pi_n(SK(\mathbf{Z},\mathbf{Z}_p);\mathbf{Z}_p)$ are rationally non-zero in exactly degrees $2,6,10$, etc. The referee noted that the computation of

rational stable completed homology should follow precisely from this description using the Milnor–Moore theorem, which shows that (for simply connected H-spaces) that the homology is (rationally) the universal enveloping algebra of the rational homotopy classes (and so, in particular, the Hurewicz map is rationally injective). One consequence is that the rational homotopy groups are precisely the primitive classes in rational homology. To orient the reader, this is exactly the theorem which allowed Borel to compute the rational K-groups of (rings of integers) of number fields from his computation of stable homology over $\mathbf{Q}$. Now I was a little worried about this, because the Milnor–Moore theorem does not literally apply, since one is comparing here homotopy groups with coefficients in $\mathbf{Z}_p$ and continuous homology (the latter is just the inverse limit of homology groups modulo p^n). However, having looked at the argument in Milnor–Moore and then having Paul Goerss explain it to me, the argument does indeed seem to simply work in this case. (Warning, this is a weaker statement than saying I checked the details.)

To be more precise, suppose that G is a simply connected infinite loop space, and suppose that G has the property that the groups $\pi_n(G; \mathbf{Z}/p^k)$ are finite for all n and k, so $\pi_n(G; \mathbf{Z}_p)$ is the inverse limit of these groups. There is a pairing

$$[,] : \quad \pi_r(G, \mathbf{Z}/p^k) \otimes \pi_s(G, \mathbf{Z}/p^k) \to \pi_{r+s}(G, \mathbf{Z}/p^k),$$

which, after taking inverse limits in k and tensoring with $\mathbf{Q}$, makes $\pi_*(G, \mathbf{Z}_p) \otimes \mathbf{Q}$ into a Lie algebra over $\mathbf{Q}_p$, then the Hurewicz map will induce an isomorphism

$$U(\pi_*(G, \mathbf{Z}_p) \otimes \mathbf{Q}) \to H_*^{\mathrm{cont}}(G, \mathbf{Q}_p) := \lim H_*(G, \mathbf{Z}/p^k) \otimes \mathbf{Q}$$

of Hopf algebras. The key technical point required here is to define the appropriate pairing on homotopy groups with coefficients, which is done by Neisendorfer. (If G is simply connected infinite loop space, one doesn't have to worry about the issue of homotopy groups with coefficients in very low degree exhibiting certain pathologies.)

As another example of this, one can take $G = SK(\mathbf{Z}_p)$. In this case, the rational continuous homology reduces, by work of Lazard, to lie algebra cohomology, and gives an exterior algebra in odd degrees ≥ 1. So $SK_n(\mathbf{Z}_p; \mathbf{Z}_p) \otimes \mathbf{Q}$ has dimension one in odd degrees ≥ 1 and is zero for all even positive degrees. This is a result of Wagoner. In fact, Wagoner proves something slightly stronger, also capturing some information away from p. To do this, he also proves a version of the Milnor–Moore theorem, but his assumptions are more stringent than what we discuss above.

65. Higher direct images of canonical extensions

Sun, 04 Jan 2015

I like Kai-Wen's talks; he gives lots of examples, writes big with big chalk, and clearly explains the key points of the argument. I'm not sure I would classify his thesis as light reading material, but if he produced a video series explaining all the details in lecture format, I would buy the DVD. Speaking of different ideas for disseminating mathematics, I have some thoughts on that, but they will have to wait for another time. For now, I just wanted to make the smallest remark concerning Kai-Wen's lecture at the Harris conference.

As all my readers surely know (this is code for I am not going to explain why), a key ingredient in the Harris–Lan–Taylor–Thorne [HLTT16] argument is the fact the

the higher direct images of the of subcanonical automorphic vector bundles under the projection from the toroidal compactification to the minimal compactification of quite general classes of Shimura varieties vanish. In contrast, this does not hold for the higher direct images of the canonical extensions, and when this was first being discussed, it was not entirely clear (at least to me) what was going on. But Kai-Wen's talk actually does make the situation very clear! That is what I want to talk about.

Let X be the open Shimura variety, let Y be a minimal compactification, and let Z be a toroidal compactification. To avoid silliness, assume that $Y \smallsetminus X$ has codimension at least two. Let W be an automorphic vector bundle on X, and let W^{can} and W^{sub} denote the canonical and subcanonical extensions of W to Z. There's a short exact sequence

$$0 \to W^{\mathrm{sub}} \to W^{\mathrm{can}} \to Q \to 0.$$

Take the pushforward of this to Y. We know that the higher direct images of the first sheaf vanish, and so we obtain an exact sequence

$$0 \to \pi_* W^{\mathrm{sub}} \to \pi_* W^{\mathrm{can}} \to \pi_* Q \to 0.$$

The last sheaf is supported on $Y \smallsetminus Z$, which has fairly small dimension, so its cohomology groups vanish in high degree by Grothendieck. Now let us assume that the higher direct images also vanish for W^{can}. It follows that the Leray spectral sequence degenerates (for both W^{sub} and W^{can}), and so we obtain isomorphisms

$$H^*(Z, W^{\mathrm{sub}}) = H^*(Z, W^{\mathrm{can}})$$

in sufficiently high degree. Now the canonical bundle on Z is also an automorphic vector bundle, and so Serre duality relates the cohomology of W^{sub} to the cohomology of V^{can} for another automorphic vector bundle V, and relates the cohomology of W^{can} to V^{sub}. For example, for modular curves, the Serre dual of ω^k is $\omega^{2-k}(\infty)$, because the canonical sheaf of the modular curve is $\Omega^1 \simeq \omega^2(\infty)$. Hence (using the assumption on codimensions made above so the numerology works out) we end up with the isomorphism

$$H^0(Z, V^{\mathrm{sub}}) = H^0(Z, V^{\mathrm{can}}).$$

But this formula says that all modular forms of weight V are cuspidal! So this gives an easy proof of:

Lemma 65.1. *If there exists at least one form of weight V which is not cuspidal, then at least one of W^{sub} or W^{can} has non-trivial higher direct images under π.*

Of course, we know from [HLTT16] that it will be the second (because the higher direct images of the first vanish), but we didn't prove that. Now I just chatted with Kai-Wen, who did one better than this lemma. First of all, remember that there is an automorphic line bundle ω on X (corresponding to "parallel weight") which is ample, and the corresponding canonical extension to Z descends to an ample on Y, which we also call ω. What's nice about this is that, using the projection formula, one can replace the question about the vanishing of the higher direct images of W by the vanishing of W under twists by powers of this bundle. But that means one can translate the problem of asking whether there exists a non-cusp form in the dual weight V to whether there exists a non-cusp form in weight $V \otimes \omega^n$ for some arbitrarily large n. Now as before, we have an exact sequence:

$$0 \to \pi_* V^{\mathrm{sub}} \otimes \omega^n \to \pi_* V^{\mathrm{can}} \otimes \omega^n \to \pi_* R \otimes \omega^n \to 0.$$

twisted by some arbitrarily high power of ω, where we have used the vanishing of $R^1\pi_* V^{\mathrm{sub}}$ and the projection formula. Here R is just $V^{\mathrm{can}}/V^{\mathrm{sub}}$. On the other hand, because ω is ample on Y, we know that

(1) $H^1(Y, \pi_* V^{\mathrm{sub}} \otimes \omega^n)$ vanishes for sufficiently large n,
(2) $\pi_* R \otimes \omega^n$ is generated by global sections for sufficiently large n, and so, for such n, we have $H^0(Y, \pi_* R \otimes \omega^n) \neq 0$ as long as $\pi_* R \neq 0$.

So if one shows that $\pi_* R$ is non-zero then one is done. Certainly R is non-zero, but analyzing $\pi_* R$ is a bit more subtle (I jumped the gun a little on the first version of this post, but Kai-Wen told me I needed to be a little more careful). On the other hand, there are many classical examples where one can explicitly construct non-cuspidal forms. For example, one can take $X = \mathcal{A}_g$ with $g \geq 2$ to be the Siegel moduli space, and take W to be the line bundle ω^k. Then Siegel himself constructed the so-called Siegel Eisenstein series for high enough k. Kai-Wen also tells me the non-vanishing of $\pi_* R$ can be proved more generally for $X = \mathcal{A}_g$, and so one has:

Lemma 65.2 (Kai-Wen)**.** *Let $g \geq 2$, let $X = \mathcal{A}_g$, and let W be an automorphic bundle. Then at least one of higher direct images $R^i \pi_* W^{\mathrm{can}}$ with $i \geq 0$ must be non-zero.*

In fact, Kai-Wen also tells me he had a proof of (a more general version of) this last result even before HLTT knew about the vanishing of $R^i \pi_* V^{\mathrm{sub}}$, but this argument gives a completely transparent proof of why they can't *both* vanish.

66. Abelian spiders

Sun, 11 Jan 2015

This is a blog post about the thesis of my student Zoey Guo, who is graduating this year.

Let Φ be a finite graph. Associated to Φ is an adjacency matrix M such that the largest eigenvalue λ is totally real. Let us call Φ *abelian* if the extension $\mathbf{Q}(\lambda^2)$ is abelian. For example, all the Dynkin diagrams are abelian.

Several years ago, Scott Morrison and Noah Snyder (of secret blogging seminar fame) asked the following question. Given a finite graph Γ, let Γ_n be the graph obtained by adjoining a 2-valent tree of length n to some fixed vertex v of Γ. Then can one classify all n for which Γ_n is abelian? It turns out Γ_n can be abelian for only finitely many n, unless the graphs Γ_n happen to be one of the two infinite families of Dynkin diagrams. The argument was effective, although not effectively effective. (We did, however, prove a slightly weaker theorem which was sufficient for the intended application which was effectively effective.)

What Zoey does in her thesis is consider the following generalization. Let Γ be a finite graph, and choose k vertices v_i of Γ. Now adjoin k two-valent graphs of varying lengths $\underline{n} = \{n_i\}$ to latex v_i. Call the resulting graph a *k-spider graph*. The main result of her thesis is the following:

Theorem 66.1 ([CG18b])**.** *For any Γ and any fixed k, only finitely many of the corresponding spiders are both abelian and not Dynkin diagrams.*

What is more, the theorem is effectively effective. One key ingredient in the finiteness results for $k = 1$ was the fact that, for characteristic polynomials $P_n(X)$ of

the graphs Γ_n, one can control the factors corresponding to Chebyshev polynomials. Or, if one writes

$$Q_n(t) = P_n(t + t^{-1}),$$

one can control the cyclotomic factors of $Q_n(t)$. This follows from a theorem of Hironaka–Gross–McMullen [GHM09], who exploit results by Mann on vanishing sums for roots of unity. However, when $k \geq 2$, this breaks down completely. In fact, in many examples, the corresponding polynomials $P_n(t + t^{-1})$ will be divisible by cyclotomic polynomials of arbitrarily large degree. Here's an example which Zoey pointed out to me. Consider the disconnected graph consisting of a two copies of the Dynkin diagrams A_{n-1} and a third component A_m for any integer m (in fact, the construction is much more general, but we will be very explicit here). Since $\zeta + \zeta^{-1}$ for $\zeta^n = 1$ is an eigenvalue of each connected component corresponding to A_{n-1}, it will be an eigenvalue of multiplicity (at least) two of their union. Now join the three graphs by adding a vertex which is connected to the end of all three graphs; this will be the 3-spider on a point corresponding to the triple $(n-1, n-1, m)$. By the interlacing lemma for graph eigenvalues, $\zeta + \zeta^{-1}$ will be an eigenvalue of the resulting connected graph. So finding all the cyclotomic factors even for the explicit polynomials coming from 3-spiders on a point seems like a real pain, and so one needs a new argument to deal with roots of unity.

The proof of the theorem comes down to two key steps. First, for any sequence of spiders, all but a uniformly bounded number of eigenvalues will lie in the interval $[-2, 2]$, and all the eigenvalues will lie in some uniform interval $[-M, M]$. Of course, this means that the squares of the eigenvalues lie always in $[0, M]$ for some M and mostly in $[0, 4]$. Now imagine the largest such number λ^2. We know that it is algebraic, and we are assuming that it is abelian. The key quantity to control turns out to be the normalized trace of $\gamma := (\lambda^2 - 2)^2$. Work of Cassels shows that, if λ^2 is cyclotomic, one can classify all cyclotomic integers for which the normalized trace of γ is small. What kind of an upper bounds do we have? Well, if the degree of λ is very large, then the bulk of the contribution has to come from $\lambda \in [-2, 2]$, or $\lambda^2 - 2 \in [-2, 2]$ (this is the reason why $\lambda^2 - 2$ occurs above — it is a Chebyshev polynomial). The worst case scenario is that all of the conjugates of λ^2 are near 4, which will give an estimate on the normalized trace of γ of 4 plus a quantity that goes to zero with the degree. However, it is hard for an algebraic number to have too many of its conjugates near any particular integer (in this case, 4). To exploit this, one can note, for example, that

$$3 - x - \log(4 - x) \geq 0, \quad x \in [0, 4].$$

If we denote the conjugates of $\gamma := (\lambda^2 - 2)^2$ by $\sigma\gamma$ and suppose that this has degree n, then we deduce that

$$3n - n\mathrm{Tr}(\gamma) - \log\left(\prod(4 - \sigma\gamma)\right) \geq O(1).$$

The $O(1)$ term (which is completely explicit) comes from the fact that finitely many of the $\sigma\gamma$ are outside $[0, 4]$ and so one can not use the previous inequality. Let us consider the inequality. As long as $\gamma \neq 4$, the logarithmic factor is a norm of the algebraic integer $4 - \gamma$, and hence non-negative. So we get an upper bound for the normalized trace which is now 3 plus some explicit error term which tends to zero as the degree goes to infinity. This type of idea was first used by Chris Smyth when studying the trace problem of Siegel (see this post). Now 3 is still too big to apply

the results of Cassels. So one has to also exploit that $(\lambda^2-2)^2$ is not too close to 3 either (this is the where the inequality above is an equality). In fact, one ends up using not just 4 and 3 but 43 different algebraic integers which are all of the form $2+\zeta+\zeta^{-1}$, and where one uses logarithms weighted by various real constants. The precise constants and algebraic integers were optimized by simulated annealing. In the end, one gets an upper bound of the form 2.4 plus a very explicit error term, which is enough for the Cassels machine. This gives a complete answer as long as the degree of λ is big enough.

If λ has small degree, then one is also in good shape — given a bound for λ and a bound for the degree, there are only finitely many such algebraic integers, which is enough to prove the theorem. However, this last step — whilst effective — is not at all effectively effective. So another argument is required to make everything work in practice. Note that the degree of the polynomial defining λ certainly goes to infinity, but it may be reducible, and in particular divisible by many cyclotomic (Chebyshev) factors all of whose roots are in $[-2,2]$. Let's explain how to overcome this issue in the case of 3-spiders coming from the trivial graph (which is typical of the general case). If you take the 3-spider with legs of length (a,b,c), then, for (a,b,c) big enough, one finds that

$$\lambda \to \frac{3}{\sqrt{2}}.$$

(The limit of the largest eigenvalues of a sequence of infinite spiders will always be an algebraic number.) Importantly, this convergence is exponential. However, it's easy to see that any algebraic integer all of whose conjugates are uniformly bounded cannot be extremely close to any fixed algebraic number, and it easy to give effective bounds to this effect. So one wins in high degree by Cassels type arguments and in low degree by the fact that the eigenvalues converge rapidly to computable algebraic numbers. One annoying issue is that the convergence requires all of the (a,b,c) to tend to infinity, so one has to inductively reduce to the case of 2-spiders with some finite list of possible c, which entails a certain amount of combinatorial explosion. However, as a complete worked example, one has the following:

Theorem 66.2 ([CG18b])**.** *The complete list of abelian 3-spiders on a point is given by:*

(1) The Dynkin diagrams A_n, D_n, E_6, E_7, E_8, $\widetilde{E}_6$, $\widetilde{E}_7$, $\widetilde{E}_8$, whose largest eigenvalue is of the form $\zeta+\zeta^{-1}$,

(2) The 3-spiders $(3,3,3)$, $(2,4,4)$, and $(2,3,7)$ with

$$\lambda^2 = \frac{5+\sqrt{13}}{2},$$

(3) The 3-spiders $(3,3,7)$, $(2,8,8)$, and $(2,7,11)$ with $\zeta^{13}=1$ and

$$\lambda^2 = \zeta^{11}+\zeta^{10}+\zeta^3+\zeta^2+2,$$

(4) The 3-spiders $(4,4,4)$, $(3,5,5)$, and $(3,4,9)$ with

$$\lambda^2 = 3+\sqrt{2}.$$

Now all of this is quite amusing, but you may complain that it doesn't really have any practical application. However, as it happens, Scott Morrison asked me whether it was possible to find all abelian 2-spiders for some very explicit graph (omitted here), in order to further the classification of finite index subfactors, because the

all the current non-number theoretic obstructions could not rule out this family of examples as coming from subfactors. Zoey's method could be applied to show that every 2-spider in the corresponding family was not abelian. So Zoey's results have *already* been used outside her field (see forthcoming work of Morrison and his collaborators) to complete the classification of subfactors of index between 5 and $3+\sqrt{5}$. All this, of course, while having the thesis with the best title.

Comment 66.3 (Scott Morrison)**.** Happily, we've found yet another application of Zoey's results in the classification of subfactors, and indeed our almost finished paper on the classification up to index $3+\sqrt{5}$ will make two independent, and rather different uses, of this work.

The first is the example you talk about above and that Zoey used as an illustration of the method.

The second is even more exciting — and we haven't yet had a chance to see if this method could help pushing the classification even further.

The basic idea is that we have a family of graphs as follows: fix some finite graph Γ, and mark one vertex "x", and name all the vertices at the maximal radius from x "Y". We now consider adding a chain of n edges to x, and gluing on some arbitrary finite graph to Y.

For the Γ we're interested in (and possibly 'often') we can show by the theory of "connections" that any graph in this family which is the principal graph of a subfactor must have graph norm satisfying some particular polynomial depending just on n (and not on what we glue to Y).

Now we don't know that this polynomial has anything to do with any particular finite graph (in fact, it's a multiple of the minimal polynomial for the norm of a certain infinite graph) so we have to work a bit harder in places to apply Zoey's method, but happily it all works out, and one can show that the graph norm cannot possibly be cyclotomic, thereby ruling out all graphs in the family as principal graphs of a subfactor.

Notes 66.4. The paper which relies on the main theorem of [CG18b] is [AMP23]. But that paper only cites the arXiv version of [CG18b], not the published ones, despite the latter appearing in 2023 and the former in 2018.

67. Inverse Galois problems II

Tue, 13 Jan 2015

David Zywina was in town today to talk about a follow up to his previous results (see § 17) discussed previously on this blog. This time, he talked about his construction of Galois groups which were simple of orthogonal type, in particular, the simple groups

$$\Omega(V) \subset \mathrm{SO}(V) \subset \mathrm{O}(V)$$

where V is a vector space V over $\mathbf{F}_l$ of odd dimension at least five. The group $\Omega(V)$ here is a simple group of index two inside $\mathrm{SO}(V)$. In the special case when $n=5$, there is an exceptional isomorphism

$$\Omega(V) \simeq \mathrm{PSp}_4(\mathbf{F}_l).$$

In contrast to his constructions of number fields with Galois group $\mathrm{PSp}_2(\mathbf{F}_l)$, Zywina actually constructs a family of compatible families whose residual image is

generically $\Omega(V)$. When David told me about this construction (scribbled on a piece of paper) in Frankfurt airport on the way back from Oberwolfach, I was troubled by something which I shall now explain. Without saying so much about the construction (you can read about it here), the compatible families of Galois representations of interest occur inside $H^2(X_T, \mathbf{Q}(1))$ for a carefully chosen family of non-isotrivial elliptic surfaces X. As David explained in his talk today, the zero section and the fibres of bad reduction contribute a large Galois trivial summand to $H^{1,1}$, and the remaining piece is five dimensional. What disturbed me at the time was that this construction was surely liftable to a compatible family of *four* dimensional representations with generalized symplectic image. After all, Tate's result on $H^2(G_{\mathbf{Q}}, \mathbf{C}^\times)$ guarantees that one can lift any projective representation with image in $\mathrm{PSp}_4(\mathbf{F}_l)$ to a genuine generalized symplectic representation. This representation should then come from a Siegel modular form, since all oddness conditions should be automatic. On the other hand, if you want a family of Galois representations giving rise to a family of Siegel modular forms, especially one for which the maximal difference between any two Hodge–Tate weights in $\wedge^2 W$ is two, then you expect that they have to come from a family of abelian surfaces, or at least abelian varieties A of dimension $2n$ with endomorphisms by the ring of integers in a totally real field of degree n. However, there is an obstruction to making this work — the corresponding Galois representations will have Hodge–Tate weights $[0, 0, 1, 1]$, and they will have similitude character that is an even finite order character times the cyclotomic character. It's easy to see that for such a family, the residual representations will (at least half the time) land in $\mathrm{PGSp}_4(\mathbf{F}_l)$ and not in the simple index two subgroup, similar to what happens for modular forms of weight two. I thought at the time that I must have been making some group theory error, so after today's talk we sorted out the details.

In the process of this computation, however, I realized what my error *actually* was. I was imagining that the original compatible family of Galois representations in H^2 had Hodge–Tate weights $[0, 1, 1, 1, 2]$, but they could equally have had Hodge–Tate weights $[0, 0, 1, 2, 2]$. And in this latter case, the Galois representation (up to twist) of the corresponding Siegel modular form in GSp_4 will have Hodge–Tate weights $[-1, 0, 0, 1]$. In particular, we are not looking for classical Siegel modular forms of low weight, but the nasty Siegel modular forms which do not contribute to holomorphic limits of discrete series and only occur in coherent cohomology via H^1 or H^2. (A reference for this fact is George Boxer's talk in Barbados.) And now everything makes sense! That is, if you have a family of Galois representations with Hodge Tate weights $[-1, 0, 0, 1]$ and quadratic similitude character, then (with some good luck) you can really have projective representations which land in the right simple group for all but finitely many l.

A related point: when lifting projective representations using Tate's theorem, one may have to increase the size of the residue field. In fact, when $\ell \equiv 3 \mod 4$, it will not be possible to lift an odd $\mathrm{PSp}_4(\mathbf{F}_l)$ representation to one in $\mathrm{GSp}_4(\mathbf{F}_l)$ (there is an obstruction at infinity). Indeed, the natural lift is the group $\mathrm{Sp}_4(\mathbf{F}_l)$ together with a scalar matrix I with $I^2 = -1$. This suggests what the picture should be motivically: there should be an eight dimensional piece of $H^2(Y)$ (for some Y) which admits an involution breaking the representation up into two four dimensional pieces, and these pieces will have coefficients in $\mathbf{Q}(\sqrt{-1})$. Can one find

such a Y explicitly? This does remind me of the motivic lifting problems that Stefan Patrikis knows about.

From this analysis, it also becomes clearer why Zywina could find a family of compatible families with residual image $\Omega(V)$ when $\dim(V)$ is odd and at least five, but only isolated examples of compatible families with residual projective image $\mathrm{PSL}_2(\mathbf{F}_l) \simeq \Omega(V)$ with $\dim(V) = 3$. In the latter case, the corresponding modular forms will be forced to have odd weight $k \geq 1$, and so the Hodge–Tate weights will differ by at least two, and so Griffiths' theorem implies that they should not deform in a family. On the other hand, if you want to look for Siegel modular forms which could possibly correspond to geometric families, and you want the similitude character to be an even power of the cyclotomic character times a finite character, then it is possible to escape the spectre of Griffiths on your shoulder, but only barely — you will be pretty much forced to work with forms whose HT weights are $[-1, 0, 0, 1]$. Of course, I'm not sure I can prove that any Siegel modular forms of this kind actually exist! (insisting the Mumford–Tate group is big, naturally). My proposal in the previous post to look for these representations using Siegel modular forms would also have only found sporadic compatible families, because to ensure computability and the determinant condition I suggested looking in weights where the Galois representation was regular and had HT weights something like $[0, 1, 3, 4]$, — the gap being necessary to make the multiplier character a square of a Hodge–Tate character.

There is one check left on these musings (though I'm sure it must be correct), namely, that for the surface X in 1.4 of Zywina's paper (also see [Zyw23]), one should have

$$h^{2,0}(X) = 2.$$

Proof. The Hodge diamond of a minimal elliptic surface $\pi : X \to C$ was computed by Miranda, see [Mir89, IV.1.1] here; I'll try to give a self contained argument. Let ω_E be the Euler characteristic of $\mathcal{O}_X$. Let L^{-1} be the bundle $L^{-1} = R^1\pi_*\mathcal{O}_X$ on X; it is a line bundle because the fibres are elliptic curves, so it makes sense to talk about L. The bundle L has positive degree if and only if the fibration is not isotrivial (this is not so hard, but let me give the proof of III.1.6 of Miranda as a reference); let us assume this is the case. From the Leray spectral sequence, there is an exact sequence

$$0 \to H^1(C, \pi_*\mathcal{O}_X) \to H^1(X, \mathcal{O}_X) \to H^0(C, L^{-1}) \to H^2(C, \pi_*\mathcal{O}_X)$$

Since $\pi_*\mathcal{O}_X = \mathcal{O}_C$, the first term has dimension g, the genus of C. Since we are assuming that L has positive degree, the third term is also zero, and hence the irregularity of a non-isotrivial elliptic surface is

$$H^1(X, \mathcal{O}_X) = g.$$

It follows that

$$\chi(\mathcal{O}_X) = h^{0,0} - h^{1,0} + h^{2,0} = 1 - g + h^{2,0}.$$

In our particular case, the genus of C is zero. On the other hand, as noted in 2.4 of Zywina's paper, the degree of the minimal discriminant is $12 \cdot \chi_E = 12 \cdot \chi(\mathcal{O}_X)$. In the example at hand, Zywina computes (see section 8) that $\chi_E = 3$, and so

$$h^{2,0} = 3 - (1 - g) = 3 - 1 = 2.$$

□

68. $H_2(\Gamma_N(p), \mathbf{Z})$

Fri, 30 Jan 2015

In this post (which is a follow-up to § 64, I wanted to compute the group $H_2(\Gamma_N(p), \mathbf{Z})$, where $\Gamma_N(p)$ is the congruence subgroup of $\mathrm{SL}_N(\mathbf{Z})$ for large enough N and p is prime. In fact, to make my life easier, I will also assume that $p \geq 3$, and in addition, ignore 2-torsion. The first problem is to compute the prime to p torsion. By Charney's theorem, this will come from the cohomology of the homotopy fibre X of the map

$$SK(\mathbf{Z}) \to SK(\mathbf{F}_p).$$

The relevant part of the Serre long exact sequence is, using classical computations of the first few K-groups of the integers together with Quillen's computation of $K_*(\mathbf{F}_p)$,

$$0 \to \pi_3(X) \to \mathbf{Z}/48\mathbf{Z} \to \mathbf{Z}/(p^2-1)\mathbf{Z} \to \pi_2(X) \to \mathbf{Z}/2\mathbf{Z} \to 0.$$

Here is where it is convenient to invert primes dividing 6; from Hurewicz theorem and Charney's theorem we may deduce that, where $\sim$ denotes an equality up to a finite group of order dividing 48,

$$H_2(\Gamma_N(p), \mathbf{Z}[1/p]) \sim \mathbf{Z}/(p^2-1)\mathbf{Z}.$$

In order to deal with 3-torsion, then we also have to show that the map $K_3(\mathbf{Z}) \otimes \mathbf{Z}/3\mathbf{Z} \to K_3(\mathbf{F}_p)$ is injective for $p \neq 3$. I have a sketch of this which I will omit from this discussion but it is not too hard (assuming Quillen–Lichtenbaum). It remains to compute the homology with coefficients in $\mathbf{Z}_p$. I previously computed that there was an isomorphism

$$H_2(\Gamma_N(p), \mathbf{F}_p) = \wedge^2 \mathfrak{g} \oplus \mathbf{F}_p = H_2(G_N(p), \mathbf{F}_p) \oplus \mathbf{F}_p,$$

where $G_N = \mathrm{SL}_N(\mathbf{Z}_p)$ and $\mathfrak{g} = H_1(\Gamma_N(p), \mathbf{F}_p)$ is the adjoint representation.

68.1. **Some facts concerning the cohomology of $G_N(p)$.** There are short exact sequences:

$$0 \to H_2(G_N(p), \mathbf{Z}_p)/p \to H_2(G_N(p), \mathbf{Z}/p\mathbf{Z}) \to H_1(G_N(p), \mathbf{Z}_p)[p] \to 0,$$

$$0 \to H_2(G_N(p), \mathbf{Z}_p)/p^2 \to H_2(G_N(p), \mathbf{Z}/p^2\mathbf{Z}) \to H_1(G_N(p), \mathbf{Z}_p)[p^2] \to 0.$$

Since $H_1(G_N(p), \mathbf{Z}_p) = \mathfrak{g}$ is annihilated by p, we may deduce that

$$H_2(G_N(p), \mathbf{Z}_p)/p = H_2(G_N(p), \mathbf{Z}_p)/p^2$$

as long as

$$|H_2(G_N(p), \mathbf{Z}/p^2\mathbf{Z})| = |H_2(G_N(p), \mathbf{Z}/p\mathbf{Z})|.$$

Such an equality (for any group) is a claim about the Bockstein maps having a big an image as possible. Indeed, for any group Φ, there is an exact sequence:

$$H_3(\Phi, \mathbf{Z}/p\mathbf{Z}) \to H_2(\Phi, \mathbf{Z}/p\mathbf{Z}) \to H_2(\Phi, \mathbf{Z}/p^2\mathbf{Z}) \to H_2(\Phi, \mathbf{Z}/p\mathbf{Z}) \to H_1(\Phi, \mathbf{Z}/p\mathbf{Z})$$

The first and last maps here are the Bockstein maps β_2 and β_1. Since p is odd, $\beta_1 \circ \beta_2 = 0$. On the other hand, we see that the orders of the cohomology groups with coefficients in $\mathbf{Z}/p\mathbf{Z}$ and $\mathbf{Z}/p^2\mathbf{Z}$ will have the same order if and only if

$$\ker(\beta_1) = \mathrm{im}(\beta_2).$$

Hence we have reduced to the following claim. Take the complex

$$H_*(G_N(p), \mathbf{F}_p) = \wedge^* \mathfrak{g}$$

where the differentials are given by the Bockstein maps. Then we have to show that the cohomology of this complex vanishes in degree two. But what are the Bockstein map is in this case? Note that since $H_1(G_N(p), \mathbf{Z}_p) = \mathfrak{g}$ is annihilated by p, the Bockstein map β_1 will be a surjective map:

$$\beta_1 : \wedge^2 \mathfrak{g} \to \mathfrak{g}.$$

To compute this explicitly, recall that the isomorphism $H_1(G_N(p), \mathbf{Z}_p) = \mathfrak{g}$ comes from the identification of $\mathfrak{g}$ with $G_N(p)/G_N(p^2)$. Then, *computation omitted due to laziness,* we find that the Bockstein is precisely the Lie bracket. Moreover, since the (co-)homology is generated in degree one, the higher Bockstein maps can be computed from the first using the cup product formula. So the Bockstein complex above is, and I haven't checked this because it must be true, the complex computing the mod-p Lie algebra cohomology of $\mathfrak{g}$. And this cohomology vanishes in degrees one and two, so we are done. One consequence of this computation is that

$$H_2(G_N(p), \mathbf{Z}_p) = H_2(G_N(p), \mathbf{Z}_p)/p$$

is annihilated by p. Moreover, the last term can be identified with the kernel of the Lie bracket (Bockstein) on $H_2(G_N(p), \mathbf{F}_p) = \wedge^2 \mathfrak{g}$.

68.2. **Returning to the main computation.** From the Hochschild–Serre spectral sequence and the computation of stable completed cohomology, one has an exact sequence:

$$0 \leftarrow H_2(G_N(p), \mathbf{Z}_p) \leftarrow H_2(\Gamma_N(p), \mathbf{Z}_p) \leftarrow \mathbf{Z}_p \leftarrow H_3(G_N(p), \mathbf{Z}_p).$$

From known results in characteristic zero, we immediately deduce that there is some α such that there is an exact sequence

$$0 \leftarrow H_2(G_N(p), \mathbf{Z}_p) \leftarrow H_2(\Gamma_N(p), \mathbf{Z}_p) \leftarrow \mathbf{Z}/p^\alpha \mathbf{Z} \leftarrow 0.$$

we also deduce that there is an exact sequence:

$$0 \leftarrow H_2(G_N(p), \mathbf{Z}/p^n\mathbf{Z}) \leftarrow H_2(\Gamma_N(p), \mathbf{Z}/p^n\mathbf{Z}) \leftarrow \mathbf{Z}/p^{\min(\alpha,n)}\mathbf{Z} \leftarrow 0,$$

There are spectral sequences:

$$H_i(\mathrm{SL}_N(\mathbf{F}_p), H_j(\Gamma_N(p), A)) \Rightarrow H_{i+j}(\Gamma_N, A)$$

for $A = \mathbf{Z}/p^n\mathbf{Z}$ and $A = \mathbf{Z}_p$. For both of these rings, we have

$$H_1(\Gamma_N(p), A) = \mathfrak{g}, \qquad H_0(\Gamma_N(p), A) = A.$$

Moreover, for sufficiently large N, we have

$$H_i(\mathrm{SL}_N(\mathbf{F}_p), A) = 0,$$

this follows from and is equivalent to Quillen's computation which implies that the K-groups of finite fields have order prime to p. Since $H_2(\mathrm{SL}_N(\mathbf{Z}), \mathbf{Z}_p)$ is trivial for $p \geq 2$, we deduce that

$$H_0(\mathrm{SL}_N(\mathbf{F}_p), H_2(\Gamma_N(p), A)) = H_2(\mathrm{SL}_2(\mathbf{F}_p), \mathfrak{g}) = \mathbf{F}_p,$$

where the last equality was already used in my paper. The compatibility of the spectral sequence above for different A implies that we also get an isomorphism

$$H_0(\mathrm{SL}_N(\mathbf{F}_p), H_2(\Gamma_N(p), \mathbf{Z}_p)) = H_0(\mathrm{SL}_N(\mathbf{F}_p), H_2(\Gamma_N(p), \mathbf{Z}/p\mathbf{Z})).$$

On the other hand, the invariant class must be an element of order p in

$$\mathbf{Z}/p^\alpha\mathbf{Z};$$

and hence the reduction map

$$\mathbf{Z}/p^\alpha\mathbf{Z} \to \mathbf{Z}/p\mathbf{Z}$$

sends an element of order p to an element of order p, and so $\alpha = 1$.

68.3. **Putting things back together.** Assembling all the pieces, we see that we have proven the following:

Theorem 68.4. *Let $p \geq 3$, and let N be sufficiently large. Let $\mathfrak{g}$ be the Lie algebra $\mathfrak{sl}_N$ over $\mathbf{F}_p$. Then, up to a finite group of order dividing 48, we have*

$$H_2(\Gamma_N(p), \mathbf{Z}) \sim \mathbf{Z}/(p^2-1)\mathbf{Z} \oplus \mathbf{Z}/p\mathbf{Z} \oplus \ker\left([\ ,\ ] \wedge^2 \mathfrak{g} \to \mathfrak{g}\right).$$

Moreover, still with $p \geq 3$, then up to a group of order dividing 16, we have the same equality with p^2-1 replaced by $(p^2-1)/3$.

69. Review of Buzzard–Gee

Wed, 04 Mar 2015

This is a review of the paper "Slopes of Modular Forms" submitted for publication in a Simons symposium proceedings volume (see [BG16]).

tl;dr: This paper is a nice survey article on questions concerning the slopes of modular forms. Buzzard has given a (very explicit) conjecture which predicts the slopes of classical modular (p-stabilized) eigenforms of level prime to p, at least under a certain regularity hypothesis. One consequence is that, under favourable circumstances, all the slopes are integers. The current paper describes the link between this and related problems to the p-adic Langlands program, as well as raising several further intriguing questions concerning the distributions of these slopes. The paper is well written, and is a welcome addition to the literature. I strongly recommend that this paper be accepted.

Review: Buzzard's slope conjectures live somewhere in the world between 19th and 21st century mathematics. Suppose that one considers the space of over-convergent cusp forms of level $N = 1$ for $p = 2$. Then, using nothing more than classical identities between modular functions, one may prove that the smallest eigenvalue of the compact operator U_2 is at most $\|2^3\|_2 = 1/8$. On the other hand, it is now a "folklore" conjecture (Conjecture 4.1.1 of the paper under review) that, if p is odd and f is a classical modular form of level $\Gamma_0(N)$ prime to p, then the residual representation:

$$\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_2(k)$$

is irreducible locally on the decomposition group at p whenever the valuation of a_p is not an integer. This problem seems to be a deep question in the p-adic Langlands program for $\mathrm{GL}_2(\mathbf{Q}_p)$. The two cases where this is known, $v(a_p) \leq 1$ and $v(a_p)$ sufficiently large, both require machinery from p-adic Hodge theory — in the former case, one needs the full local Langlands correspondence for $\mathrm{GL}_2(\mathbf{Q}_p)$.

Comments: Here are some comments given in some order that bears little relation to the actual paper.

- I don't like the table on page 6, in particular, because certain ranges of numbers are bunched together, the output looks a little strange. Can one improve this in some way? Perhaps finish at 3^{10}? Perhaps include only selected powers bigger than 3^{10}? Perhaps normalize for the length of the range?
- Corollary 5.1.2. Do you want to speculate on what happens in the reducible case? In some sense, in Buzzard's conjecture, one doesn't see the fact that the residual representations are globally reducible or not. On the other hand, weird stuff certainly happens for $p = 2$, as previously mentioned here. What happens in the reducible case for $p = 3$?
- Conjecture 4.1.1 demands that k is even, but that is a consequence of the level being of the form $\Gamma_0(N)$ — something which is noted immediately after the statement. So why include the condition in the statement of the conjecture? Also, perhaps it's also worth remarking upon the case when a_p is a unit.
- The authors (in Remark 4.1.3) point to the origins of this Conjecture 4.1.1 to around 2005. However, I feel like I remember some discussion of this conjecture in the Durham symposium of 2004. There were certainly hints of this conjecture on «*la serviette de Kisin*», upon which Mark gave a heuristic local argument for why the Eigencurve was proper — although the argument was slightly dodgy in that it collapsed if the napkin was rotated 90 degrees. (Of course, Mark was proved right when Hansheng Diao and Ruochuan Liu did indeed prove this result using local methods here, [DL16]) Also, isn't Conjecture 4.1.1 a consequence of Buzzard's original conjecture as modified by Lisa Clay? Somehow it seems to me that what Remark 4.1.3 is referring to is the idea that Conjecture 4.1.1 is a consequence of a purely *local* conjecture, and refers to the period (2005?) when Breuil was formulating the first versions of the p-adic Langlands program.
- For Conjecture 4.2.1, wouldn't it make more sense to normalize the valuation in terms of the coefficient field $\mathbf{Q}_p(\chi)$, so the statement once more becomes that a_p has integral valuation?
- Why is the condition on Buzzard's conjecture different when $p = 2$? (I understand it has to be modified in order to have a chance of being true, but I am asking if there is any explanation for why this is necessary.)
- The authors remark (p.3) that it is not known whether there are infinitely many Buzzard irregular primes. Here is a short argument to prove that this is a consequence of standard conjectures of prime values of polynomials. We start with the observation that the first Buzzard irregular prime is $p = 59$, and that the offending representation:

$$\rho_{Q\Delta} : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{F}_{59})$$

has exceptional image in the context of Serre and Swinnerton-Dyer (*On l-adic representations and congruences for coefficients of modular forms*, Antwerp III, [SD73]). Indeed, this particular example features prominently in that paper. I always thought this was not an entirely random coincidence, and since it seems relevant here, I thought I would finally bother to figure out what is going on. (For the next prime, $p = 79$, the corresponding representation has image containing $\mathrm{SL}_2(\mathbf{F}_{79})$, so it is somewhat of an accident.) The mod-59 representation above has projective image S_4. Now suppose

that $p \equiv 3 \mod 4$ is a prime such that $H/\mathbf{Q}$ is an S_4-extension which is unramified away from p. Such a representation will give rise (following an argument of Tate) to a mod-p representation of $\mathbf{Q}$ which is unramified away from p. The congruence condition on p implies that it will be odd, and hence modular, by Langlands-Tunnell. Now let us suppose, in addition, that 4 divides the ramification index e_p. Under this assumption, the representation cannot be locally reducible, because the ramification index of any power of the cyclotomic character divides $p-1 \equiv 2 \mod 4$. Hence, if there are infinitely many such fields, there are infinitely many $\mathrm{SL}_2(\mathbf{Z})$-irregular primes. Consider the following fields studied by Darrin Doud [Dou99]:

$$K = \mathbf{Q}[u]/f(u), \quad f(u) = (u+x)^4 - p^* u,$$

$$(-1)^{(p-1)/2} p = p^* = \frac{256x^3 - y^2}{27}, \quad p^* \neq 1 + 4x.$$

Doud shows that K has discriminant $(p^*)^3$ and Galois group S_4, and it is easy to see that the splitting field $H/\mathbf{Q}$ has all the required properties needed above as long as $p \equiv 3 \mod 4$. The formula relating tame ramification and the discriminant implies that $e_p(K/\mathbf{Q}) = 4$. Standard conjectures now predict that there are infinitely many such primes of this form — if y is odd, then $p^* \leq 0$. The first few such primes which are $3 \mod 4$ are

$$59, 107, 139, 283, \ldots$$

which compares with the first few Buzzard irregular primes (taken from Buzzard's paper):

$$59, 79, 107, 131, 139, 151, 173 \ldots$$

On the other hand, an unconditional proof by these means seems out of reach, because any modular S_4-extension unramified outside $p \equiv 3 \mod 4$ forces $\mathbf{Q}(\sqrt{-p})$ to have class number divisible by 3, and we don't know if there exist infinitely many such primes.

70. Chenevier on the Eigencurve

Fri, 24 Apr 2015

Today I wanted to mention a theorem of Chenevier about components of the Eigencurve. Let $\mathcal{W}$ denote weight space (which is basically a union of discs), and let

$$\pi : \mathcal{E} \to \mathcal{W}$$

be the Coleman–Mazur eigencurve together with its natural map to $\mathcal{W}$. It will do well to also consider the versions of the eigencurve corresponding to quaternion algebras $D/\mathbf{Q}$ as well.

Theorem 70.1 (Chenevier)**.** *Suppose that*

1. *$\mathcal{E}$ has "no holes" — that is, a family of finite slope forms over the punctured disc extends over the missing point,*
2. *The "halo" of $\mathcal{E}$ is given by a union of finite flat components whose slope tends to zero as $x \in \mathcal{W}$ tends to the boundary of the disc.*

Then every non-ordinary component of $\mathcal{E}$ has infinite degree.

In particular, since both of these theorems are now known in many cases (properness by Hansheng Diao and Ruochuan Liu [DL16], and haloness by Ruochuan Liu, Daqing Wan, and Liang Xiao [LWX17], at least in the definite quaternion algebra case), the conclusion is also known.

The proof is basically the following. Given a component C of finite degree, the first assumption implies that it actually is proper and finite. One may then consider the norm of U_p on C to the Iwasawa algebra to obtain a bounded (hence Iwasawa) function $F = \mathrm{Norm}(U_p)$. This function cannot have any zeros (again by properness), and hence, by the Weierstrass preparation theorem, it is a power of p times a unit. But that implies that F has constant valuation near the boundary, which contradicts the fact that the slopes are tending to zero (except in the ordinary case).

Naturally one may ask whether $\mathcal{E}$ has only *finitely* many components, although this seems somewhat harder to prove.

Notes 70.2. For the general halo conjecture for GL(2), see here, [DY23].

71. 144169

Tue, 12 May 2015

The space of classical modular cuspforms of level one and weight 24 has dimension two — the smallest weight for which the dimension is not zero or one. What can we say about the Hecke algebra acting on this space without computing it?

Formally, the Hecke algebra $\mathbf{T}$ is a rank two $\mathbf{Z}$-algebra, which is either an order in the ring of integers of a real quadratic field, or a subring of $\mathbf{Z} \oplus \mathbf{Z}$. Let's investigate the completion of this algebra at various primes p.

Let's first consider the prime $p = 23$. The curve $X_0(23)$ has genus two, and the corresponding Hecke algebra in weight two is $\mathbf{Z}[\phi]$, where ϕ is the Golden Ratio. The prime $p = 23$ does not split in this field, and hence modulo p there is a pair of conjugate eigenforms with coefficients in $\mathbf{F}_{p^2}$. Multiplying by the Hasse invariant, we see that this eigenform also occurs at level one and weight 24 over $\mathbf{F}_p$. It follows that:

$$\mathbf{T} \otimes \mathbf{Z}_{23} = W(\mathbf{F}_{23^2}).$$

In particular, $\mathbf{T} = \mathbf{Q}(\sqrt{D})$ for some square-free integer $D \geq 0$.

Now let us consider primes $p \leq 23$. Any Galois representation modulo such a prime will occur — possibly up to twist — in lower weight. Yet all the spaces in lower weight have dimension at most one, and hence it follows that the residue fields of $\mathbf{T}$ are all of the form $\mathbf{F}_p$. Suppose further that $5 \leq p \leq 23$. Then, using theta operators, we may find *two* distinct eigenforms in weight 24, from which it follows that $\mathbf{T}$ has two distinct residue fields of characteristic p, and so, for $5 \leq p \leq 23$, we have:

$$\mathbf{T} \otimes \mathbf{Z}_p = \mathbf{Z}_p \oplus \mathbf{Z}_p.$$

One expects at level one that $a_2(f)$ always generates the Hecke field. This is still a conjecture, but we may deduce this unconditionally in weight 24 because the dimension of the cuspforms is two, and so this follows automatically from the Sturm bound! Hence we may write:

$$\mathbf{T} = \mathbf{Z}[a_2(f)], \quad a_2(f) = \frac{a + b\sqrt{D}}{2} \in \mathbf{Z}\left[\frac{1+\sqrt{D}}{2}\right]$$

where $b \neq 0$. Even better, using Hatada's Theorem — giving congruences for a_2 and a_3 for eigenforms of level one modulo 8 and 3 respectively — we may write

$$a_2(f) = 12(a + b\sqrt{D}), \quad a, b \in \mathbf{Z}$$

where $b \neq 0$. This gives an upper bound on D in light of the Deligne bound $|a_2| \leq 2 \cdot 2^{23/2}$. More precisely, we obtain the bound $b^2 D \leq 2^{27}/24^2$, and hence that $D \leq 233017$.

Let's now think more carefully about $p = 2$ and 3. For these primes, there will be a unique Coleman family of slope $v(-24) = 3$ for $p = 2$ and $v(252) = 2$ for $p = 3$. I can't quite see a pure thought way of proving this, but at least this would be a consequence of the strong form of the GM-conjecture as predicted by Buzzard. So we should expect that, in these cases

$$\mathbf{T} \otimes \mathbf{Z}_p \hookrightarrow \mathbf{Z}_p \oplus \mathbf{Z}_p.$$

In addition to congruences for small primes, there will also be congruences between the unique cusp form with an Eisenstein series modulo the numerator of B_{24}, which is

$$B_{24} = \frac{-1}{2 \cdot 3 \cdot 5 \cdot 7 \cdot 13} \times 103 \times 2294797.$$

I claim that these primes will also have to split in $\mathbf{T}$. For example, it is impossible for b to be divisible by 2294797, because that would violate the inequality on $b^2 D$ above, and hence it follows that $p = 2294797$ must also split in $\mathbf{T} \otimes \mathbf{Q}$. The same argument works for $p = 103$ having ruled out some very small D. To summarize, we have the following:

The primes $5 \leq p \leq 23$, $p = 103, 2294797$ split in $K = \mathbf{Q}(\sqrt{D})$, but $p = 23$ does not split, and $D \leq 233017$. Moreover, we expect that 2 and 3 also split.

This is enough to determine D completely up to 72 possibilities, and 9 with the unproven assumption at 2 and 3. On the other hand, all of these D are quite large (the smallest are 3251 and 15791 respectively), which forces b to be very small. But we also have the congruence

$$12(a + b\sqrt{D}) \equiv 1 + 2^{23} \mod 2294797.$$

For the remaining D, we can determine, with $|b|$ satisfying the required inequality, whether there exists such a congruence with $|a| \leq 2^{27/2}/24 \sim 483$. A simple check shows that is a unique solution (with the assumption on two or three or not), and hence, by (something close to) pure thought, we have shown that $D = 144169$, and moreover (using Deligne's bound again) that

$$a_2(f) = 12(45 \pm \sqrt{144169}), \qquad \mathbf{T} = \mathbf{Z}[12\sqrt{144169}].$$

One can indeed check this is the case directly, if you like. Curiously enough, this Hecke eigenvalue is quite close to the Deligne bound — the probability it is (in absolute value) this big is, assuming a Sato–Tate distribution, slightly under 5%.

Extra Credit Problem: Hack Ken Ribet's Yelp password by using the fact that 144169 is his favorite prime number.

72. Counting solutions to $a_p = \lambda$

Fri, 15 May 2015

We know that the eigenvalue of T_2 on Δ is 24. Are there any other level one cusp forms with the same Hecke eigenvalue? Maeda's conjecture in its strongest form certainly implies that there does not. But what can one prove along these lines? Conjecturally, one would certainly predict the following:

Conjecture 72.1. *Fix a tame level N prime to p. If $\lambda \neq 0$, there are finitely many eigenforms of level N an arbitrary weight such that $a_p = \lambda$. If $\lambda = 0$, there are finitely many eigenforms with the additional condition that they do not have CM by a quadratic field in which p is inert.*

I have no idea how to prove this conjecture. If one counts the number of such forms of weight $\leq X$, then the trivial bound for eigenforms with $a_p = \lambda$ is $O(X^2)$. When I visited Princeton a few weeks ago, Naser Sardari, a student of Sarnak, showed me a short preprint he is writing which improves this bound by a power saving (additionally, it gives a power saving for each individual weight as well). The most interesting case of this result is when $\lambda = 0$, but today I want to talk about the much easier case when $\lambda \neq 0$, where, via some p-adic tricks, one can obtain a substantial improvement on the trivial bound. Let's start from the following:

Proposition 72.2. *Let $S_\lambda(X)$ denote the number of cuspforms of level N and weights $\leq X$ such that $a_p = \lambda$. Assume that $\lambda \neq 0$. Then*

$$S_\lambda(X) = O(X).$$

Proof. Since $\lambda \neq 0$, the p-adic valuation of λ is finite. However, all forms with bounded slope belong to one of finitely many Coleman families, so the number of such forms in any weight is bounded. Using Wan's explicit results, one can even give an explicit bound here that depends only on N, p, and the valuation of λ. □

The point of this post, however, is to give an improvement on this bound.

Theorem 72.3. *Let $S_\lambda(X)$ denote the number of cuspforms of level N and weight $\leq X$ such that $a_p = \lambda$. Assume that $\lambda \neq 0$. Then, as $X \to \infty$,*

$$S_\lambda(X) \ll_\lambda \log\log\log\log\log\log\log X.$$

The argument will (obviously) allow for an arbitrary number of logs. But then the statement would become more cumbersome.

Proof. As in the proof of the previous result, we may reduce to the case where we are considering a single Coleman family $\mathcal{F}$. Over this family, the function U_p is continuous, and hence so is $U_p(U_p - \lambda)$. More importantly, over a small enough disc, it is an Iwasawa function. Let Σ denote an infinite set of integral weight such that, for the relevant points of $\mathcal{F}$, we have $T_p = \lambda$, or

$$U_p(U_p - \lambda) = -p^{k-1}.$$

If s is a limit point of Σ, then certainly $U_p(U_p - \lambda)$ will vanish at s. Since this function is a non-zero bounded function on a disc, it has only finitely many zeros, and so the set of weights Σ will have only finitely many limit points. Thus, we may reduce to the case where the set of weights has a single limit point. In particular, if $S_\lambda(X)$ is not bounded, we may imagine that the set Σ consists of a sequence of integers

(which we may assume to be increasing in the Archimedean norm): $k_0, k_1, k_2, \ldots$ which converge p-adically to k_∞, and, at the relevant point of $\mathcal{F}$, correspond to an eigenform which satisfies the equation

$$U_p(U_p - \lambda)(k_i) = -p^{k_i - 1}.$$

Around a zero η, any Iwasawa function F has an asymptotic expansion of the form

$$F(\eta + \epsilon) \simeq A \cdot \epsilon^m + \ldots$$

where the LHS has the same valuation as the leading term of the RHS for sufficiently small ϵ. If $F = U_p(U_p - \lambda)$ and $\eta = k_\infty$, we deduce that, for sufficiently large integers k_i in our list,

$$-p^{k_n - 1} = F(k_n) = F(k_\infty + (k_n - k_\infty)) = A(k - k_\infty)^m + \ldots,$$

and so, taking p-adic valuations,

$$k_n - 1 = v(A) + m \cdot v(k_n - k_\infty),$$

where $m > 0$ is constant. This certainly implies that $v(k_n - k_{n+1}) = v(k_n - k_\infty)$ by the triangle inequality, and so

$$v(k_n - k_{n+1}) = ak_n + b,$$

for constants a, b with $a > 0$. But two integers whose valuation are very close are very far apart, and indeed we deduce that

$$k_{n+1} - k_n \geq Cp^{ak_n}$$

for some $a > 0$ and some constant $C > 0$. This iterated exponential growth proves the result. □

The argument also shows that if the set Σ is infinite, the limit roots of $U_p - \lambda = 0$ will be transcendental Liouville numbers, which seems unlikely. The result also applies if one replaces λ by a sufficiently continuous function without zeros, say $a_2 = 24(1 + 2(k - 12)^2)$. On the other hand, I don't think these analytic methods will ever be enough to prove the conjectural bound, which is $O(1)$.

Notes 72.4. I still think that for $\lambda \neq 0$ this is the best known bound. But the $\lambda = 0$ case now has a non-effective but asymptotically optimal result by [CTS21].

73. Hilbert modular forms of partial weight one, Part III

Sat, 17 Oct 2015

My student Richard Moy is graduating! Richard's work has already been discussed in § 3 and § 8 on this blog before, where we discussed his joint work with Joel Specter showing that there *existed* non-CM Hilbert modular forms of partial weight one. Today I want to discuss a sequel of sorts to that paper, which also forms part of Richard's thesis (I should note that he already has five publications and will have 7 or 8 papers by the time he graduates.) The starting observation is as follows. Fix a real quadratic field F. From the perspective of Galois representations, the Hilbert modular forms of partial weight one fall under the case $\ell_0 = 1$ in the notation of my paper with David Geraghty (this is in the context of **coherent** cohomology). To orient the reader, let us discuss three classes of such forms:

(1) Hilbert modular forms of weight $[2k + 1, 1]$ for a real quadratic field F.

(2) Regular algebraic cuspidal automorphic forms for $\mathrm{GL}(3)/\mathbf{Q}$.
(3) Regular algebraic cuspidal automorphic forms for $\mathrm{GL}(2)/F$ for an imaginary quadratic field F.

Suppose one fixes a tame level N and then looks at the space of such forms as the weights vary. In both of the latter cases, the problem has been raised (or even conjectured, for $N = 1$ and GL(3) by Ash and Pollack here [AP08]), of whether all but finitely many such forms arise via functoriality from a smaller group. More explicitly, one can ask whether:

(1) If $G = \mathrm{GL}(2)/F$, then all but finitely many cuspidal regular algebraic forms of conductor N either arise (up to twist) via base change from $\mathrm{GL}(2)/\mathbf{Q}$, or are induced from a quadratic CM extension E/F.
(2) If $G = \mathrm{GL}(3)/\mathbf{Q}$, then all but finitely many cuspidal regular algebraic forms of conductor N arise up to twist as the symmetric square of a form from $\mathrm{GL}(2)/\mathbf{Q}$.

Naturally enough, one can make the same conjecture whenever $\ell_0 > 0$, appropriately formulated. There does not seem to be any case of this conjecture which is known, although there are analogous results (where one fixes the weight and varies the level) in both weight one (where it is almost trivial) and for imaginary quadratic fields (in the work of Calegari–Dunfield [CD06] and Boston–Ellenberg [BE06]). Still, the conjectures in varying weight seem pretty hard even for $N = 1$. In that context, Richard proves the following nice complementary pair of theorems below. Let $F = \mathbf{Q}(\sqrt{7})$. The field F has narrow class number 2 and there is a unique odd everywhere unramified quadratic character χ of G_F with fixed field $E = F(\sqrt{-1})$.

Theorem 73.1 (Moy)**.** *Let F and χ be as above. Every Hilbert modular form over F of weight $[2k+1, 1]$ and level $N = 1$ is CM, and in particular is induced from E.*

Theorem 73.2 (Moy)**.** *Let F and χ be as above. Let M be a strongly compatible family of two dimensional Galois representations of F with determinant χ, level $N = 1$, and Hodge–Tate weights $[0, 0]$ and $[k, -k]$. Then M is induced from E.*

The first theorem is an almost immediate corollary of the second, with the caveat that one doesn't quite have complete local-global compatibility for partial weight one modular forms (though results and methods of Luu, Jorza, and Newton get close). Theorem 73.2 on the other hand is a consequence of the following:

Theorem 73.3 (Moy)**.** *Let F and χ be as above. Let*

$$\rho : G_F \to \mathrm{GL}_2(\overline{\mathbf{Q}}_3)$$

be a continuous irreducible representation with determinant χ that is unramified at all finite places except for one prime $v|3$. Then ρ is induced from a character of G_E.

The argument in this case is (roughly) the following. Using a Tate-style argument (with discriminant bounds), one proves that the residual representation $\overline{\rho}$ must have semi-simplification $\chi \oplus 1$. The restriction of ρ to G_E then has the property that its image is pro-3 and unramified outside the fixed prime $v|3$. Yet one shows by a class field theory computation that the largest abelian 3-extension unramified outside $v|3$ is cyclic, which (by consideration of the Frattini quotient) immediately implies that the image of ρ restricted to G_E factors through a cyclic quotient as well, and one is done.

Note that to deduce Theorem 73.1, one first has to prove (using a congruence argument) that at the other prime $w|3$, either:

(1) The representation ρ is unramified at w,
(2) The representation ρ restricted to D_w has unramified semi-simplification. In particular, the generalized eigenvalues of Frob_w for $\overline{\rho}$ are both the same.

To finish, one rules out the second possibility by computing all the modular residual representations explicitly by doing computations in low weight (this can ultimately be reduced to a computation on the definite quaternion side, although Richard had to write his own programs to do this since the current magma implementation required trivial character for non-parallel weight.)

It is true that these arguments will not suffice for the more general conjecture, but then, I haven't seen a viable strategy to prove those conjectures either!

74. Ventotene, Part II

Fri, 18 Mar 2016

I promised to return to a more mathematical summary of the conference in Ventotene, and indeed I shall do so in the next two posts.

One of the themes of the conference was bounding the order of the torsion subgroup in arithmetic lattices. Tsachik Gelander gave a number of talks (in part) on the seven author paper. One nice result was a uniform bound of the shape

$$\log |H^*(\Gamma, \mathbf{Z})^{\mathrm{tors}}| \leq C \cdot \mathrm{Vol}(\Gamma),$$

where Γ ranges (say) over all lattices in $\mathrm{SL}_n(\mathbf{R})$ for a fixed $n \geq 3$. (The key result here is the uniformity — this result is much easier for covers of a fixed manifold.) Two natural questions that came up (in conversation at least) during the conference are as follows:

(1) Can one do better in low degree?
(2) What is the true expectation for the size of this group for (say) congruence subgroups of $\mathrm{SL}_n(\mathbf{Z})$?

Let's consider the first question. For (congruence) subgroups of $\mathrm{SL}_n(\mathbf{Z})$, one can certainly say quite a bit more. For example, H^1 is essentially trivial, by the congruence subgroup property. However, in the stable range of cohomology (in particular, when the completed cohomology groups become stable), the groups H^* are finite over $\mathbf{Z}_p$, and so contribute very little. One does, at least, have the following soft arguments for general groups.

Proposition 74.1. *Let G be a semi-simple group over $\mathbf{Q}$ with $\mathbf{Q}$-rank $r = r_{\mathbf{Q}}$. Then $\widetilde{H}_i$ is a torsion $\Lambda = \mathbf{Z}_p[[G(\mathbf{Z}_p)]]$-module for $i \leq r_{\mathbf{Q}}$.*

Proof. The proof is as follows: the boundary terms are also torsion, so it suffices to show that all the $\widetilde{H}_i^{BM}$ in the appropriate range are also torsion, where we consider Borel-Moore homology. Assume otherwise. Let $\dim G(\mathbf{R})/K(\mathbf{R}) = d$. From the spectral sequence $\mathrm{Ext}^i(\widetilde{H}_j^{BM}, \Lambda) \Rightarrow \widetilde{H}_{d-i-j}$, we deduce that there is at least one $i \leq r_{\mathbf{Q}}$ such that $\widetilde{H}_{d-i} \neq 0$. Yet the homological dimension of $\Gamma\backslash G(\mathbf{R})/K(\mathbf{R})$ is, by Borel-Serre, equal to $d - r_{\mathbf{Q}}$, and so all the homology in these degrees (and hence certainly the completed homology) vanishes. □

One can do better in certain algebraic cases, where one can deduce vanishing of the completed cohomology in certain degrees by perfectoid technology (as in Corollary 4.2.3 of [Sch15b]).

The answer to the second question, even conjecturally, is more mysterious. There are some speculations related to this question in Bergeron–Venkatesh. But it seems a little tricky to formulate a precise guess (for a good upper bound).

75. Tensor products

Tue, 29 Mar 2016

Let W be an irreducible representation of a finite group G. Say that W is tensor indecomposable if any isomorphism $W = U \otimes V$ implies that either U or V is a character. In conversations with Matt and Toby (which permeate the rest of this post as well), the following problem came up:

Problem 75.1. Let G be a finite group. Let V be an irreducible representation of G. Is there a unique decomposition

$$V = V_1 \otimes V_2 \ldots \otimes V_k$$

of V as a tensor product of tensor indecomposable representations (up to re-ordering and twist)? I don't think this can be too hard, but I confess I don't see how to do it. (Since we didn't really need this, we didn't think about it too hard.)

(**edit**: when I say I don't think this can be too hard, I don't mean to imply that I think it is true; just that I think either a proof or counterexample should not be too hard to find — hopefully not both.)

One can ask an analogous problem for Lie groups. Actually, the problem for Lie algebras is actually quite simple (and the answer is positive). It is related to the following:

Lemma 75.2. *Let V and W be irreducible non-trivial representations of a simple Lie group $\mathfrak{g}$. Then $V \otimes W$ is reducible.*

Proof. Assume that $V \otimes W$ is irreducible. In particular, it is determined by its highest weight. Let the highest weights of V and W be λ and μ respectively. Then the highest weight of $V \otimes W$ must be $\lambda + \mu$. But now, by the Weyl character formula, we deduce that

$$1 = \frac{\dim(V)\dim(W)}{\dim(V \otimes W)} = \prod_{\Phi^+} \frac{\langle \rho, \alpha \rangle \langle \rho + \lambda + \mu, \alpha \rangle}{\langle \rho + \lambda, \alpha \rangle \langle \rho + \mu, \alpha \rangle}.$$

The product term can also be written as:

$$1 + \frac{\langle \lambda, \alpha \rangle \langle \mu, \alpha \rangle}{\langle \rho + \lambda, \alpha \rangle \langle \rho + \mu, \alpha \rangle}.$$

In particular, since the pairing is non-negative between positive roots and highest weights, we deduce a contradiction unless

$$\langle \lambda, \alpha \rangle \langle \mu, \alpha \rangle = 0$$

for all $\alpha \in \Phi^+$. The assumption that $\mathfrak{g}$ is irreducible, however, is equivalent to saying that Φ^+ has a maximal root β, and for such a maximal root, we have

$$\langle \lambda, \beta \rangle = 0 \Rightarrow \langle \lambda, \alpha \rangle = 0, \ \forall \alpha \in \Phi^+ \Leftrightarrow \lambda = 0.$$

□

Note that this lemma is actually a special case of a theorem of Rajan [Raj04], who proved that, for simple $\mathfrak{g}$, the factors of a (not necessarily irreducible) tensor product are determined by the representation. In particular, the tensor product of two non-trivial irreducible representations cannot be irreducible.

The problem with the initial question is that it's hard to construct tensor products of irreducible representations which are irreducible. Or rather, it is easy, simply by taking $U \otimes V$ where U is an irreducible representation of H and V is an irreducible representation of G and the tensor is irreducible for $H \times G$. Yet the interesting case is something closer to assuming that U and V are faithful. Actually, this motivates the following question:

Problem 75.3. Do there exist irreducible non-trivial representations U and V of a finite simple non-abelian group G such that $U \otimes V$ is irreducible?

The argument above for Lie groups suggests that this may not happen for Chevelley groups (although it certainly doesn't prove this). It also suggests (relating the representation theory of A_n and S_n to GL_n) that it doesn't happen for the alternating groups either. It almost surely doesn't happen for the sporadic groups either. So my guess that the answer to the problem above is no, and that this is probably known, and probably requires classification. (Please comment if you know the answer.) Actually, this also reminds me of a similar problem which I think is open.

Problem 75.4. Fix N. Does there exist a non-trivial representation V of a finite group G of dimension N such that $\mathrm{Hom}^0(V,V)$ (of dimension N^2-1) is irreducible?

This question came up in my paper with Barry, where I was surprised to find very few examples. I seem to remember that the Mathieu group M_{12} has an 11-dimensional representation whose corresponding 120 dimensional adjoint is irreducible. Can one classify all such examples coming from simple groups?

Comment 75.5 (Persiflage)**.** I'm was confused why the list doesn't contain M_{12}, and then checked that it is false for M_{12}, and I must have read the character table wrong, oops! For those playing at home, Katz' paper can be found here (see [Kat04]).

76. Report from Berkeley

Fri, 22 Apr 2016

My recent trip to Berkeley did not result in a chance to test whether the Cheeseboard pizza maintained its ranking, but did give me the opportunity to attend the latest Bay Area Number Theory and Algebraic Geometry day, on a (somewhat disappointingly) rainy Saturday in Evans Hall. The weather was somewhat better on Sunday, however, allowing myself to make the trip to Mint Plaza for the following cup, which should bear some resemblance to the banner picture on this site. (Unfortunately, they were no longer serving their mini-Brioche buns.) But now on to the good stuff, a report on some of the talks:

Jaclyn Lang gave a talk on her work concerning the image of big Galois representations (see [Lan16]). The setup is roughly as follows. Let

$$\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{F}_p)$$

be an absolutely irreducible odd Galois representation over a finite field (hence modular). Suppose this Galois representation was the residual representation associated to an ordinary modular form that lived inside a Hida family that was smooth over weight space. Then one might expect that the corresponding representation

$$\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{Z}_p[[T]])$$

to have image which was as big as possible, namely, containing $\mathrm{SL}_2(\mathbf{Z}_p[[T]])$. This can't always be the case, however; for example, the residual representation (or the entire family) could be dihedral. However, if the residual representation contains $\mathrm{SL}_2(\mathbf{F}_p)$, and one additionally assumes that the image of inertia at p is sufficiently large, then this is indeed the case (probably this assumes that the residue characteristic is at least 5). There have been a number of generalizations of this result due to Hida and others which improves the result by weaken the various hypotheses; for example, allowing coefficients, allowing the residual representation to be dihedral, and weakening the ramification hypothesis at p. For these results, one can't expect that the image is full, but rather that the image contains an appropriate congruence subgroup of SL_2. I like to think of this as follows: at classical specializations, one knows that the image (if it is not CM or weight one) will have open image; the results of this talk and of previous work show that index can be controlled in families. Actually, this is not quite true, because another obstruction to having open image even classically is the existence of inner twists. The main result of the talk was to deal with this issue of inner twists, and hence also allow for a generalization of the results not only to smooth Hida families but to any irreducible component of any Hida family. (More details to be found here.)

A natural question: one output of Lang's result is to give an ideal $\mathfrak{b}$ of the Hida family for which the image of these Galois representations contains the $\mathfrak{b}$-congruence subgroup (after accounting for inner twists). In characteristic zero, my impression from the talk was that one can identify the support of this ideal as coming from CM points and classical weight one modular forms. On the other hand, apparently there is also a version of this result in the reducible case (due to Hida and with extra hypotheses); in that case the zeros should correspond to the reducible locus, or equivalently, the zeroes of the p-adic zeta function. However, a stronger result is true, namely, that $\mathfrak{b}$ can essentially be identified with this p-adic zeta function. So, returning back to the residually irreducible case, the natural question is: can the support of $\mathfrak{b}$ contain the prime p?

Kęstutis Česnavičius gave a talk on the Manin–Stevens and Manin constants for elliptic curves, with emphasis on the prime $p = 2$. He raised the following question: Suppose that N is odd. Is there a surjection from the space of weight 2 classical modular cusp forms of level $\Gamma_0(N)$ with coefficients in $\mathbf{Z}_2$ to the space of weight 2 Katz cusp forms of the same level over $\mathbf{F}_2$? The issue here is that the latter space is really the cohomology of the associated stack, not the course moduli space. Unfortunately, this question distracted me a little as I tried to find a counter-example (I failed). A result of Serre and Carayol basically implies that the result can only fail after localizing at a non-Eisenstein maximal ideal $\mathfrak{m}$ of the Hecke algebra $\mathbf{T}$ *if* the corresponding representation $\overline{\rho}_{\mathfrak{m}}$ is induced from $\mathbf{Q}(\sqrt{-1})$. (Analogously, for $p = 3$, when the representation is induced from $\mathbf{Q}(\sqrt{-3})$.) This is related to the classic failure of the first version of Serre's conjecture for $p = 3$ at level $\Gamma_1(13)$. However, as Serre quickly realized, this failure ultimately comes from a failure to lift mod-p forms as above, except in this case from the intermediate curve

$X_H(13)$, not from $X_0(13)$. I ultimately convinced myself that lifting was always possible unless $\mathfrak{m}$ was not only Eisenstein but also the ideal containing T_p for all odd p not dividing N. I think this must be related to Ken's result on component groups of Neron models, and how the non-Eisenstein parts arise for $X_H(N)$ but not for $X_0(N)$ or $X_1(N)$. (More details here and [Č18])

The final speaker of the day was Daquin Wan. The key question that arose in his talk was the following. Suppose that $D(k,T)$ is the characteristic power series of the U operator on the space of overconvergent p-adic modular forms in integral weight k. Can one show that

$$L(k,T) = \frac{D(k+2,T)}{D(k,pT)}$$

has infinitely many zeros and infinitely many poles? One actually has to assume that $k \neq 0$ here, since otherwise the result is false, as this will be a polynomial of dimension the space of weight two forms. One feels that p-adic Langlands should be able to say enough about slopes in these weights to obtain a contradiction, but I don't unfortunately see how to do it. The main point of the talk was two-fold. There is an argument of Coleman that shows that $D(k,T)$ is not itself a polynomial. This argument can be generalized to prove that $L(k,T)$ is not a rational function. Second, the product $L(k,T)L(-k,p^kT)$ is actually a rational function because of the properties of the theta operator. So one deduces that at least one of these functions had infinitely many poles and the other had infinitely many zeroes. This also relies on a previous result of Wan that these functions are meromorphic. (Oh, I should mention that this was joint work with ... and here I didn't take notes for a talk two weeks ago ... Liang Xiao? Please correct me if I'm wrong)

(Romyar Sharifi also talked, but since I am actively trying to understand something about that talk on a more technical level, so I will have to return to a discussion of it later.)

77. Central extensions and weight one forms

Mon, 23 May 2016

As mentioned in the comments to a previous post (not on math), Kevin Buzzard and Alan Lauder have made an extensive computation of weight one modular forms in characteristic zero (see also [BL17]). Thinking about what that data might contain, I wondered about the following question: what are the images of the Galois representations associated to the weight one forms of type A_5?

Let us take a step back. Consider a projective representation

$$\psi : G_{\mathbf{Q}} \to \mathrm{PGL}_2(\mathbf{C})$$

with image A_5, and assume that it is odd. (That is, complex conjugation has order 2.) According to Tate, there exists a lift

$$\rho : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{C}).$$

This lift is unique up to twisting. Since the Schur multiplier $H_2(A_5, \mathbf{Z})$ of A_5 is $\mathbf{Z}/2\mathbf{Z}$, there is a unique minimal lift up to twist whose image is a central extension $\widetilde{A}_5$ by a cyclic group Δ of 2-power order. Note that Δ is not trivial, since A_5 does not have any two-dimensional representations. If $|\Delta| = 2$, then the determinant of the corresponding 2-dimensional representation of $\widetilde{A}_5$ is trivial, which contradicts

the assumption that ψ is odd. (Equivalently, there is an obstruction at ∞ to lifting to the central extension by $\mathbf{Z}/2\mathbf{Z}$.) Hence 4 divides $|\Delta|$. What is the expected distribution of Δ as one runs over all odd A_5-extensions?

My first guess (without any prior thought or computation) was that this might obey some form of Cohen–Lenstra heuristic, suitably interpreted.

Note that the image of the determinant has order $|\Delta|/2$. The corresponding determinant representation is a character of $\mathbf{Q}$ of 2-power order. Since $\mathbf{Q}$ has trivial class number, the order $|\Delta|/2$ is equal to the maximal ramification degree e_p of this representation over all primes p.

Over $\mathbf{Q}$, Tate's lifting theorem has the following stronger form: one may choose a lift ρ_p of $\psi_p := \psi|_{D_p}$ and insist that $\rho|_{I_p} = \rho_p|I_p$; that is, they agree on inertia. This is essentially a consequence of the fact that $\mathbf{Q}$ has trivial class group. For convenience, suppose that ψ is unramified at 2 and 3. Suppose that ψ is ramified at p. There are three possibilities:

(1) The image of ψ_p at a ramified prime p is cyclic of order 2, 3, or 5.
(2) The image of ψ_p at a ramified prime p is D_6 or D_{10}.
(3) The image of ψ_p at a ramified prime p is $\mathbf{Z}/2\mathbf{Z} \oplus \mathbf{Z}/2\mathbf{Z}$.

For a fixed p, let ϵ denote the Teichmuller lift of the mod-p cyclotomic character. (Fix an isomorphism of $\mathbf{C}$ with $\overline{\mathbf{Q}}_p$ for all p.)

Let us consider the three cases in turn.

In the first case, the image factors through a cyclic quotient. One may thus take ρ_p to be a direct sum which, on inertia, has the shape $\chi \oplus 1$ up to twist. By comparing this to the projective representation, we see that χ has order 2, 3, or 5, and so, after finding the twist such that the determinant has 2-power order, we see that $e_p = 1$ or $e_p = 2$.

In the second case, the lift on inertia is (up to twist) of the form:

$$\omega_2^r \oplus \omega_2^{pr}$$

for some r. Since the order of ω_2 is $p^2 - 1$, the order of the ratio is

$$\frac{p+1}{\gcd(r, p+1)}.$$

which must be equal to 3 or 5. It follows that r is even. Yet the determinant is equal to

$$\omega_2^{(p+1)r} = \epsilon^r,$$

Since r is even, we see that, after twisting, we may take $e_p = 1$.

Finally, in the third case, the lift is of the form:

$$\omega_2^r \oplus \omega_2^{pr}$$

for some r. We now find that the order of the ratio of these characters is

$$\frac{p+1}{\gcd(r, p+1)}.$$

which must be equal to 2, and the determinant is ϵ^r. If r is even, then, as above, we may twist so that $e_p = 1$. Hence, the only way that the image after minimal twist does not have $|\Delta| = 4$ is if we are in this third situation with $p \equiv 1 \mod 4$, with r odd, and then (after twisting) we find that e_p is the largest power of 2 dividing $p - 1$.

(I confess that I originally forgot the fact that the third possibility could occur, and was only after noticing that this seemed to imply the inverse Galois problem was false thought a little bit more about the possibilities.)

To summarize:

Lemma 77.1. *Assume that ψ is unramified at 2 and 3 and has projective image A_5, and a lift with image $\widetilde{A}_5$ with minimal kernel. Then order of Δ is 4 unless there exists a prime $p \equiv 1 \mod 4$ such that the image of the decomposition group at p under ψ is $\mathbf{Z}/2\mathbf{Z} \oplus \mathbf{Z}/2\mathbf{Z}$. In this case, we have Δ to be twice the largest power of 2 dividing $p-1$ for all such primes p.*

Let $\Delta(\psi)$ denote the corresponding power of 2.

We see that $\Delta(\psi)$ is determined by purely local phenomena. This still doesn't quite answer what the distribution of the extension $\widetilde{A}_5$ will be. However, I imagine that Bhargava style heuristics should certainly be able to predict the ratio of A_5 with $\Delta(\psi) = 2^n$. Does anyone have a sense of how easy this might be to prove? Or, much more modestly, how easy it would be to compute from these quite precise heuristics the exact predicted distribution of the central extensions of A_5 coming from weight one modular forms?

(I confess, it is not even obvious to me from this construction how to prove that *all* central extensions $\widetilde{A}_5$ occur as Galois groups — but I presume this is known, and hopefully one of my readers can provide a reference.)

(According to Kevin, BTW, all the A_5 representations with $N \leq 1500$ have $\Delta = 4$.)

78. Prime divisors of polynomials

Sun, 29 May 2016

A heuristic model from the last post § 77 suggests that the "expected" order of the Galois group associated to a weight one modular form of projective type A_5 is infinite. And when one tries to solve the inverse Galois problem for central extensions of this group, one is lead to problems concerning the prime divisors of polynomials and their properties modulo 2. But I don't know how to answer this type of problems! Here is an analogous question that seems a little tricky to me:

Question 78.1. Show that there are infinitely many integers n such that all the odd prime divisors of n^2+1 are of the form 5 mod 8.

To make the problem slightly easier, one can ask:

Question 78.2. Show there exists an integer m and infinitely many integers n such such that all the odd prime divisors of n^2+1 are of the form $\not\equiv 1 \mod 2^m$.

Is this an open problem?

Comment 78.3 (Ben Green)**.** Frank, I believe this is a half-dimensional sieve problem and so one expects an answer of the form: the number of $n \leq X$ for which n^2+1 has all its odd prime factors 5 (mod 8) is asymptotic to $cX(\log X)^{-1/2}$ for some c, which you can specify explicitly. Details of the half-dimensional sieve can be found in Friedlander and Iwaniec, Opera Cribro [FI10], which I don't have to hand.

The reason this is a half-dimensional sieve problem is that you can locate such n by "sieving out" the conditions $n \equiv \pm a_p \pmod{p}$ for all $p \equiv 3 \pmod{8}$, where a_p is one of the roots of $a_p^2 \equiv -1 \pmod{p}$. So that's two congruences classes mod p for $\frac{1}{4}$ of the primes, i.e. on average you're sieving by half a residue class for each prime.

Another example of a half-dimensional sieve problem is counting sums-of-two-squares, where now you sieve by the condition $n \equiv 0 \pmod{p}$ for all $p \equiv 3 \pmod{4}$. To find such n up to X, you only need to sieve by those p with $p \leq \sqrt{X}$.

Your problem is more difficult because you need to sieve by all p up to about X, which is very large. I'd have to look at the details of the half-dimensional sieve to figure out whether this is fatal in terms of finding infinitely many n with your property — or, better, ask an expert on the subject.

Comment 78.4 (Sound)**.** As Ben says, this is a semi-linear sieve problem, and the best reference for that is probably the book of Friedlander and Iwaniec (chapter 14). They explain that the semi-linear sieve is capable of producing asymptotic formulas (as in the "sums of two squares" problem), under two basic conditions:

(1) A parity restriction: one must sieve integers in a sequence A by primes in a set P, with the property that elements of A have an even number of divisors from P — this holds for sums of two squares if A is taken to be integers $= 1 \bmod 4$ and P the primes which are 3 mod 4, but not in your case if one starts with $A = \{\text{values of } n^2+1\ \}$ and $P = \{\text{primes} = 1 \bmod 8\}$ [which are those to remove, Ben has a typo with a 3 instead of 1] . (See [FI10, 14.4]).
(2) a level of distribution condition: the sum of error terms in the sieve, up to the bound on p that you impose, must be under control.

Condition (1) seems very intrinsic to the method (e.g., I don't see how to get a lower bound with the sieve for the integers where all prime factors are 3 mod 4). Concretely, if instead of your problem one wants to have n^2+1 only be divisible by primes that are 1 mod 8, one could take $A = \{\ n^2+1$ with n divisible by $4\}$ and $P = \{$ primes $= 5 \bmod 8\}$; the parity condition is then true (since n^2+1 will be 1 mod 8). Moreover, you only need to sieve $n \leq x$ by primes up to $\sqrt{x}$; the level of distribution of A is just a bit smaller (by a log or so), so there's a chance that it works by applying the semilinear sieve up to $x^{1/2}/(\log x)$, and then separately counting the integers that remain unsieved but are not of the type you want, which are very restricted, and should be fewer in number I think.

Notes 78.5. See this paper (also [BS21]) for a solution. See also § 84

79. Serre 1: Calegari 0

Tue, 18 Oct 2016

I just spent a week or so trying to determine whether Serre's conjecture about the congruence subgroup property was false for a very specific class of S-arithmetic groups. The punch line, perhaps not surprisingly, was that I had made an error. I should note that I was pretty skeptical during the entire endeavor, so the final resolution was not a surprise, but there were still a few interesting twists along the way. (Thanks to Matt for some informative chats along the way.)

Let's start by recalling Ribet's proof of (what is one of many statements known as) Ihara's lemma. Let Γ be a congruence subgroup of $\mathrm{SL}_2(\mathbf{Z})$ of level prime to q. There is a congruence subgroup $\Gamma_0(q)$ defined in the usual way, where $q|c$. However, there is also a second copy $\Gamma^0(q)$ of this group inside Γ with $q|b$. (Well, there are $q+1$ copies of this group, but let's just consider these two for the moment.) The two groups are conjugate inside $\mathrm{GL}_2(\mathbf{Q})$, but not inside Γ. An argument of Serre now shows that the amalgam of Γ with itself along these groups (identified by conjugation by $[q,0;0,1]$) is the congruence subgroup $\Gamma[1/q]$ of $\mathrm{SL}_2(\mathbf{Z}[1/q])$. That is, the congruence subgroup where the local conditions away from q are the same as Γ. The Lyndon long exact sequence associated to an amalgam of groups shows that there is an exact sequence:

$$H_1(\Gamma_0(q), A) \to H_1(\Gamma, A)^2 \to H_1(\Gamma[1/q], A) \to 0,$$

for any trivial coefficient system A. Now the group $\Gamma[1/q]$ satisfies the congruence subgroup property, so the group on the right is easily seem to be finite and Eisenstein. By duality, there is also a map

$$H_1(\Gamma, A)^2 \to H_1(\Gamma_0(q), A),$$

and the composition of this map with the projection above is a matrix with determinant $T_q^2 - (1+q)^2$. A bookkeeping argument now gives Ribet's famous level raising theorem (taking coefficients $A = \mathbf{F}_p$.)

Fred Diamond and Richard Taylor [DT94] generalized this theorem by replacing the modular curve with both definite and indefinite quaternion algebras. The actual theorem itself at this point is probably quite easily to prove by the K-W method, but that's not relevant here. Instead, let's think a little about the proof. The more difficult and interesting case is when Γ comes from the norm one units in an indefinite quaternion algebra, which we consider from now on (the case of Shimura curves over $\mathbf{Q}$.) Morally, the proof should be exactly the same. The only wrinkle is that the corresponding group $\Gamma[1/q]$ is notoriously not known to satisfy the congruence subgroup property, although Serre conjectures that it does. Diamond and Taylor instead argued in the following way. (Let us specialize to the case of trivial weight, which is the only relevant case here.) Suppose that p is a prime greater than two and different from q. Then instead of working with Betti cohomology, one can instead, via a comparison theorem, use de Rham cohomology. The Hodge filtration consists of two pieces, one of which is $H^0(X, \Omega^1)$, and the other is $H^1(X, \mathcal{O}_X)$. They then investigate the kernel of the map:

$$H^1(X, \Omega^1)^2 \to H^1(X_0(q), \Omega^1)$$

where everything is now over $\mathbf{F}_p$. Here the two maps are the two pullbacks under the two projections $X_0(q) \to X$. They now show that element in the kernel gives rise to a differential ω which vanishes at all the supersingular points or does not vanish at all. The first is impossible by a degree argument when $p \geq 3$, and the second is always impossible. They conclude that, returning to etale cohomology, any kernel of the map

$$H^1(X, \mathbf{F}_p)^2 \to H^1(X_0(q), \mathbf{F}_p)$$

must lie entirely in one filtered piece, from which they deduce it must be Eisenstein. But let's look at this argument a little more closely. Even in Ribet's case, the conclusion is really much stronger than level raising for non-Eisenstein primes; there is a very precise description of the kernel (or cokernel in homology) in terms of the

homology of Γ coming from congruence quotients, which one can compute quite explicitly. So Ribet's theorem also gives level raising for Eisenstein representations in some contexts. In particular, for a suitable choice of congruence subgroup (with $p \geq 2$) one can make the group $H_1(\Gamma[1/q], \mathbf{F}_p)$ vanish identically. Let's now return to the argument of Diamond and Taylor when $p = 3$. All the comparison theorems are still valid, so the only issue is that the map

$$H^1(X, \Omega^1)^2 \to H^1(X_0(q), \Omega^1)$$

does have a kernel, namely, if one takes the "Hasse Invariant" A which vanishes to degree one at all supersingular points, then the two pullbacks of A to $X_0(q)$ coincide up to a scalar, and so the kernel is at least one dimensional. In fact, the argument of Diamond-Taylor shows that the kernel is at most one dimensional. But what does this mean in the proof of Ihara's Lemma? It means that, assuming X has good reduction at the prime $p = 3$, the level raising map **always** has a kernel, and thus $H_1(\Gamma[1/q], \mathbf{F}_3)$ is always non-trivial.

This now seems suspicious: all we need to do is find a quaternion algebra which doesn't have any congruence homology of degree 3. If the quaternion algebra $D/\mathbf{Q}$ is ramified at a prime r, then the congruence homology coming from this prime (for $p \neq 2r$) is a subgroup of the norm one elements of $\mathbf{F}_{r^2}^{\times}$, which has order $r + 1$. So it makes sense to take a quaternion algebra ramified at $7 \cdot 13$, since these are the two smallest primes different from 3 which are congruent to 1.

Because this seemed to contradict Serre's conjecture, I decided for fun to explicitly compute a presentation for the amalgam $\Gamma[1/2]$ to help work out what was going on. To first start, one needs a presentation for Γ. John Voight (friend of the blog) has written a very nice magma package to do exactly this. (More precisely, it's trivial to write down a presentation — Γ is torsion free, and hence a surface group $\pi_1(\Sigma_g)$ for a genus g one can compute via other means to be $g = 7$; the point is that one also wants an explicit representation as well as an explicit identification with the norm one units of the corresponding quaternion algebra.)

I then took an embarrassingly long time to compute the subgroup $\Gamma_0(2)$. The main issue was finding a suitable element in D to play the role of $\eta = [2, 0; 0, 1]$ in $M_2(\mathbf{Q})$. There certainly exists such a unit in $D \otimes \mathbf{Q}_2$, so in real life one just has to find an actual norm 2 unit which is sufficiently close 2-adically to this. However, I am absolute rubbish at mathematica and so repeatedly made the following error: when you define suitable quaternions i, j, k in $D \otimes_{\mathbf{Q}} E$ for some quadratic splitting field $E/\mathbf{Q}$, and then compute with the matrix $a + bi + cj + dk$, mathematica helpfully interprets "a" here as $[a, a; a, a]$ rather than a multiple of the identity, a programming decision which makes a lot of sense, said no one ever. I did this more times than I care to admit. Then, using John's program, one can find the subgroup $\Gamma_0(2)$, and then write down a presentation for the amalgam by conjugating this subgroup by η and identifying the corresponding elements via a solution to the word problem as words in the original generators, and then substitute the names for these generators for the second copy of Γ. The result is a group with $14 + 14$ generators and $2 + 38$ relations (corresponding to the 2 surface relations and the fact that $\Gamma_0(2)$ has $3(14 - 2) + 2 = 38$ generators.) Finally, one takes this group, plugs it into magma, and finds:

```
AbelianQuotientInvariants(G);
```

$$\geq [168]$$

There are known congruence factors coming from $7+1$ and $13+1$, but here one sees that the factor of three survives!

And then, shortly after this point, I realized that $\mathrm{SL}_2(\mathbf{F}_3)$ has a quotient of order 3, because it is A_4. So that degree three quotient is congruence after all... Oops! Still, it's nice to see that mathematics is consistent.

However, at this point one might just ask why can't one replace the quaternion algebra $D/\mathbf{Q}$ by (say) a real quadratic field in which 3 is unramified and inert. Serre got away with it above because $\mathrm{SL}_2(\mathbf{F}_3)$ is solvable, but $\mathrm{SL}_2(\mathbf{F}_9)$ has the good manners not to have any such quotients. So why can't one now run the same argument as above and disprove Serre's conjecture? That's a good question, and the entire argument works, up to the issue of defining the Hasse invariant. Quaternion algebras over fields other than $\mathbf{Q}$ are a bit of a disaster, because they don't have nice moduli theoretic descriptions. That doesn't mean they don't have Hasse invariants, however. But now what happens, which at this point in the game I suspected but was confirmed and explained to be by George Boxer (Keerthi also suggested a computation which would lead to the same conclusion): the Hasse invariant is no longer a section of $\Omega^1 = \omega^{\otimes 2} = \omega^{p-1}$, but rather a section of ω^{p^2-1}, and this has too large a degree to contribute to the cohomology of Ω^1. Since $2^2 - 1 \geq 2$, it still has too large a degree when $p = 2$, which is good, because otherwise working at this prime could have given rise to a counter-example to Serre's conjecture because $\mathrm{SL}_2(\mathbf{F}_4) = A_5$ is perfect. (One would have to be slightly more careful with $p = 2$ about comparison theorems, but at least one is dealing with curves.) So the conclusion is that Serre's conjecture still stands, but only because various Hasse invariants in low weight are exactly accounted for by the solvability of $\mathrm{SL}_2(\mathbf{F})$ when $|\mathbf{F}| = 2, 3$.

(Also, completely randomly and apropos of nothing, this link is now the top hit on the web to the search "Fred Diamond's Beard.")

Notes 79.1. It still is (the first hit). The youtube link in the blog no longer works, here's a direct link to Fred's talk.

80. $\mathbf{Z}_p$-extensions of number fields, Part I

Thu, 24 Nov 2016

In the next few posts, I want to discuss a problem that came up when I wrote a paper with Barry Mazur. We had a few observations and remarks that we discussed as part of a possible sequel but which we never wrote up; mostly because we never could quite prove what we wanted to prove. But some of those remarks might be worth sharing.

The basic problem is as follows. Let $E/\mathbf{Q}$ be a number field of signature (r, s). Let p be a prime that splits completely in E (this is not strictly necessary, but it makes things cleaner). Let S be a set of primes above p. If S includes all the primes above p, then the Leopoldt Conjecture for E and p is the statement that

$$r_S := \dim_{\mathbf{Q}_p} \mathrm{Gal}(E^S/E)^{\mathrm{ab}} \otimes \mathbf{Q} = 1 + s.$$

The question is then to predict what happens when S is a strict subset of the primes above p. This leads to the following minimalist definition:

Definition 80.1. The field E is rigid at p if

$$r_S := \dim_{\mathbf{Q}_p} \mathrm{Gal}(E^S/E)^{\mathrm{ab}} \otimes \mathbf{Q} = \begin{cases} \#S - (r+s-1), & \#S \geq r+s-1, \\ 0, & \text{otherwise.} \end{cases}$$

Note that, for any field E, the right hand side is always a lower bound. So rigid pairs (E,p) are those which have no "unexpected" $\mathbf{Z}_p$-extensions. If E is totally real, the Leopoldt Conjecture at p is equivalent to E being rigid. However, one does not predict that all fields E are rigid. The following is elementary:

Proposition 80.2. *If E is a totally imaginary CM field, then complex conjugation acts naturally on the set S. There are inequalities $r_S \geq [E^+ : \mathbf{Q}] + 1$ if S consists of all primes above p, and*

$$r_S \geq \frac{1}{2}\#(S \cap cS)$$

otherwise. If Leopoldt's conjecture holds, then these inequalities are equalities.

It follows that if E is a CM field of degree at least 4, then E is not rigid for any prime p, because when S consists of two primes conjugate to each other under complex conjugation, then

$$r_S \geq 1 \geq 2 - (r+s-1) = 2 - s.$$

The "extra" extensions are coming from algebraic Hecke characters. Our expectation is that this is the only reason for a pair (E,p) to be rigid. For example:

Conjecture 80.3. *Suppose that E does not contain a totally imaginary CM extension F of degree at least 4. Then (E,p) is rigid for any prime p that splits completely in E.*

(When I say conjecture here, I really mean a guess; it could be false for trivial reasons.) Naturally these conjectures are hard to prove, since they imply Leopoldt's Conjecture. Even if one *assumes* Leopoldt's Conjecture, this conjecture still seems tricky. It makes sense, however, to see what can be proven under further "genericity" hypotheses on the image of the global units inside the local units. To this end, let me recall the **Strong Leopoldt Conjecture** which Barry and I formulated our original paper. Let $F/\mathbf{Q}$ be the splitting field of $E/\mathbf{Q}$. Let G be the Galois group of $F/\mathbf{Q}$. There is a G-equivariant map

$$\mathcal{O}_F^\times \otimes \mathbf{Q}_p \to \prod_{v|p} \mathcal{O}_{F,v}^\times \otimes \mathbf{Q}_p.$$

The right hand side is isomorphic as a G-module to $\mathbf{Q}_p[G]$. However, more is true; for a fixed prime $v|p$, there is an isomorphism

$$\mathbf{Q}_p[G] = \mathbf{Q}[G] \otimes \mathbf{Q}_p$$

which is well-defined up to a scalar in $\mathbf{Q}_p$ coming from a choice of p-adic logarithm for the given place at p. It makes sense to talk about a rational subspace V of the right hand side, namely, a space of the form $V = V_{\mathbf{Q}} \otimes \mathbf{Q}_p$ for some $V_{\mathbf{Q}} \subset \mathbf{Q}[G]$. The strong Leopoldt conjecture (of [CM09]) asserts that the intersection of the global units which such a rational subspace is as small as it can possibly be subject to the constraints of the G-action on both V and the units, together with Leopoldt's

conjecture that the map from the units tensor $\mathbf{Q}_p$ is injective. Let $H = \mathrm{Gal}(F/E)$. By inflation-restriction, there is an isomorphism

$$H^1_S(E, \mathbf{Q}_p) = H^1_T(F, \mathbf{Q}_p)^H,$$

where the subscript denotes classes "unramified outside S," and where T denotes the set of primes in F above S. By class field theory, this may be identified with the H-invariants of the cokernel of the map

$$\mathcal{O}_F^\times \otimes \mathbf{Q}_p \to \prod_T \mathcal{O}_{F,v}^\times \otimes \mathbf{Q}_p.$$

The cokernel is larger than expected if and only if the kernel is bigger than expected. In particular, $r_S = \dim H^1_S(E, \mathbf{Q}_p)$ is bigger than expected only if

$$\left(\mathcal{O}_F^\times \otimes \mathbf{Q}_p \bigcap \prod_{\neg T} \mathcal{O}_{F,v}^\times \otimes \mathbf{Q}_p \right)^H$$

is bigger than expected. Note that the product over any subset T of primes in the right hand side is a rational subspace. Certainly the Strong Leopoldt Conjecture determines the dimension of the intersection $U \cap V$ of the unit group with a rational subspace. What is slightly less clear is that the intersection $(U \cap V)^H$ for any subgroup H is also determined by the strong Leopoldt Conjecture, but this is true (and we prove it). As a consequence, one has:

Lemma 80.4. *Assuming the Strong Leopoldt Conjecture, the dimension r_S depends only on G, H, and S.*

This "reduces" the computation of r_S to an intersection problem in a certain Grassmannian. But this is a computation we were never really able to do!

This is the problem: One knows very well the structure of the unit group of F as a G-module. So to compute the relevant intersections, one only has to compute the intersection with a "generic" rational subspace. Paradoxically, it seems very difficult in general to give explicit examples of rational subspaces which are generic enough to obtain the correct minimal value. So while, for formal reasons, almost any rational subspace will do, none of the nice subspaces which allow us to compute the intersection tend to be good enough.

Instead, to compute these intersections, we somewhat perversely look at actual number fields and their unit groups. This seems like a bad idea, since even verifying Leopoldt for a particular K and p is not so easy to do. So instead, we start with a totally real number field K of a certain form. Then, *under the assumption of Leopolodt's conjecture* we can (non-constructively) find subspaces of rational subspaces V which provably minimize various intersections $\dim(W \cap V)$ for various unit-like submodules W. We then *deform* the field K to other fields L of different signature, and use this construction (as well as the Strong Leopoldt Conjecture) to make deductions about L. In the next post, we explain how this led Barry and me to a proof of the following:

Theorem 80.5. *Let $E/\mathbf{Q}$ be a degree n field with whose Galois closure F has Galois group $G = S_n$. Assume the Strong Leopoldt Conjecture. Then (E, p) is rigid for any prime p which splits completely in p.*

I will explain the details next time. But to unwind the serpentine argument slightly, we do not prove the result by finding rational subspaces in $\mathbf{Q}_p[G]$ whose

intersection with the units of F has the a dimension which we can compute to be the expected value, but only rational subspaces whose dimension we can compute *contingent* on Leopoldt's conjecture for some auxiliary totally real number field. In other words, we would like to compute the generic dimension of some intersection inside some G-Grassmannian, a problem which has nothing to do with number theory, and we compute it using Leopoldt's conjecture. More next time!

81. $\mathbf{Z}_p$-extensions of number fields, Part II

Sat, 26 Nov 2016

This is continuation of § 80. We claimed there that we were going to deform a totally real number field of degree n into a field with signature (r, s) with $r+2s = n$, and pass information about Leopoldt's conjecture from one field to the other.

How does one "deform" a number field? One natural way is to think of a finite etale map of varieties $X \to Y$ defined over $\mathbf{Q}$, and then consider the fibres. More prosaically, write down some family of polynomials and then vary the coefficients. Most of the time, the unit group doesn't behave so well in such families. For example, consider the equation:

$$x^2 - D = 0.$$

If one varies D, even with some local control at primes dividing infinity (that is, keeping D positive), then it is not at all clear how the fundamental unit varies. In fact, one knows that the height of the fundamental unit is very sensitive to the size of the class number, which changes somewhat irregularly with D. On the other hand, consider the equation:

$$x^2 - Dx = 1.$$

Here one is in much better shape: as D varies, the element x will always be a unit, and moreover always generates a finite index subgroup of the full unit group. How might one use this for arguments concerning Leopoldt's conjecture? The idea is to consider (as above) a family of number fields in which some finite index subgroup of the full unit group is clearly visible, and is deforming "continuously" in terms of the parameters. Then, by Krasner's Lemma, we see that Leopoldt's conjecture for one number field (and a fixed prime p) will imply the same for all sufficiently close number fields. To start, however, one needs to have such nice families.

81.1. **Ankeny–Brauer–Chowla Fields.** One nice family of number fields that deforms nicely is the class of so-called Ankeny–Brauer–Chowla fields (from their 1956 paper [ABC56]):

$$\prod(x - a_i) - 1 = 0$$

It is manifestly clear that in the field $\mathbf{Q}(x)$, the elements $x - a_i$ are all units, and that (generically) there is only one multiplicative relation, namely that the product over all such units is trivial. In this way, we get a family of number fields (with generic Galois group) S_n and with a family of units generating a free abelian group of rank $n - 1$. With a little tweak, we can also ensure that the prime p splits completely. Concretely, consider the equations

$$\prod(X - a_i) - \prod(X - b_i) = 1,$$

where X is a formal variable. The corresponding variety Y is connected of dimension n, and the projection to $\mathbf{A}^n$ given by mapping to $\{b_i\}$ is a finite map, and so, generically, the values of b_i on Y are all distinct. In particular, for sufficiently large primes p, Y has points over $\mathbf{F}_p$ where all the b_2 are distinct modulo p. Fix such a point $\{a_i, b_i\}$ over $\mathbf{F}_p$. Lift the a_i in $\mathbf{F}_p$ to arbitrary integers in $\mathbf{Z}$. Then, by Hensel's lemma, there exist p-adic integers v_i congruent to b_i mod p such that

$$\prod(x - a_i) - 1 = \prod(x - v_i),$$

and so p splits completely in our field as long as the a_i satisfy some suitable non-empty congruences mod$p$.

81.2. **Deforming the signature.** Suppose we assume that, for a fixed choice $A = (a_1, .., a_n)$, the corresponding field F satisfies Leopoldt's conjecture. Then we see that, in a sufficiently small neighbourhood of A, we obtain many other fields which are totally real with Galois group S_n that also satisfy Leopoldt's conjecture. On the other hand, our goal is to study fields of signature (r, s) with $r + 2s = n$. So we want to deform our fields to have non-trivial signature. I learnt this trick by reading a paper of Bilu: we deform the fields in a slightly different way, by making the replacement

$$(x - a_i)(x - a_j) \Rightarrow (x - a_i)(x - a_j) + u,$$

where u has very small p-adic valuation, and yet is a large positive integer. The corresponding field no longer has n obvious units (whose product is one), but now only $n - 1$ obvious units (whose product is one), where one of the units is now the quadratic polynomial above. On the other hand, one can also see that the signature of the number field is now $(n-2, 1)$. So we still have a nice finite index subgroup of the unit group. Moreover, p-adically, if our original units are written as $\{u_i\}$, then we get (p-adically) something very close (by Krasner), except now u_i and u_j have been replaced by $u_i + u_j$. By combining other pairs of units in the same way, we can reduce the signature to (r, s) with any $r + 2s = n$ and still have a nice p-adically continuous finite index family of global units.

Proposition 81.3. *Suppose that Leopoldt's conjecture holds for the original field K at p. Then, by deforming suitably chosen pairs of roots, we obtain a (infinitely many) fields L with Galois group S_n and signature(r, s) with $r + 2s = n$ such that, for a choice of $r + s - 1$ primes above p in L, the p-adic regulator of the units at those $r + s - 1$ primes is non-zero.*

As a consequence, for that choice of $r + s - 1$ primes, the corresponding maximal $\mathbf{Z}_p$-extension has rank zero. This proves that (L, p) is rigid for this choice of S. However, since S_n is n-transitive, the same result applies for any such choice of $r + s - 1$ primes. It's an elementary lemma to see that this also implies the result for sets S which are either larger or smaller than $r + s - 1$.

Proof. The argument is exactly as you expect: Given the original field K, the assumption of Leopoldt's conjecture for K implies that at least one of the corresponding $(r + s - 1) \times (r + s - 1)$ minors must be non-zero. We then deform the field globally so that the corresponding units in L of signature (r, s) are related to this minor, which (by Krasner) will still be non-zero. □

Question 81.4. The starting point of this construction was the assumption that K satisfied Leopoldt's conjecture. Can one prove this directly? That is, can one find a choice of a_i such that the field

$$\prod(x - a_i) - 1 = 0.$$

satisfies Leopoldt at p?

This seems quite plausible, after all, we have seen above that there are n nice units of finite index in the unit group whose regulator varies p-adically. So, it suffice to show that the regulator is not zero in the entire family. This certainly *seems* like an easier problem, because it's easier to prove a function is non-zero rather than the special value of a function (for example, by looking at the derivative). Still, I confess that I don't know how to prove this.

82. Artin no-go Lemma

Tue, 13 Dec 2016

The problem of constructing Galois representations associated to Maass forms with eigenvalue $1/4$ is, by now, a fairly notorious problem. The only known strategy, first explained by Carayol, is to first transfer the representation to a unitary group over an imaginary quadratic field, where one can realize the corresponding transfer in the coherent cohomology of a related "Griffiths–Schmid" variety X. Then one hopes to study the action of Hecke operators on this space and relate it to some (hopefully existing) rational structure on the cohomology. The wrinkle is that X is not algebraic but merely a complex manifold, so it's not so easy to see how to impose any rational structure on the (higher) coherent cohomology. I have nothing intelligent to say about whether this approach will work. However, suppose one is as optimistic as possible, and thinks about what one might *hope* to be true — not only to construct Galois representations but also prove the converse (Artin). Then, following [CG18a], one might hope to find an *integral* structure on this cohomology (with interesting torsion) on which to study congruences and then glue together torsion classes using Taylor–Wiles to produce a patched complex of the right length. What is the invariant l_0 in this case? One might (generally) hope in this context to study conjugate self-dual representations

$$\rho : G_E \to \mathrm{GL}_3(A)$$

for an imaginary quadratic field E (in which p splits) for local Artinian rings $(A, \mathfrak{m})$ with $A/\mathfrak{m}$ of characteristic p which are unramified at p. The difference in dimension between the ordinary local deformation ring and the unramified deformation ring appears to be 3, and thus we expect $l_0 = 3$. Correspondingly, we expect cohomology to occur in a range of cohomological degrees $[q_0, q_0+3]$ for some q_0. Moreover, in the presence of cohomology in characteristic zero, we expect to see cohomologies in all such degrees. Yet this doesn't happen for X; in fact, the cohomology only occurs (in characteristic zero) in degrees 1 and 2 (according to Richard). This suggests not only that we *won't* be able to prove modularity using integral cohomology of X, but even that — in the most naive sense — we should not expect an integral structure at least with the usual properties. Namely, if we patch a complex of integral cohomology of length 1, then the corresponding patched modules in cohomology will be too big

for any unramified deformation ring to act. So it appears that the best possible scenario is too good to be true.

On a different matter, there is another pressing issue I would like to bring to my readers. In the papers I have written with Matt and David (and some by myself), we have used the notation l_0 — which has its origins in the book of Borel and Wallach. There is, however, a competing notation in some of Akshay's papers, namely δ. One argument for the latter is that l_0 specifically comes from a particular calculation in $(\mathfrak{g}, K)$-cohomology, and is not compatible with other situations in which one might want to consider the problem of cohomology in various degrees. (For example, for weight one modular forms, the Galois $l_0 = 1$ whereas GL(2)/Q has discrete series.) My argument is that there will never be any confusion when using l_0, and that it has the property of being unlikely to every conflict with any other notation. Moreover, the phenomenology in both coherent and Betti cohomology both depend on l_0 in exactly the same way. Dear reader: what is your opinion?

Notes 82.1. I am happy to say that l_0 has won the battle of the notations.

83. Correspondance Serre–Tate, Part I

Sun, 25 Dec 2016

Reading the correspondence between Serre and Tate (see [Col15a, Col15b, Col17]) has been as delightful as one could expect. What is very nice to see — although perhaps not so surprising — is the utter delight that both Serre and Tate find in discussing numerical examples. One of the beautiful aspects of number theory is that there is an abundance of examples, each of which exhibit both special cases of a vast general theory and yet each delighting with their own idiosyncrasies: $\mathbf{Q}(\sqrt{-23})$, $X_0(11)$, 691, 144169, etc. (It is precisely the absence of such examples, or at least any discussion of them, why geometric Langlands tends to leave me completely cold.) Take, for example, the following:

> **Letter from Tate to Serre, Dec 8, 1958:**
>
> Are you aware that the class number of the field of 97th roots of 1 is divisible by 3457 and 118982593? And that 3457 = 36 * 96 + 1 and 118982593 = 1239402 * 96 + 1?

If reading that doesn't give you just a little thrill, then you have no soul. Does it have any significance mathematically? The class number is large, of course, which relates to the fact (proved by Odlyzko) that there are only finitely many Galois CM fields with bounded class number. (The reason why one can access class numbers of CM fields F/F^+ is that the unit group of F and F^+ are the same up to finite index, so the *ratio* of zeta values $\zeta_F(1)/\zeta_{F^+}(1)$ is directly related to the minus part of the class group h^- uncoupled from any regulator term, so one can access this analytically.) Alternatively, one might be interested in the congruences of the primes q dividing the class number. In this case, we see a reflection of the conjectures of Cohen and Lenstra. Namely, we expect that there is a strong preference for the class group to be "more cyclic," especially for larger primes. The class group also has an action of $(\mathbf{Z}/97\mathbf{Z})^\times$ which is cyclic of order 96. Since one expects the plus part h^+ to be very small (and indeed in this case it is trivial), this means that complex conjugation should act non-trivially, which means that the group of order 96 should (at least) act through a quotient of order at least 32. So if the class group is actually

cyclic, this forces the prime divisors q of h_F to be 1 mod 32, and even 1 mod 96 if the class group of F doesn't secretly come from the degree 32 subfield of F (which it doesn't). (Not entirely irrelevant is Rene Schoof's nice paper [Sch03] on computing class groups of real cyclotomic fields.)

Both Serre and Tate are unfailingly polite to each other. As a running joke, the expression "talking through one's hat" occurs frequently, as for example the letter of Nov 14, 1961, where the subtle issue of the failure of $B \otimes_A C \to B \widehat{\otimes}_A C$ is discussed. (Another amusing snippet from that letter "Even G. himself makes mistakes when he thinks causally.") The correspondence is also fascinating from the perspective of mathematical history — one sees the progress of many ideas as they are created, including the Honda-Tate theorem and the Tate conjecture over finite fields. The first time the latter appears (as a very special case) it actually turns out to be an argument of Mumford, who shows Tate an argument (using Deuring) why when two elliptic curves have the same zeta function they are isogenous. This elicits the following reaction from Tate:

> **Letter from Tate to Serre, May 9, 1962:**
> "Damn! The result is certainly new to me, and it frankly makes me mad that I never noticed it"

We have all been there, although, to be fair, most of us have the excuse of not being Tate!

84. Central extensions, updated

Mon, 16 Jan 2017

I previously mentioned in § 78 a problem concerning polynomials, whose motivation came from thinking about weight one forms and the inverse Galois problem for finite subgroups of $\mathrm{GL}_2(\mathbf{C})$. I still like the polynomial problem, but I realized that I was confused about the intended application. Namely, given a weight one form with projective image A_5, there is certainly a unique minimal lift up to twist, but the images of the twists *also* automatically have image given by a central extension A_5. So, just by twisting, one can generate all such groups as Galois groups by starting with a minimal lift. More prosaically, every central extension of A_5 by a cyclic group is either a quotient of $A_5 \times \mathbf{Z}$ or of $\widetilde{A}_5 \times \mathbf{Z}$ where $\widetilde{A}_5$ is the Darstellungsgruppe of A_5 (which is $\mathrm{SL}_2(\mathbf{F}_5)$). So, to solve the inverse Galois problem for central extensions of A_5, it suffices to solve it for $\mathrm{SL}_2(\mathbf{F}_5)$. That is not entirely trivial, but it is true.

I still think it's an interesting problem to determine which extensions of A_5 by cyclic groups occur as the Galois groups of *minimally ramified up to twist* extensions, but that is not the same as the inverse Galois problem.

85. The class number 100 problem

Thu, 19 Jan 2017

Some time ago, Mark Watkins busted open the "class number n" problem for smallish n, finding all imaginary quadratic fields of class number at most 100 (the original paper is here, see also [Wat04]) Although the paper describes the method in detail, it does not actually give the complete list of imaginary quadratic fields which occur (for fairly obvious reasons given the size of the list). I've occasionally

wanted to consult the actual list, and most of the time I have just emailed Mark to find out the answer. But now it is available online! Here is the link. (Maybe someone could put this on the LMFDB?)

Consulting the table one immediately notices a number of beautiful facts, such as the fact that $(\mathbf{Z}/3\mathbf{Z})^3$ does not occur as a class group. (Our knowledge of p-parts of class groups, following Gauss, Pierce, Helfgott, Venkatesh, and Ellenberg, is enough to show that $(\mathbf{Z}/2\mathbf{Z})^n$ and $(\mathbf{Z}/3\mathbf{Z})^n$ for varying n only occur finitely often [similarly these groups plus any fixed group A], but those results are not effective.) One also sees that $D = -5519$ and $D = -1842523$ are the first and last IQF discriminants of class number 97. It's the type of table that immediately bubbles up interesting questions which one can at least try to give heuristic guesstimates. For example, let $\mu(A)$ denote the number of imaginary quadratic fields with class group A. Can one give a plausible guess for the rough size of $\mu(A)$? One roughly wants to combine the Cohen–Lenstra heuristics with the estimate $h \sim \Delta^{1/2}$. To do this, I guess one would roughly want to have an estimate for

$$\sum_{x^{1/2-\epsilon} \leq |A| \leq x^{1/2+\epsilon}} \frac{1}{|\mathrm{Aut}(A)|}.$$

I wouldn't be surprised if someone has already carried out this analysis (thought I don't know any reference). As a specific examples:

Question 85.1. What is the expected growth rate of $\mu(\mathbf{Z}/q\mathbf{Z})$ over primes q?

Question 85.2. Is there a finitely generated abelian group which provably does not occur as the first homology of a congruence arithmetic hyperbolic 3-manifold?

At any rate, this is a result that Gauss would have appreciated.

Comment 85.3 (Emmanuel Kowalski)**.** I vaguely remember some work of Sound on the asymptotic number of imaginary quadratic field with given class number, that might have discussed also specifying the group... Searching leads to this paper (see [Sou07]) and then to the follow-up here (see [HJK+19]) by Holmin, Jones, Kurlberg, McLeman and Petersen that seems to address exactly the question with the group structure taken into account.

And in fact there's a new preprint this morning by Y. Lamzouri that's also related (counting imaginary quadratic fields with odd class number $< H$): here (see [Lam17]).

86. Virtual coherent cohomology

Wed, 22 Feb 2017

I gave a talk yesterday where I attempted to draw parallels between the cohomology of (arithmetic) 3-manifolds and weight one modular forms. It was natural then to think about whether there was an analogue of the virtual Betti number conjecture. Recall the following:

Theorem 86.1 (Agol, [Ago13])**.** *Let M be a compact hyperbolic 3-manifold. Then $\dim H^1(N, \mathbf{Q})$ is unbounded as N ranges over all finite covers $N \to M$.*

(There's an analogous version for finite volume hyperbolic manifolds with cusps.) What is the corresponding conjecture in coherent cohomology? Here is a first attempt at such a question.

Question 86.2. Let X be a proper smooth curve of genus $g \geq 2$ defined over $\mathbf{Q}$. Let $\mathscr{L}$ denote a line bundle such that $\mathscr{L}^{\otimes 2} = \Omega^1_X$. As one ranges over all (finite etale) covers $\pi : Y \to X$, are the groups

$$H^0(Y, \pi^* \mathscr{L})$$

of unbounded dimension?

One might ask the weaker question as to whether there is a cover where this space has dimension at least one (and in fact this is the first question which occurred to me). However, there are some parity issues. Namely, Mumford showed the dimension of $H^0(X, \mathscr{L})$ is locally constant in $(X, \mathscr{L})$, and this dimension is odd for precisely $2^{g-1}(2^g + 1)$ choices of $\mathscr{L}$ (there are 2^{2g} such choices and the choices are a torsor for 2-torsion in the corresponding Jacobian). But I think this means that one can always make $\pi^*\mathscr{L}$ effective for some degree 2 cover, and thus produce at least one dimensions worth of sections. For example, when $g = 1$, then $\Omega^1_X = \mathcal{O}_X$, and $\mathscr{L} = \mathcal{O}_X$ has global sections whereas the other square-roots correspond literally to 2-torsion points. But those sections become trivial after making the appropriate 2-isogeny.

Another subtlety about this question which is worth mentioning is that I think the result will have to be false over the complex numbers, hence the deliberate assumption that X was defined over $\mathbf{Q}$, or at least over a number field. Specifically, I think it should be a consequence of Brill–Noether theory that the set of X in $\mathscr{M}_g$ such that

$$\dim H^0(Y, \pi^* \mathscr{L}) \geq 1$$

for any choice of $\mathscr{L}$ and any cover $\pi : Y \to X$ of degree bounded by a fixed constant D will be a finite union of proper varieties of positive dimension. And now the usual argument shows that, as D increases, any countable union of varieties cannot exhaust $\mathscr{M}_g$. But it *can*, of course, exhaust all the rational points, and even all the algebraic points.

There's not really much evidence in favor of this question, beyond the following three very minor remarks.

(1) The only slightly non-trivial case one can say anything about is when X is a Shimura curve over $\mathbf{Q}$, and then the answer is positive because there exist lots of weight one forms (which one can massage to have the right local structures after passing to a finite cover).
(2) The analogy between $H^0(X, \mathscr{L})$ and $H^1(M, \mathbf{Q})$ is fairly compelling in the arithmetic case, so why not?
(3) There doesn't seem to be any *a priori* reason why the virtual Betti number conjecture itself was true, and it is certainly false in for related classes of groups (groups with the same number of generators and relations, word hyperbolic groups), so, by some meta-mathematical jiu-jitsu, one can view the lack of a good heuristic in the hyperbolic case as excusing any real heuristic in the coherent case.

Comment 86.3 (Felipe Voloch)**.** If X is defined over a field of characteristic two, then there is a natural line bundle like in your question that you can take. Namely,

in terms of divisors, the divisor of an exact differential dx is of the form $2D$. Use the line bundle corresponding to D and consider first etale covers of X with degree a power of 2. If X is ordinary your H^0's will be trivial and if X is not ordinary they will grow linearly with the degree. The proof is not difficult but is too long for a blog comment. Now, if X is ordinary but has a non-ordinary etale cover, then you get unbounded H^0 by taking covers of that cover. Does that work for all X? Maybe, I don't know. Unfortunately, this does not help in characteristic zero even if X has good reduction at 2, as the inequalities for H^0 go the wrong way.

Comment 86.4 (Jordan Ellenberg)**.** And of course the comment about what goes wrong over $\mathbf{C}$ reveals that this is yet another example of a charming kind of question, "when can a "natural" countable union of proper subvarieties cover all the $\overline{\mathbf{Q}}$-points of a variety" I scratched my chin about this one too here but I think your example goes even beyond the class of examples I discussed there!

87. A non-liftable weight one form modulo p^2

Fri, 10 Mar 2017

I once idly asked Richard (around 2004ish) whether one could use Buzzard–Taylor arguments to prove that any representation:

$$\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\mathbf{Z}/p^2\mathbf{Z})$$

which was unramified at p and residually irreducible (and modular) was itself modular (in the Katz sense). Galois representations of this flavour are obviously something I've thought about (and worked on with David Geraghty) quite a lot since then. But I have never actually seen any examples of mod p^2 forms which didn't lift to characteristic zero. I asked George Schaeffer about it once, but his computations were only set up to detect the primes for which non-liftable forms existed rather than to compute the precise structure of the torsion in $H^1(X, \omega)$. But just today I stumbled across an example in relation to a pairing I learned about from Akshay (which I will tell you all about some other time).

The particular form (or rather pair, since it comes with a twist by the nebentypus character) occurs at level $\Gamma_0(103) \cap \Gamma_1(3)$, and is defined over the ring $\mathbf{Z}/11^2\mathbf{Z}$. It doesn't lift to a weight one form mod11^3. The nebentypus character is the only one it could be at this level and weight: the odd quadratic character of conductor 3. When I looked again at Schaeffer's thesis, he does indeed single out this particular level as a context where computations suggested there might exist a mod p^2 form. (Literally, he says that a computation "seems to imply the existence" of such a form.) I guess this remark was not in any previous versions of the document I had, so I hadn't seen it. Here are the first few terms of the q-expansion(s):

$$g = q + 16q^2 + 20q^3 + 15q^4 + 58q^5 + 78q^6 + 22q^7 + \ldots + 91q^{11} + \ldots + 104q^{103} + \ldots$$

$$f = q + 105q^2 + 115q^3 + 15q^4 + 63q^5 + 96q^6 + 22q^7 + \ldots$$

Some remarks. Note that the coefficients of g and f satisfy $a(g, n) = \chi(n)a(f, n)$ for all $(n, 3) = 1$ and where χ is the quadratic character of conductor 3 (the nebentypus character). On the other hand, at the prime 3, we have

$$\rho_f = \begin{pmatrix} \chi\psi^{-1} & 0 \\ 0 & \psi \end{pmatrix}, \qquad \rho_g = \chi \otimes \rho_f = \begin{pmatrix} \chi\psi & 0 \\ 0 & \psi^{-1} \end{pmatrix},$$

and so the eigenvalue of U_3 is the image of Frobenius at 3 under ψ or ψ^{-1}, and hence satisfies the equality

$$a(g,3)a(f,3) = \psi(\mathrm{Frob}_3)\psi^{-1}(\mathrm{Frob}_3) = 20 \cdot 115 = 1 \mod 121.$$

I was temporarily confused about the fact that $a_q = 1 + q$ for the Steinberg prime $q = 103$ rather than $a_q = \pm 1$, and thought for a while I had made an error or mathematics was wrong. But then I realized this was weight one not weight two, and so one should have instead that $(a_q)^2 = q^{-1}$ (note that $\chi(103) = 1$.) And it just so happens that the equation

$$(1+q)^2 = q^{-1}$$

in a weird coincidence has a solution very close to 103 (this is a solution $\mod 11^3$, in fact). It's easy enough to see that the image of rho and its twist contains $\mathrm{SL}_2(\mathbf{Z}/11^2\mathbf{Z})$ with index two, and so has degree 3513840. (At this level, the only real alternative is that the form is Eisenstein, which it isn't.) The root discriminant is not particularly small, it is

$$103^{1-1/11^2} \cdot 3^{1/2} = 171.6970\ldots$$

Finally, the Frobenius eigenvalues at the prime $p = 11$ are distinct, which is easy enough to see because otherwise the coefficient of q^{11} would have to be twice the squareroot of $\chi(11) = -1$, which isn't even a square $\mod 11$.

88. Pseudo-representations and the Eisenstein Ideal

Wed, 29 Mar 2017

Preston Wake is in town, and on Tuesday he gave a talk on his recent joint work with Carl Wang Erickson [WWE20]. Many years ago, Matt and I studied Mazur's Eisenstein Ideal paper from the perspective of Galois deformation rings. Using some subterfuge (involving a choice of auxiliary ramification line at the prime N following an idea of Mark Dickinson), we proved an $R = \mathbf{T}$ theorem. One satisfactory aspect (to us, at least) of our paper was that we were able to reconstruct from a purely Galois theoretic perspective some of the thorny geometric issues in Barry's paper, particularly at the prime 2. Another problem of Barry's that we studied was the question of determining for which N and p the cuspidal Hecke algebra was smooth (equivalently, whether the cuspidal Hecke algebra completed at a maximal Eisenstein ideal $\mathfrak{m}$ of residue characteristic p was equal to $\mathbf{Z}_p$). Our theorem showed this was equivalent to the existence of certain Galois deformations to $\mathrm{GL}_2(\mathbf{F}[e]/e^3)$. Although we were able to give a precise account of what happens for $p = 2$, for larger p we could only prove the following:

Theorem 88.1. *Let $p \geq 3$ be prime, and let $N \equiv 1 \bmod p$ be prime. If the rank of the cuspidal Hecke algebra of level $\Gamma_0(N)$ localized at the Eisenstein prime is greater than one, then*

$$K = \mathbf{Q}(N^{1/p})$$

has non-cyclic p-class group.

Note that there is always trivial p-torsion class in the class group of K coming from the degree p extension inside the Nth roots of unity. In our paper, we speculated that this was actually an equivalence. To quote the relevant passage:

> We expect (based on the numerical evidence) that the condition that the class group of K has p-rank [at least] two is equivalent to the existence of an appropriate group scheme, and thus to [the rank being greater than one].

Not a conjecture, fortunately, as it turns out to be false, already for $p = 7$ and $N = 337$. Oops! In fact, this had already been observed by Emmanuel Lecouturier in [Lec18]. Wake and Wang Erickson, however, give a complete characterization of when the rank is greater than one, namely

Theorem 88.2 ([WWE20])**.** *Let $a \in H^1(\mathbf{Z}[1/Np], \mathbf{F}_p(1))$ be the Kummer class corresponding to N. Let $b \in H^1(\mathbf{Z}[1/Np], \mathbf{F}_p(-1))$ be the (unique up to scalar) non-trivial class which is unramified at p. Then the rank of the Hecke algebra is greater than one if and only if the cup product $a \cup b$ vanishes.*

They prove many other results in their paper as well. The main theoretical improvement of their method over the old paper was to work with pseudo-representations rather than representations. On the one hand, this requires some more technical machinery, in particular to properly define exactly what it means for a pseudo-representation to be finite flat. On the other hand, it avoids certain tricks that Matt and I had to make to account properly for the ramification at N as well as to make the deformation problem representable. Our methods would never work as soon as N is not prime, whereas this is not true for their new results. In particular, there is real hope that there method can be applied to much more general N.

Let me also note that Merel in the '90s found a completely different geometric characterization of when the cuspidal Hecke algebra had rank bigger than one; explicitly, for $p \geq 3$ and $N = 1 \bmod p$, it is bigger than one when the slightly terrifying expression:

$$\prod_{i=1}^{(N-1)/2} i^i$$

is a pth power modulo N. So now there are a circle of theorems relating three things: vanishing of cup products, ranks of Eisenstein Hecke algebras, and Merel's invariant above. It turns out that one can directly relate Merel's invariant to the cup product using Stickelberger's Theorem. On the other hand, Wake and Wang Erickson also have a nice interpretation of the expression above as it relates to Mazur–Tate derivatives (possibly this observation is due to Akshay), and they also prove some nice results in this direction. And I haven't even mentioned their other results relating to higher ranks and higher Massey products, and many other things. Lecouturier's paper is also a good read, and considers the problem from another perspective.

In Preston's talk, he sketched the relatively easy implication that the vanishing of the cup product $a \cup b$ above implies that the class group of $\mathbf{Q}(N^{1/p})$ has non-cyclic p-part. The main point is that the vanishing of cup products is exactly what is required for a certain extension problem, and in particular the existence of a Galois representation of the form:

$$\begin{pmatrix} 1 & a & c \\ 0 & \chi^{-1} & b \\ 0 & 0 & 1 \end{pmatrix},$$

where χ is the mod-p cyclotomic character. The class c gives the requisite extension (after some adjustment). Curiously enough, both the classes a and b exist for primes $N \equiv -1 \bmod p$. On the other hand, the corresponding H^2 group vanishes in this case, and so the pairing is always zero. Hence one deduces the following very curious corollary:

Theorem 88.3. *Let $p \geq 3$, and let $N \equiv -1 \bmod p$ be prime. Then the class number of $\mathbf{Q}(N^{1/p})$ is divisible by p.*

Question 88.4. Is there a direct proof of this theorem? In particular, is there an easy way to construct the relevant unramified extension of degree p for all such primes N? I offer a beer to the first satisfactory answer.

Notes 88.5. The beer question has been answered by Lang and Wake in this paper [LW22], although both authors are yet to claim their prize.

89. Who proved it first?

Wed, 26 Apr 2017

During Joel Specter's thesis defense, he started out by remarking that the q-expansion:

$$f = q \prod_{n=1}^{\infty} (1-q^n)(1-q^{23n}) = \sum a_n q^n$$

is a weight one modular forms of level $\Gamma_1(23)$, and moreover, for p prime, a_p is equal to the number of roots of

$$x^3 - x + 1$$

modulo p minus one. He attributed this result to Hecke. But is it really due to Hecke, or is this more classical? Let's consider the following claims:

(1) The form f is a modular form of the given weight and level.
(2) If p is not a square modulo 23, then $a_p = 0$.
(3) If p is a square modulo 23, and $x^3 - x + 1$ has three roots modulo p, then $a_p = 2$.
(4) If p is a square modulo 23, and $x^3 - x + 1$ is irreducible modulo p, then $a_p = -1$.

At when point in history could these results be proved?

Let's first start with Euler, who proved that

$$\prod_{n=1}^{\infty} (1-q^n) = \sum_{-\infty}^{\infty} q^{(3n^2+n)/2}(-1)^n$$

Using this, one immediately sees that

$$f = \sum \sum q^{\frac{1}{24}\left((6n+1)^2+23(6m+1)^2\right)}(-1)^{n+m}$$

This exhibits f as a sum of theta series. With a little care, one can moreover show that

$$2f = \sum \sum q^{x^2+xy+6y^2} - \sum \sum q^{2x^2+xy+3y^2}.$$

This is not entirely tautological, but nothing that Gauss couldn't prove using facts about the class group of binary quadratic forms of discriminant -23. The fact that f is a modular form of the appropriate weight and level surely follows from

known results about Dedekind's η function, which covers (1). From the description in terms of theta functions, the claim (2) is also transparent. So what remains? Using elementary number theory, we are reduced to showing that a prime p with $(p/23) = +1$ is principal in the ring of integers of $\mathbf{Q}(\sqrt{-23})$ if and only if p splits completely in the Galois closure H of $x^3 - x + 1$.

Suppose that $K = \mathbf{Q}(\sqrt{-23}) \subset H$. What is clear enough is that primes p with $(p/23) = +1$ split in K, and those which split principally can be represented by the form $x^2 + xy + 6y^2$ in essentially a unique way up to the obvious automorphisms. Moreover, the class group of $\mathrm{SL}_2(\mathbf{Z})$ equivalent forms has order 3, and the other $\mathrm{GL}_2(\mathbf{Z})$ equivalence class is given by $2x^2 + xy + 3y^2$. In particular, the primes which split non-principally in K are represented by the binary quadratic form $2x^2 + xy + 3y^2$ essentially uniquely. From Minkowski's bound, one can see that H has trivial class group. In particular, if $x^3 - x + 1$ has three roots modulo p, then the norm of the corresponding ideal to K is also principal and has norm $p = x^2 + xy + 6y^2$. This is enough to prove (3).

So the only fact which would not obviously be easy to prove in the 19th century is (4), namely, that *if* $p = x^2 + xy + 6y^2$, then p splits completely in H. The most general statement along these lines was proved by Furtwängler (a student of Hilbert) in 1911 — note that this is a different (and easier?) statement than the triviality of the transfer map, which was not proved until 1930 (also by Furtwängler), after other foundational results in class field theory had been dispensed with by Tagaki (another student of Hilbert!). Yet we are not dealing with a general field, but the much more specific case of an imaginary quadratic field, which had been previously studied by Kronecker and Weber in connection with the Jugendtraum. I don't know how much Kronecker could actually prove about (for example) the splitting of primes in the extension of an imaginary quadratic field given by the singular value $j(\tau)$. Some of my readers surely have a better understanding of history than I do. Does this result follow from theorems known before 1911? Who proved it first?

90. Elementary class groups updated

Sun, 11 Jun 2017

In § 88, I gave a short argument showing that, for odd primes p and N such that $N \equiv -1 \mod p$, the p-class group of $\mathbf{Q}(N^{1/p})$ is non-trivial. This post is just to remark that the same argument works under weaker hypotheses, namely:

Proposition 90.1. *Assume that N is p-power free and contains a prime factor of the form $q \equiv -1 \mod p$, and that p is at least 5. Then the p-class group of $K = \mathbf{Q}(N^{1/p})$ is non-trivial.*

The proof is pretty much the same. If N has a prime factor of the form $1 \mod p$, then the genus field is non-trivial. Hence we may assume there are no such primes, from which it follows that $H^1_S(\mathbf{F}_p)$ has dimension one and $H^2_S(\mathbf{F}_p)$ is trivial, where S denotes the set of primes dividing Np. The prime q gives rise to a non-trivial class $b \in H^1_S(\mathbf{F}_p(-1))$ which is totally split at p (this requires that p be at least 5), and the field K itself gives rise to a class $a \in H^1_S(\mathbf{F}_p(1))$. But now the vanishing of H^2

implies that $a \cup b = 0$ and hence there exists a representation of G_S of the form:

$$\rho : G_S \to \begin{pmatrix} 1 & a & c \\ 0 & \chi^{-1} & b \\ 0 & 0 & 1 \end{pmatrix},$$

where χ is the mod-p cyclotomic character. The class c gives the requisite extension (after possibly adjusting by a class in the one-dimensional space $H^1_S(\mathbf{F}_p)$). The main point is that the image of inertia at primes away from p is tame and so cyclic, but any unipotent element of $\mathrm{GL}_3(\mathbf{F}_p)$ has order p if p is at least three. This ensures c is unramified over K away from the primes above p. On the other hand, the class b is totally split at p. This implies that the class c is locally a homomorphism of the Galois group of $\mathbf{Q}_p$, and so after modification by a multiple of the cyclotomic class in $H^1_S(\mathbf{F}_p)$ may also be assumed to be unmarried at p. The fact that $b \neq 0$ ensures that $c \neq 0$, and moreover the fact that p is at least 5 implies that the kernel of c is distinct from that of a, completing the proof. (This result was conjectured in the paper *Class numbers of pure quintic fields* by Hirotomo Kobayashi [Kob16], which proves the claim for $p = 5$.)

Comment 90.2 (Franz Lemmermeyer)**.** The proposition follows, as Iimura remarked in [Iim86], from results due to Jaulent [Jau81]. I guess that it can be constructed classically by taking a subfield of the ray class group modulo q in the field of p-th roots of unity.

91. New results In modularity, Part I

Fri, 23 Jun 2017

I usually refrain from talking directly about my papers, and this reticence stems from wishing to avoid any appearance of tooting my own horn. On the other hand, nobody else seems to be talking about them either. Moreover, I have been involved recently in a number of collaborations with multiple authors, thus sufficiently diluting my own contribution enough to the point where I am now happy to talk about them.

The first such paper [ACC+23] I want to discuss has 9(!) co-authors, namely Patrick Allen, Ana Caraiani, Toby Gee, David Helm, Bao Le Hung, James Newton, Peter Scholze, Richard Taylor, and Jack Thorne. The reason for such a large collaboration is a story of itself which I will explain at the end of the second post. But for now, you can think of it as a polymath project, except done in a style more suited to algebraic number theorists (by invitation only).

In this first post, I will start by giving a brief introduction to the problem. Then I will state one of the main theorems and give some (I hope) interesting consequences. In the next post, I will be a little bit more precise about the details, and explain more precisely what the new ingredients are.

Like all talks in the arithmetic of the Langlands program, we start with:

The Triangle:

Let F be a number field, let p be a prime, and let S be a finite set of places containing all the infinite places and all the primes above p. Let G_S denote the absolute Galois group of the maximal extension of F unramified outside S. In many talks in the Langlands program, one encounters the triangle, which is a conjectural correspondence between the following three objects:

- **A:** Irreducible pure motives M/F (with coefficients) of dimension n.
- **B:** Continuous irreducible n-dimensional p-adic representations of G_S (for some S) which are de Rham at the places above p.
- **C:** Cuspidal algebraic automorphic representations π of $\mathrm{GL}(n)/F$.

In general, one would like to construct a map between any two of these objects, leading to six possible (conjectural) maps, which we can describe as follows:

- **A** $\to$**B:** This is really the only map we understand, namely, etale cohomology. (I'm being deliberately vague here about what a motive actually is, but whatever.)
- **B** $\to$**A:** This is the Fontaine–Mazur conjecture, and maybe some parts of the standard conjectures as well, depending on exactly what a motive is.
- **B** $\to$**C:** This is "modularity."
- **C** $\to$**B:** This is the existence of Galois representations associated to automorphic forms.
- **A** $\to$**C:** We really think of this as **A** $\to$**B** $\to$**C** and also call this modularity.
- **C** $\to$**A:** Again, this is a souped up version of **C** $\to$**B**. But note, we still don't understand how to do this even in cases where **C** $\to$**B** is very well understood. For example, suppose that π comes from a Hilbert modular form with integer coefficients of trivial level over a totally real field F of even degree. We certainly have an associated compatible family of Galois representations, and we even know that its symmetric square is geometric. But it should come from an elliptic curve, and we don't know how to prove this. The general problem is still completely open (think Maass forms). On the other hand, often by looking in the cohomology of Shimura varieties, one proves **C** $\to$**A** and uses this to deduce that **C** $\to$**B**.

This triangle is also sometimes known as "reciprocity." The other central tenet of the Langlands program, namely functoriality, also has implications for this diagram. Namely, there are natural operations which one can easily do in case **B** which should then have analogs in **C** which are very mysterious.

91.1. **Weight zero.** For all future discussions, I want to specialize to the case of "weight zero." On the motivic/Galois side, this corresponds to asking that the representations are regular and which Hodge–Tate weights which are distinct and consecutive, namely, $[0, 1, 2, \ldots, n-1]$. The hypotheses that the weights are distinct is a restrictive but crucial one — already the case when $F = \mathbf{Q}$ and the Hodge–Tate weights are $[0, 0]$ is still very much open (specifically, the case of even icosahedral representations). On the automorphic side, the weight zero assumption corresponds to demanding that the π in question contribute to the cohomology of the associated locally symmetric space with constant coefficients.

For example, if $n = 2$, then we are precisely looking at abelian varieties of GL(2) type over F (e.g. elliptic curves). This is an interesting case! We know they are modular if F is $\mathbf{Q}$, or even a quadratic extension of $\mathbf{Q}$. More generally, we know that if F is totally real, then such representations are at least *potentially* modular, that is, their restriction to some finite extension F'/F is modular. This is often good enough for many purposes. For example, it is enough to prove many cases of (some version of) **B** $\to$**A**. In this case, we have quite complete results, although still short of the optimal conjectures, especially in the case when the residual representation is reducible.

There are many other modularity lifting results generalizing those for $n = 2$, but they really involve Galois representations whose images have extra symmetry properties. In particular, they are either restricted to representations which preserve (up to scalar) some orthogonal or symplectic form, or they remain unchanged if one conjugates the representation by an outer automorphism of G_F (for example when F/F^+ is CM and one conjugates by complex conjugation). There were basically no unconditional results which applied *either* in the situation that $n \geq 2$ or that F was not completely real, and the representation did not otherwise have some restrictive condition on the global image. Our first main theorem is to prove such an unconditional result. Here is such a theorem (specialized to weight zero):

Theorem 91.2 ([ACC+23])**.** *Let F be either a CM or totally real number field, and p a prime which is unramified in F. Let*

$$\rho : G_S \to \mathrm{GL}_n(\overline{\mathbf{Q}_p})$$

be a continuous irreducible representation which is crystalline at $v|p$ with Hodge–Tate weights $[0, 1, .., n-1]$. Suppose that

1. *The residual representation $\overline{\rho}$ has suitably big image.*
2. *The residual representation is "modular" in the sense that there exists an automorphic form π_0 for $\mathrm{GL}(n)/F$ of weight zero and level prime to p such that $\overline{r}(\pi_0) = \overline{\rho}$.*

Then ρ is modular, that is, there exists an automorphic representation π of weight zero for $\mathrm{GL}(n)/F$ which is associated to ρ.

One could be more precise about what it means to have big image. In fact, I can do this by saying that it has enormous image after restriction to the composite of the Galois closure of F with the pth roots of unity. Here enormous is a technical term, of course. There is also a version of this theorem with an ordinary (rather than Fontaine–Laffaille) hypothesis (more on this next time).

Let me now give a few nice theorems which can be deduced from the theorem above:

Theorem 91.3 ([ACC+23])**.** *Let E be an elliptic curve over a CM field F. Then E is potentially modular.*

When I had a job interview at MIT in 2006, I was asked by Michael Sipser, the chair at the time, to come up with a theorem which (in a best case scenario) I would hope to prove in 10 years. I said that I wanted to prove that elliptic curves over imaginary quadratic fields were modular. (Reader, I got the job . . . then went to Northwestern.) It is very gratifying indeed that, roughly 10 years later, this result has actually been proved and that I have made some contribution towards its eventual resolution. (OK, we have potential modularity rather than modularity, but that is splitting hairs. . .). It is also amusing to note that a number of co-authors were still in high school at this time! (**Fact Check:** OK, just one . . .)

In fact, one can improve on the theorem above:

Theorem 91.4 ([ACC+23])**.** *Let E be an elliptic curve over a CM field F. Then $\mathrm{Sym}^n(E)$ is potentially modular for every n. In particular, the Sato–Tate conjecture holds for E.*

Finally, for an application of a different type, suppose that π is a weight zero cuspidal algebraic automorphic representation for $\mathrm{GL}(2)/F$. For each prime v of good

reduction, one can associate to π_v a pair of Satake parameters $\{\alpha_v, \beta_v\}$ satisfying $|\alpha_v\beta_v| = N(v)$. The Ramanujan conjecture says that one has

$$|\alpha_v| = |\beta_v| = N(v)^{1/2}.$$

An equivalent formulation is that the sum a_v of these two eigenvalues satisfies $|a_v| \leq 2N(v)^{1/2}$. We prove the following:

Theorem 91.5 ([ACC+23])**.** *: Let F be a CM field, and let π be a weight zero cuspidal algebraic automorphic representation for* $\mathrm{GL}(2)/F$*. Then the Ramanujan conjecture holds for π.*

If F is totally real, then the Ramanujan conjecture follows from Deligne's theorem. One can associate to π a motive, whose Galois representation is either $\rho = \rho(\pi)$ or $\rho^{\otimes 2}$. Then, by applying purity to these geometric representations, one deduces the result. (Of course, this was famously proved by Deligne himself in the case when $F = \mathbf{Q}$. The case of a totally real field, especially in cases where one has to go via a motive associated to $\rho^{\otimes 2}$, is due (I think) to Blasius.) This is decidedly not the way we prove this theorem. In fact, we do not know how to prove the Fontaine–Mazur conjecture for the representation ρ associated to π, even in the weak sense of showing that ρ or even $\rho^{\otimes 2}$ appears inside the cohomology of some projective variety. Instead, we prove that $\mathrm{Sym}^n\rho$ is *potentially modular*, then use the weaker convexity bound to prove the inequality:

$$|\alpha_v|^n \leq N(v)^{n/2+1/2}.$$

Taking n sufficiently large, we deduce that $|\alpha_v| \leq N(v)^{1/2}$, which (by symmetry) proves the result. Experts will recognize this as precisely Langlands' original strategy for proving Ramanujan using functoriality! In a certain sense, this is the first time that Ramanujan has been proved without a direct recourse to purity. I say "in some sense", because there is also the ambiguous case of weight one modular forms. Here the Ramanujan conjecture (which is $|a_p| \leq 2$ in this case) was deduced by Deligne and Serre as a consequence of showing that ρ has finite image so that α_v and β_v are roots of unity. On the other hand, that argument does also simultaneously imply that the representations are motivic. So our theorem produces, I believe, the only cuspidal automorphic representations for $\mathrm{GL}(n)/F$ for which we know to be tempered everywhere and yet for which we do not know are directly associated in any way to geometry.

Question 91.6. Suppose I'm sitting in my club, and Tim Gowers asks me to say what is really new about this paper. What should I say?

Answer 91.7. The distinction (say) between elliptic curves over imaginary quadratic fields and real quadratic fields, while vast, is quite subtle to explain to someone who hasn't thought about these questions. You could explain it, but the club is hardly a place to do so. Instead, go with this narrative: We generalize Wiles' modularity results for 2-dimensional representations of Q to n-dimensional representations of **Q**. If you are pressed on previous generalizations, (especially those due to Clozel–Harris–Taylor), say that Wiles is the case GL(2), Clozel–Harris–Taylor is the case $\mathrm{GSp}(2n)$, and our result is the case $\mathrm{GL}(n)$.

If you had slightly more time, and the port has not yet arrived, you might also try to explain how the underlying geometric objects involved for $\mathrm{GSp}(2n)$ are all algebraic varieties (Shimura varieties), but for $\mathrm{GL}(n)$ they involve Riemannian

manifolds which have no direct connection to algebraic geometry. Here is a good opportunity to name drop Peter Scholze, and explain how this is the first time that the methods of modularity have been combined with the new world of perfectoid spaces.

Notes 91.8. Of course there have been many updates since this paper. To mention just one, there is the work of Caraiani–Newton [CN23] on the modularity of elliptic curves over imaginary quadratic fields.

92. New results in modularity, Part II

Fri, 23 Jun 2017

This is part two of series on work in progress with Patrick Allen, Ana Caraiani, Toby Gee, David Helm, Bao Le Hung, James Newton, Peter Scholze, Richard Taylor, and Jack Thorne.

It has been almost 25 years since Wiles first announced his proof of Taniyama–Shimura, and, truthfully, variations on his method have been pretty much the only game in town since then (this paper included). In all generalizations of this argument, one needs to have some purchase on the integral structure of the automorphic forms involved, which requires that they contribute in some way to the cohomology of an arithmetic manifold (locally symmetric space). This is because it is crucial to be able to exploit the integral structure to study congruences between modular forms. Let's briefly recall Wiles' strategy. One starts out with a residual representation

$$\overline{\rho} : G_S \to \mathrm{GL}_2(\mathbf{F}_p)$$

which one assumes to be modular, that is, is the mod-p reduction of a representation associated to a modular form which is assumed to have some local properties similar to rho. One then considers a deformation ring R which captures all deformations of the residual representation which "look modular" of the right weight and level (some aspects of Serre's conjecture due to Ribet are employed here, although Skinner-Wiles came up with a base change trick to circumvent some of these difficulties). On the automorphic side, one looks at the cohomology groups $M = H^1(X, Z_p)_{\mathfrak{m}}$ of modular curves ($X = X_0(N)$) localized at a maximal ideal $\mathfrak{m}$ of the Hecke algebra $\mathbf{T}$ associated to rhobar, and proves that there is a surjective map:

$$R \to \mathbf{T}_{\mathfrak{m}}.$$

Already many deep theorems have been used to arrive at this point. To begin, one needs Galois representations associated to modular forms, but moreover, one needs to know that these representations satisfy all of the expected local-global compatibilities at the primes in S. In the case of modular forms, all of these facts were basically known before Wiles.

The next step, which lies at the heart of the Taylor–Wiles method, is to introduce certain auxiliary sets Q of carefully chosen primes, and consider the spaces $M_Q = H^1(X_1(Q), Z_p)_{\mathfrak{m}}$ which relate to spaces of modular forms of larger level. If $\mathbf{T}_Q$ is the associated Hecke algebra, and R_Q is the corresponding deformation ring in which ramification is allowed not only at S but now also at Q, there are compatible maps as follows:

$$\begin{array}{ccc} R_Q & \twoheadrightarrow & \mathbf{T}_Q \subset \mathrm{End} M_Q \\ \downarrow & & \downarrow \\ R & \twoheadrightarrow & \mathbf{T} \subset \mathrm{End} M \end{array}$$

The key point concerning how one chooses the sets Q is to ensure that, even though R_Q may get bigger, its infinitesimal tangent space does not. Hence all the R_Q are quotients of some fixed ring $R_\infty = \mathbf{Z}_p[[X_1, \dots, X_q]]$ (Here q is chosen so that $q = |Q|$.) In this process, all the rings also have an auxiliary action of a ring $S_\infty = \mathbf{Z}_p[[T_1, \dots, T_q]]$ of diamond operators, coming from the Galois group of $X_1(Q)$ over $X_0(Q)$ on the automorphic side, and the inertia groups at $\mathbf{Q}$ on the Galois side. The action of S_∞ on these modules factors through R_Q by construction, by local global compatibility. After throwing away the Galois representations almost entirely (but keeping the diamond operators), one can patch the modules M_Q/p^n for *different* sets of primes Q, and arrive at a patched module M_∞ for R_∞ and S_∞ such that:

- The module M_∞ has positive rank as an S_∞ module.
- If $\mathfrak{a}$ is the augmentation ideal of S_∞, then $R_\infty/\mathfrak{a} = R$, and $M_\infty/\mathfrak{a} = M$.

The first statement may be viewed as saying that there are "lots" of automorphic forms. On the other hand, the fact that R_∞ has the same dimension of S_∞ says that there are not "too many" Galois representations. Indeed, this friction is enough in this context to prove that M_∞ is free over R_∞, and then to deduce the same claim for M over R, from which $R = \mathbf{T}$ follows. (Already included here is an innovation due to Diamond where one deduces freeness as a consequence rather than building it in as an assumption.) The argument I have very briefly sketched above is really only a proof of modularity in the *minimal* case. The general case requires a completely separate argument to bootstrap from minimal to non-minimal level using two further ingredients: Wiles' numerical criterion, and a lower bound on the congruence ideal necessary to apply the numerical criterion, which ultimately follows from Ihara's Lemma.

The "first generation" of improvements to Wiles consisted of understanding enough integral p-adic Hodge theory to make the required arguments on the Galois side. Notable papers here include the work of Conrad–Diamond–Taylor and Breuil–Conrad–Diamond–Taylor [CDT99, BCDT01] (but let us also not forget here the contribution of the Hawk [Sav04]). Improvements along these lines continue to today, and are very closely intertwined with p-adic Langlands program and work of Breuil, Colmez, Kisin, Emerton, Paškūnas, and many others.

The "second generation" of improvements consisted of relaxing the assumption that R_∞ is smooth, by allowing instead R_∞ to have multiple components (but still of the same dimension) associated to different components in the local deformation rings at primes in S (at p and away from p). This innovation was due to Kisin, who also introduced the notion of framing to handle this.

The "third generation" of improvements (somewhat orthogonal to the second) comes from replacing 2-dimensional representations with n-dimensional representations, but still under some very restrictive assumptions on the image of rho. One key consequence of these assumptions is that the spaces of modular forms $M_Q = H^*(X_1(Q), Z_p)_{\mathfrak{m}}$ all occur inside a *single* cohomology group, which allows one to control the growth of these spaces when patching. Here one thinks of the work of

Clozel–Harris–Taylor [CHT08]. Also pertinent is that the analog of Ihara's Lemma is open for higher rank groups; Taylor came up with a technique to bypass it when proving modularity lifting theorems now known as "Ihara avoidance."

(Of course there were many other developments less directly relevant to this post, including but not limited to Skinner–Wiles and Khare–Wintenberger.)

The problem with considering general representations for $\mathrm{GL}(n)$ for $n \geq 2$, even over $\mathbf{Q}$, is that the automorphic forms are spread over a number of different cohomology groups, in fact in some range $[q_0, q_0+1, \ldots, q_0+l_0]$ for specific invariants q_0 and l_0. This manifests itself in two ways:

(1) There are not enough automorphic forms; the patched modules M_∞ will not be free over S_∞.
(2) There are not enough Galois representations: the ring R_∞ does not have the same dimension as S_∞ but rather $\dim R_\infty = \dim S_\infty - l_0$.

Of course these problems are related! My work with David Geraghty was precisely about showing how to make these problems cancel each other out. The rough idea is as follows. The cohomology groups $H^*(X_1(Q), \mathbf{Z}_p)_{\mathfrak{m}}$ which contain interesting classes in characteristic zero occur in the range $[q_0, \ldots, q_0+l_0]$. Suppose one knows this to be true integrally as well, even with coefficients over $\mathbf{F}_p$ instead of $\mathbf{Z}_p$. Then instead of patching the cohomology groups M_Q themselves, one instead patches complexes P_Q of length l_0. The result is a complex P_∞ of finite free S_∞ modules of length l_0, with an action of R_∞on the cohomology of this complex. But the only way the cohomology of this complex can be small enough to admit an action of R_∞ is if the complex is a free resolution of the patched module M_∞ of cohomology groups in the extreme final degree, and moreover it also implies that M_∞ is big enough as in Wiles' original argument to give an $R = \mathbf{T}$ theorem. Note that it is crucial here that one work with the torsion in integral cohomology. It is quite possible that, at all auxiliary levels Q, there are no more automorphic forms at level Q than are were at level one. (This can only happen for $l_0 > 0$, and the idea that torsion should be a suitable replacement is the moral of my paper with Barry Mazur.) These argument is also compatible with the improvements to the method including Taylor's "Ihara Avoidance" argument.

On the other hand, there is a big problem. This argument required many inputs which were completely unknown at the time we worked this out, so our results were very conditional. To be precise, our results were conditional on the following desiderata:

(1) The existence of Galois representations on Hecke rings $\mathbf{T}$ which acted as endomorphisms of $H^*(X, \mathbf{Z}/p^n\mathbf{Z})$ for locally symmetric spaces X associated to $\mathrm{GL}(n)/F$.
(2) The stronger claim that the Galois representations constructed in part one satisfied the correct "local-global" compatibility statements for all v in S (including $v|p$).
(3) The vanishing of the cohomology groups $H^i(X, \mathbf{Z}/p^n\mathbf{Z})_{\mathfrak{m}}$ outside the range $i \in [q_0, \ldots, q_0 + l_0]$, for a non-Eisenstein ideal $\mathfrak{m}$.

A different approach to some of these questions (which Matt and I discussed, see § 26) involves first passing to completed cohomology, where one expects (or hopes!) that all the cohomology groups except in degree q_0 should vanish after localization at a non-maximal ideal.

The first big breakthrough was the result of Scholze, who proved part 1 above, at least up to issues concerning a nilpotent ideal (this was discussed § 23). Another innovation appeared in Khare–Thorne, where it was observed that one can sometimes drop the third assumption under the strong condition that there existed global automorphic forms with the exact level structure corresponding to the original representation. (Unfortunately, in the $l_0 > 0$ setting, there is no way to produce such forms.)

So this is roughly where we stood in 2016. The key new ingredient which led to this project was the new result of Caraiani and Scholze [CS24] proving vanishing theorems for the cohomology of *non-compact* Shimura varieties in degrees above the middle dimension (localized at $\mathfrak{m}$) under the assumption of certain genericity hypotheses on $\mathfrak{m}$. Since the cohomology of the boundary (for suitably chosen Shimura varieties) is precisely related to the cohomology of arithmetic locally symmetric spaces for $\mathrm{GL}(n)$ over CM fields, this allowed for the first time a new construction of the Galois representations for $\mathrm{GL}(n)$ which directly related them to the Galois representations coming from geometry. (I say "directly related," but perhaps I mean simply more direct than Peter's original construction.) In particular, it was clear to Caraiani and Scholze that this result should have implications for the required local-global compatibility result above. Meanwhile, the IAS had just started a new series of workshops on emerging topics. I guess that Richard must have had conversations with Ana about her work with Peter, which led them to choosing this as the theme, namely:

> Ana Caraiani and Peter Scholze are hopeful of extending the methods of their joint paper (see [CS17]) to non-compact Shimura varieties. This would give a new way to attack local-global compatibility at p for some of the Galois representations Scholze attached to torsion classes in the cohomology of arithmetic locally symmetric spaces. The aim of this workshop will be to understand how much local-global compatibility can be proved and to explore the consequences of this, particularly for modularity questions.

So now (1) was available, there was an approach to (2), and a technique for avoiding (3). One issue with the Khare-Thorne trick, however, was that it involved localizing at some prime ideal of characteristic zero, and so did not interact so well with Ihara Avoidance, which was crucial for any sort of applicable theorem. Here's the subtlety, which can be described even in the case when $l_0 = 0$. The usual Ihara avoidance game is to compare deformation rings R and R' at Steinberg level and ramified principal series level respectively (after making a base change to ensure that the prime v at the relevant prime q satisfies $N(v) = 1 \bmod p$). Let M and M' be the corresponding modules. One has that $M/p = M'/p$ and $R/p = R'/p$. Suppose, however, that M behaved perfectly as expected, so that M_∞ was free (even of rank one say) over S_∞ and free over R_∞. What could happen, if one doesn't have vanishing of cohomology outside a single degree, is that $M'_\infty/p = M_\infty/p$ is free over S_∞/p, but that M'_∞ is the cohomology of a non-trivial complex $S_\infty \to S_\infty$ given by multiplication by p. So M'_∞ is trivial in characteristic zero, even though $M'_\infty/p = M_\infty/p$. So this is a problem. But it is exactly a problem which was resolved during the workshop. The point, very loosely speaking, is that even though the complexes "S'_∞" and "$[p] : S_\infty \to S_\infty$" have the same H^0 after reducing modulo p and taking cohomology, their intersection with S_∞/p are quite different

on the derived level, so if one can formulate a version of *derived* Ihara avoidance, then one is in good shape.

So what remained? First, there were a number of technical issues, some of which could be dealt with individually, and one had to make sure that all the fixes were compatible. For example, it is straightforward to modify the original strategy in my paper with David to handle the issue of only having Galois representations up to nilpotence ideals of fixed nilpotence, but one had to make sure this would not interfere with the more subtle derived Ihara avoidance type arguments. Relevant here was the work of Newton and Thorne which placed some of the arguments with complexes more naturally in the derived category. Second, there was the issue of really proving local-global compatibility from the new results of Caraiani-Scholze. A particularly interesting case here was the ordinary case. The rough problem one has to deal with here is deducing that rho is ordinary from knowing that $\rho \oplus \rho^\vee$ is ordinary. But be careful — the latter representation is reducible and so really a pseudo-representation — so it's not even clear what ordinary this means (though see work of Wake and Wang Erickson, as well as of my student Joel Specter). It turns out that some interesting and subtle things turn up in this case which were found by the "team" of people who wrote up this section. (Although we achieved quite a lot in a week, there were obviously a list of details to be worked out, and we divided ourselves up into certain groups to work on each part of the paper.) But I think we were fairly confident at this point that everything would work out. What was my role in the writing up process you ask? I was selected as the ENFORCER, who goes around harassing everybody else to work and write up their sections of the paper while sipping on Champagne. Presumably I was less selected for my organizational skills and more for my ability to tell Richard Taylor what to do (see [Tay12]).

So there we have it! It was clear even during the workshop that some improvements to our arguments were possible, but since the paper is already going to be quite long, we did not try to be completely comprehensive. I expect a number of improvements will follow shortly. I would not be surprised to see in a few years a modularity result for regular weight compatible systems over CM fields which are as complete as the ones (say) in [BLGGT14].

Comment 92.1 (David Loeffler)**.** Just out of curiosity, how important is the "weight 0" hypothesis for what you're doing? Is there some specific step that definitely breaks down when the Hodge–Tate weights are distinct but non-consecutive; or is that restriction just imposed to keep the project manageable?

Comment 92.2 (Persiflage)**.** The problem is that in potential automorphy results you are always comparing to some geometric family (in this case, the Dwork family studied in [HSBT10]), and such families have consecutive Hodge–Tate weights. You can get around this with Hida theory (but this only works if you know that you have lots of ordinary primes, which we don't know in any generality), or the Harris tensor product trick — but unfortunately (and perhaps slightly unexpectedly) there are technical problems with getting this trick to work for CM fields.

So something new is needed to change weight in this setting; to the best of my knowledge this hasn't been worked out yet, but people have promising ideas. If I was a betting man, I would put money on [BLGGT14]-style theorems being proved over CM fields in the next couple of years, but at the moment, Ramanujan is open in weight ≥ 2.

Notes 92.3. We are still some distance from [BLGGT14]-style results, but we do know the Ramanujan conjecture in (many) higher weights by [BCG+23b]. See also [Mat23].

93. Schaefer and Stubley on class groups

Sun, 29 Oct 2017

I talked 88 about work of Wake and Wang-Erickson on deformations of Eisenstein residual representations. In that post, I also mentioned a paper of Emmanuel Lecouturier who has also proved some very interesting theorems. Today, I wanted to talk about some complementary results by my student Eric Stubley in collaboration with Karl Schaefer (a student of Matthew Emerton) (see [SS19]). To duplicate slightly from that previous post, recall that Matt and I proved the following:

Theorem 93.1. *Let $p \geq 3$ be prime, and let $N \equiv 1 \bmod p$ be prime. If the rank of the cuspidal Hecke algebra of level $\Gamma_0(N)$ localized at the Eisenstein prime is greater than one, then*

$$K = \mathbf{Q}(N^{1/p})$$

has non-cyclic p-class group.

Using work of Merel, one can dispense with the discussion of Hecke algebras and instead give an equivalent reformulation of the first condition, namely, $e \geq 1$ if and only if M_1 is a p-th power, where

$$M_1 = \prod_{k=1}^{p-1} (Mk)!^k \in \mathbf{F}_N^\times, \qquad M = \frac{N-1}{p}$$

We followed up this result with the comment:

> We expect (based on the numerical evidence) that the condition that the class group of K has p-rank [at least] two is equivalent to the existence of an appropriate group scheme, and thus to [the rank being greater than one].

As noted previously, there are counter-examples, already for $p = 7$ and $N = 337$. However, there was still clearly *some* relationship between these quantities beyond the one-way implication above. In particular, the numerical evidence still stubbornly supported the hope that the converse *may* indeed be true for $p = 5$. This is the first theorem that Schaefer and Stubley prove. More precisely, they *completely determine* the rank of the class group of $\mathbf{Q}(N^{1/5})$ for primes N which are 1 mod 5.

Theorem 93.2 (Schaefer–Stubley)**.** *Let $N \equiv 1 \mod 5$ be prime. Then the 5-rank r_K of the class group of $K = \mathbf{Q}(N^{1/5})$ is either 1, 2, or 3 Moreover:*

1. *$r_K = 1$ if and only if the Merel invariant M_1 is not a perfect 5th power.*
2. *$r_K = 2$ if and only if M_1 is a perfect 5th power, and $\alpha = \dfrac{\sqrt{5}-1}{2}$ is not a perfect 5th power modulo N.*
3. *$r_K = 3$ if and only if M_1 and α are both 5th powers modulo N.*

This also answers a conjecture of Lecouturier. Their argument greatly clarified (to me) the exact relationship between the class group of K and a number of other related quantities in this picture. To recall, a third reformulation of

whether the Hecke algebra has non-trivial deformations can be given (as in Wake–Wang-Erickson) by whether a certain pairing between specific classes b and c_{-1} in $H^1_{Np}(\mathbf{Q}, \epsilon)$ and $H^1_{Np}(\mathbf{Q}, \epsilon^{-1})$ vanish or not. The point is that the vanishing of a cup product ensures the existence of an extension

$$\begin{pmatrix} 1 & b & c_0 \\ 0 & \epsilon^{-1} & c_{-1} \\ 0 & 0 & 1 \end{pmatrix}$$

and one can show (after some massaging) that c_0 gives rise to something in the p-class group of K. Conversely, if one starts with a class in the p-class group of K, and then takes the Galois closure over $\mathbf{Q}$, then (sometimes) one arrives with a Galois extension $M/\mathbf{Q}$ with a Galois representation to GL(3) of the above form. The problem is, in other circumstances, one arrives at a representation which has a much larger Galois group and a map to the Borel subgroup in higher dimension, which looks something like this:

$$\begin{pmatrix} 1 & \epsilon^{-1} \cdot b & \epsilon^{-2} \cdot b^2/2 & \epsilon^{-3} \cdot b^3/6 & & \dots & c_0 \\ 0 & \epsilon^{-1} & \epsilon^{-2} \cdot b & \epsilon^{-3} \cdot b^2/2 & & \dots & c_{-1} \\ & & \ddots & & & & \\ \dots & & & & \epsilon^{1-m} & \epsilon^{-m} \cdot b & c_{1-m} \\ \dots & & & & & \epsilon^{-m} & c_{-m} \\ \dots & & & & & & 1 \end{pmatrix}$$

Suppose one now tries to construct a representation of this form in order to find a non-trivial class in the p-class group of K. First, one can start by finding a suitable class $c_{-m} \in H^1_{Np}(\mathbf{Q}, \epsilon^{-m})$ which cups trivially with lb. The vanishing of a generalized Merel invariant (under a regularity hypothesis) is exactly what guarantees the existence of such a suitable class c_{-m}, at least when m is odd. However, one is then faced with an increasing sequence of obstruction problems in order to climb the ladder and get all the way to the full representation of the form above. Here one has to deal with not only cup products, but also (implicitly) higher Massey products. Ultimately, the relation between the quantity r_K and the deformation rings of Hecke algebras is most precise only when $p = 5$. It turns out that there is still something one can say for $p = 7$, however. Consider the higher Merel invariant

$$M_n = \prod_{k=1}^{p-1} (Mk)!^{k^n} \in \mathbf{F}_N^\times, \qquad M = \frac{N-1}{p}$$

for odd values of n. Suppose that p is a regular prime. One can show that if $r_K \geq 2$, then at least *one* of these quantities M_n is a perfect pth power for an odd $n \leq p-4$. When $p = 5$, this is a weaker version of the theorem above. So an optimistic variation on the conjecture above is that $r_K \geq 2$ if and only if M_n is a perfect pth power of for at least *one* odd $n \leq p-4$. The description of the relationship between these classes (which also come up in Lecouturier, they arise via an explicit analysis of Gauss sums and Stickelberger's theorem) suggests that this conjecture is too optimistic in general, and indeed there are counter-examples for $p = 11$. But, Schaefer and Stubley do prove the following:

Theorem 93.3 (Schaefer, Stubley)**.** *Let $p = 7$, and let $N = 1 \bmod p$ be prime. Then the 7-class group of $K = \mathbf{Q}(N^{1/p})$ has rank $r_K \geq 2$ if and only if either M_1 or M_3 is a perfect 7th power modulo N.*

For example, consider the previous "counter-example" for $N = 337$ and $p = 7$. Here the non-trivial class group is explained by the fact that M_3 is a perfect 7th power modulo N.

One thing I especially like about this result is that there are three groups of people (Wake–Wang-Erickson, Lecouturier, and Schaefer–Stubley) are all working around a similar problem, but their results are complementary to each other. I believe that all five people will be at the upcoming IAS workshop, so I hope to hear more about this then.

94. Mathieu magic

Tue, 17 Oct 2017

I previously mentioned (in Comment 75.5) that I once made (in a footnote) the false claim that for a 11-dimensional representation V of the Mathieu group M_{12}, the 120 dimensional representation $\mathrm{Ad}^0(V)$ was irreducible. I had wanted to write down representations W of large dimension n such that $\mathrm{Ad}^0(W)$ of dimension n^2-1 was irreducible. In the comments, Emmanuel Kowalski pointed to a paper of Katz where he discusses actual examples (including the 1333 dimensional representation of the Janko group J_4). On the other hand, I recently learned from Liubomir Chirac's thesis. that it's an open problem to determine whether there exists such a representation for all n (although he does write down infinitely many examples in prime power dimension). Chirac's thesis also lead me to the paper of Magaard, Malle, and Tiep, who do classify all such examples for (central extensions of) simple groups. Turns out that I *could* have used M_{12} after all, or rather the 10-dimensional representation of the double cover $2.M_{12}$, which *does* have the required property (the 99-dimensional representation factors through M_{12}, naturally).

One reason (amongst many) that (either of the) 11-dimensional representations V of M_{12} do not have $\mathrm{Ad}^0(V)$ irreducible is that they are self-dual (oops). On the other hand, if you eyeball the character table, you will find that there *is* an irreducible representation W of dimension 12. Moreover, let me write down the characters of $[V \otimes V^*] - [1]$ and $[W]$:

$$\begin{aligned} [V \otimes V^*] - [1] : &\quad 120, 0, \quad 8, 3, 0, 0, 8, 0, 0, -1, 0, 0, 0, -1, -1; \\ [W] : &\quad 120, 0, -8, 3, 0, 0, 0, 0, 0, \quad 1, 0, 0, 0, -1, -1. \end{aligned}$$

These seem surprisingly close to me! So now the question is, as one ranges over (some class perhaps all) finite groups G, what is the minimum number of conjugacy classes for which

$$\chi = [V \otimes V^*] - [1] - [W]$$

can be non-zero for irreducible V and W, assuming that it is non-zero? Since V is irreducible, by Schur's Lemma, this virtual representation is orthogonal to [1] (unless $[W] = [1]$ which would be silly). So $\langle \chi, 1 \rangle = 0$, which certainly implies that there must be at least *two* non-zero entries of opposite signs. I don't see any immediate soft argument which pushes that bound to 3. I admit, this is a slightly silly question. But still, a beer to anyone who proves the example above is either optimal or comes up with an example with only two non-zero terms. (To avoid silliness, say that the dimension of V has to be at least 5.) More precisely:

Problem 94.1. Classify all pairs (G, V, W) of a finite group G and irreducible representations V and W with V faithful such that $\chi = [V \otimes V^*] - [1] - [W]$ is zero on all but at most 2 conjugacy classes but χ itself is non-zero.

The characters in the example above look strikingly similar to me, and it does make we wonder if there is any reason for why they are so close. Perhaps if I knew more about groups, I could feel more confident in just chalking up the resemblance above to a law of small numbers.

Probably a more sensible question is to ask for how small the number of non-zero entries of $[V] - [W]$ can be for two distinct irreducibles. That question has surely been studied!

95. Abelian surfaces are potentially modular

Sat, 11 Nov 2017

Today I wanted (in the spirit of this post) to report on some new work in progress with George Boxer, Toby Gee, and Vincent Pilloni. (see [BCGP21])

Recall that, for a smooth projective variety X over a number field F unramified outside a finite set of primes S, one may write down a global Hasse-Weil zeta function:

$$\zeta_{X,S}(s) = \prod \frac{1}{1 - N(x)^{-s}}$$

where the product runs over closed points of a smooth integral model. From the Weil conjectures, the function $\zeta_{X,S}(s)$ is absolutely convergent for s with real part at least $1 + m/2$, where $m = \dim(X)$. One has the following well-known conjecture:

Conjecture 95.1 (Hasse–Weil Conjecture)**.** *The function $\zeta_{X,S}(s)$ extends to a meromorphic function on the complex plane. Moreover, there exists a rational number A, a collection of polynomials $P_v(T)$ for v dividing S, and infinite Gamma factors $\Gamma_v(s)$ such that*

$$\xi_X(s) = \zeta_{X,S}(s) \cdot A^{s/2} \cdot \prod_{v|\infty} \Gamma_v(s) \cdot \prod_{v|S} \frac{1}{P_v(N(v)^{-s})}$$

satisfies the functional equation $\xi_X(s) = w \cdot \xi_X(m + 1 - s)$ with $w = \pm 1$.

Naturally, one can be more precise about the conductor and the factors at the bad primes. In the special case when $F = \mathbf{Q}$ and X is a point, then $\zeta_{X,S}(s)$ is essentially the Riemann zeta function, and the conjecture follows from Riemann's proof of the functional equation. If F is a general number field but X is still a point, then $\zeta_{X,S}(s)$ is (up to some missing Euler factors at S) the Dedekind zeta function $\zeta_F(s)$ of F, and the conjecture is a theorem of Hecke. If X is a curve of genus zero over F, then $\zeta_{X,S}(s)$ is $\zeta_F(s)\zeta_F(s-1)$, and one can reduce to the previous case. More generally, by combining Hecke's results with an argument of Artin and Brauer about writing a representation as a virtual sum of induced characters from solvable (Brauer elementary) subgroups, one can prove the result for any X for which the l-adic cohomology groups are potentially abelian. This class of varieties includes those for which all the cohomology of X is generated by algebraic cycles.

For a long time, not much was known beyond these special cases. But that is not to say there was not a lot of progress, particularly in the conjectural understanding of what this conjecture really was about. The first huge step was the discovery and

formulation of the Taniyama-Shimura conjecture, and the related converse theorems of Weil. The second was the fundamental work of Langlands which cast the entire problem in the (correct) setting of automorphic forms. In this context, the Hasse-Weil zeta functions of modular curves were directly lined to the L-functions of classical weight 2 modular curves. More generally, the Hasse-Weil zeta functions of all Shimura varieties (such as Picard modular surfaces) should be linked (via the trace formula and conjectures of Langlands and Kottwitz) to the L-functions of automorphic representations. On the other hand, these examples are directly linked to the theory of automorphic forms, so the fact that their Hasse-Weil zeta functions are automorphic, while still very important, is not necessarily evidence for the general case. In particular, there was no real strategy for taking a variety that occurred "in nature" and saying anything non-trivial about the Hasse-Weil zeta function beyond the fact it converged for real part greater than $1+m/2$, which itself requires the full strength of the Weil conjectures.

The first genuinely new example arrived in the work of Wiles (extended by others, including Breuil-Conrad-Diamond-Taylor), who proved that elliptic curves E/Q were modular. An immediate consequence of this theorem is that Hasse-Weil conjecture holds for elliptic curves over $\mathbf{Q}$. Taylor's subsequent work on *potentially modularity,* while not enough to prove modularity of all elliptic curves over all totally real fields, was still strong enough to allow him to deduce the Hasse-Weil conjecture for any elliptic curve over a totally real field. You might ask what have been the developments since these results. After all, the methods of modularity have been a very intense subject of study over the past 25 years. One problem is that these methods have been extremely reliant on a regularity assumption on the corresponding motives. One nice example of a regular motive is the symmetric power of any elliptic curve. On the other hand, if one takes a curve X over a number field, then $h^{1,0} = h^{0,1} = g$, and the corresponding motive is regular only for $g = 0$ or 1. The biggest progress in automorphy of non-regular motives has actually come in the form of new cases of the Artin conjecture — first by Buzzard–Taylor and Buzzard, then in the proof of Serre's conjecture by Khare–Wintenberger over Q, and more recently in subsequent results by a number of people (Kassaei, Sasaki, Pilloni, Stroh, Tian) over totally real fields. But these results provide no new cases of the Hasse-Weil conjecture, since the Artin cases were already known in this setting by Brauer. (It should be said, however, that the generalized modularity conjecture is now considered more fundamental than the Hasse-Weil conjecture.) There are a few other examples of Hasse-Weil one can prove by using various forms of functoriality to get non-regular motives from regular ones, for example, by using the Arthur-Clozel theory of base change, or by Rankin-Selberg. We succeed, however, in establishing the conjecture for a class of motives which is non-regular in an essential way. The first corollary of our main result is as follows:

Theorem 95.2 ([BCGP21])**.** *Let X be a genus two curve over a totally real field. The Hasse-Weil conjecture holds for X.*

It will be no surprise to the experts that we deduce the theorem above from the following:

Theorem 95.3 ([BCGP21])**.** *Let A be an abelian surface over a totally real field F. Then A is potentially modular.*

In the case when A has trivial endomorphisms (the most interesting case), this theorem was only known for a finite number of examples over $\mathbf{Q}$. In each of those cases, the stronger statement that A is modular was proved by first explicitly computing the corresponding low weight Siegel modular form. For example, the team of Brumer–Pacetti–Tornaría–Poor–Voight–Yuen prove [BPP+19] that the abelian surfaces of conductors 277, 353, and 587 are all modular, using (on the Galois side) the Faltings-Serre method, and (on the automorphic side) some really quite subtle computational methods developed by Poor and Yuen. A paper of Berger–Klosin [BK20] handles a case of conductor 731 by a related method that replaces the Faltings–Serre argument by an analysis of certain reducible deformation rings.

The arguments of our paper are a little difficult to summarize for the non-expert. But George Boxer did a very nice job presenting an overview of the main ideas, and you can watch his lecture online (posted below, together with Vincent's lecture on higher Hida theory). The three sentence version of our approach is as follows. There was a program initiated by Tilouine to generalize the Buzzard–Taylor method to GSp(4), which ran into technical problems related to the fact that Siegel modular forms are not directly reconstructible from their Hecke eigenvalues. There was a second approach coming from my work with David Geraghty, which used instead a variation of the Taylor–Wiles method; this ran into technical problems related to the difficulty of studying torsion in the higher coherent cohomology of Shimura varieties. Our method is a synthesis of these two approaches using Higher Hida theory as recently developed by Pilloni. Let me instead address one or two questions here that GB did not get around to in his talk:

Question 95.4. What is the overlap of this result with [ACC+23]?

Answer 95.5. Perhaps surprisingly, not so much. For example, our results are independent of the arguments of Scholze (and now Caraiani–Scholze [CS24]) on constructing Galois representations to torsion classes in Betti cohomology. We do give a new proof of the result that elliptic curves over CM fields are potentially modular, but that is the maximal point of intersection. In contrast, we don't prove that higher symmetric powers of elliptic curves are modular. We do, however, prove potentially modularity of all elliptic curves over all quadratic extensions of totally real fields with mixed signature, like $\mathbf{Q}(2^{1/4})$. The common theme is (not surprisingly) the Taylor–Wiles method (modified using the ideas in my paper with David Geraghty).

Question 95.6. What's new in this paper which allows you to make progress on this problem?

Answer 95.7. George explains this well in his lecture. But let me at least stress this point: Vincent Pilloni's recent work on higher Hida Theory (see [Pil20]) was absolutely crucial. Boxer, Gee, and I were working on questions related to modularity in the symplectic case, but when Pilloni's paper first came out, we immediately dropped what we were doing and started working (very soon with Pilloni) on this problem. If you have read the Calegari–Geraghty [CG20] paper on GSp(4) and are not an author of the current paper (hi David!), and you look through our manuscript (currently a little over 200 pages and [optimistically?!] ready by the end of the year), then you also recognize other key technical points, including a more philosophically satisfactory doubling argument and Ihara avoidance in the symplectic case, amongst other things.

Question 95.8. So what about modularity?

Answer 95.9. Of course, we deduce our potential modularity result from a modularity lifting theorem. The reason we cannot deduce that Abelian surfaces are all modular, even assuming for example that they are ordinary at 3 with big residual image, is that Serre's conjecture is not so easy. Not only is $\mathrm{GSp}_4(\mathbf{F}_3)$ not a solvable group, but — and this is more problematic — Artin representations do not contribute to the coherent cohomology of Shimura varieties in any setting other than holomorphic modular forms of weight one. Still, there are some sources of residually modular representations, including the representations induced from totally real quadratic extensions (for small primes, at least). We do, however, prove the following (which GB forgot to mention in his talk, so I bring up here):

Proposition 95.10. *There exist infinitely modular abelian surfaces $A/\mathbf{Q}$ (up to twist) with $\mathrm{End}_{\overline{\mathbf{Q}}}(A) = \mathbf{Z}$.*

This is proved in an amusing way. It suffices to show that, given a residual representation

$$\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GSp}_4(\mathbf{F}_3)$$

with cyclotomic similitude character (or rather inverse cyclotomic character with our cohomological normalizations) which has big enough image and is modular (plus some other technical conditions, including ordinary and p-distinguished) that it comes from infinitely many abelian surfaces over $\mathbf{Q}$, and then to prove the modularity of those surfaces using the residual modularity of $\overline{\rho}$. This immediately reduces to the question of finding rational points on some twist of the moduli space $\mathcal{A}_2(3)$. And this space is rational! Moreover, it turns out to be a very famous hypersurface much studied in the literature — it is the Burkhardt Quartic. Now unfortunately — unlike for curves — it's not so obvious to determine whether a twist of a higher dimensional rational variety is rational or not. The problem is that the twisting is coming from an action by $\mathrm{Sp}_4(\mathbf{F}_3)$, and that action is not compatible with the birational map to $\mathbf{P}^3$, so the resulting twist is not a priori a Severi-Brauer variety. However, something quite pleasant happens — there is a degree six cover

$$\mathcal{A}_2^-(3) \stackrel{6:1}{\to} \mathcal{A}_2(3)$$

(coming from a choice of odd theta characteristic) which is not only still rational, but now rational in an *equivariant* way. So now one can proceed following the argument of Shepherd-Barron and Taylor [SBT97] in their earlier paper on mod-2 and mod-5 Galois representations.

Question 95.11. What about curves of genus $g > 2$?

Answer 95.12. Over $\mathbf{Q}$, there is a tetrachotomy corresponding to the cases $g = 0$, $g = 1$, $g = 2$, and $g > 2$. The $g = 0$ case goes back to the work of Riemann. The key point in the $g = 1$ case (where the relevant objects are modular forms of weight two) is that there are two very natural ways to study these objects. The first (and more classical) way to think about a modular form is as a holomorphic function on the upper half plane which satisfies specific transformation properties under the action of a finite index subgroup of $\mathrm{SL}_2(\mathbf{Z})$. This gives a direct relationship between modular forms and the coherent cohomology of modular curves; namely, cuspidal modular forms of weight two and level $\Gamma_0(N)$ are exactly holomorphic differentials on the modular curve $X_0(N)$. On the other hand, there is a second interpretation of

modular forms of weight two in terms of the *Betti* (or etale or de Rham) cohomology of the modular curve. A direct way to see this is that holomorphic differentials can be thought of as smooth differentials, and these satisfy a duality with the homology group $H_1(X_0(N), \mathbf{R})$ by integrating a differential along a loop. And it is the second description (in terms of etale cohomology) which is vital for studying the arithmetic of modular forms. When $g = 2$, there is still a description of the relevant forms in terms of coherent cohomology of Shimura varieties (now Siegel 3-folds), but there is no longer any direct link between these coherent cohomology groups and etale cohomology. Finally, when $g > 2$, even the relationship with coherent cohomology disappears — the relevant automorphic objects have some description in terms of differential equations on locally symmetric spaces, but there is no longer any way to get a handle on these spaces. For those that know about Maass forms, the situation for $g > 2$ is at least as hard (probably much harder) than the notorious open problem of constructing Galois representations associated to Maass forms of eigenvalue 1/4. In other words, it's probably very hard! (Of course, there are special cases in higher genus when the Jacobian of the curve admits extra endomorphisms which can be handled by current methods.)

Finally, as promised, one can find the videos here and here.

Notes 95.13. The "over 200 pages" became 349 pages, I guess. There is also progress on the question of modularity in genus $g = 2$, this is work in progress of the same four authors [BCGP24].

96. The ABC conjecture has (still) not been proved

Mon, 18 Dec 2017

The ABC conjecture has (still) not been proved.

Five years ago, Cathy O'Neil laid out a perfectly cogent case for why the (at that point recent) claims by Shinichi Mochizuki should not (yet) be regarded as constituting a proof of the ABC conjecture. I have nothing further to add on the sociological aspects of mathematics discussed in that post, but I just wanted to report on how the situation looks to professional number theorists today. The answer? It is a complete disaster.

This post is not about making epistemological claims about the truth or otherwise of Mochizuki's arguments. To take an extreme example, if Mochizuki had carved his argument on slate in Linear A and then dropped it into the Mariana Trench, then there would be little doubt that asking about the veracity of the argument would be beside the point. The reality, however, is that this description is not so far from the truth.

Each time I hear of an analysis of Mochizuki's papers by an expert (off the record) the report is disturbingly familiar: vast fields of trivialities followed by an enormous cliff of unjustified conclusions. The defense of Mochizuki usually rests on the following point: The mathematics coming out of the Grothendieck school followed a similar pattern, and that has proved to be a cornerstone of modern mathematics. There is the following anecdote that goes as follows:

> The author hears the following two stories: Once Grothendieck said that there were two ways of cracking a nutshell. One way was to crack it in one breath by using a nutcracker. Another way was to

> soak it in a large amount of water, to soak, to soak, and to soak, then it cracked by itself. Grothendieck's mathematics is the latter one.

While rhetorically expedient, the comparison between Mochizuki and Grothendieck is a poor one. Yes, the Grothendieck revolution upended mathematics during the 1960's "from the ground up." But the ideas coming out of IHES immediately spread around the world, to the seminars of Paris, Princeton, Moscow, Harvard/MIT, Bonn, the Netherlands, etc. Ultimately, the success of the Grothendieck school is not measured in the theorems coming out of IHES in the '60s but in how the ideas completely changed how everyone in the subject (and surrounding subjects) thought about algebraic geometry.

This is not a complaint about idiosyncrasy or about failing to play by the rules of the "system." Perelman more directly repudiated the conventions of academia by simply posting his papers to the arXiv and then walking away. (**Edit:** Perelman **did** go on an extensive lecture tour and made himself available to other experts, although he never submitted his papers.) But in the end, in mathematics, ideas always win. And people were able to read Perelman's papers and find that the ideas were all there (and multiple groups of people released complete accounts of all the details which were also published within five years). Usually when there is a breakthrough in mathematics, there is an explosion of new activity when other mathematicians are able to exploit the new ideas to prove new theorems, usually in directions not anticipated by the original discoverer(s). This has manifestly not been the case for ABC, and this fact alone is one of the most compelling reasons why people are suspicious.

The fact that these papers have apparently now been accepted by the Publications of the RIMS (a journal where Mochizuki himself is the managing editor, not necessary itself a red flag but poor optics none the less) really doesn't change the situation as far as giving anyone a reason to accept the proof. If anything, the value of the referee process is not merely in getting some reasonable confidence in the correctness of a paper (not absolute certainty; errors do occur in published papers, usually of a minor sort that can be either instantly fixed by any knowledgeable reader or sometimes with an erratum, and more rarely requiring a retraction). Namely, just as importantly, it forces the author(s) to bring the clarity of the writing up to a reasonable standard for professionals to read it (so they don't need to take the same time duration that was required for the referees, amongst other things). This latter aspect has been a complete failure, calling into question both the quality of the referee work that was done and the judgement of the editorial board at PRIMS to permit papers in such an unacceptable and widely recognized state of opaqueness to be published. We now have the ridiculous situation where ABC is a theorem in Kyoto but a conjecture everywhere else. (**edit:** a Japanese reader has clarified to me that the newspaper articles do not definitively say that the papers have been accepted, but rather the wording is something along the lines of "it is planned that PRIMS will accept the paper," whatever that means. This makes no change to the substance of this post, except that, while there is still a chance the papers will not be accepted in their current form, I retract my criticism of the PRIMS editorial board.)

So why has this state persisted so long? I think I can identify three basic reasons. The first is that mathematicians are often very careful (cue the joke about a

sheep *at least one side of which is black*). Mathematicians are very loath to claim that there is a problem with Mochizuki's argument because they can't point to any definitive error. So they tend to be very circumspect (reasonably enough) about making any claims to the contrary. We are usually trained as mathematicians to consider an inability to understand an argument as a failure on our part. Second, whenever extraordinary claims are made in mathematics, the initial reaction takes into account the past work of the author. In this case, Shinichi Mochizuki was someone who commanded significant respect and was considered by many who knew him to be very smart. It's true (as in the recent case of Yitang Zhang) that an unknown person can claim to have proved an important result and be taken seriously, but if a similarly obscure mathematician had released 1000 pages of mathematics written in the style of Mochizuki's papers, they would have been immediately dismissed. Finally, in contrast to the first two points, there are people willing to come out publicly and proclaim that all is well, and that the doubters just haven't put in the necessary work to understand the foundations of inter-universal geometry. I'm not interested in speculating about the reasons they might be doing so. But the idea that several hundred hours at least would be required even to scratch the beginnings of the theory is either utter rubbish, or so far beyond the usual experience of how things work that it would be unique not only in mathematics, but in all of science itself.

So where to from here? There are a number of possibilities. One is that someone who examines the papers in depth is able to grasp a key idea, come up with a major simplification, and transform the subject by making it accessible. This was the dream scenario after the release of the paper, but it becomes less and less likely by the day (and year). But it is still possible that this could happen. The flip side of this is that someone could find a serious error, which would also resolve the situation in the opposite way. A third possibility is that we have (roughly) the status quo: no *coup de grâce* is found to kill off the approach, but at the same time the consensus remains that people can't understand the key ideas. (I should say that whether the papers are accepted or not in a journal is pretty much irrelevant here; it's not good enough for people to attest that they have read the argument and it is fine, someone has to be able to explain it.) In this case, the mathematical community moves on and then, whether it be a year, a decade, or a century, when someone ultimately does prove ABC, one can go back and compare to see if (in the end) the ideas were really there after all.

Comment 96.1 (Jordan Ellenberg)**.** Thanks for posting this, Frank.

Comment 96.2 (William Stein)**.** Thanks for posting this Frank! A fourth possibility is that the fairly strong form of ABC that Mochizuki claims to have proved turns out to not be true ...

Comment 96.3 (Andrew Sutherland)**.** I haven't been following this closely — can one extract from Mochizuki's arguments an explicit effective inequality (not an asymptotic) that is falsifiable?

Comment 96.4 (Persiflage)**.** I believe that this paper by Vesselin Dimitrov shows that that one can formally extract such a quantity, although completely explicit bounds are not found in that paper.

Comment 96.5 (Matthew Emerton)**.** Great post! It does indeed do an excellent job of summarizing the situation from the perspective of professional number theorists.

Comment 96.6 (Toby Gee)**.** Than you very much for posting this!

Comment 96.7 (Akshay Venkatesh)**.** I couldn't agree more.

Comment 96.8 (Terry Tao)**.** Thanks for this. I do not have the expertise to have an informed first-hand opinion on Mochizuki's work, but on comparing this story with the work of Perelman and Yitang Zhang you mentioned that I am much more familiar with, one striking difference to me has been the presence of short "proof of concept" statements in the latter but not in the former, by which I mean ways in which the methods in the papers in question can be used relatively quickly to obtain new non-trivial results of interest (or even a new proof of an existing non-trivial result) in an existing field. In the case of Perelman's work, already by the fifth page of the first paper Perelman had a novel interpretation of Ricci flow as a gradient flow which looked very promising, and by the seventh page he had used this interpretation to establish a "no breathers" theorem for the Ricci flow that, while being far short of what was needed to finish off the Poincare conjecture, was already a new and interesting result, and I think was one of the reasons why experts in the field were immediately convinced that there was lots of good stuff in these papers. Yitang Zhang's 54 page paper spends more time on material that is standard to the experts (in particular following the tradition common in analytic number theory to put all the routine lemmas needed later in the paper in a rather lengthy but straightforward early section), but about six pages after all the lemmas are presented, Yitang has made a non-trivial observation, which is that bounded gaps between primes would follow if one could make any improvement to the Bombieri–Vinogradov theorem for smooth moduli. (This particular observation was also previously made independently by Motohashi and Pintz, though not quite in a form that was amenable to Yitang's arguments in the remaining 30 pages of the paper.) This is not the deepest part of Yitang's paper, but it definitely reduces the problem to a more tractable-looking one, in contrast to the countless papers attacking some major problem such as the Riemann hypothesis in which one keeps on transforming the problem to one that becomes more and more difficult looking, until a miracle (i.e. error) occurs to dramatically simplify the problem.

From what I have read and heard, I gather that currently, the shortest "proof of concept" of a non-trivial result in an existing (i.e. non-IUTT) field in Mochizuki's work is the 300+ page argument needed to establish the abc conjecture. It seems to me that having a shorter proof of concept (e.g. ≤ 100 pages) would help dispel skepticism about the argument. It seems bizarre to me that there would be an entire self-contained theory whose only external application is to prove the abc conjecture after 300+ pages of set up, with no smaller fragment of this setup having any non-trivial external consequence whatsoever.

Comment 96.9 (Alon Amit)**.** Thank you so much for weighing in so clearly and unambiguously on the situation. The mathematical community needs to speak up more clearly about it.

Comment 96.10 (Dick Gross)**.** This is an excellent post. Terry's comment (from the outside of number theory) is particularly telling. For those of us inside of it, the

situation is infuriating. Shortly after Faltings announced his proof of Tate's isogeny conjecture and the Mordell conjecture, he lectured on it at the Arbeitstagung, explaining the new tools he had introduced. Everyone in the audience who had thought about the problem was immediately convinced. Instead of producing 300+ pages of manuscript, Mochizuki needs to give one or two lectures (in Bonn, or Paris, or Boston, or ...) clearly explaining the new ideas in his argument and showing how they lead to a proof of ABC. This shouldn't be difficult — I have no idea why he refuses to do so.

Comment 96.11 (Peter Scholze)**.** Thanks for the wonderful post! I agree with everything that was said.

One small thing I would like to add is that most accounts indicate that no experts have been able to point to a place where the proof would fail. This is in fact not the case; since shortly after the papers were out I am pointing out that I am entirely unable to follow the logic after Figure 3.8 in the proof of Corollary 3.12 of Inter-universal Teichmüller theory part III: "If one interprets the above discussion in terms of the notation introduced in the statement of Corollary 3.12, one concludes [the main inequality]." Note that this proof is in fact the *only* proof in parts II and III that is longer than a few lines which essentially say "This follows from the definitions". Those proofs, by the way, are completely sound, very little seems to happen in those two papers (to me). Since then, I have kept asking other experts about this step, and so far did not get any helpful explanation. In fact, over the years more people came to the same conclusion; from everybody outside the immediate vicinity of Mochizuki, I heard that they did not understand that step either. The ones who do claim to understand the proof are unwilling to acknowledge that more must be said there; in particular, no more details are given in any survey, including Yamashita's, or any lectures given on the subject (as far as they are publicly documented). [I did hear that in fact all of parts II and III should be regarded as an explanation of this step, and so if I am unable to follow it, I should read this more carefully... For this reason I did wait for several years for someone to give a better (or any) explanation before speaking out publicly.]

One final point: I get very annoyed by all references to computer-verification (that came up not on this blog, but elsewhere on the internet in discussions of Mochizuki's work). The computer will not be able to make sense of this step either. The comparison to the Kepler conjecture, say, is entirely misguided: In that case, the general strategy was clear, but it was unclear whether every single case had been taken care of. Here, there is no case at all, just the claim "And now the result follows".

Comment 96.12 (Brian Conrad)**.** PS, thank you so much for writing in such specificity about your experience. In the spirit of stating things in public that have been known among some experts in arithmetic geometry for quite a while, I'd like to now share something in public (I think for the first time) concerning Corollary 3.12 in IUT3 that I have been bringing to the attention of many mathematicians in private during the past 2 years. Soon after I posted my essay on Cathy O'Neil's blog summarizing my impressions about the Oxford IUT workshop in December 2015, I received unsolicited emails from people whom I knew in quite distant parts of the world (one in Europe, one in Asia, and one in North America). Each of them told me that they had worked through the IUT papers on their own and were able to more-or-less understand things up to a specific proof where they had become

rather stumped. For each of these people, the proof that had stumped them was for 3.12 in IUT3. It was striking to get three independent unsolicited emails in a matter of days which all zeroed in on that same proof as a point of confusion.

A focus of concern on the proof of 3.12 in IUT3 never came up in discussions during the Oxford IUT workshop; my first awareness about it was from those three unsolicited emails. Since that time, the number of people whom I know that have invested tremendous effort reading the IUT papers (some giving talks at IUT workshops) and became stumped by that proof has grown further. (I will not reveal the identities of any of these people, since they communicated their concerns to me in private. It is also entirely unnecessary, since PS's comment addresses the matter quite well.)

One reason that I have never before discussed this experience in "public" (= Internet posting) is that I assumed the referee process would ultimately lead to a revision that completely clarified the proof of 3.12 in IUT3 and thereby made the matter disappear (so the earlier concerns would be rendered moot). I know from much experience as an editor at various journals that it is very common that papers submitted to math journals has errors that are caught during the refereeing process and then fixed by the author(s) before acceptance. Thus, there is generally no purpose in publicizing such matters; we are all human, after all, and (as Frank notes) part of the referee's job is to make a reasonable attempt to ferret out mistakes.

I was therefore very surprised when I heard recently (incorrectly, as it turns out) that the IUT papers had been accepted, since the public version of IUT3 still did not have a revision to the proof of 3.12 that cleared up the matter (as I immediately confirmed with several who have invested a lot of time on the IUT papers). Of course, referees are human too and may sometimes overlook something; this is why authors sometimes publish an erratum afterwards, and it is also why it is imperative for papers to be written with a degree of clarity about the techniques so that other can explore the ideas further. Clarity of communication is an essential part of progress in mathematical research. I sincerely hope that wider awareness of the genuine concern about the proof of 3.12 in IUT3 will finally lead to greater collective understanding about what is going on there.

Comment 96.13 (Harald Helfgott)**.** I am very glad that someone of note has put in writing, and rather articulately at that, what many have long said or suspected.

Notes 96.14. The papers have been published [Moc21a, Moc21b, Moc21c, Moc21d], but the community has certainly accepted that as things currently stand, to put it generously, there is nothing there.

97. Abandonware

Mon, 25 Dec 2017

For a young mathematician, there is a lot of pressure to publish (or perish). The role of for-profit academic publishing is to publish large amounts of crappy mathematics papers, make a lot of money, but at least in return grant the authors a certain imprimatur, which can then be converted into reputation, and then into job offers, and finally into pure cash, and then coffee, and then back into research. One great advantage of being a tenured full professor (at an institution not run by bean

counters) is that I don't have to play that game, and I can very selective in what papers I choose to submit. In these times — where it is easy to make unpublished work available online, either on the arXiv, a blog, or a webpage — there is no reason for me to do otherwise. Akshay and I are just putting the finishing touches on our manuscript on the torsion Jacquet–Langlands correspondence (a project begun in 2007!), and approximately 100 pages of the original version has been excised from the manuscript. It's probably unlikely we will publish the rest, not because we don't think it's interesting, but because it can already be found online. (Although we might collect the remains into a supplemental "apocrypha" to make referencing easier.) Sarnak writes lots of great letters and simply posts them online. I wrote a paper a few years ago called "Semistable modularity lifting over imaginary quadratic fields." (see [Cal]) It has (IMHO) a few interesting ideas, including one strategy for overcoming the non-vanishing of cohomology in multiple degrees in an $l_0 = 1$ situation, a way of proving a non-minimal modularity lifting theorem in an (admittedly restricted) $l_0 = 1$ situation without having to use Taylor's Ihara Avoidance or base change (instead using the congruence subgroup property), and an argument explaining why the existence of Nilpotent ideals in Scholze's Galois representation is no obstruction to the modularity lifting approach in my paper with David. But while I wrote up a detailed sketch of the argument, gave a seminar about it, and put the preprint on my webpage, I never actually submitted it. One reason was that David and I were (at the time, this was written in 2014-2015 or so) under the cosh by an extremely persnickety referee (to give you some idea, the paper was submitted in 2012 and was only just accepted), and I couldn't stomach the idea of being raked over the coals a second time merely to include tedious details. (A tiny Bernard Woolley voice at the back of my head is now saying: *excuse me minister, you can't be raked over by a cosh, it doesn't have any teeth.* Well done if you have any idea what I am talking about.) But no matter, the paper is on my webpage where anyone can read it. As it happens, the paper [ACC+23] has certainly made the results of [Cal] entirely redundant, but there are still some ideas which might be useful in the future someday. But I don't see any purpose whatsoever in subjecting an editor, a reviewer, and (especially) myself the extra work of publishing this paper.

So I am all in favor of avoiding publishing all but a select number of papers if you can help it, and blogging about math instead. So take a spoon, pass around the brandy butter and plum pudding, and, for the rest of this post, let us tuck in to something from the apocrypha.

97.1. **Galois Extensions Unramified Away From One Place.** I learned about one version of this question in the tea room at Harvard from Dick Gross. Namely, does there exist a non-solvable Galois extension $K/\mathbf{Q}$ unramified at all primes except p? Modular forms (even just restricting to the two eigenforms of level one and weights 12 and 16) provide a positive answer for p greater than 7. On the other hand, Serre's conjecture shows that this won't work for the last three remaining primes. Dick explained a natural approach for the remaining primes, namely to consider instead Hilbert modular forms over a totally real cyclotomic extension ramified at p (once you work out how to actually compute such beasts in practice). And indeed, this idea was successfully used to find such representations by Lassina Dembélé in [Dem09] and also this paper (with Greenberg and Voight [DGV11]). But there is something a little unsatisfactory to me about this, namely, these extensions

are all ramified at p *and* ∞. What if one instead asks Gross' question for a *single* place?

Minkowski showed there are no such extensions when $v = \{\infty\}$, but I don't see any obstruction to there being a positive answer for a finite place. The first obvious remark, however, is that Galois representations coming from Hilbert modular forms are not going to be so useful in this case at least when the residual characteristic is odd, for parity reasons.

On the other hand, conjecturally, the Langlands program still has something to say about this question. One could ask, for example, for the smallest prime p for which there exists a Galois representation:

$$\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_2(\overline{\mathbf{F}}_p)$$

whose image is big (say not only irreducible but also not projectively exceptional) and is unramified at all places away from p including infinity. (This is related to § 2. Here is how one might go about finding such a representation, assuming the usual suite of conjectures. First, take an imaginary quadratic field F, and then look to see if there is any extra mod-p cohomology of $\mathrm{GL}_2(\mathcal{O}_F)$ in some automorphic local system which is not coming from any of the "obvious" sources. If you find such a class, you could then try to do the (computationally difficult) job of computing Hecke eigenvalues, or alternatively you could do the same computation for a *different* such imaginary quadratic field E, and see if you find a weight for which there is an "interesting" class simultaneously for both number fields. If there are no such classes for any of the (finitely many) irreducible local systems modulo p, then there are (conjecturally) no Galois representations of the above form.

There are some heuristics (explained to me by Akshay) which predict that the number of Galois representations of the shape we are looking for (ignoring twists) is of the order of 1/p. On the other hand, no such extensions will exist for very small p by combining an argument of Tate together with the Odlyzko bounds. So the number of primes up to X for which there exist such a representation might be expected to be of the form

$$\log\log X - \log\log C$$

for some constant C to account for the lack of small primes (which won't contribute by Tate + Odlyzko GRH discriminant bounds). This is unfortunately a function well-known to be constant, and in this case, with the irritating correction term, it looks pretty much like the zero constant. Even worse, the required computation becomes harder and harder for larger p, since one needs to compute the cohomology in the corresponding local system of weight (k, k) for k up to (roughly) p. Alas, as it turns out, these things are quite slippery:

Lemma 97.2. *Suppose $\overline{\rho}$ is absolutely irreducible with Serre level 1 and Serre weight k and is even. Assume all conjectures. Then:*

1. *The prime p is at least 79.*
2. *The weight k is at least 33.*
3. *If $\overline{\rho}$ exists with $k \leq 53$, then $p \geq 1000$.*
4. *If $\overline{\rho}$ exists with $k = 55$, then $p \geq 200$, or $p = 163$, and $\overline{\rho}$ is the unique representation with projective image A_4.*

Of course the extension for $p = 163$ (which is well-known) does not have big image in the sense described above. The most annoying thing about this computation

(which is described in the apocrypha) is that it can only be done once! Namely, someone who could actually program might be able to extend the computation to (say) $p \leq 200$, but the number of extensions which one would expect to see is roughly $\log\log 200 - \log\log 79$, which is smaller than a fifth. So maybe an extension of this kind will never be found! (Apologies for ruining it by not getting it right the first time.)

98. The paramodular conjecture is false for trivial reasons

Mon, 15 Jan 2018

(This is part of a series of occasional posts discussing results and observations in [BCGP21]).

Brumer and Kramer made a conjecture (see [BK14]) positing a bijection between isogeny classes of abelian surfaces $A/\mathbf{Q}$ over the rationals of conductor N with $\mathrm{End}_{\mathbf{Q}}(A) = \mathbf{Z}$ and paramodular Siegel newforms of level N with rational eigenvalues (up to scalar) that are not Gritsenko lifts (Gritsenko lifts are those of Saito–Kurokawa type). This conjecture is closely related to more general conjectures of Langlands, Clozel, etc., but its formulation was made more specifically with a view towards computability and falsifiability (particularly in relation to the striking computations of Poor and Yuen), see [PY15].

The recognition that the "optimal level" of the corresponding automorphic forms is paramodular is one that has proved very useful both computationally and theoretically. Moreover, it is almost certain that something very close to this conjecture is true. However, as literally stated, it turns out that the conjecture is false (though easily modifiable). There are a few possible ways in which things could go wrong. The first is that there are a zoo of cuspidal Siegel forms for GSp(4); it so happens that the forms of Yoshida, Soudry, and Howe–Piatetski-Shapiro type never have paramodular eigenforms (as follows from a result of Schmidt), although this depends on the accident that the field $\mathbf{Q}$ has odd degree and no unramified quadratic extensions (and so the conjecture would need to be modified for general totally real fields). Instead, something else goes wrong. The point is to understand the relationship between motives with $\mathbf{Q}$-coefficients and motives with $\overline{\mathbf{Q}}$-coefficients which are invariant under the Galois group (i.e. Brauer obstructions and the motivic Galois group.)

It might be worth recalling the (proven) Taniyama–Shimura conjecture which says there is a bijection between cuspidal eigenforms of weight two with rational eigenvalues and elliptic curves over the rationals. Why might one expect this to be true from general principles? Let us imagine we are in a world in which the Fontaine–Mazur conjecture, the Hodge conjecture, and the standard conjectures are all true. Now start with a modular eigenform with rational coefficients and level $\Gamma_0(N)$. Certainly, one can attach to this a compatible family of Galois representations:

$$\mathcal{R} = \{\rho_p\}, \qquad \rho_p : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\overline{\mathbf{Q}}_p).$$

with the property that the characteristic polynomials $P_q(T) = T^2 - a_q T + q$ of Frobenius at any prime q not dividing Np have *integer* coefficients, and the representations are all de Rham with Hodge–Tate weights [0,1]. But what next? Using the available conjectures, one can show that there must exist a corresponding simple abelian variety $E/\mathbf{Q}$ which gives rise to $\mathcal{R}$. The key to pinning down this abelian

variety is to consider its endomorphism algebra over the rationals. Because it is simple, it follows that the endomorphism algebra is a central simple algebra D/F for some number field F. From the fact that the coefficients of the characteristic polynomial are rational, one can then show that the number field F must be the rationals. But the Albert classification puts some strong restrictions on endomorphism rings of abelian varieties, and the conclusion is the following:

Either:

(1) $E/\mathbf{Q}$ is an elliptic curve.
(2) $E/\mathbf{Q}$ is a fake elliptic curve; that is, an abelian surface with endomorphisms over $\mathbf{Q}$ by a quaternion algebra $D/\mathbf{Q}$.

The point is now that the second case can never arise; the usual argument is to note that there will be an induced action of the quaternion algebra on the homology of the real points of A, which is impossible since the latter space has dimension two. (This is related to the non-existence of a general cohomology theory with rational coefficients.) In particular, we do expect that such modular forms will give elliptic curves, and the converse is also true by standard modularity conjectures (theorems in this case!). A similar argument also works for all totally real fields. On the other hand, this argument does *not* work over an imaginary quadratic field (more on this later). In the same way, starting with a Siegel modular form with rational eigenvalues whose transfer to GL(4) is cuspidal, one should obtain a compatible family of irreducible 4-dimensional symplectic representations $\mathcal{R}$ with cyclotomic similitude character. And now one deduces (modulo the standard conjectures and Fontaine–Mazur conjecture and the Hodge conjecture) the existence of an abelian variety A such that:

Either:

(1) $A/\mathbf{Q}$ is an abelian surface.
(2) $A/\mathbf{Q}$ is a fake abelian surface; that is, an abelian fourfold with endomorphisms over $\mathbf{Q}$ by a quaternion algebra $D/\mathbf{Q}$.

There is now no reason to suspect that fake abelian surfaces cannot exist. Taking D to be indefinite, the corresponding Shimura varieties have dimension three, and they have an abundance of points — at least over totally real fields. But it turns out there is a very easy construction: take a fake elliptic curve over an imaginary quadratic field, and then take the restriction of scalars!

You have to be slightly careful here: one natural source of fake elliptic curves comes from the restriction of certain abelian surfaces of GL(2)-type over $\mathbf{Q}$, and one wants to end up with fourfolds which are simple over $\mathbf{Q}$. Hence one can do the following:

Example 98.1. Let $B/\mathbf{Q}$ be an abelian surface of GL(2)-type which acquires quaternion multiplication over an imaginary quadratic field K, but is not potentially CM. For example, the quotient of $J_0(243)$ with coefficient field $\mathbf{Q}(\sqrt{6})$ with $K = \mathbf{Q}(\sqrt{-3})$. Take the restriction to K, twist by a sufficiently generic quadratic character χ, and then induce back to $\mathbf{Q}$. Then the result will be a (provably) modular fake abelian surface whose corresponding Siegel modular form has rational eigenvalues. Hence the paramodular conjecture is false.

Cremona (in his papers) has discussed a related conjectural correspondence between Bianchi modular forms with rational eigenvalues and elliptic curves over K. His original formulation of the conjecture predicted the existence of a corresponding

elliptic curve over K, but one also has to allow for fake elliptic curves as well (as I think was pointed out in this context by Gross). The original modification of Cremona's conjecture was to only include (twists of) base changes of abelian surfaces of GL(2)-type from $\mathbf{Q}$ which became fake elliptic curves over K, but there is no reason to suppose that there do not exist fake elliptic curves which are autochthonous to K, that is, do not arise after twist by base change. Indeed, autochthonous fake elliptic curves do exist! We wrote down a family of such surfaces over $\mathbf{Q}(\sqrt{-6})$, for example. (We hear through Cremona that Ciaran Schembri, a student of Haluk Sengun, has also found such curves) On the other hand, the examples coming from base change forms from $\mathbf{Q}$ have been known in relation to this circle of problems for 30+ years, and already give (by twisting and base change) immediate counter-examples to the paramodular conjecture, thus the title.

It would still be nice to find fake abelian surfaces over $\mathbf{Q}$ (rather than totally real fields) which are geometrically simple. I'm guessing that (for $D/\mathbf{Q}$ ramified only at 2 and 3 and a nice choice of auxiliary structure) the corresponding 3-fold may be rational (one could plausibly prove this via an automorphic form computation), although that still leaves issues of fields of rationality versus fields of definition. But let me leave this problem as a challenge for computational number theorists! (The first place to look would be Jacobians of genus four curves [one might be lucky] even though the Torelli map is far from surjective in this case.)

Let me finish with one fake counter example. Take any elliptic curve (say of conductor 11). Let $L/\mathbf{Q}$ be any Galois extension with Galois group Q, the quaternion group of order 8. The group Q has an irreducible representation V of dimension 4 over the rationals, which preserves a lattice Λ. If you take

$$A = E^4 = E \otimes_{\mathbf{Z}} \Lambda,$$

then A is a simple abelian fourfold with an action of an order in D, (now the definite Hamilton quaternions) and so gives rise to compatible families $\mathcal{R}$ of 4-dimensional representations which are self-dual up to twisting by the cyclotomic character. However, the four dimensional representations are only symplectic with respect to a similitude character which is the product of the cyclotomic character and a non-trivial quadratic character of $\mathrm{Gal}(L/\mathbf{Q})$, and instead they are orthogonal with cyclotomic similitude character. So these do not give rise to counterexamples to the paramodular conjecture. A cursory analysis suggests that the quaternion algebra associated to a fake abelian surface which gives rise to a symplectic $\mathcal{R}$ with cyclotomic similitude character should be indefinite.

Comment 98.2. The title is a reference to this paper.

Notes 98.3. Ciaran Schembri's paper is [Sch19].

99. The boundaries of Sato–Tate, part I

Mon, 09 Apr 2018

A caveat: the following questions are so obvious that they have surely been asked elsewhere, and possibly given much more convincing answers. References welcome!

The Sato–Tate conjecture implies that the normalized trace of Frobenius $b_p \in [-2, 2]$ for a non-CM elliptic curve is equidistributed with respect to the pushforward of the Haar measure of SU(2) under the trace map. This gives a perfectly good

account of the behavior of the unnormalized $a_p \in [-2\sqrt{p}, 2\sqrt{p}]$ over regions which have positive measure, namely, intervals of the form $[r\sqrt{p}, s\sqrt{p}]$ for distinct multiples of $\sqrt{p}$.

If one tries to make global conjectures on a finer scale, however, one quickly runs into difficult conjectures of Lang–Trotter type. For example, given a non-CM elliptic curve E over $\mathbf{Q}$, if you want to count the number of primes $p \leq X$ such that $a_p = 1$ (say), an extremely generous interpretation of Sato–Tate would suggest that probability that $a_p = 1$ would be

$$\frac{1}{4\pi\sqrt{p}},$$

and hence the number of such primes $\leq X$ should be something like:

$$\frac{X^{1/2}}{2\pi \log(X)},$$

except one *also* has to account for the fact that there are congruence obstructions/issues, so one should multiply this factor by a (possibly zero) constant depending one adelic image of the Galois representation. So maybe this does give something like Lang–Trotter.

But what happens at the other extreme end of the scale? Around the boundaries of the interval $[-2, 2]$, the Sato–Tate measure converges to zero with exponent one half. There is a trivial bound $a_p \leq t$ where t^2 is the largest square less than $4p$. How often does one have an equality $a_p^2 = t^2$? Again, being very rough and ready, the generous conjecture would suggest that this happens with probability very roughly equal to

$$\frac{1}{6\pi p^{3/4}},$$

and hence the number of such primes $\leq X$ should be something like:

$$\frac{2X^{1/4}}{3\pi \log(X)}.$$

Is it at all reasonable to expect $X^{1/4\pm\epsilon}$ primes of this form? If one takes the elliptic curve $X_0(11)$, one finds a_p^2 to be as big as possible for the following primes:

$$a_2 = -2 \geq -2\sqrt{2} = -2.828\ldots,$$
$$a_{239} = -30 \geq -2\sqrt{239} = -30.919\ldots,$$
$$a_{6127019} = 4950 \leq 2\sqrt{p} = 4950.563\ldots,$$

but no more from the first 500000 primes. That's not completely out of line for the formula above!

Possibly a more sensible thing to do is to simply ignore the Sato–Tate measure completely, and model $E/\mathbf{F}_p$ by simply choosing a randomly chosen elliptic curve over $\mathbf{F}_p$. Now one can ask in this setting for the probability that a_p is as large as possible. Very roughly, the number of elliptic curves modulo p up to isomorphism is of order p, and the number with $a_p = t$ is going to be approximately the class number of $\mathbf{Q}(\sqrt{-D})$ where $-D = t^2 - 4p$; perhaps it is even exactly equal to the class number $H(t^2 - 4p)$ for some appropriate definition of the class number. Now the behaviour of this quantity is going to depend on how close $4p$ is to a square. If $4p$ is very slightly — say $O(1)$ — more than a square, then $H(t^2 - 4p)$ is pretty much a constant, and the expected probability going to be around 1 in p. On the

other hand, for a generic value of p, the smallest value of $t^2 - 4p$ will have order $p^{1/2}$, and then the class group will have approximate size $p^{1/4\pm\epsilon}$, and so one (more or less) ends up with a heuristic fairly close to the prediction above (at least in the sense of the main term being around $X^{1/4\pm\epsilon}$).

But why stop there? Let's push things even closer to the boundary. How small can $a_p^2 - 4p$ get relative to p? For example, let us restrict to the set $S(\eta)$ of prime numbers p such that

$$S(\eta) := \{p \mid p \in (n^2, n^2 + n^{2\eta}) \text{ for some } n \in \mathbf{Z}\}.$$

For such primes, the relative probability that $a_p = \lfloor \sqrt{4p} \rfloor = 2n$ is approximately $n^\eta/p \sim n^{2\eta-1}$. So the expected number of primes with this property will be infinite providing that

$$\sum \frac{n^{3\eta}}{n^2 \log(n)}$$

is infinite, or, in other words, when $\eta \geq 1/3$. So this leads to the following guess (don't call it a conjecture!):

Question 99.1. Let $E/\mathbf{Q}$ be an elliptic curve without CM. Is

$$\liminf \frac{\log(a_p^2 - 4p)}{\log(p)} = \frac{1}{3}?$$

Of course, one can go crazy with even more outrageous guesses, but let me stop here before saying anything more stupid.

100. Chicago seminar roundup

Sat, 28 Apr 2018

Here are two questions I had about the past two number theory seminars. I haven't had the opportunity to think about either of them seriously, so they may be easy (or more likely stupid).

100.1. **Anthony Várilly-Alvarado.** Tony gave a talk on his joint work with Dan Abramovich [AVA17] about the relation between Vojta's conjecture and the problem of uniform bounds on torsion for abelian varieties. (Spoiler: one implies the other.) More specifically, assuming Vojta's conjecture, there a universal bound on m (depending only on g and K) beyond which no abelian variety of dimension g over K can have full level structure.

If one wanted to prove this (say) for elliptic curves, and one was allowed to use any conjecture you pleased, you could do the following. Assume that $E[m] = \mu_m \oplus \mathbf{Z}/m\mathbf{Z}$ for some large integer m. One first observes (by Neron–Ogg–Shafarevich plus epsilon) that E has to have semi-stable reduction at primes dividing N_E. Then the discriminant Δ must be an mth power, and then Szpiro's Conjecture (which is the same as the ABC conjecture) implies the desired result.

If you try to do the same thing in higher dimensions, you similarly deduce that A must have semi-stable reduction at primes dividing N_E.One then gets implications on the structure of the Neron model at these bad primes, which one can hope to parlay in order to get information about local quantities associated to A analogous to the discriminant being a perfect power. But I'm not sure what generalizations of

Szpiro's conjecture there are to abelian varieties. A quick search found one formulation attributed to Hindry in terms of Faltings height, but it was not immediately apparent if one could directly deduce the desired result from this conjecture, nor what the relationship was with these generalizations to either ABC or to Vojta's conjecture.

100.2. **Ilya Khayutin.** Ilya mentioned Linnik's theorem that, if one ranges over imaginary quadratic fields in which a fixed small prime is split, the CM j-invariants become equidistributed. The role of the one fixed prime is to allow one to use ergodic methods relative to this prime. My naive question during the talk: given p is split, let $\mathfrak{p}$ be a prime above p. Now one can take the subgroup of the class group corresponding to the powers of $\mathfrak{p}$. Do these equidistribute? The speaker's response was along the lines that it would probably be quite easy to see this is false, but I didn't have time after the talk to follow up. It's certainly the case that, most of the time, the prime $\mathfrak{p}$ will itself generate a subgroup of small index in the class group (the quotient will look like the random class group of a real quadratic field), but sometimes it will be quite large. For example, I guess one can take

$$D = 2^n - 1, \qquad \mathfrak{p}^{n-2} = \left(\frac{1+\sqrt{-D}}{2}\right),$$

and the subgroup generated by this prime has order $\log(D)$ compared to $D^{1/2+\epsilon}$. So I decided (well, after writing this line in the blog I decided) to draw a picture for some choice of Mersenne prime. And then, after thinking a little how to draw the picture, realized it was unnecessary. The powers of $\mathfrak{p}$ in this case are given explicitly by

$$\mathfrak{p}^m = \left(2^m, \frac{1+\sqrt{-D}}{2}\right),$$

It is transparent that for the first half of these classes, the first factor is much smaller than the second, but since the second term also has small real part, the ratio already lies inside the (standard) fundamental domain. Hence the corresponding points will lie far into the cusp. Similarly, the second half of the classes are just the inverses in the class group of the first half, and so will consist of the reflections of those points in $x = 0$ and so also be far into the cusp. So I guess the answer to my question is, indeed, a trivial no. So here is a second challenge: suppose that 2 AND 3 both split. Then do the CM points generated by $\mathfrak{p}$ for primes above 2 AND 3 equidistribute? Actually, in this case, it's not clear off the top of my head that one can easily write down discriminants for which the index of this group is large. But even if you can, sometimes $\mathbf{Z}^2$ subgroups get you much closer to equidistribution than $\mathbf{Z}$!

Comment 100.3 (Bisi Agboola)**.** There's a generalisation of Szpiro's conjecture to jacobians of hyperelliptic curves (due to Paul Lockhart) that might be more along the lines of what you're looking for [Loc94].

101. Update on Sato–Tate for abelian surfaces

Thu, 19 Jul 2018

Various people have asked me for an update on the status of the Sato–Tate conjecture for abelian surfaces in light of recent advances in modularity lifting

theorems. My student Noah Taylor has exactly been undertaking this task, and this post is a summary of his results. (Which have now appeared in [Tay20].)

First, let me recall the previous status of this conjecture. An explicit form of this conjecture (detailing all the 52 possible different Sato–Tate groups which could occur for abelian surfaces over number fields — 34 of which occur over $\mathbf{Q}$) was given in a paper of Fité, Kedlaya, Rotger, and Sutherland (I recommend either reading these slides or especially watching this video for the background and some fun animations, also see [FKRS12]). Christian Johansson [Joh17] gave proofs of this conjecture over totally real fields in many of the possible cases in which the abelian surface had various specific types of extra endomorphisms over the complex numbers by exploiting modularity results that had been used in the proof of the Sato–Tate conjecture for elliptic curves. Over totally real fields, this left essentially four remaining cases:

(1) The case when the Galois representations associated to A decomposes over a quadratic extension L/F into two representations which are Galois twists of each other, and L/F is not totally real.
(2) The case when the Galois representations associated to A decomposes over a quadratic extension L/F into two representations which are not Galois twists of each other, *and* L/F *is CM.*
(3) The case when the Galois representations associated to A decomposes over a quadratic extension L/F into two representations which are not Galois twists of each other, and L/F is *neither totally real nor CM.*
(4) The case when the geometric endomorphism ring of A is $\mathbf{Z}$.

Noah has something to say about each of these cases.

Case 1: Noah completed the proof of Sato–Tate in this case using only the methods from the paper [BLGGT14], by exploiting the fact that the corresponding two-dimensional representations — while possibly only defined over a field L which need not be totally real or CM — in fact give rise to *projective* representations which extend to F. By a theorem of Tate, each of these representations can be extended to F after twisting by a character, and so the original 4-dimensional representation looks like the tensor product of a 2-dimensional representation over F (which is potentially modular) and an Artin representation. At this point one is in good shape.

Case 2: The Sato–Tate conjecture is proved in this case. This case required the least amount work, because it is pretty much an immediate consequence of the modularity results proved in [ACC$^+$23].

If the totally real field is $\mathbf{Q}$ this implies the Sato–Tate conjecture for all abelian surfaces except those of type (4).

Cases 3 & 4: In these cases, one can apply the potentially modularity results proved in my (very close to being finished) paper with Boxer, Gee, and Pilloni [BCGP21]. It is too much to expect a full proof of Sato–Tate at this point. However, knowing potential modularity allows one to obtain partial results, similar to those of Serre and Kim–Shahidi for the case of elliptic curves (after Wiles but before Clozel–Harris–Taylor). Here is a sample result:

Theorem 101.1 (Noah Taylor)**.** *Let C be a genus two curve over a totally real field F. Then, for any $\epsilon \geq 0$, there exists a positive density of primes $\mathfrak{p}$ (with*

$N(\mathfrak{p}) = p$), *one has*

$$\#C(\mathcal{O}/\mathfrak{p}) - p - 1 \geq \left(\frac{2}{3} - \epsilon\right)\sqrt{p}.$$

Compare this to the Hasse bounds, which imply that the quantity on the LHS has absolute value at most $4\sqrt{p}$. Of course this theorem is much weaker than the Sato–Tate conjecture. But even the weaker version of this theorem which says that $\#C(\mathbf{F}_p) \geq p+1$ for infinitely many primes was *completely open* before such curves were known to be potentially modular. Similarly, I don't think one can prove the corresponding result for elliptic curves without either using something very close to modularity (in the non-CM case) or the equidistribution theorems of Hecke in the CM case. I think the following example is instructive: take the elliptic curve $y^2 = x^3 - x$ which admits CM by the Gaussian integers. One has a formula for the difference $a_p = 1 + p - \#E(\mathbf{F}_p)$ as follows: for a prime which is $1 \mod 4$, one may write $p = a^2 + b^2$ uniquely in integers by imposing the additional congruence

$$(a + bi) \equiv 1 \mod (1+i)^3.$$

Then one has the formula $a_p = 2a$. The problem then becomes: do there exist infinitely many primes $p \equiv 1 \bmod 4$ such that one has $a > 0$? This seems suspiciously like something that can be proven using Cebotarev, but it is not. The problem is that the infinite places of $F = \mathbf{Q}(\sqrt{-1})$ are all complex, so there is no choice of "conductor" which differentiates between complex numbers with positive or negative real part at the infinite places in $\mathbf{A}_F^\times$.

Noah's proof of the theorem above exploits the following idea. Potential modularity not only gives meromorphy of the L-function, but more importantly (in this case) holomorphy and non-vanishing in the (analytically normalized) halfplane $\Re(s) \geq 1$. Moreover, again using functorialities, potential automorphy, and results of Shahidi, one obtains similar results not only for the degree 4 L-function, but also the degree 5 L-function, and also crucially the Rankin-Selberg L-functions of degrees 16, 20, and 25. From this one can obtain various "prime number theorem" estimates for quantities involving the Frobenius eigenvalues, and then one has to massage these into an inequality. A simple version of this argument is as follows: given some infinite set of real numbers $a_n \in [-2, 2]$ such that

$$\frac{1}{n}\sum_{i=1}^{n} a_i \to 0, \qquad \frac{1}{n}\sum_{i=1}^{n} a_i^2 \to 1,$$

One can draw the conclusion that $a_n \geq 1/2 - \epsilon$ infinitely often, by (for example) considering the average of the quantity $(2a_n - 1)(a_n + 2)$. Moreover, this is the best possible bound given these constraints.

Note that since the Sato–Tate conjecture is known in all other cases, one only has to consider cases (3) and (4), which behave slightly differently in this argument. In fact, in case (3), one can do much better:

Theorem 101.2 (Noah Taylor)**.** *Let C be a curve over a totally real field F such that $A = \mathrm{Jac}(C)$ is of type (3). Then there exists a positive density of primes $\mathfrak{p}$ (with $N(\mathfrak{p}) = p$), such that*

$$\#C(\mathcal{O}/\mathfrak{p}) - p - 1 \geq 2.47\sqrt{p}.$$

(Note that once this result is known in case (3) it is known for all curves whose Jacobian is not of type (4), that is, those whose Jacobians admit a non-trivial

endomorphism over $\mathbf{C}$.) The point is that, in this case, one knows not just the potential automorphy of A, but also the potential automorphy of the corresponding two-dimensional representations over the quadratic extension L, and so one can also exploit the automorphy of symmetric powers of the corresponding GL(2)-automorphic representations (and further analyticity results for higher symmetric powers) as well as a zoo of Rankin-Selberg L-functions coming from pairs of low symmetric powers. (As for the constants involved in both of these theorems, they are essentially optimal given the automorphic input.)

These results tie in to problems raised in various talks of Nick Katz (see for example this talk). Noah's result above implies that, given an curve C over a totally real field, one *can tell* that it doesn't have genus one from the distribution of the traces of Frobenius *except* possibly in the case when its Jacobian has no non-trivial geometric endomorphism (the "typical" case, of course). It's a little sad that the modularity results are not sufficient to handle that last case as well — showing that the support of the normalized trace of Frobenius extends beyond $[-2, 2]$ would require knowing something close to functoriality of the map $\mathrm{Sym}^2 : \mathrm{GL}(4) \to \mathrm{GL}(10)$, and this is currently out of reach, unfortunately. Oh well, that's a shame: wow I dearly would have loved to give a talk entitled *Simple things that Nick Katz doesn't know (but I do)*.

102. Mazur's program B on abelian surfaces

Fri, 07 Sep 2018

In the book "More mathematical people," there is an interview with Robin Wilson with the following quote:

> At the meal I found myself sitting next to Alistair Cooke who was very charming, and absolutely fascinating to listen to. The very next Sunday when I was back in England I turned on his "Letter from America" on the radio — he started off by saying, "I went to a very boring dinner at the White House. There was no one interesting to talk to." That amused me a lot.

So let me start off by saying that even though this post is about one or two things I learnt at Oberwolfach, it is deliberately **not** about anything I learnt in the talks, lest my choosing some talks over others leading to false inferences on what I thought interesting. For example, the title of this post alludes to David Zureick-Brown's talk, which I will not mention again.

Let g be a non-negative integer and p a prime. Suppose one starts with a representation

$$\rho : G_{\mathbf{Q}} \to \mathrm{GSp}_{2g}(\mathbf{F}_p)$$

with cyclotomic similitude character. To avoid later circumlocutions, let me (most of the time) assume it is absolutely irreducible. One can ask whether this representation arises infinitely often from the p-torsion on an abelian variety — perhaps additionally assuming that it does arise from at least *one* such variety, or perhaps not.

This problem is very well-studied in the case $g = 1$, where we know that the answer is positive exactly for the primes $p = 2$, 3, and 5, where the corresponding moduli space $X(p)$ has genus zero, and the associated twists $X(\rho)$ are Brauer–Severi

varieties that also turn out to be rational over $\mathbf{Q}$ under the given hypothesis on the similitude character. When $p \geq 5$, the curves $X(p)$ have genus at least 3, and so their twists always have at most finitely many points over any field, by Faltings. So there is certainly a satisfactory answer in this case. (Of course, there are many more subtle versions of this question — for example, replacing "infinitely" by "at least twice" — and those variations are open in general.)

If we move to genus $g = 2$, then the case of $p = 2$ is also straightforward — the 2-torsion of the Jacobian of $y^2 = f(x)$ for a degree 6 polynomial with Galois group G just comes from the isomorphism $G \hookrightarrow S_6 \simeq \mathrm{GSp}_4(\mathbf{F}_2)$. (One needs to be a little bit careful here because the outer automorphism of S_6 means there are two non-conjugate such maps and one has to choose the right one.) Given that one can write down families of sextics where the etale $\mathbf{Q}$-algebra $\mathbf{Q}[x]/f(x)$ is constant, it's easy to see that the answer is positive in this case without any restrictions. For example, given an S_6 extension, there's a six dimensional family of polynomials one can write down whose splitting field generically gives this extension, and so after accounting for the action of PGL_2 on the roots, this still gives a three-dimensional rational family of genus two curves whose two torsion comes from this extension.

In my paper [BCGP21] with Boxer, Gee, and Pilloni, we will also give a similarly conclusive answer for $p = 3$, although there are some unexpected surprises, as well as some complementary results recently proved by my student Shiva Chidambaram. But more on this in a post coming up soon!

When $p > 3$, then the corresponding 3-folds obtained by taking full level p-structure of the corresponding Siegel 3-fold $\mathcal{A}_2$ are of general type. (Note that it is essentially known when $\mathcal{A}_g(n)$ is either geometrically rational of general type, see for example Theorem II.2.1 and the surrounding comments in this paper, see [HS02].) Of course, unlike the case of curves, varieties of dimension greater than one of general type can have many rational points. For example, it's obvious that there are many abelian surfaces over Q whose 5-torsion has the form $(\mathbf{Z}/5\mathbf{Z} \oplus \mu_5)^2$, because one can take A to be $E+E$ where E is an elliptic curve whose 5-torsion has the form $(\mathbf{Z}/5\mathbf{Z} \oplus \mu_5)$, and there are infinitely many such E because the classical modular curve of full level 5 is rational over Q. To put it a different way, the 3-fold $\mathcal{A}_2(5)$ corresponding to abelian surfaces with fixed 5-torsion will contain a number of rational Shimura subvarieties coming both from Hilbert modular surfaces and from modular curves, even though it itself is of general type. This can happen even if the mod-p representation ρ is irreducible. For example, given an elliptic curve over a quadratic field $K/\mathbf{Q}$, there will once more be a rational curve of elliptic curves with the same mod-5 representation, and so the restriction of scalars will give a rational curve on some twist $\mathcal{A}_2(\rho)$ of $\mathcal{A}_2(5)$. On the other hand, one might at least start off by making the following naive minimal guess.

Question 102.1. Suppose that ρ is surjective for $g = 2$ and $p \geq 5$. Then are there only finitely many points on $\mathcal{A}_2(\rho)$?

An even more extreme version of this question would be to ask if there is at *most one* such point. This seems a little unlikely even by comparison with the case of $g = 1$. I learnt the following nice example talking to John Cremona during the hike through the Black Forest: for $g = 1$ and $p = 7$ and varying E, the twist $X(E[7])$ has genus 3 (it is a twist of the Klein quartic). This twist is still geometrically a plane quartic. By considering the tangent to the point of $X(E[7])$ corresponding to E, the line has two further intersections with the curve, and one obtains two

further points over $X(E[7])$ which now (in general) lie over a quadratic extension. But one can parametrize the E for which these points are actually *rational* and this turns out to be the rational cover of the j-line corresponding to asking that the invariant c_4 is a square. So there are infinitely many elliptic curves A (even with $A[7]$ surjective) for which there exist at least a pair of non-isogenous elliptic curves A, B with $A[7] = B[7]$ as symplectic Galois representations. So a better question is the following:

Question 102.2. Can one find examples of non-isogenous abelian surfaces A and B with $A[5] = B[5]$ and such that the corresponding representation has a surjective Galois representation?

This is the type of question where it is useful to have Andrew Sutherland nearby with a laptop. Within an hour or two, he sent me the following examples (using the [LMF24])

$$C_1 : y^2 = -120x^6 - 264x^5 + 186x^4 + 276x^3 - 201x^2 + 24x$$
$$C_2 : y^2 = 16x^5 - 33x^4 + 60x^3 - 42x^2 + 36x - 9$$

both of conductor $2^{10} \cdot 3^7$ with surjective and isomorphic mod-5 Galois representations which are not isogenous. Nice!

Naturally, the question turned to the existence of a pair with $A[7] = B[7]$. That proved a tougher challenge, but not an insurmountable one, and here is such a pair (again found by Andrew the same day):

$$C_1 : y^2 + (x^3 + x)y = -x^6 + 2x^4 + 2x^3 + 16x^2 + 4x + 16$$
$$C_2 : y^2 + (x^2)y = 14x^5 - 44x^4 + 46x^3 - 23x^2 + 12x - 3$$

this time of conductor $2^7 \cdot 3^2 \cdot 7^4$. Any guesses as to whether there are any such pairs for $p = 11$? I'm not sure I have any idea.

102.3. **Other news from Oberwolfach.** I do appreciate being invited to the Oberwolfach conference on computational number theory — it pushes me outside my usual range of interests. It's also the conference I have attended most often, now 8 times since 2003, although even that is far fewer than some of the regular participants. The conference is also chance to see a bunch of people I pretty much never get to see anywhere else. Even better, they are all nice enough to still invite me after this post. On the other hand, every time I give a talk I think that *this is the time that I finally have something interesting to say to this audience*, and it never quite seems to work out that way. I was certainly convinced that this was going to be the year, but then during my talk I managed to catch three people asleep in the front row. To be fair, it was the third last talk of the conference. On the other hand, Mike Bennett talked directly after me and completely failed to rise to my level of soporificness, despite his best efforts and his own predictions he would do otherwise.

There was a lunar eclipse on the final night of our stay. Most of us took to the roof to observe it, but the tall mountains of the Schwarzwald obscured our view until the final moment. Mike Bennett took the following photo, which he describes as the "best of a bad bunch"

Notes 102.4. it was an auspicious moon — I found out the very next day that my wife was pregnant with our daughter

FIGURE 11. The distance of the Moon

Notes 102.5. Tom Fischer points out a reference for Cremona's comment is [HK03, Remarque 6.2].

103. MORE OR LESS OPAQUE

Wed, 17 Oct 2018

I recently talked with Lynnelle Ye (a soon to be graduating student of Mark Kisin) for a few hours about her thesis and related mathematics. In her thesis, she generalizes (in part) the work Liu-Wan-Xiao [LWX17] on the boundary (halo) of the eigencurve to unitary groups. One of her main results gives a precise asymptotic growth rate of the Newton Polygon of U_p as one moves towards the boundary. Turning this around, this leads to estimates for the function $N_\lambda(X)$ which counts the number of eigenvalues λ of U_p (with multiplicity) of valuation at most X.

I have always had a soft spot for counting slopes, although I haven't really done anything in this business for many years. It is already interesting to estimate this growth function for classical overconvergent modular forms in the centre of weight space. Precise estimates were first obtained by Wan in his work on the Gouvea–Mazur conjectures.

Suppose we fix a tame level Γ, and let $X = X(\Gamma)$ denote the relevant modular curve. Then it turns out that, conjecturally at least, that:

$$N_\lambda(X) \sim^? \frac{\mathrm{Vol}(X_0(p))}{4\pi} X.$$

But this is precisely the growth estimate in Weyl's law for the Laplacian on $X_0(p)$! This suggests an analogy between the spectrum of the compact operator U_p in the p-adic case and the spectrum of the Laplacian operator in the complex case which was first suggested to me by Don Blasius and which I always hoped but never quite managed to extract anything from (see section 5 of these notes, which also contain more precise details about Wan's results and related results towards the conjecture above, as well as many further speculations on Overconvergent p-adic Quantum Unique Ergodicity, if you were wondering about the title).

What growth rate should one expect for the Unitary group $U(n)$? Lynnelle exploits the fact (as do Liu–Wan–Xiao) that one can work on a compact form of

the group which is zero dimensional. However, the eigenvariety is (or should be) essentially the same as the eigenvariety for other forms of the group. Following the analogy above, we can consider the growth rate of Weyl's law for $U(n-1,1)$, which, since the Shimura variety for $U(n-1,1)$ has complex dimension $n-1$, grows like X^{n-1}. However, the exponent in Lynnelle's work turns out to be

$$X^{n(n-1)/2}.$$

If I understood correctly, this one can even predict (if not prove) by simply counting the dimension of certain classical spaces of regular algebraic automorphic forms as one ranges over local systems of appropriate weights (proving it requires more work, of course). However, this seems to spoil the very precise (up to the level of constants) analogy for the complex dimension $n = 1$ case above. Is there something one can do to massage these results so they look more similar or was the $n = 1$ case simply misleading?

104. Irregular lifts, Part I

Fri, 19 Oct 2018

This post motivated in part by the recent preprint of Fakhruddin, Khare, and Patrikis (see [FKP22]) and also by Matt's number theory seminar at Chicago this week. (If you are interested in knowing what the calendar is for the Chicago number theory seminar this quarter, then that makes two of us. Actually, if you are **giving** a number theory seminar at Chicago this quarter, please leave a comment on this post with the day you are visiting, because several readers of this blog would be interested in finding out who is coming and what they are talking about.)

Let

$$\overline{\rho} : G_{\mathbf{Q}_p} \to \mathrm{GL}_n(\overline{\mathbf{F}}_p)$$

be a continuous representation. We now know, by the work of Emerton–Gee [EG23], that this representation admits a lift to characteristic zero representation of regular weight which is de Rham (and is even potentially diagonalizable). On the other hand, can it be the case that there do not exist *any* de Rham lifts in *non-regular* weight?

In the most extreme case, where we demand that all the Hodge–Tate weights are zero, then there are obstructions to lifting. In this case, the image of inertia on any lift must have finite image, but the image of inertia of $\overline{\rho}$ may already be sufficiently large to preclude this possibility. (This was exploited in the proof of Theorem 5.1 here, [Cal12].) So this answers the case when n = 2.

But what happens (for example) for $n > 2$ and Hodge–Tate weights $= [0, \dots, 0, 1]$? Or even $n = 2$ and replacing $\mathbf{Q}_p$ by a finite extension K? The first remark is that even when the residual image lands inside the Borel, there will certainly be obstructions to finding lifts inside the Borel, which means that inductive arguments will not be sufficient. On the other hand, this definitely smells like a tractable problem.

I offer an **Aperol Spritz** to an answer to this problem — let me do so even in the constrained version in weight $[0, 0, 1]$ and $K = \mathbf{Q}_p$. To recap:

Problem 104.1. Find a representation $\overline{\rho} : G_{\mathbf{Q}_p} \to \mathrm{GL}_3(\overline{\mathbf{F}}_p)$ such that there is no lift ρ of $\overline{\rho}$ which is de Rham with Hodge–Tate weights $[0, 0, 1]$ or prove that no such $\overline{\rho}$ exists.

Comment 104.2 (Pierre Colmez)**.** Some rule of thumb computation suggest that de Rham liftings of representations of $G_{\mathbf{Q}_p}$ with HT weights $0, 0, 1$ come in (non-empty) varieties of dimension 2 (ignoring twists by an unramified character). Since the mod p representation has a lifting in characteristic 0, the space of such liftings V should be of dimension 10 (or just 9 if one ignores unramified twists). The theory of (φ, Γ)-modules (or Fontaine's extension of Sen's theory) produces a module D^+_{dif} of rank 3 over $F_n[[t]]$ for n big (and F_n is the n-th layer in the cyclotomic tower) with a connexion ∇ (with $\nabla t = t$) and V is de Rham if and only if the connexion is trivial (after inverting t). So, if all weights are 0, for example, this is equivalent to $\nabla = 0$ on $D_{\mathrm{Sen}} = D^+_{\mathrm{dif}}/t$, which gives 9 conditions and explains why de Rham representations with HT weights $0, 0, 0$ are isolated. Now, in the case $0, 0, 1$, you need ∇ to be of trace and rank 1 on D_{Sen} which gives you 4 parameters (a line in P^2, 3 vectors in this line, and a relation implying that the trace is 1), and each such ∇ produces a plane W on which ∇ is trivial, and a line $K_n[[t]]f$ with $\nabla f - f \in tD^+_{\mathrm{dif}}$ and the connexion is trivial if and only if $\nabla f - f$ has no component in tW/t^2W, which gives 2 extra conditions. So you are left with 2-dimensional families (if you do the same thing with Hodge–Tate weights $0, 1, 2$, you end up with 3-dimensional families). If one wants to turn the above into a rigorous argument, I am afraid that some work is needed ... (One needs facts about the map sending a representation to its Sen operator which are maybe not in the literature.)

Comment 104.3 (Persiflage)**.** I certainly agree with the numerology here, having recently done a similar calculation — at a weight corresponding to a dominant cocharacter λ, the stabilizer of this cocharacter inside the Weyl group will be of the form

$$S_{a_1} \times S_{a_2} \times \dots S_{a_k} \subset S_n$$

with $\sum a_i = n$ giving rise to a Levi subgroup $\mathrm{GL}(a_1) \times \dots \mathrm{GL}(a_k)$, and then the "expected" dimension of the deformation ring with fixed determinant should be

$$\begin{aligned} &\dim(B(n)) - \sum_{i=1}^{k} \dim(B(a_i)) \\ &= \frac{n(n+1)}{2} - \sum_{i=1}^{k} \frac{a_i(a_i+1)}{2} \\ &= \frac{n(n-1)}{2} - \sum_{i=1}^{k} \frac{a_i(a_i-1)}{2}, \end{aligned}$$

which (for $3 = 2 + 1 = 1 + 1 + 1$) gives 0, 2, and 3 respectively. And this number certainly positive unless one is in trivial weight. What I'm not sure of, however, if you are arguing that this is a *heuristic* for the existence of lifts or a *strategy* for proving so. The heuristic doesn't entirely convince me — in part because the heuristic still *somewhat* suggests there should still be lifts of weight zero (because $0 \geq -1$). If you are saying this is a strategy, then the fact that these Sen maps are only locally analytic certainly makes me nervous, combined with the fact that p-adic balls are not projective spaces so if you want to impose A conditions in B dimensions then you still have work to do when $B \geq A$ to show there are solutions ... of course you know this better than me, so you might have a better feeling for how worried one should be.

Comment 104.4 (Pierre Colmez)**.** That was more a heuristic than a strategy, but I agree that your observation that $0 > -1$ makes this heuristic somewhat shaky ...

105. Irregular lifts, Part II

Sun, 28 Oct 2018

This is the global counterpart to § 104. I was going to write this post in a more general setting, but the annoyances of general reductive groups got the better of me.

Suppose we fix the following:

(1) A number field F and a prime $p \geq 2$.
(2) A conjugacy class of involutions (possibly trivial) c_v of $\mathfrak{gl}_n$ for all real places of F.

Problem 105.1. Does **every** representation

$$\overline{\rho} : G_F \to \mathrm{GL}_n(\overline{\mathbf{F}}_p),$$

with complex conjugation acting on the adjoint by c_v for each real place of F

(1) Have a de Rham lift?
(2) Have a de Rham lift of non-regular weight?

I have basically come to the conclusion that the answer to this question is, almost always, no. Namely, the only time the answer is yes is when F is totally real and all the complex conjugations are totally odd. (With one caveat that comes later.)

Most of the theoretical evidence — slim that it is — is in favour of this minimalist conjecture. Namely:

(1) When $n = 2$ and $F = \mathbf{Q}$ and c is non-scalar, there is a global obstruction to lifting to a weight one modular form, since the image of such forms is a finite subgroup of $\mathrm{GL}_2(\mathbf{C})$, and this can already be precluded from making the image of $\overline{\rho}$ contain a large Borel subgroup.
(2) When $n = 2$ and p is totally split in F, there are also even local obstructions to lifting to non-regular weight. (There may be local obstructions in other cases as well, although I'm not sure (see § 104).
(3) When F has a real place, the usual conjectures imply that, when c_v is not the "odd" involution, there are obstructions to lifting to regular weights. In the extreme case when c_v is trivial, all lifts should be of trivial weight, and then one can prevent this happening by local (or conjecturally global) reasons similar to those mentioned above.

One can also extend this conjecture to other settings, where one still might conjecture the answer is always no unless one is in a context (regular weight) where $l_0 = 0$.

One caveat is that the case of GL(1) doesn't quite work out. In this case, oddness is automatic and regularity is automatic, but even when F is not totally real there still exist lifts. I think this is too degenerate to really be so persuasive.

The first real case of this conjecture is when F is an imaginary CM field, and then the claim is that there should be representations

$$\overline{\rho} : G_F \to \mathrm{GL}_2(\overline{\mathbf{F}}_p),$$

with no de Rham lifts. To be honest, I don't have anything intelligent to say about how to prove this, I merely wanted to put on the record that I think I used to believe that such lifts might always exist, and now I'm willing to go on the record and conjecture that they don't always exist. And, as I tell my group theory class, half the battle to answering a question in mathematics is determining what you think the right answer should be!

106. A strange continuity

Sun, 25 Nov 2018

Returning to matters opaque (§ 103), , here is the following problem which may well now be approachable by known methods. Let me phrase the conjecture in the case when the prime $p = 2$ and the level $N = 1$.

As we know from Buzzard-Kilford (see [BK05]), in every classical weight κ "close enough" to the boundary of weight space, the slopes of the space of overconvergent forms are given by the arithmetic progression nt where t depends only on the 2-adic valuation of $\kappa(5) - 1$. Now, for each of these overconvergent forms, one obtains a Galois representation

$$\rho_n : G_{\mathbf{Q}} \to \mathrm{GL}_2(\overline{\mathbf{Q}}_2)$$

for every positive integer n. This gives a map from the integers $\mathbf{N}$ considered as a discrete set to $\mathrm{Spf}(R)$ for a deformation ring R (there is only one residual representation in this setting. Yes, it is residually reducible, but ignore this for the moment).

Problem 106.1. Show that this map from $\mathbf{N}$ extends to a **continuous** map from $\mathbf{Z}_2$.

I've never done any computations in these weights, but my spidey senses says it should be true. Naturally, one should also try to work out the most precise statement where N and p are now arbitrary.

I don't have any sense about is whether, for a fixed weight κ, there is actually a representation

$$\rho : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathcal{O}[[T]])$$

(for some $\mathcal{O}$ containing enough roots of unity) whose specialization to $T = n$ for a non-negative integer n is ρ_n, or whether the continuity is not so strong. That might be interesting to check.

More natural questions:

(1) Once one has the correct formulation in fixed weight κ, explain what happens over the entire boundary, and at the halo.
(2) I'm pretty sure that ρ_0 will just be the Eisenstein series, or more accurately the Galois representation $1 \oplus \chi$, where χ is determined from κ in some easy way involving normalizations which I don't want to get wrong. But what is ρ_{-1}? I'm not sure if it is interesting or not. But is there any way of parametrizing this family of Galois representations so that the potentially crystalline points transparently correspond to non-negative integers?

All of this is just to say that, even for $N = 1$ and $p = 2$, there's a lot we don't know about the eigencurve over the boundary of weight space.

Comment 106.2 (Summary)**.** As pointed out in the comments, the paper this paper in particular [Von18, §3.3] is certainly relevant. (Jan was in my AWS group which touched on these questions)

107. Local-global compatibility for imaginary quadratic fields

Thu, 17 Jan 2019

One of the key steps in [ACC+23] is to prove results on local-global compatibility for Galois representations associated to torsion classes. The results proved in that paper, unfortunately, fall well-short of the *optimal* desired local-global compatibility statement, because there are very restrictive conditions on how the relevant primes interact with the corresponding CM field F/F^+. This is not a difficulty when it comes to modularity lifting providing one can replace F by a solvable CM extension H/F where all the required hypotheses hold. However, there are certainly other circumstances where one would like to work with a fixed F without making such a base change. One particularly interesting case is the case when the maximal totally real subfield F^+ is the rational numbers, or equivalently when F is an imaginary quadratic field. There are many reasons to be interested in this case in particular; it relates to classically studied objects (Bianchi groups) and it's one of the very few contexts in which we have optimal results about which homology groups can have interesting torsion (in this case, you only have torsion in degree one). So how restrictive are the local-global theorems in this case? The answer is pretty restrictive — that is, they **never** apply directly. If one is happy to restrict to *residual* representations, however, then there are cheats in some cases.

For example:

Lemma 107.1. *Let F be an imaginary quadratic field in which $p \geq 3$ splits, and suppose that Γ is a congruence subgroup of $\mathrm{GL}_2(\mathcal{O}_F)$ of level N prime to p. Let*

$$\overline{\rho} : G_F \to \mathrm{GL}_2(\overline{\mathbf{F}}_p)$$

be a semi-simple Galois representation associated to a Hecke eigenclass in $H_1(\Gamma, \overline{\mathbf{F}}_p)$. Assume that the image of this representation contains $\mathrm{SL}_2(\mathbf{F}_p)$. Then $\overline{\rho}$ is finite flat at primes dividing p.

The point is as follows. One wants to apply Theorem 4.5.1 of the 10-author paper, but not all the conditions are satisfied. First consider the decomposed generic condition. This is guaranteed (a tedious lemma) by the big image assumption. (In fact, this hypothesis is no doubt much too strong, and possibly — in this setting where F is an imaginary quadratic field — something close to irreducibility should be enough, but I don't really want to bother checking that now.) The more serious hypothesis in 4.5.1 is that a certain inequality holds for the degrees of various local extensions at primes dividing p in F. This inequality **never** holds unless there are at least three primes above p, not something that usually happens for imaginary quadratic fields. But it *is* possible to achieve this via a cyclic extension. For characteristic zero forms, we can appeal to cyclic base change, but this doesn't apply for torsion classes. On the other hand, we see that we *can* achieve a transfer of Galois representations in the case of a cyclic extension of *degree* p, by the main result of this paper (see [TV16]) (I checked with at least one of the authors this preserves the property of having level prime to p). We still have to assume that p

splits in F because another condition of 4.5.1 is that F contains an imaginary field in which p splits, and one can't force this to happen after a cyclic extension H/F of (odd) degree p unless it was true to begin with. So this hypothesis will always be required if one wants to use the results of Venkatesh–Treumann in this way.

It's an intriguing question to ask to what extent this argument could also be applied to $\mathbf{T}/I$ valued representations, where $\mathbf{T}$ is the Hecke algebra acting on mod-p classes and I is some nilpotent ideal with nilpotence of some fixed (absolute) order. This boils down to the corresponding question of how much of $\mathbf{T}$ one sees after the cyclic degree p extension through the Venkatesh–Treumann argument. I don't know the answer to this, but possibly a reader will. (Having done that, there are further tricks available in which one might hope to access the ring $\mathbf{T}$ corresponding to all of $H_1(\Gamma, \mathbf{Z}_p)$ rather than just the p-torsion.)

108. Jacquet–Langlands and a new $R = \mathbf{T}$ conjecture

Wed, 09 Jan 2019

It is somewhat mysterious how one should formulate the Jacquet–Langlands correspondence integrally, particularly in the presence of torsion classes. Even the classical case has many subtleties including for example some results in this paper (see [Rib90]) of Ribet.

In the case of imaginary quadratic fields, Akshay and I (see [CV19]) observed a number of new pathologies that don't occur in the classical case. One of the confusing aspects was how to define a "space of newforms" which might match (in some vague sense) the cohomology of some inner form. I want to discuss here a new conjecture which is very speculative and for which I have absolutely no computational evidence. It started off as a troubling example in my mind where things seemed to go wrong in the setting of my work with Akshay, and this is the result of me trying to put down those concerns in written form. My guiding principle is that $R = \mathbf{T}$ in every situation, so if this doesn't seem to work, you have to find the right definition of R (or $\mathbf{T}$).

Let F be a fixed imaginary quadratic field, say of class number 1, and let P and W be primes (of residue characteristic different from p). Suppose that

$$H_1(\Gamma_0(P), \mathbf{Z}_p)_{\mathfrak{m}} = \mathbf{Z}_p,$$

where localization is done with respect to a non-Eisenstein maximal ideal of the Hecke algebra (assume all Hecke algebras are anemic for now). It can (and does) totally happen that one might have

$$H_1(\Gamma_0(PQ), \mathbf{Z}_p)_{\mathfrak{m}} = \mathbf{Z}_p^2,$$

That is, at level PQ there are two old forms but nothing new either in characteristic zero *or* at the torsion level. In this setting, there are apparently no "newforms" of level PQ, and so one might predict that, on the quaternionic side ramified at PQ, there is no cohomology at this maximal ideal. This is certainly true in characteristic zero by classical Jacquet–Langlands. But it is false integrally! In particular, suppose that the corresponding residual representation

$$\overline{\rho} : G_F \to \mathrm{GL}_2(\mathbf{F}_p)$$

has the property that the image of Frobenius at Q has eigenvalues with ratio N(Q). Then one indeed expects a contribution on the non-split side. Akshay and I managed

to find an interpretation of this result by giving a "better" definition of the space of newforms as the cokernel of a transfer map:

$$\Phi^\vee : H_1(\Gamma_0(P), \mathbf{Z}_p)^2_{\mathfrak{m}} \to H_1(\Gamma_0(PQ), \mathbf{Z}_p)_{\mathfrak{m}},$$

and this can have interesting torsion even in the context above. In fact, by a version of Ihara's Lemma, one can (and we did) compute that the order of the cokernel in this case will be exactly the order of

$$\mathbf{Z}_p/(a_Q^2 - 1 - N(Q))\mathbf{Z}_p,$$

and (again in this precise setting) Akshay and I predicted that this should have the same order as the corresponding localization at the same maximal ideal on the non-split side. (In the Eisenstein case, this is not true, and one sees contributions from various K_2 groups). We even prove a few theorems which prove results of this form taking a product over all maximal ideals of the Hecke algebra.

But even in this example, something a little strange can happen. In particular, I want to argue in this post that **there are two natural definitions of the appropriate global deformation ring, and in order to have a consistent theory, one should consider both of them**. To remind ourselves, we now have two modules, one, defined in terms of the cokernel above, call it M, and then the cohomology localized at the appropriate maximal ideal on the non-split side, which we call M'.

What should we predict about M? The first prediction is that the image of the Hecke algebra should be precisely the universal deformation ring R_Q which records deformations that are Steinberg at Q (and what they should be at the other places). But what does Steinberg at Q even mean for torsion representations? There are basically two types of guesses for the corresponding local deformation ring, and correspondingly two guesses for the associated global deformation ring.

(1) A deformation ring defined in terms of characteristic polynomials. In particular, the maximal quotient of R_Q which corresponds to classes unramified at Q is the unramified deformation ring where the characteristic polynomial of Frob_Q is $(X-1)(X-N(Q))$.
(2) A more restrictive ring in which (on this same unramified quotient) the image of Frob_Q must actually fix a line.

These certainly will have the same points in characteristic zero, but they need not *a priori* coincide integrally. And this will save us below.

Returning to the corresponding global deformation rings (which should be framed, but now ignore the framing), call the corresponding rings R_Q and R'_Q. There is a surjection from R_Q to R'_Q.

Now we make the following conjecture on the smell of an oily rag:

Conjecture 108.1. *The Hecke action on M has image R_Q while the Hecke action on M' has image R'_Q.*

I base this conjecture entirely on the following thought experiment. Let's suppose for convenience that $N(Q)$ is not $-1 \bmod p$. This implies that a_Q is congruent to precisely one of $1 + N(Q)$ or its negative — assume the former. Then the "space of newforms" M as we define it (under all the hypotheses above) will be actually be isomorphic to

$$\mathbf{Z}_p/(a_Q^2 - 1 - N(Q))\mathbf{Z}_p =: \mathbf{Z}_p/\eta\mathbf{Z}_p,$$

because one of the factors will be a direct summand. (The case when $N(Q) = -1 \bmod p$ is no problem but one has to break things up more using the Hecke operator at U_Q which I am ignoring.) So the claim in this case is that R_Q is isomorphic to this ring. What about R'_Q? Let us consider two possibilities. Note that if $N(Q) \neq 1 \bmod p$ then $R_Q = R'_Q$, so we are assuming that $N(Q) \equiv 1 \mod p$ in the examples below.

Example 108.2. Suppose that $a_Q - 1 - N(Q)$ is exactly divisible by p^2, and that

$$\rho(\mathrm{Frob}_Q) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \mod p.$$

In this case, the non-split property implies that the corresponding matrix modulo p^2 will **always** have 1 as an eigenvalue, so the prediction is that $R_Q = R'_Q$.

Example 108.3. Suppose that $a_Q - 1 - N(Q)$ is exactly divisible by p^2, and that

$$\rho(\mathrm{Frob}_Q) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \mod p.$$

In this case, the split condition and the assumption that $a_Q - 1 - N(Q)$ is *exactly divisible by* p^2 force the lift to be of the form

$$\rho(Frob_Q) = \begin{pmatrix} 1 + ap & pb \\ 0 & 1 + cp \end{pmatrix} \mod p^2.$$

where a and c are non-zero. In particular, 1 will **never** be an eigenvalue. So in this case, one predicts that $R_Q = \mathbf{Z}/p^2\mathbf{Z}$ but $R'_Q = \mathbf{Z}/p\mathbf{Z}$.

So how do we see this in terms of $R = \mathbf{T}$ and Jacquet–Langlands and our Conjecture above? First of all, my paper with Akshay suggests indeed that $|M'| = |M| = p^2$, and certainly M' should be an R_Q-module. But now the following should happen:

(1) In Example 108.2, we should have multiplicity one, and so M' should be free of rank 1 over $R_Q = R'_Q$.
(2) In Example 108.3, we should have multiplicity **two**, following Ribet (Helm, Cheng, Manning... [Hel07, Man21]), since multiplicities should be determined by local conditions, and in particular multiplicities should arise exactly when primes which ramify in the quaternion algebra are split and such that the image of the corresponding Frobenius is scalar. Hence M' should be free of rank 2 over R'_Q in this case.

In particular, the Hecke action on M' should factor through R'_Q in both cases, and R_Q does not act faithfully. Perhaps this conjecture is worth a computation!

Notes 108.4. This post (and the next!) are wrong; one should read Note 109.3 below to see what is going on.

109. Jacquet–Langlands and an old $R = \mathbf{T}$ conjecture

Sat, 12 Jan 2019

Edit: this is still incorrect and there should have been a part 3, but I've been distracted ... in conversations with Boxer, Emerton, and Gee shortly after this post, all issues were resolved. Jeff Manning also independently found the correct formulation. (See Note 109.3 below.)

I feel that I should preface this post with the following psychological remark. Occasionally you have the germ of an idea at the back of your mind that you sense is in conflict with your world view. Perhaps you try subconsciously to banish it from your mind, or perhaps you are drawn towards it. But inevitably, the idea breaks through your consciousness and demands to be addressed. The game is now winner takes all — either you can defeat the challenge to your world view, or you will be swallowed up by this new idea and emerge a new person. This is how I came face to face with the non-trivial multiplicities in cohomology for non-split forms of GL(2) over an imaginary quadratic field. Part of me somehow, unconsciously, worried about the conflict between extra multiplicities on the one hand and, on the other hand, the "numerical" equality between the space of "newforms" on the split side with the corresponding space on the non-split side (this equality is not known for each maximal ideal of the Hecke algebra, but rather the "averaged" version over all maximal ideals is the topic of. Then, earlier this week, I turned my face directly towards the problem and admitted its existence, which lead to the previous post. But now ... there may be a way to defeat the beast after all!

Here is the issue. I talked last time about two types of local framed Steinberg deformation rings at $l = 1 \mod p$. The first was defined by imposing conditions on characteristic polynomials, but the second was a more restrictive quotient which demanded the existence of an eigenvalue which was genuinely equal to 1. This modification seemed to pass some consistency checks, and more importantly resolved the compatibility issue between having both the equality |M| = |M'| but also having M be cyclic whilst M' was not. Then I went away for a few days and was distracted by other math, until I flew back to Chicago this evening. While on the plane, I tried to flesh out the argument a little more by writing down more carefully what these two deformation rings R (and its smaller quotient R') were like. And here's the problem. It started to seem as though this quotient R' didn't really exist — after all, demanding the existence of an eigenvector without pinning it down in the residual representation is a dangerous business, and runs into exactly the same issues one sees when trying to give an integral definition of the ordinary deformation ring for $l = p$. Then I thought a little more about the ring R, and it turns out that, for all the natural integral framed deformation rings one writes down, the ring R is a Cohen–Macaulay normal integral domain! In particular, since R' has to be of the same dimension of R, this pretty much forces R to equal R'. So it seems that my last post is completely bogus.

So what then is going on? When you have eliminated the impossible, whatever remains, however improbable, must be the truth. It is impossible that R does not equal $\mathbf{T}$, so I can only conclude the improbable — that even when the representation $\overline{\rho}$ is unramified at l and the image of Frobenius at l under rhobar is scalar, the multiplicity on the quaternionic side ramified at l will **still have multiplicity one** (See Note 109.3!). In other words, the local multiplicity behavior will be sensitive to the archimedean places. This is not what I would (or did) guess, but I cannot see another way around it. So, at the very least, we should investigate this assumption more closely.

Let's talk about two situations where multiplicity two occurs. The first is in the Jacobian $J_1(Np)$ for mod-p representations which are ramified *at* p. In this case, the source of multiplicities is coming from the fact that the local deformation ring R is Cohen–Macaulay but not Gorenstein. On the other hand, the structure of the Tate module is well understood to be of the form $\mathbf{T} \oplus \mathrm{Hom}(\mathbf{T}, \mathbf{Z}_p)$, and so the multiplicity can (ultimately) be read off from the dialyzing module of R. This is what happens in my paper with David Geraghty. The second, which is something I should have paid more attention to last time, is in the work of Jeff Manning (I can't find a working link to either the paper or to Jeff!). The setting of Manning's work is precisely as above: one has $l = 1 \bmod p$ and one is looking at the cohomology of an inner form of $\mathrm{GL}(2)/F$. The only difference is that F is totally real and the geometric object is a Shimura curve. The corresponding local deformation ring R — which is basically the corresponding ring R above — is Cohen–Macaulay but not Gorenstein. On the other hand, one doesn't now know what the structure of the Jacobian is as a module over the Hecke ring. Manning's idea is to exploit the fact that, in his setting, the module M is reflexive (and generically of rank one), and then by studying the class group of R, pin down M exactly. But here is the thing. The reflexivity of M is coming, ultimately, from the fact that the cohomology group H^1 for Shimura curves is **self-dual**. And this is fundamentally **not** the case for these inner forms for $\mathrm{GL}(2)$ over an imaginary quadratic field, where the cohomology is spread between H^1 and H^2. So this is where the archimedean information can change the structure. At this point, I am pretty much obligated to make the following conjecture.

Conjecture 109.1. *For inner forms of* $\mathrm{GL}(2)$ *over an imaginary quadratic field, and for a minimal rhobar which is irreducible and finite flat at primes dividing* $p \geq 2$*, the multiplicity of rhobar in cohomology is one. Moreover, the corresponding module* M' *of this cohomology group localized at this maximal ideal is isomorphic (as* R*-modules and so as* $\mathbf{Z}_p$*-modules) to the space of newforms on the split side, as defined in the last post.*

To put it another way, in Example 108.3 of the previous post, I am now forced to say that $M' = \mathbf{Z}/p^2\mathbf{Z}$ rather than $(\mathbf{Z}/p\mathbf{Z})^2$.

To reiterate from last time — perhaps this conjecture is worth a computation! I guess we shall have to wait a few days to see whether there will be a part 3!

Comment 109.2. (There was some discussion between myself and Aurel Page but in light of the note below there is no need to include it here.)

Notes 109.3. WARNING! There is a fundamental confusion going on in these posts. Imagine a situation where on the split side the space of newforms M was free and of rank one over R, then $M = R$ and $|M| = |R|$ (imagine that R is finite). Now let M' denote the space of forms on the quarternion algebra side. Let's suppose that $|R| < \infty$. The conflict in the past two posts was coming from the following desideratum:

(1) We should have $|M| = |M'|$, following [CV19] (which proves a non-Hecke equivariant version of this equality).
(2) We should often have, under suitable local conditions, that M' has multiplicity > 1 and so M' is not free over R.
(3) The Hecke algebra on the quaternion algebra is still R, so M' is a faithful R-module.

But the point is that these are not in conflict — it's certainly possible that a finite ring R has a module M' which is faithful, has $|R| = |M'|$, and yet M' is not free. For example, $M' = \mathrm{Hom}(R, \mathbf{Q}/\mathbf{Z})$ is faithful with $|M'| = |R|$ and M' is only free if R is Gorenstein. In particular, the mod-p multiplicities should be determined by the local behavior as I first believed (see [Man21]). In fact, M' will more or less be self-dual (when it is finite) by the linking pairing.

110. Jean-Marc Fontaine, 1944-2019

Mon, 04 Mar 2019

The results which generate the most buzz in mathematics are usually those which can be expressed in an elementary (or at least pithy) way to a general mathematical audience. It is certainly true that such results may be profound (see Wiles, Andrew), but this is not always the case. An indirect consequence of this phenomenon is that there are mathematicians who are considered absolute titans of their own field, but who are less well-known by the broader mathematical community. Fontaine, who died this year, might be considered one of these people. Fontaine will forever be associated with p-adic Hodge theory, a subject which is absolutely central to algebraic number theory today. While the initial seed of this subject came from Tate's paper on p-divisible groups, a huge part of its development was due to Fontaine over a period of 30 years (both in his solo papers and in his joint work). The usual audience for my posts is experts, but on the rare chance that someone who knows less p-adic Hodge theory than me reads this post, let me give the briefest hint of an introduction to the subject.

For a smooth manifold M, de Rham's Theorem gives an isomorphism

$$H^n_{\mathrm{dR}}(M) \to H^n(M, \mathbf{R}) = H_n(M, \mathbf{R})^\vee$$

which can more naturally be phrased as that the natural pairing between (classes of) closed forms $[\omega]$ and (classes of) paths $[\gamma]$ given by

$$\langle [\omega], [\gamma] \rangle = \int_\gamma \omega$$

induces a perfect pairing on the corresponding (co-)homology groups. The class of paths in homology has a very natural integral basis coming from the paths themselves. For a general M, the de Rham cohomology has no such basis. On the other hand, if M is (say) the complex points of an algebraic variety over the rational numbers, then there are more algebraic ways to normalize the various flavours of differential forms. To take an example which doesn't quite fit into the world of compact manifolds, take X to be the projective line minus two points, so M is the complex plane minus the origin. There is a particularly nice closed form dz/z on this space which generates the holomorphic differentials. But now if one pairs the rational multiples of this class with the rational multiples of the loop γ around zero, the pairing does *not* land in the rational numbers, since

$$\int_\gamma \frac{dz}{z} = 2\pi i.$$

In particular, to compare de Rham cohomology over the rationals with the usual Betti cohomology over the rationals, one first has to tensor with a bigger ring such

as $\mathbf{C}$, or at least with a ring big enough to see all the integrals which arise in this form. Such integrals are usually called periods, so in order to have a comparison theorem between de Rham cohomology and Betti cohomology over $\mathbf{Q}$, one first has to tensor with a ring of periods.

It is too simplistic to say that p-adic Hodge theory (at least rationally) is a p-adic version of this story, but that is not the worst cartoon picture to keep in your mind. Returning to the example above, note that the period is a purely imaginary number. This is a reflection of the fact that some arithmetic information is still retained, namely, an action of complex conjugation on the complex points of a variety over the rationals is compatible (with a suitable twist) with the de Rham pairing. A fundamental point is that, in the local story, something similar occurs where now the group $\mathrm{Gal}(\mathbf{C}/\mathbf{R}) = \mathbf{Z}/2\mathbf{Z}$ generated by complex conjugation is replaced by the much bigger and more interesting group $\mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$. Very (very) loosely, this is related to the fact that p-adic analysis behaves much better with respect to the Galois group, for example, the conjugate of an infinite (convergent) sum of p-adic numbers is the sum of the conjugates. In particular, there is a Galois action on the ring of all p-adic periods. So now there is a much richer group of symmetries acting on the entire picture. Moreover, the structure of the p-adic differentials can be related to how the variety X looks like when reduced modulo-p, because smoothness in algebraic geometry can naturally be interpreted in terms of differential forms.

So now if one wants to make a p-adic comparison conjecture between (algebraic) de Rham cohomology on the one side, and etale cohomology (the algebraic version of Betti cohomology) on the other side, one (optimally) wants the comparison theorem to respect (as much as possible) all the extra structures that exist in the p-adic world, in particular, the action of the local Galois group on etale cohomology, and the algebraic structures which exist on de Rham cohomology (the Hodge filtration and a Frobenius operator), and secondly, involve tensoring with a ring of periods B which is "as small as possible".

Identifying the correct mechanisms to pass between de Rham cohomology and etale cohomology in a way that is compatible with all of this extra structure is very subtle, and one of the fundamental achievements of Fontaine was really to identify the correct framework in which to phrase the optimal comparison (both in this and also in many related contexts such as crystalline cohomology). (Of course, his work was also instrumental in proving many of these comparison theorems as well.) I think it is fair to say that often the most profound contributions to mathematics come from revealing the underlying structure of what is going on, even if only conjecturally. (To take another random example, take Thurston's insight into the geometry of 3-manifolds.) Moreover, the reliance of modern arithmetic geometry on these tools can not be overestimated — studying global Galois representations without p-adic Hodge theory would be like studying abelian extensions of $\mathbf{Q}$ without using ramification groups.

Two further points I would be remiss in not mentioning: One sense in which the ring B_{dR} is "as small as possible" is the amazing conjecture of Fontaine–Mazur which captures which *global* Galois representations should come from motives. Secondly, Fontaine's work on *all* local Galois representations in terms of (φ, Γ) modules which is crucial even in understanding Motivic Galois representations though p-adic deformations, the fields of norms (with his student Wintenberger, who also sadly died recently), the proof of weak admissibility implies admissibility (with Colmez,

another former student, who surprisingly to me only wrote this one joint paper with Fontaine), and the Fargues-Fontaine curve. (I guess this is more than two.)

Probably the first time I talked with Fontaine was at a conference in Brittany (Roscoff) in 2009. That was the first time I ever gave a talk on my work on even Galois representations and the Fontaine–Mazur conjecture, about which Fontaine had some very kind words to say. (One of the most rewarding parts of academia is getting the respect of people you admire.) I never got to know him too well, due (in equal parts) to my ignorance of the French language and p-adic Hodge theory. But he was always a regular presence at conferences at Luminy with his distinct sense of humour. Over a long career, his work continued to be original and deep. He will be greatly missed.

Comment 110.1 (Persiflage)**.** See also this much more substantial (at every level) remembrance by Colmez and this note in the notices of the AMS.

111. The stable cohomology of $\mathrm{SL}(\mathbf{F}_p)$

Wed, 19 Jun 2019

Today's problem is the following: compute the cohomology of $\mathrm{SL}(\mathbf{F}_p)$ for a (mod-p) algebraic representation.

Step 0 is to say what this problem actually is. It makes sense to talk about certain algebraic representations of $\mathrm{SL}_n(\mathbf{F}_p)$ as n varies (for example, the standard representation or the adjoint representation, etc.). For such representations, one can prove stability phenomena for the corresponding cohomology groups. But my question is whether one can actually **compute** these groups concretely.

The simplest case is the representation $\mathbf{L} = \mathbf{F}_p$ and here one has a complete answer: these cohomology groups are all zero in higher degree, a computation first done by Quillen and which is closely related to the fact that the $K_n(\mathbf{F}_p) \otimes \mathbf{F}_p = 0$. Most of the references I have found for cohomology computations of special linear groups in their natural characteristic consider the case were p is very large compared to n, but let me remind the reader that we are exactly the opposite situation. One of the few references is a paper of Evens and Friedlander from the '80s which computes some very special cases in order to compute $K_3(\mathbf{Z}/p^2\mathbf{Z})$.

Note, however, that p should still be thought of as "large" compared to the partition which defines the corresponding stable local system(s).

In order to get started, let us make the following assumptions:

ANSATZ 111.1. There exists a space X with a $\mathrm{SL}(\mathbf{Z}_p)$ pro-cover such that:

(1) The corresponding completed cohomology groups with $\mathbf{F}_p$ coefficients are $\mathbf{F}_p$ for $i = 0$ and vanish otherwise.
(2) If $\mathbf{L}$ is the mod-p reduction of (an appropriately chosen) lattice in a (**added** non-trivial irreducible) algebraic representation of $\mathrm{SL}(\mathbf{Z}_p)$, then $H^i(X, \mathbf{L}) = 0$ for i small enough compared to the weight of $\mathbf{L}$.

Some version of this is provable in some situations and it may be generally true, but let us ignore this for now. (One explicit example is given by the locally symmetric space for $\mathbf{SL}(\mathbf{Z}[\sqrt{-2}])$ and taking the cover corresponding to a prime $\mathfrak{p}$ of norm p satisfying certain global conditions.) The point is, this anzatz allows us

to start making computations. From the first assumption, one deduces by Lazard that

$$H^i(X(p), \mathbf{F}_p) = \wedge^i M,$$

where M is the adjoint representation. But now one has a Hochschild–Serre spectral sequence:

$$H^i(\mathrm{SL}(\mathbf{F}_p), \mathbf{L} \otimes \wedge^j M) \Rightarrow 0.$$

The point is now that one can now start to unwind this (even knowing nothing about the differentials) and make some conclusions, for example:

(1) $H^1(\mathrm{SL}(\mathbf{F}_p), \mathbf{L}) = 0$.
(2) $H^2(\mathrm{SL}(\mathbf{F}_p), \mathbf{L}) = H^0(\mathrm{SL}(\mathbf{F}_p), \mathbf{L} \otimes M)$.

In particular, the first cohomology always vanishes, and the second cohomology is non-zero only for the adjoint representation where it is one dimensional. (One can see the non-trivial class in H^2 in this case coming from the failure of the tautological representation to lift mod p^2.) Note of course I am not claiming that the first cohomology vanishes for all representations, but only the "algebraic" ones, and even then with p large enough (compared to the weight). Note also that one has to be careful about the choice of lattices, but that is somehow built into the stability — for n fixed, the dual of M is given by trace zero matrices in $M_n(\mathbf{F}_p)$ and so (from the cohomology side) "$\mathbf{L} = M$" is the correct object to consider rather than its dual since the dual is not stable even in degree zero. But I think you can secretly imagine that p is big enough and the weight small enough so that you can choose n so that all these representations are actually irreducible).

The first question is whether 1 & 2 are known results — I couldn't find much literature on these sort of questions (they are certainly consistent with the very special cases considered by Evens and Friedlander).

The second question is what about degrees bigger than 2? For H^3 things start getting a little murkier, but it seems possible that H^3 always vanishes. Beyond that (well even before that) I am just guessing. But one might hope to even come up with a guess the answer which is consistent with the spectral sequence above.

Notes 111.2. There were quite a few useful and relevant remarks in the comments by Will Sawin. But I later gave a talk on this problem at the BIRS workshop *Cohomology of Arithmetic Groups: Duality, Stability, and Computations* which can be viewed here. This lead to some conversations with Oscar Randall-Williams, who then resolved all of these questions in this paper (see [RW22]), see also § 137

112. I asked. . . and you responded!

Mon, 30 Sep 2019

I often ask mathematical questions on this blog that I don't know how to answer. Sometimes my smart readers are able to answer the questions I ask. Surely they deserve some recognition for this? Here are two such occasions which come to mind (one very recent):

In § 78, I asked whether there are infinitely many integers n such that all the odd divisors of $(n^2 + 1)$ *not* of the form 1 mod 2^m for large enough fixed m, and asked whether that was an open problem. The answer: it was then, but no longer! It has

now been answered by Soundararajan and Brüdern in Theorem 4 of this preprint (see [BS21]). Problem solved!

In § 34, I was looking at tables of George Schaeffer at non-liftable weight 1 modular forms of level $\Gamma_1(N)$ for various quadratic characters, and noting that often there were forms with large prime factors. I said:

> I said "However, something peculiar happens in the range of the tables, namely, there is not a single example with N prime. This leads to the (incredibly) vague question: can this be predicted in advance?"

But later George Boxer pointed out to me that when N is not prime, and the corresponding quadratic character (in the tables) is not divisible by a prime q, then the Galois representation at the auxiliary prime q need not be unramified (it can be of Steinberg type) and the corresponding Galois representations can have have significantly larger root discriminant — the ramification index at those primes is $e_q = \ell$ for the residue characteristic ℓ rather than $e_q = 2$. And indeed, looking more closely at the tables, most of the big primes ℓ for which there exist non-liftable forms of level (N, χ) occur when the conductor of χ strictly smaller than N.

113. Read my NSF proposal

Wed, 09 Oct 2019

Since this is NSF season, I took the opportunity to go back and look at some of my old proposals. I am definitely too shy to put my *most recent* proposal online, but I thought it might be interesting to share the very first proposal I ever submitted back in 2006. You can find it here. Honestly, it's not as bad as I might have imagined. Here are some first impressions:

- The first thing that strikes me is that there is no "results from prior support section." In particular, there is a pretty limited discussion of my previous work. It looks like I don't even try to name drop my paper with Matt in Inventiones [CE05] which I been recently accepted before writing this grant; how virtuous.
- I attribute a theorem to "Taylor" which is really a theorem of Taylor and Harris–Soudry–Taylor. Sorry Michael! (I do reference [HSBT10] later on in the proposal.)
- What is claimed in Theorem 3 is not entirely accurate — this was later fixed by my student Vlad Serban in this paper (see [Ser22]), see also § 37.
- It's less than the full 15 pages — Possibly this is an incomplete draft?
- Already in 2006, I had started thinking about the modularity of elliptic curves over imaginary quadratic fields. Many ideas are missing. There is at least one reasonable idea here, however, namely, that if one can prove that the "half" Hida families (taking limits for one prime above $p = \pi\pi'$ but not the other) are flat over $\mathbf{Z}_p$), then one is effectively in an $\ell_0 = 0$ situation. Of course, even today, nobody has any idea how to prove this flatness. The problem is that one can sometimes show that it is pure of co-dimension one over the Iwasawa ring, but then one has to deal with a μ-invariant type question proving that the support over Λ does not contain (p). George Boxer and I occasionally discussed whether it was reasonable

even to conjecture this. I think I am more bullish that it should always be flat, but the question remains open.

- Using poles of (as yet unconstructed) p-adic L-functions to prove lifting criteria from smaller groups is a great idea! I'm sure I discussed this with Matt. If you don't want to find it in the PDF, here is the basic idea. Given an automorphic form π, Langlands explains how (morally) to determine whether it arises via functoriality from a smaller group by considering $L(\pi, \rho, s)$ for every representation ρ and determining the order of vanishing (or the order of poles) of this L-function at $s = 1$. This is the automorphic analog of the group theoretic fact that one can determine a representation V of a group G by knowing not only the dimension of the invariant subspace of V but also of $S(V)$ for every Schur functor applied to V. Actually, it's more than just an analogy, since both are just consequences of the Tannakian formalism (which only conjecturally applies to automorphic forms). For example, a completely concrete example of this is that a cuspidal π for GSp(4) should arise as an induction from $\mathrm{GL}(2)/F$ for a quadratic extension F if and only if $L(\pi \otimes \chi, \rho, s)$ has a pole at $s = 1$ where ρ is the standard 5-dimensional representation and χ is the quadratic character of F. I believe this is even a theorem in this case. The point made in the proposal is that this formalism should apply equally to ordinary Siegel modular forms of non-classical weight, where the consequence of course is the weaker claim that π comes via induction from a non-classical ordinary form ϖ for GL(2). Here is a nice example which suggests that this picture is consistent. Start with a classical ordinary ϖ for GL(2) over an imaginary quadratic field (with some Galois invariance condition on the central character). After inducing, we obtain an ordinary Siegel modular form π such that $L(\pi \otimes \chi, \rho, s)$ has a pole. This should also be true more or less for the p-adic L-function, defined correctly. But now as we vary π over the ordinary family, the locus where the p-adic L-function has a pole should have codimension one. Thus the philosophy predicts a one-dimensional family of ordinary deformations of ϖ. And this is indeed something that Hida proved. But everything we know strongly suggests that this will be a non-classical family in general, so this lifting criterion is something that is really completely different from the classical analog. It also suggests and even partially implies corresponding results for lifting torsion classes as well. I think that this project is definitely something worth pursuing, but I've never learnt enough about p-adic L-functions to do so. Whenever I have talked to someone who has constructed such functions, they are always working in some context where normalizations have been made to ensure that the L-functions are Iwasawa functions and certainly don't have poles. Anyway, I think this remains the most attractive open problem in this proposal.
- Question 2 has been answered (and much more) by Ian Agol [Ago13]. Agol (et. al.) pretty much put an end to the cottage industry of using number theory to answer various special cases of these Thurston conjectures. Interesting problems still remain, of course.
- I haven't had anything really interesting to say about the geometry of the Eigencurve since writing this proposal. But Hansheng Diao and Ruochuan Liu did end up proving that the Eigencurve is indeed proper in this paper.

- I redacted some stuff! There's an idea in this proposal that I might want to give to a graduate student — so I blacked it out (no peeking using secret technologies)
- The broader impact section suffers from the fact that this was my first year as a tenure track assistant professor. But the panel understands that there is only so much you can do at this point. The more senior you are, the more you should be doing.

In the end, I think this was not a bad proposal from a young researcher. There are some good ideas and some good problems. Probably the part on the geometry of the eigencurve is the weakest bit, and that is not unrelated to the fact that I stopped thinking about these types of questions. I think I accomplished less of what I set out to do than for some of my more recent proposals. This is not entirely surprising from looking at the proposal — a (forgivable) weakness is that it's somewhat speculative and optimistic. What did I end up doing instead? Probably my most interesting result in the next cycle was my result with Matt on bounds for spaces of tempered automorphic forms using completed cohomology. This proposal was (in the end) funded — I think I certainly must have benefited from the fact that panels look generously on proposals from people within 5 years (or is it six?) from their PhD ("early career researchers").

Comment 113.1 (Will Sawin)**.** Thanks for posting this! Are you sure that the "fact that one can determine a representation V of a group G by knowing not only the dimension of the invariant subspace of V but also of $S(V)$ for every Schur functor applied to V is, in fact, a fact?

My understanding that this is not true, for any reasonable interpretation of it as a precise mathematical statement, and this poses a difficulty to Langlands' program to develop a Tannakian theory of non-algebraic automorphic forms (though not a difficulty anyone will have to deal with anytime soon, as there are many more pressing issues, like various basic cases of functoriality). For instance I believe that there are two distinct groups G_1, G_2 such that for every (!) irreducible representation of G_1, there is a corresponding irreducible representations of G_2, such that the invariants of all Schur functors on the two representations are equal.

I learned (maybe a slightly different version of) this non-fact from PS (who hasn't commented on this blog).

I said: Ha! I knew when writing these words I was being sloppy and that what I was saying might not literally be true. (I should have made it more vague to cover such a possibility.) But fortunately I have readers to keep me honest!

I do at least know that if V is a n-dimensional irreducible representation and $V \otimes V$ contains at least two one-dimensional summands, then V is induced. (For example, if V preserves a generalized symplectic form of dimension 4 and the corresponding 5-dimensional representation inside $\wedge^2 V$ contains an invariant one-dimensional summand.) This is because the assumptions imply that $\mathrm{Hom}(V, V)$ contains (at least) two one-dimensional summands. By Schur's Lemma, at most one of these one-dimensional summands can be trivial, and thus $V \simeq V \otimes \chi$ for some non-trivial χ, which then implies V is induced.

114. A homework exercise for Oaxaca

Sat, 12 Oct 2019

Here's a homework problem for those coming to Oaxaca who have a facility for working with Breuil-Kisin modules and finite flat group schemes. Let $\mathbf{F}$ be a finite field of characteristic p, and consider a Galois representation:

$$\rho : G_{\mathbf{Q}_p} \to \mathbf{GL}_2(\mathbf{F}).$$

which (one should imagine) is the local restriction of a global representation coming from a modular form. By a standard global argument, one can find a congruent form in weight 2, and thus a lift to a representation which is de Rham with Hodge–Tate weights $[0, 1]$. For almost all such representations one can ensure that lift is potentially crystalline and hence comes from a representation which is potentially Barsotti-Tate. An immediate consequence is that the representation ρ itself is – after restriction to some finite extension K – the generic fibre of a finite flat group scheme. Without any other conditions this is obvious, since one can take K to be the splitting field of ρ. However, the global argument gives a further restriction that one can take $K/\mathbf{Q}_p$ Galois with the property that, for some 2-dimensional representation V_K lifting the restriction of ρ to G_K, there is a representation:

$$\varrho : \Gamma := \mathrm{Gal}(K/\mathbf{Q}_p) \to \mathrm{GL}(D_{\mathrm{cris}}(V_K)) \to \mathrm{GL}_2(\overline{\mathbf{Q}}_p)$$

which is faithful on the inertia subgroup. In particular, this forces Γ and K to be "small" in some sense. One can prove this result directly without recourse to any global arguments. For example, consider the case when ρ is reducible, and, if the ratio of characters is cyclotomic, then additionally assume the extension is not très ramifée. In this case, I claim that one can take K to be the (unramified extension) of $\mathbf{Q}_p(\zeta_p)$ which contains the fixed field of the characters on the diagonal. The restriction of ρ to K is then the extension of the generic fibre of the trivial group scheme by the multiplicative group scheme. But our assumptions imply that the Kummer extension that arises will come from the pth power of a unit and hence come from a finite flat group scheme over K. The (abelian) group $\mathrm{Gal}(K/\mathbf{Q}_p)$ has no problem admitting a representation of small dimension which is faithful on inertia.

When $n = 3$, the automorphic picture would suggest that one can find de Rham lifts with Hodge–Tate weights $[0, 1, 2]$, and this is the type of thing that I guess one knows now in full generality by Emerton–Gee (but probably earlier in this case). But suppose we are still interested in whether there exist lifts of ρ which are potentially Barsotti-Tate. We can ask the weaker question: does ρ come (after restriction to K) from the generic fibre of a finite flat group scheme for a Galois extension $K/\mathbf{Q}_p$ which admits a representation:

$$\varrho : \Gamma := \mathrm{Gal}(K/\mathbf{Q}_p) \to \mathrm{GL}_3(\overline{\mathbf{Q}}_p)$$

which is faithful on inertia? This seems like a question which one should be able to answer. In particular, suppose that ρ is some representation with upper-triangular image. It seems possible that if $K/\mathbf{Q}_p$ is any extension such that ρ is the generic fibre of a finite flat group scheme over K then K might be "too big" to admit such a ϱ. If that were true, this would give a direct proof that ρ does not admit lift which are potentially crystalline with Hodge–Tate weights $[0, 0, 1]$, which would (essentially) answer the final question in this post. (I say "essentially" because one should also consider potentially semistable lifts as well. Certainly one should be

able to address this by similar methods, but for now, perhaps just assume that the ratio of any two consecutive characters occurring in ρ is not cyclotomic.)

This seems to be an eminently answerable question to someone who knows what they are doing, and there are certainly some experts in this sort of computation who will be in Oaxaca in a few weeks time. So maybe one of you can work out the answer (calling the Hawk!).

Comment 114.1 (Anonymous)**.** "the automorphic picture would suggest" —could you clarify how does it suggest that?

Comment 114.2 (Persiflage)**.** via the standard argument: (Globalize to some cuspidal essentially self-dual situation; Generalizations of Serre's conjecture then predict a corresponding eigenclass in $H^*(X, V)$ for some mod-p representation and some suitable X, which, after increasing the level, gives a class in $H^*(X, \mathbf{F}_p)$, which in non-Eisenstein situations should be concentrated in a single degree and so give rise to an appropriate characteristic zero class. You can even take X to be zero dimensional. (I could have given more or less detail but when requests for clarification are anonymous it is hard to respond appropriately.)

Notes 114.3. The homework was too hard, as usual. These problems seem beyond current technology.

115. Appropriate citations

Sun, 27 Oct 2019

Once I wrote a paper (two, in fact) on even Galois representations. The second paper in particular (see [Cal12]) proved what I thought was a fairly definitive result ruling out the existence of a wide class of even de Rham representations with distinct Hodge–Tate weights. It turns out that almost nobody seems to cite these results, probably because they aren't particularly *useful* — at least in any obvious sense. On the other hand, almost everyone who does cite the paper seems to cite it for a specific proposition (3.2) which is an easy consequence of the results of Moret-Bailly. The proposition, more or less, is a *potential* inverse Galois problem with (any finite collection) of local conditions. The main application of such a proposition (both in my paper and in papers which cite it) is that, given a local mod-p representation which *looks* like it could come (say) from the localization of a global representation associated to an automorphic form, the proposition often allows one to produce such a form at the cost of making a finite totally real extension in which p splits completely. This suffices for many purposes.

It turns out, however, that the lemma (pretty much in an equivalent form by an equivalent argument) was already proved by Moret-Bailly himself in this paper (see [MB90]). This means that if you cite my paper for this particular lemma, you should definitely cite the paper of Moret-Bailly. Of course, if you are *also* applying it in a context similar to my paper (say in order to construct automorphic forms with certain local properties), you should certainly feel free to continue to cite my paper as well.

116. En Passant VI

Tue, 29 Oct 2019

I just learnt (from a comment on this blog) that Pierre Colmez hosts a wonderful page on Fontaine and Wintenberger here. I particularly recommend reading both the personal recollections of their friends and collaborators (sample quote from Mark: *These p-adic Hodge theorists seemed to me like an order of monks, who were able to reveal the hidden design of a tapestry by examining it one thread at a time*), as well as this article by Colmez which gives a beautiful introduction to Fontaine's work (rather than my own somewhat superficial summary).

One can't mention the early work of Fontaine in p-adic Hodge theory without also mentioning the recent passing of John Tate (my mathematical grandfather). Tate's enormous contributions to mathematics are very well-known by readers of this blog, many of whom certainly knew him personally much better than me. I first met him at the 2000 Arizona Winter School, where there was an impromptu celebration for his 75th birthday. We crossed paths a few times since then, chatting about a number of things from p-adic modular forms to smoked trout (his wife made a particularly tasty version of the latter for some Harvard holiday party). I last saw him at the banquet for Barry's 80th birthday when he called out my name in a friendly way to say hello, and I felt the flutter of satisfaction that comes when one of your idols remembers who you are. Instead of trying to write a summary of his work, however, let me instead recommend (again) that you purchase for yourself a copy of the Serre–Tate correspondence ([Col15a, Col15b]), also § 83.

117. New results in modularity, Christmas update II

Mon, 30 Dec 2019

Just like last year, once again saint Nick has brought us a bounty of treasures related to Galois representations and automorphic forms in the final week of the year.

First there was this paper by Newton and Thorne [NT21a, NT21b], proving, among other things, the modularity of symmetric powers for a large range of holomorphic modular forms, including Δ and any newform associated to a semistable elliptic curve. There is a lot to enjoy about this paper, not least of which is the nice application of an old computation of Buzzard and Kilford. But there are also some very nice new results on Selmer groups and reducible modularity lifting proved in the substantial related papers by Newton–Thorne [NT23] and Allen-Newton–Thorne respectively [ANT20]. (Also this paper by Thorne and Christos Anastassiades as well [AT22]) It's often hard for the non-specialist to appreciate "technical" improvements on previous theorems, but in this case, they are all wrapped up neatly with a bow by such a clean application: $\mathrm{Sym}^n(\Delta)$ is modular!

Moving on, we have this paper (monograph?) by Liu, Tian, Xiao, Zhang, and Zhu [LTX+22] on the Bloch-Kato conjecture for a very general class of motives associated to Rankin-Selberg convolutions of forms on GL_n and GL_{n-1}. I remember a few years ago talking to Yifeng during his interview at Northwestern (reader, we hired him) about this beautiful paper, giving a totally new argument to study questions of Selmer groups using cycles and level raising congruences. The current paper seems to be not only a version of that on steroids but also with a nice hot cup

of tea with 3 lumps of potassium. It's an amazing achievement which pulls together a lot of wonderful ideas, including Xiao–Zhu's work on the Tate conjecture, not to mention all the previous work on the Gan–Gross–Prasad conjecture.

Well done to both groups of authors!

(In different times I would have given more details as to what these papers actually do, but as my free time nowadays consists of brief moments like this at 5:00AM in the morning you will have to forgive me, and anyway, these papers all seem to be very well written with nice introductions. That said, there will be some more technical mathematics posts coming up, not least of which relates to work of my own students. Stay tuned, Persiflage intends to keep posting!)

Comment 117.1 (David Loeffler)**.** Happy new year from me as well. Just a remark on the paper by Liu et al: the "level-raising congruences" approach to bounding Selmer groups didn't emerge entirely from nowhere in Liu's 2015 paper – it is recognisably a generalisation of earlier work of Bertolini and Darmon [BD05], which does a version of this for Heegner points on modular curves.

Liu was the first to make this idea work in a higher-dimensional case, and now he and his coauthors have pushed this far beyond anything that one could have dreamed possible in 2005 or even in 2015 – it's a magnificent piece of work.

118. The last seven words of Kedlaya–Medvedovsky

Tue, 14 Jan 2020

New paper by my student Noah Taylor! (see [Tay22]) It addresses some conjectures raised by Kedlaya and Medvedovsky in this paper [KM19]. Let $\mathbf{T}$ denote the Hecke algebra acting on modular forms of weight two and prime level N generated by Hecke operators T_p for p prime to N and 2 (the so-called "anemic" Hecke algebra). If $\mathfrak{m}$ is a maximal ideal of $\mathbf{T}$ of residue characteristic two, and $\mathbf{T}/\mathfrak{m} = k$, there exists a corresponding Galois representation:

$$\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{T}/\mathfrak{m}) = \mathrm{GL}_2(k).$$

If S denotes the space of modular forms modulo 2, then certainly $\dim_k(S[\mathfrak{m}]) \geq 1$. Since there can exist congruences between modular forms, it is certainly possible that the generalized $\mathfrak{m}$-eigenspace of S has dimension greater than one. Kedlaya and Medvedovsky observe that if one assumes that $\overline{\rho}$ has (projectively) *dihedral* image, then one can systematically predict lower bounds for this generalized eigenspace contingent on various properties of $\overline{\rho}$. They prove a number of such results, but they finish the paper with what amounts to six more conjectures. Actually, one of the conjectures splits into two completely different cases, and so I like to think of it as seven conjectures.

Before stating the conjectures, first note that $\overline{\rho}$ (when projectively dihedral) is necessarily induced from the field $\mathbf{Q}(\sqrt{\pm N})$. The corresponding representation may or may not be ordinary at the prime 2. Also, let $h(N)$ denote the even part of the class number of $\mathbf{Q}(\sqrt{N})$. Now we can state the conjectures, which are now all proved by Noah:

Theorem 118.1 (Noah Taylor)**.**

(1) Suppose that $N \equiv 1 \pmod 8$. If $\mathfrak{m}$ is $\mathbf{Q}(\sqrt{N})$-dihedral, then the generalized $\mathfrak{m}$-eigenspace has dimension at least 4.

(2) Suppose that $N \equiv 1 \pmod 8$. If $\mathfrak{m}$ is $\mathbf{Q}(\sqrt{-N})$-dihedral, then the generalized $\mathfrak{m}$-eigenspace has dimension at least $h(-N)$.
(3) Suppose that $\mathfrak{m}$ is Eisenstein. Then the generalized $\mathfrak{m}$-eigenspace has dimension at least $(h(-N)-2)/2$.
(4) Suppose that $N \equiv 5 \pmod 8$. If $\mathfrak{m}$ is ordinary-$\mathbf{Q}(\sqrt{N})$-dihedral, then the generalized $\mathfrak{m}$-eigenspace has dimension at least 4.
(5) Suppose that $N \equiv 5 \pmod 8$. If $\mathfrak{m}$ is $\mathbf{Q}(\sqrt{-N})$-dihedral, then the generalized $\mathfrak{m}$-eigenspace has dimension at least 2.
(6) Suppose that $N \equiv 3 \pmod 4$. If $\mathfrak{m}$ is $\mathbf{Q}(\sqrt{N})$-dihedral, then the generalized $\mathfrak{m}$-eigenspace has dimension at least 2.
(7) Suppose that $N \equiv 3 \pmod 4$. If $\mathfrak{m}$ is ordinary-$\mathbf{Q}(\sqrt{-N})$-dihedral, then the generalized $\mathfrak{m}$-eigenspace has dimension at least 2.

Noah uses quite a number of different arguments to prove this theorem. One basic idea is that the extra dimensions are related to deformations of $\overline{\rho}$, but only in some of the proofs is this connection transparent. More directly, Noah exploits the following:

(1) The existence of weight one dihedral representations. When $\overline{\rho}$ is unramified at 2 it is natural to look to such forms. However, even when $\overline{\rho}$ is ramified at two, the weight one forms, after giving rise via congruences to weight two forms, can often be level-lowered to level N using an argument similar to that employed by me and Matt in our paper on the modular degree of elliptic curves.
(2) Known properties of the real points of the Jacobian $J_0(N)$, in particular the connectedness of $J_0(N)(\mathbf{R})$ for prime N as proved by Merel. This can be used to give a lower bound of 2 when $\overline{\rho}$ is totally real. In order to get a better bound in the even case (if necessary) one has to combine this with other arguments.
(3) The difference between the Hecke algebra $\mathbf{T}$ and the Hecke algebra where the operator T_2 is also included. If this Hecke algebra is strictly larger than $\mathbf{T}$ after localization at $\mathfrak{m}$, then one can show that the $\mathfrak{m}$-torsion of S has to be at least two, and moreover one can make this argument work nicely with some of the other methods for producing non-trivial lower bounds.

Concerning the third point: the difference between the Hecke algebra $\mathbf{T}$ and the full Hecke algebra is the addition of the operators T_2 *and* T_N. Noah's arguments crucially use this in the case of T_2 but not of T_N. But this is also explained in the paper: once you add the Hecke operator T_2, it turns out that you have the full Hecke algebra! The fact that the Hecke algebra is integrally generated by T_p for p prime to the level is not true for general levels N but just happens to be true for N prime. It suffices to prove the result after localizing at any maximal ideal $\mathfrak{m}$. Mazur proved it in the Eisenstein case by a somewhat subtle argument (it's false in general for Eisenstein primes at non-prime level). Second, in the non-Eisenstein case, the argument uses the result that all irreducible representations modulo 2 are ramified at N. If there were such a representation, it would be an absolutely irreducible and unramified away from 2, and Tate prove that no such representations exist!

Of course, apropos of the title, this post must finish with the following:

The last seven words of Christ

119. Vesselin Dimitrov on Schinzel–Zassenhaus

Mon, 10 Feb 2020

Suppose that $P(x) \in \mathbf{Z}[x]$ is a monic polynomial. A well-known argument of Kronecker proves that if every complex root of $P(x)$ has absolute value at most 1, then $P(x)$ is cyclotomic. It trivially follows that, for a non-cyclotomic polynomial, the largest root α in absolute value satisfies $|\alpha| > 1$. Elementary considerations imply that this can be improved to

$$|\alpha| > 1 + c_n$$

for some real constant $c_n > 0$ that only depends on the degree. What is the true rate of decay of this parameter as the degree increases? By considering the example $x^n - 2$, the best one can hope for is that c_n can be taken to have the form c/n for some constant c. This is exactly what is predicted by the Schinzel–Zassenhaus conjecture:

Conjecture 119.1 (Schinzel–Zassenhaus)**.** *there is an absolute constant c and a bound*

$$|\alpha| > 1 + \frac{c}{n}$$

for the largest root of all non-cyclotomic polynomials.

In fact, Schinzel–Zassenhaus don't *actually* make this conjecture. Rather, they first prove a bound where c_n has the form 2^{-n} up to a constant, and then go on to say that they "cannot disprove" the claim. And of course, this then gets turned into a conjecture named after them! The best bounds were rapidly improved from exponential to something much better, but the original conjecture remained open. That is, until Vesselin Dimitrov in this paper [Dim19] proved the following:

Theorem 119.2 (Vesselin Dimitrov)**.** *The Schinzel-Zassenhaus conjecture is true.*

Vesselin's result is completely explicit, and gives the effective bound $|\alpha| \geq 2^{1/4n}$, or

$$c_n = 2^{1/4n} - 1 \sim \frac{\log(2)}{4n}.$$

The actual proof is very short. Step 0 is to assume the polynomial is reciprocal, which is a quite reasonable assumption because the conjecture (and much more, including Lehmer's conjecture) was already known by work of Smyth the non-reciprocal case (see [Smy71]). I'm not sure this step is even needed, since the conjecture is certainly true for polynomials whose constant term is not plus or minus one, and so one can simply replace the polynomial by the reciprocal polynomial in what comes below. Step 1 is to show the inclusion

$$\sqrt{\prod (1 - \alpha_i^2/X)(1 - \alpha_i^4/X)} \in \mathbf{Z}[[1/X]].$$

The argument here is elementary (the only prime to worry about is $p = 2$). If the original polynomial is cyclotomic, then this squareroot is actually a polynomial, but otherwise it is a power series which is not rational. But now one has a power series which has an analytical continuation outside a very specific region in the plane, namely the "hedgehog" (I would have called it a spider) consisting of rays from 0 to α^4 and α^2 in $\mathbf{C}$. These rays may overlap, but that only improves the final bound. The complement of the Hedgehog is a simply connected region, and know one wants to say that any power series with integer coefficients that has an analytic

continuation to such a region with sufficiently large transfinite diameter *has to be* rational. Step 2 is then to note that such theorems exist! The transfinite diameter of the region in question can be computed from results already computed in the literature, and the consequent bounds are enough to prove the main theorem, all in no more than a couple of pages! It is a very nice argument indeed. As a comparison, to orient the reader not familiar with Bertrandias' theorem (which is used to deduce rationality of the power series in question), it might be useful to give the following elementary variation. Suppose that instead of the hedgehog, one instead took the complement of the *entire* disc or radius r for some $r < 1$. (Importantly, this does not contain the hedgehog above which has spikes outside the unit circle.) Replacing X by $1/X$, one ends up with a power series on the complement which is a disc of radius greater than one. Now one can apply the following

Theorem 119.3 (Trivial Theorem)**.** *A power series in* $\mathbf{Z}[X]$ *with radius of convergence bigger than one is a polynomial.*

Plugging this into Dimitrov's setup, one deduces a new proof of Kronecker's theorem! So the main technical point is that the "trivial theorem" above can be replaced by a more sophisticated version (to due Bertrandias and many others) where the region of analytic continuation can be taken to be something other than a disc. (For an exposition of some of these rationalization/algebraization theorems, a good point to start is this post of Matt Baker.)

120. Counting solutions to $a_p = \lambda$, Part II

Tue, 03 Mar 2020

This is a sequel to § 72, where the problem of counting eigenforms with $a_p = \lambda$ and $\lambda \neq 0$ was considered. Here we report on recent progress in the case $\lambda = 0$.

It is a somewhat notorious conjecture attributed to Lehmer (who merely asked the question, naturally) that the coefficients of

$$\Delta = q \prod_{n=1}^{\infty} (1 - q^n)^{24} = \sum \tau(n) q^n = q - 24q^2 + 252q^3 + \dots$$

never vanish. One problem with this conjecture is that there really isn't any compelling reason it should be true except (basically) on probability grounds given the growth of the coefficients. As with a number of problems concerning "horizontal" questions about modular eigenforms (fixing the form and varying the prime p), it is often easier to consider the analogous "vertical" question where one fixes p and varies the weight. Namely: fix a tame level, say $\Gamma = \Gamma_1(N)$, fix a p prime to the level, and then consider the eigenforms of level Γ and varying weight with $a_p = 0$. Unlike in the case of Lehmer's conjecture, is certainly can happen that $a_p = 0$, for example:

(1) If f is associated to a modular elliptic curve E with supersingular reduction at $p \geq 5$.
(2) If f has CM by an imaginary quadratic field K in which p is inert.

Let $S(X)$ denote the number of cuspforms of level Γ and weight $\leq X$ such that $a_p = 0$. Consider bounds on this function. The trivial bound, given by counting all cuspforms, is $S(X) \ll X^2$. If you try to improve this bound using analytic

techniques, namely via the trace formula, you can only do very slightly better, say $S(X) \ll X^2/\log(X)$. The problem is that the condition $a_p = 0$ within the space of all spherical representations π_p which could possibly occur has measure zero, so any trace formula approach will have to use a test function where a_p has support in some non-trivial interval $[-\varepsilon, \varepsilon]$ depending on the weight. This is the same problem (more or less) which prevents the analytic approach from giving optimal bounds to the number of weight one modular forms (where now the measure zero condition is being imposed at the infinite prime instead of at the prime p). One approach to improving these bounds is to use non-commutative Iwasawa theoretic methods as employed in my paper with Matt and then used by Simon Marshall to give the first non-trivial bounds for spaces of modular forms for GL(2) over imaginary quadratic fields of fixed level and varying weight. This approach should lead, in principle, to a power saving over the trivial bound.

On the other hand. the best *possible* bound on $S(X)$ will have the shape $S(X) \ll X$, because the number of CM forms of each weight will be bounded independently of the weight, and there is no reason to imagine that the other exceptions will contribute anything of this order. Indeed, in § 72, I conjectured that there should only be *finitely* many such forms of fixed level which are not CM as the weight varies.

When I last visited Madison in 2018, Naser Sardari was working on this problem, and in a preprint from late 2018, he proved exactly a bound of the optimal shape $S(X) \ll X$, with the slight caveat that one should restrict to even weights. Quomodocumque blogged about it here.

Just a few weeks ago, Naser was in town in Chicago, and we got to talking about this problem again. Happily, we were able to come up with one more extra ingredient to push the original result to a (close to) optimal conclusion, and prove the aforementioned conjecture:

Theorem 120.1 (C, Sardari, [CTS21]). *Fix a prime $p > 2$ and a tame level $\Gamma_1(N)$. Then there are only finitely many eigenforms of level Γ and even weight with $a_p = 0$ which are* not *CM.*

This establishes a vertical version of Lehmer's conjecture, up to a congruence on the weight, which arises for a technical reason discussed more below.

The first main idea of the proof is as follows. The p-adic Galois representation associated to ρ_f for a modular form can be very complicated viewed as a representation of the Galois group of $\mathbf{Q}_p$. However, if $a_p = 0$, then the local representation has a very special form: it is induced from an unramified extension $K/\mathbf{Q}_p$. Breuil gave a precise formula for the representation, but a fairly soft argument shows that it is induced – 2-dimensional irreducible crystalline representations over $\mathbf{Q}_p$ are determined by a_p, and twisting by an unramified character fixes both the determinant and the condition $a_p = 0$, hence $V = V \otimes \eta_K$ is induced. That means that one can capture the locus of such representations by a local deformation condition. It is not the case that locally induced implies globally induced, as can be seen from the example of supersingular elliptic curves. This is related to the fact that the map

$$\pi : R^{\mathrm{loc}} \to R^{\mathrm{glob}}$$

of (unrestricted at p) local to global deformation rings is not a surjection. On the other hand, we know in some generality that this is a *finite* map. This was explored in §22, and then more properly taken up in [AC14]. The argument to this point

is now enough to prove the original result of Sardari. Let $R^{\mathrm{loc,ind}}$ denote the local deformation ring of induced representations. If $R = R^{\mathrm{glob}} \otimes_{R^{\mathrm{loc}}} R^{\mathrm{loc,ind}}$ denotes the global deformation ring of locally induced representations, we know that the forms with $a_p = 0$ and a fixed weight are the points of this deformation ring which lie in the fibre over some fixed point in local deformation space. Hence the finiteness of π gives a uniform bound on the number of points in this fibre, and hence a uniform bound over the number of such modular forms in any fixed weight. BTW, for those wondering why there is a restriction on the parity of the weight, it is only really there to prevent the residual representation from being *globally* reducible, a setting in which one doesn't quite yet know the finiteness of π. (When the optimal $R = \mathbf{T}$ theorems become available in the reducible case, our methods should apply without any restrictions.)

Now comes the second ingredient. In order to explain it, let me describe the ring $R^{\mathrm{loc,ind}}$ in more detail, or at least the part coming from inertia. This local deformation ring is basically equal to the deformation ring of the trivial character of G_K, and in particular the ring has the form

$$\mathbf{Z}_p[[\mathcal{O}_K^\times(p)]]$$

where $A(p)$ denotes the maximal pro-p subgroup of A. This ring is isomorphic (at least for odd p) to the Iwasawa algebra $\mathbf{Z}_p[[X,Y]]$ after (via the p-adic logarithm) fixing a choice of multiplicative basis for $\mathcal{O}_K^\times(p)$. Imagine that some component of the global deformation ring (with a locally induced condition) has infinitely many points which correspond to classical non-CM modular forms of level prime to p. The points in weight space correspond to the algebraic characters of the following form:

$$\mathcal{O}_K \to K^\times, \qquad z \mapsto z^n$$

We now exploit the following fact which might (at first) be surprising: any infinite collection of these weights are Zariski dense! To make things a little more concrete, suppose we choose a basis of $\mathcal{O}_K^\times(p)$ of the form $1+p$ and $(1+p)^\eta$, for a suitable $\eta \in \mathcal{O}_K$ which will not be in $\mathbf{Z}_p$, for example, $\sqrt{u}$ for some non-quadratic-residue. The corresponding points with respect to the usual Iwasawa parameters have the shape:

$$X \mapsto (1+p)^n - 1, \qquad Y \mapsto (1+p)^{\eta n} - 1.$$

Instead of proving here why these are Zariski dense, it might be more useful to explain a very close analogy that Naser brought up with Lang's Conjecture: if you take an infinite set of pairs of points of the form $(\exp(x), \exp(\eta x)) \subset (\mathbf{C}^\times)^2$, then they will be Zariski dense whenever $\eta \notin \mathbf{Q}$. In other words, the group subvarieties of the formal torus going through $(X,Y) = (0,0)$ basically all have to be of the form $(1+X)^\eta = (1+Y)$ for $\eta \in \mathbf{Z}_p$. (Coincidentally, the arithmetic applications of Lang's conjecture was the subject of the recent Ahlfors lecture by Peter Sarnak which you can watch here. Our result is yet another application!)

Once your non-CM points are Zariski dense, you are home and hosed: using an idea due to Ghate–Vatsal [GV04], you now specialize at lots of points which are inductions of *finite order* characters. The corresponding Galois representations have finite image on inertia and so are classical by known results. But then (apart from finitely many exceptions) they have to all be CM, because they are classical weight one forms, and the image of inertia is sufficiently large to rule out them having exceptional image.

One might ask whether the results are *effective*. I'm not so sure because of the following issue. Suppose you take $p = 79$ and level one (I'm not sure this case will exhibit the required behavior but it might.) Then you might be able to prove that the global locally-induced deformation ring is (now over all weights) $\mathbf{Z}_p = \Lambda/\mathfrak{P}$. But it might be very hard to tell if that weight $\mathfrak{P}$ corresponds to a classical weight or a random weight, simply because $\mathbf{Z}$ is dense in $\mathbf{Z}_p$. This is not unlike the problem of showing that the zeros of the Kubota–Leopoldt zeta function are not in arithmetic weights.

121. NSF proposal, graduate fellowship edition

Tue, 17 Mar 2020

Note: I feel as a service to the number theory entertainment complex that I should blog more often in these times, even if it means being less coherent than usual. I might even try to get a few guest posts since I won't be going to any conferences any time soon.

I recently linked to my first NSF proposal in § 113, but just today I stumbled upon my *graduate* NSF fellowship application from 1998. There is really only one page which involves any proposal (rather than a list of courses I took or references), and I include the mathematical portion here in full (the only changes from the original are one or two spelling errors and getting the latex to compile. Here we go:

My research interests center mainly around the study of two dimensional Galois representations, the connection of such representations to Modular forms, and application of these connections to the arithmetic of Elliptic curves. Here are several possible questions which are of interest to me.

(1) Serre's conjectures predict that for any odd, absolutely irreducible Galois representation ρ into $\mathrm{GL}_2(\mathbf{F}_q)$, there exists a modular form f which gives rise to ρ, in the sense of Deligne/Shimura/Deligne-Serre. The characteristic zero representation ρ_f, however, need not be defined into $\mathrm{GL}_2(W(\mathbf{F}_q))$, ($W(\mathbf{F}_q)$ = Witt–vectors of $\mathbf{F}_q$), but perhaps in $\mathrm{GL}_2(\mathcal{O})$, for some ramified extension $\mathcal{O}$ of $W(\mathbf{F}_q)$. Is there any sense in which one can quantify the ramification of $\mathcal{O}$ in advance? Is there perhaps a clear cohomological obstruction to a modular lift over $W(\mathbf{F}_q)$? Can one quantify this in terms of some $H^2(G_{\mathbf{Q}}, *)$, or perhaps in terms of R^{univ}, where R^{univ} is the universal deformation ring of ρ? Perhaps if this is too difficult, some qualitative result in this direction?

(2) Applications of the above ideas to rational cuspidal eigenforms of Level 1. Such forms are only known to exist if $\dim S_{2k}(1) = 1$. Can one use ideas from representations to show that no other cuspidal eigenform can be defined over $\mathbf{Q}$?

My first thought is "I guess I haven't changed that much as a mathematician over 22 years" followed by "not a bad problem but too optimistic." The funny thing is that I do think of myself as a number theorist with a certain amount of breadth (despite protestations to the opposite from my most dyspeptic collaborator), so I guess I have to claim that I work on a large circle of ideas and sometimes return to very similar points on the circle. There are also echos in the first proposal of

future work with Matt where we studied the ramification of $\mathcal{O}$ for the reducible representation occurring in weight 2 and prime level N, as studied by Mazur. The most definitive result in that paper was computing the exact ramification degree when $p = 2$, where in the case that the ring was a DVR one had $e = 2^{h-1} - 1$, where 2^h was the order of the 2-part of the class group of $\mathbf{Q}(\sqrt{-N})$. Other progress on this problem more in the spirit of the formulation above has been done by Lundell and Ramakrishna (see [LR11]), although I still think there are many open questions around this problem which are of interest.

On the other hand, the second problem is too optimistic. One reason is related to Buzzard's observation that, in high weight with p fixed, the representations all seemed to be defined over rings with very little ramification. (He goes on to make a conjecture along these lines for which nobody has made any progress.) So it seems unlikely to rule out forms of large weight with coefficients in $\mathbf{Q}$ by showing that there are no such forms over $\mathbf{Q}_p$ because the latter seems to be false in a strong way. The problem of showing there are no more eigenforms over $\mathbf{Q}$ when the weight is at least 24, which is very close to Maeda's Conjecture, is something on which virtually no progress has been made since my proposal, so I guess I don't have to feel bad for not making any progress myself. On the other hand, I don't actually think it is an impossible problem. Maybe I should work on it!

122. More on Lehmer's conjecture

Sat, 21 Mar 2020

Lehmer said it was a "natural question" whether there existed an integer such that $\tau(n) = 0$ or not. I've wondered a little bit recently about how reasonable this is. The historical context is presumably related to the fact that, by the multiplicativity of coefficients, the vanishing of $\tau(p)$ for one prime guarantees that a positive proportion of other coefficients vanish. From the perspective of Galois representations, however, I'm a little confused as to whether we expect any sort of "automorphic" Lehmer's conjecture to hold. To recall, we have

$$\Delta = q \prod_{n=1}^{\infty} (1 - q^n)^{24} = \sum_{n=1}^{\infty} \tau(n) q^n.$$

Deligne's bound says that $|\tau(p)| \leq p^{11/2}$, so a probabilistic argument suggests that there should only be finitely many primes for which $\tau(p) = 0$. Since there aren't any such primes in the first few thousand primes, that's a fairly convincing heuristic for why it might be true. But it's basically impossible to prove anything this way, so one might hope to formulate a more general conjecture which is true for a more systematic reason.

A first attempt might be to ask that $a_p(f) \neq 0$ for any eigenform f of weight $k \geq 3$ and level prime to p which is not CM. (When $k = 2$, of course, there are plenty of modular elliptic curves without CM, and (thanks to Noam) there are plenty of primes p with $a_p(f) = 0$). At first thought this seems a little strong; after all, if we just work in weight 12 (say) then we know that $|a_2(f)| < 2^{11/2} < 46$, so surely if you take enough such forms you should find one with $a_2(f) = 0$. However, this secretly assumes that there are many weight 12 forms with coefficients in $\mathbf{Z}$, and it seems that there are only finitely many such forms. So, for most forms, the

coefficients would lie in (presumably) larger and larger number fields, and there would be more possibilities for $a_2(f)$.

For those who did the computation and might be worried, note that the probabilistic heuristic above only applies when the weight $k \geq 4$. On the other hand, an easy exercise shows that when the weight is odd and the coefficients are integral then the form has CM. The conjecture that there only finitely many non-CM forms with rational coefficients in large even weight is certainly made in this paper, although Dave seems to be hesitant on numerical grounds to make the conjecture for $k = 4$. There seem to be enough forms of weight $k = 4$ and integer coefficients that perhaps there exists a form of odd level with $a_2(f) = 0$. In fact, it should be easy to search for such forms if you can search the LMFDB with a fixed Hecke eigenvalue, which I remember John Voight demonstrating at the Simons Institute general meeting, but I couldn't work out just know when writing this post. Ah, but I guess one can just search for forms with coefficients in $\mathbf{Q}$ and just look at them by hand. It appears that there is a form

$$f = q + 4q^3 - 8q^4 - 5q^5 - 22q^7 - 11q^9 \ldots \in S_4(\Gamma_0(95), \mathbf{Q})$$

with $a_2(f) = 0$. Are there any examples in higher > 4 weight? (Yes, see the notes).

All of this becomes similarly confusing on the level of Galois representations. The modular forms with $a_p(f) = 0$ have the special property that the *local* p-adic Galois representation ρ_f is *induced* from the unique unramified quadratic extension of $\mathbf{Q}_p$. From this perspective, the Lehmer conjecture looks a little bit like Greenberg's conjecture that an ordinary modular form is split if and only if it has complex multiplication. But whereas that conjecture (or at least a stronger version where one assumes such a splitting at all primes of the coefficient field) does follow from plausible conjectures about motives as explained by Matt. I wonder if Matt has any more opinions on what happens if one makes the assumption for only one prime of the coefficient field. Note that if you read Matt's paper, you might be confused why you can't also use Serre–Tate theory to prove that elliptic curves with $a_p = 0$ have CM. I think Florian Herzig gives a nice explanation here of what is going on.

This is also related to the question raised in this post. While that conjecture is not unreasonable, it does skirt the border of conjectures which are actually false, for example, the conjecture that *any* exceptional splitting in a local Galois representation is caused by (more or less) some global splitting. After all, taken to its logical conclusion it would imply not only Lehmer's conjecture but also (combined with Elkies' theorem) that all elliptic curves are CM. Greenberg's conjecture excludes the case of weight one forms, since certainly any form with finite image has many primes for which the local Galois representation is split but the global representation is not if the image is of exceptional type. One can still argue, however, that these forms are *potentially CM*. On the other hand, non-CM Hilbert modular forms of partial weight one, induced to $\mathbf{Q}$, also admit some exceptional splitting on inertia. (Note that non-CM Hilbert modular forms actually exist, as follows from the computation of Moy and Specter [MS15]. While these induced forms are not of regular weight (the HT weights are, up to twist $[0, 2, 2, 4]$, the splitting of the local Galois representation is not explained by any correspondences over any finite extension.

I guess the summary is that all of this discussion points to the fact that Lehmer's conjecture is not true for any good reason beyond random probability grounds and so is kind of rubbish. Actually this reminds me of an experience one occasionally

has after giving a seminar in which you feel like you proved a snazzy result and then the questions from the audience are somewhat deflating. Rest assured, this happens to the best of us — I remember watching a talk online where Richard was talking about his (joint) proof of the Sato–Tate conjecture for Δ, and the only question from the audience was *does this have any implications for Lehmer's conjecture?*

Comment 122.1 (Emmanuel Kowalski)**.** Here's an old blog post of mine that is partially about whether Lehmer's conjecture is reasonable or not:

It contains the bolder statement that maybe the tau function is injective ...

Comment 122.2 (Aurel Page)**.** Here is how to search for a value of a_p in the LMFDB: on the classical modular forms page, first fill in the coefficient field to be 1.1.1.1, then click "Traces Table"; you will get extra search options including "Trace constraints" where you can input $a_2 = 0$,

Notes 122.3. At the time, it wasn't possible to also search for forms which had level prime to 2, since $a_2 = 0$ a lot when $4|N$. But now it is (thanks in part to a ticket request by Aurel), and there are 53 cuspidal eigenforms f of weight ≥ 4 with coefficients in $\mathbf{Z}$, level prime to 2, and $a_2 = 0$ in [LMF24], and all have weight 0. See this search. For $p = 3$, however, there is this form

$$f = q + 14q^5 - 170q^7 + \ldots \in S_6(\Gamma_0(832), \mathbf{Z})$$

with $a_3 = 0$ and weight 6. There is also a form in $S_6(\Gamma_0(66), \mathbf{Z})$ with $a_5 = 0$. Another very amusing example is $f \in S_6(\Gamma_0(390), \mathbf{Z})$, this time with $a_7 = 0$. But look at the coefficients:

$$f = q + 4q^2 - 9q^3 + 16q^4 + 25q^5 - 36q^6 + 64q^8 + 81q^9$$
$$+100q^{10} - 36q^{11} - 144q^{12} + 169q^{13} - 225q^{15} + 256q^{16} + \ldots$$

They are all squares! Alas, the coefficient of q^{17} is 866. Part of the mystery is revealed when considering the level: $2 \cdot 3 \cdot 5 \cdot 13$, by newform theory this forces the identity $a_p = \pm p^2$ for these values, so it's just $a_7 = 0$ and $a_{11} = -36$ which "luckily" turn out to be squares (the other squares come from multiplicativity).

123. Chidambaram on genus two curves, I

Wed, 15 Apr 2020

Those who study elliptic curves certainly know that if you start with an elliptic curve $E/\mathbf{Q}$, the p-torsion gives rise to a Galois representation:

$$\rho : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{F}_p)$$

with cyclotomic determinant. Conversely, if $p = 2, 3, 5$ then the converse is true, that is, any such Galois representation comes from an elliptic curve. Moreover, any such representation comes from an infinite number of curves which are parametrized by $\mathbf{P}^1_{\mathbf{Q}}$. This is intimately related to the fact that the curves $X(p)$ have genus zero for these p.

What is also true is that, given any E, one can write down explicit parameterizations of these families. This was done by Rubin and Silverberg [RS95] for $p = 3, 5$ around the time Fermat's last theorem was proved. Indeed, the idea of passing between elliptic curves with the same mod-3 Galois representation features prominently in Wiles' argument.

One might ask what happens for higher genus. First of all, there is a geometric problem over the complex numbers: when is the moduli space $\mathcal{A}_g(p)$ of PPAV of dimension g with full p-level structure a rational variety? It turns out the only possibilities when $g > 1$ are $p = 2, 3$ when $g = 2$ and $p = 2$ when $g = 3$. The case $(g, p) = (2, 3)$ arose in my work with Boxer, Gee, and Pilloni [BCGP21]. In that paper, we proved a weaker version of the result above, namely the following:

Proposition 123.1 ([BCGP21])**.** *If $\rho : G_{\mathbf{Q}} \to \mathrm{GSp}_4(\mathbf{F}_3)$ is any continuous representation with cyclotomic similitude character, then the corresponding twist $\mathcal{A}_2(\rho)$ is unirational over $\mathbf{Q}$ via a map of degree at most six. In particular, it has many rational points.*

The degree six cover is not so mysterious. When $g = 2$, PPAV are (more or less, the less being the Humbert surface) are Jacobians of genus two curves. So certainly birationally one can replace $\mathcal{A}_2(\rho)$ by $\mathcal{M}_2(\rho)$, the moduli of genus two curves whose Jacobian has the given 3-torsion. The degree six cover is then the moduli of genus two curves with *a fixed Weierstrass point*, or, more prosaically, the genus two curves of the form:

$$y^2 = x^5 + ax^3 + bx^2 + cx + d$$

whose Jacobian has the given 3-torsion. (Any fixed Weierstrass point can be moved to ∞, and then the x^4 term can be suppressed by an obvious linear transformation.) This moduli space will be discussed in more detail in the next post. But for now, this leaves open the question of whether $\mathcal{A}_2(\rho)$ itself is rational.

Over the complex numbers, things are well understood. The space $\mathcal{A}_2(3)$ has a number of compactifications, including the (singular) Satake compactification, and the various smooth toroidal compactifications. When $g = 2$, things work out extra nicely: there is a somewhat canonical compactification $\mathcal{A}_2^*(3)$ due to Igusa. It turns out that $\mathcal{A}_2(\rho)$ is birational to a very nice 3-fold known as the Burkhardt quartic. The Burkhardt quartic is given explicitly in $\mathbf{P}^5$ by the equations:

$$\sigma_1 = x_0 + x_1 + x_2 + x_3 + x_4 + x_5 = 0,$$

$$\sigma_4 = x_0x_1x_2x_3 + \ldots + x_2x_3x_4x_5 = 0.$$

Eliminating any variable using the first equation leads to a quartic in $\mathbf{P}^4$, but this is the most symmetric presentation. This variety $\mathcal{B}$ is singular and has 45-nodes — a maximal number, as it turns out [dJSBVdV90]. Not surprisingly, it also has an action by automorphisms of the simple group $G = \mathrm{PSp}_4(\mathbf{F}_3)$. Blowing up $\mathcal{B}$ at these nodes gives the smooth variety $\mathcal{A}_2^*(3)$.

Things are more subtle over $\mathbf{Q}$. It turns out that for the *trivial* level 3 structure corresponding to the representation $(\mathbf{Z}/3\mathbf{Z})^2 \oplus (\mu_3)^2$ with the obvious symplectic structure, the corresponding variety $\mathcal{A}_2^*(3)$ is still rational (e.g. see here [BN18]). But it is no longer so obvious whether the twists we are considering should be rational over $\mathbf{Q}$ or not. (There are actually some twists of a different flavor which don't have points, but all the ones we are considering do.) Note there is a big difference between what happens in higher dimensions and what happens in dimension one: In dimension one the only unirational smooth projective curve with a rational point is projective space itself, but this is completely false in higher dimensions (for example, take products of projective spaces).

We left the question of the rationality of $\mathcal{A}_2(\rho)$ open in [BCGP]. But my student Shiva Chidambaram took up the question. The first question is how can you prove

a smooth projective variety X is *not* rational over $\mathbf{Q}$ assuming that it is rational over $\mathbf{C}$ and has rational points. One obstruction was found by Manin [Man86]. If X *is* projective space, then the geometric Picard group of X is $\mathbf{Z}$. The Picard group does not always have to be $\mathbf{Z}$ for a smooth rational variety, but Manin showed that, still assuming that X is smooth and projective, if it is birational to projective space then its (geometric) Picard group is *similar* to the trivial representation in a technical sense we now explain. Here we say that two $\mathbf{Z}[G_{\mathbf{Q}}]$-modules A and B (which are free finitely generated abelian groups) are similar if there are integral permutation representations P and Q of $G_{\mathbf{Q}}$ such that

$$A \oplus P \simeq B \oplus Q.$$

(I think one should imagine a sequence of birational maps where one introduces (or removes) the class of some cycle and all of its conjugates.)

Now we can hope to apply this in practice if we can compute the $G_{\mathbf{Q}}$ action on $M = \mathrm{Pic}_{\overline{\mathbf{Q}}}(\mathcal{A}_2^*(\rho))$.

How might one go about computing M? First of all, consider the non-twisted space $\mathcal{A}_2^*(\rho)$. Using the *explicit* geometry of this space, one can hope to go about computing the Neron–Severi group completely explicitly, together with the action of $G = \mathrm{PSp}_4(\mathbf{F}_3)$. And this was indeed done by Hoffman and Weintraub (amongst other things) in this paper (see [HW01]). In particular, they show that the cohomology of this variety is all torsion free, trivial in odd degrees, and satisfies

$$H^2(X, \mathbf{Z}) \simeq H^4(X, \mathbf{Z}) = \mathbf{Z}^{61}.$$

Moreover, the cohomology is entirely generated by cycles, and these cycles are all defined over $E = \mathbf{Q}(\sqrt{-3})$ and can be written down explicitly, together with the corresponding intersection pairing, and the action of the group $G = \mathrm{PSp}_4(\mathbf{F}_3)$ on these cycles is self-evident because of their geometric nature. Clearly the Neron-Severi group of any twist will also be $\mathbf{Z}^{61}$, because the geometric object is the same — the only thing that will change is the Galois action. For this, it is more convenient to work over $E = \mathbf{Q}(\sqrt{-3})$. In this case, the G_E action will be as follows: the projective image of ρ when restricted to G_E factors through $\mathrm{PSp}_4(\mathbf{F}_3)$ given the assumption on the similitude character. Thus ρ gives a canonical map

$$G_E \to G,$$

and the action of G_E on M is simply the restriction of the action of G. Manin's obstruction says that for the variety to be rational over E, the action of G_E has to factor through a representation similar to the trivial representation. But that depends only on the image $H \subset G$. Thus the problem (at least in terms of when we can apply Manin's criterion) is "reduced" to group theory.

Some more caveats: It turns out to be pretty hard to tell if a representation is similar to the trivial representation. There is one obstruction coming from cohomology: using Shapiro's lemma, if H is acting on M by a permutation representation, then

$$H^1(P, M) = H^1(P, M^\vee) = 0, \quad \text{all } P \subset H.$$

But then it follows that the same is true of M is similar to a permutation representation. This gives a way to explicitly verify in some cases that M is *not* similar to a permutation representation by finding a subgroup P for which the group above is non-trivial. Moreover, computing cohomology is something that magma can do! So it remains to:

(1) Explicitly translate the description of Hoffman=-Weintraub into a presentation of $\mathbf{Z}^{61}$ as a $G = \mathrm{PSp}_4(\mathbf{F}_3)$-representation.
(2) Determine for what subgroups P of G one has $H^1(P, M) = H^1(P, M^\vee) = 0$.
(3) Deduce that $\mathcal{A}_2^*(\rho)$ is not rational whenever the projective image contains such a P as above.

The conclusion (see [CC22]):

Theorem 123.2 (C–Chidambaram, [CC22])**.** *For all but* 27 *of the* 116 *conjugacy classes of G, the corresponding twist $\mathcal{A}_2^*(\rho)$ is not rational over $E = \mathbf{Q}(\sqrt{-3})$ and hence not rational over $\mathbf{Q}$ either. In particular, if the projective image over E has order greater than* 20*, the twist is not rational.*

You actually get something stronger from Manin's criterion — if the variety becomes rational over some map of degree d, then the cohomology of the modules must be annihilated by d. From our computations we find, for example:

Theorem 123.3 (C–Chidambaram, [CC22])**.** *Suppose that $\rho : G_{\mathbf{Q}} \to \mathrm{GSp}_4(\mathbf{F}_3)$ is surjective with cyclotomic similitude character. Then the minimal degree of any dominant rational map $\mathbf{P}^3_{\mathbf{Q}} \to \mathcal{A}_2^*(\rho)$ is six.*

Note that from the construction of [BCGP21] we know that there is such a cover of degree six, so the six in this theorem is best possible! (It was good that the computation was consistent with the existence of this cover!)

It turns out that (in the surjective case) one can give a softer argument that only depends on the *rational* representation. The point is that there can still be an obstruction to a rational representation to being a difference of permutation representations. This is easy enough to compute using the character table; you take the group of all virtual representations over $\mathbf{Z}$ and compute the subgroup of all induced representations. For $G = \mathrm{PSp}_4(\mathbf{F}_3)$, this quotient, sometimes called the Burnside cokernel (at least this is what it is called in the magma documentation), turns out to be $\mathbf{Z}/2\mathbf{Z}$ (magma computes it). It's also not so hard to see that there exist subgroups G_{40} and G_{45} of the obvious index such that

$$[H^2(X, \mathbf{Q})] = [G/G_{40}] + [G/G_{45}] - [\chi_{24}],$$

where χ_{24} is the unique representation of G of dimension 24 which also happens to be defined over $\mathbf{Q}$ and also generates the Burnside cokernel. On the other hand, this method this gives weaker results for subgroups of $G = \mathrm{PSp}_4(\mathbf{F}_3)$ and even in the surjective case only shows the minimal cover has degree two rather than six.

123.4. **A word on the actual computation.** Shiva went off and did the task of converting the description in Hoffman–Weintraub into a form which could be used by `magma`. I also went off and tried to do this independently. We then both produced codes (mine much messier) which computed the cohomology of all the subgroups and arrived at completely different answers, which was a bit troubling. But then Shiva pointed out to me that magma automatically does something with matrices that converts right actions to left actions or something like that [could it really be that Magma treats matrices as acting from the right? that sounds crazy], and so his computation of $H^1(P, M)$ was correct, but I was computing $H^1(P, M^\vee)$. But fortunately both are useful! (Of course, one could easily also extract that data from Shiva's code which was much cleaner than mine.)

124. Chidambaram on genus two curves, II

Fri, 24 Apr 2020

We now continue a series of posts on the work of my student Shiva Chidambaram following on from § 123. Today I would like to discuss another project [CCR20] with Shiva that was also joint with David Roberts (no, not David Roberts).

We saw last time that the moduli spaces $\mathcal{A}_2(\rho)$ and $\mathcal{M}_2(\rho)$ are not in general rational over $\mathbf{Q}$. On the other hand, the degree six cover $\mathcal{M}_2^w(\rho)$ is always rational. So the next question is: what is an explicit parametrization? Slightly differently, start with a genus two curve with a Weierstrass point

$$y^2 = x^5 + ax^3 + bx^2 + cx + d$$

Problem 124.1. Parametrize all other genus two curves with a Weierstrass point which have the same 3-torsion representation.

It might be worth briefly revisiting the argument from [BCGP21] that such a parameterization exists. The key point is that there is an birational map

$$\mathcal{M}_2^w(3) \to \mathbf{P}^3$$

which is $\mathrm{PSp}_4(\mathbf{F}_3)$-equivariant. This allows one to show that the corresponding twists are Brauer–Severi varieties, and then deduce they are rational by the same group theoretic trick which appears in this paper of Shepherd-Barron and Richard Taylor. More explicitly, there are maps

$$H^1(\mathbf{Q}, \mathrm{GL}_4(\overline{\mathbf{Q}})) \to H^1(\mathbf{Q}, \mathrm{PGL}_4(\overline{\mathbf{Q}})) \to H^2(\mathbf{Q}, \overline{\mathbf{Q}}^\times)$$

Here the LHS is trivial by Hilbert 90. One shows, using the fact that the Darstellungsgruppe of $\mathrm{PSp}_4(\mathbf{F}_3)$ is $\mathrm{Sp}_4(\mathbf{F}_3)$, that the cocycle corresponding to any Galois representation $\rho : G_{\mathbf{Q}} \to \mathrm{GSp}_4(\mathbf{F}_3)$ with cyclotomic determinant lifts in an explicit way to a cocycle in the LHS and hence is trivial. The problem is to bridge the gap between a theoretical argument that a cocycle is trivial and a way to produce an equation of the corresponding twist. That amounts to the problem of taking a cocycle

$$Z^1(\mathbf{Q}, \mathrm{GL}_4(\overline{\mathbf{Q}}))$$

and writing it as a coboundary. Before going further, it's worth pointing out that the case of $(g, p) = (2, 3)$ is very similar (but more complicated) than the case of $(g, p) = (1, 3)$. In the latter case, one has that

$$\dim H^0(X(3), \omega) = 2,$$

and this simple equality leads to an identification of $X(3)$ with $\mathrm{Proj} H^0(X(3), \omega)$. So, let's talk about the problem of parametrizing elliptic curves as a warm-up case. If you start with a curve

$$E : y^2 = x^3 + ax + b$$

for which (for convenience of exposition) you assume $\rho_{E,3}$ surjective, then the splitting field $L/\mathbf{Q}$ is a $\mathrm{GL}_2(\mathbf{F}_3)$ extension L which contains $F = \mathbf{Q}(\sqrt{-3})$. There is an isomorphism of $H = \mathrm{SL}_2(\mathbf{F}_3)$-modules

$$L = F[H].$$

The group H admits a certain specific 2-dimensional representation V, and the representation ρ can be interpreted as giving an explicit map

$$V \to L.$$

OTOH, the identification above means there is an inclusion $V^2 \to L$ because $\dim(V) = 2$. The problem of solving Hilbert 90 (ignoring a certain descent from F to $\mathbf{Q}$) then becomes the question of finding the "other" extension. Now if you can write L down you can do this by linear algebra. But even in any specific example, L has degree 48, and computations with fields of that size can be pretty formidable and are at the limit of what one can do with explicit number fields in (say) pari/gp or magma.

Of course, one wants to do this over the field $\mathbf{Q}(a, b)$ with general parameters in order to have the general formula. The extension L, for example, is the Galois closure of the equation

$$27y^8 + 216by^6 - 18\Delta y^4 - \Delta^2 = 0,$$

but you probably don't want to even write down a polynomial of degree 48 in these general variables which gives L, let alone try to compute the Galois action. We did succeed in solving this problem by a certain amount of trickery — working in special cases and making the right ansatz for the general case. There were many intermediate formulas which involved polynomials with (say) 100 terms, but the final answer turns out to be perhaps surprisingly simple, namely, the general equation is given by

$$y^2 = x^3 + A(a, b, s, t)x + B(a, b, s, t),$$

where

$$\begin{aligned} 3A(a, b, s, t) =& 3as^4 + 18bs^3t - 6a^2s^2t^2 - 6abst^3 - (a^3 + 9b^2)t^4, \\ 9B(a, b, s, t) =& 9bs^6 - 12a^2s^5t - 45abs^4t^2 - 90b^2s^3t^3 + 15a^2bs^2t^4 \\ & - 2a(2a^3 + 9b^2)st^5 - 3b(a^3 + 6b^2)t^6. \end{aligned}$$

Here $[s, t]$ is the $\mathbf{P}^1$ parameter. Curiously enough this exact formula was also found before in the literature. That reflects something a little surprising about this equation. The moduli space we are looking for is a $\mathbf{P}^1$, and this has many automorphisms. On the other hand, we are starting with an E so we have a fixed point normalized here to be $[1, 0]$. But the projective line with one fixed point still has many automorphisms! However, it turns out that there is some extra hidden structure which gives rise to a second canonical point normalized here as $[0, 1]$, which is why different people would possibly end up with the same equation independently. ($\mathbf{P}^1$ with two fixed points still has a $\mathbf{G}_m$'s worth of automorphisms, but an informal consideration of the integral structure can be used to pin this down further.) The map which takes one E and spits out the *other* point therefore ends up giving a canonical rational map on $\mathbf{P}^1_j$ which has the property that it preserves the (projective) 3-torsion representation. Explicitly it is given by:

$$j \mapsto \frac{(6912 - j)^3}{27j^2}$$

I wonder if this has interesting dynamical properties?

The computation above was not so easy, even though the answer turned out to be simple enough. But for $(g, p) = (2, 3)$ things are looking pretty bad. First of all, the extension L now has degree 103680, which one is not going to write down explicitly. Even the analogue of the degree 8-polynomial above is a degree 40 polynomial in x^6 with 1673 terms.

Despite that, we found the answer:

Theorem 124.2 (C–Chidambaram–Roberts [CCR20])**.** *There exist (and we compute) explicit polynomials A, B, C, D in $\mathbf{Q}[a, b, c, d, s, t, u, v]$ which specialize to a, b, c, d at $[s, t, u, v] = [1, 0, 0, 0]$ such that*

$$y^2 = x^5 + Ax^3 + Bx^2 + Cx + D$$

is the general genus two curve with a rational Weierstrass point and fixed 3-torsion representation.

Even though we find the simplest form of these polynomials, they turn out to be quite big. As in, the number of monomial terms they contain are 14671, 112933, 515454, and 1727921 respectively. The *text* files were so big that I ran into space problems on my university account! (OK, so it's only 200MB or so, but that's a big text file!)

The reason such a computation is ultimately possible relates to an accidental fact that is common between the two cases, namely, that the groups $\mathrm{SL}_2(\mathbf{F}_3)$ and $\mathrm{Sp}_4(\mathbf{F}_3) \times \mathbf{Z}/3\mathbf{Z}$ are two of the 37 exceptional complex reflection groups as determined by Shephard and Todd. The story is explained in our paper [CCR20] so I won't discuss it here, but it might be worth mentioning two further facts:

The first is that these methods can also deal (in principle) with an analogue of this problem for $g = 3$ and $p = 2$. Just as with $g = 2$, the moduli space which admits an equivariant birational map to $\mathbf{P}^6$ is not $\mathcal{M}_3(2)$ but once more a finite cover, and this cover does not correspond to any level structure but rather some cover coming genuinely from the mapping class group. This picture relates to the isomorphism $\mathbf{Z}/2\mathbf{Z} \times \mathrm{Sp}_6(\mathbf{F}_2) \simeq W(E_7)$, another exceptional complex reflection group. There is even a less analogous version for $g = 4$ and $p = 2$ related to the fact that the largest complex reflection group $W(E_8)$ admits a description $W(E_8) \simeq 2.\mathrm{O}_8^+(\mathbf{F}_2) : 2$, and the projective version of this group $\mathrm{O}_8^+(\mathbf{F}_2) : 2$ is a subgroup of $\mathrm{Sp}_8(\mathbf{F}_2)$, although of genuine index (136) rather than as an isomorphism, which is the main reason why this is a little different to the other cases. We estimated that an explicit version of the last moduli problem would involve polynomials with approximately 100 trillion terms, so needless to say we did not try to compute it.

Second, there is an interesting story concerning the auxiliary copy of $\mathbf{Z}/3\mathbf{Z}$ that turns up in the $g = 2$ setting. The formulas that we write down actually correspond not only to projective spaces $\mathbf{P}^1$ and $\mathbf{P}^3$ but actually to affine spaces $\mathbf{A}^2$ and $\mathbf{A}^4$ which represent moduli problems related to the complex reflection groups. In these affine families, not only is the representation corresponding to $\ker(\rho)$ fixed, but the splitting field of $X^3 - \Delta$ also remains unchanged. When $g = 1$, this is not surprising, because the S_3 extension comes from the map $\mathrm{GL}_2(\mathbf{F}_3) = \widetilde{S_4} \to S_4 \to S_3$. On the other hand, that's obviously not happening in the genus two case where the group is almost simple. This is a little peculiar! However, it related to the fact that the splitting field of $X^3 - \Delta$ for genus two curves *depends on the Weierstrass equation*. If you scale the Weierstrass equation by $(x, y) \mapsto (t^2x, t^5y)$, this sends $\Delta \to t^{40}\Delta$. So the affine equation represents a moduli space for some larger group which disappears when considering the equation projectively, and you can always normalize your Weierstrass equation so that Δ is a perfect cube.

125. Picard groups of moduli stacks

Thu, 30 Apr 2020

Here are some algebraic geometry musings related to § 124, most of which is hopefully correct. Everything below is secretly over $\mathbf{Z}[1/6]$ but I think one may as well think about what is happening over $\mathbf{C}$. **Warning:** I don't know any algebraic geometry, please correct me if you see any nonsense.

As mentioned in the last post, if you fix a 3-torsion representation with cyclotomic determinant and look at the corresponding moduli space of elliptic curves with this 3-torsion, you get a $\mathbf{P}^1$ (at least accounting for cusps). A natural followup question is: what geometric object do you get over the stack $\mathcal{A}_1 = \mathcal{M}_{1,1}$? Thinking about stacks in the most naive way, we just consider

$$y^2 = x^3 + ax + b$$

for (a, b) in $\mathbf{P}(4, 6)$ minus $\Delta = 0$ in the stacky sense. But just thinking about this as an elliptic curve over $\mathbf{Q}(a, b)$, you can write down:

$$y^2 = x^3 + Ax + B$$

where

$$\begin{aligned} 3A(a,b,s,t) =& 3as^4 + 18bs^3t - 6a^2s^2t^2 - 6abst^3 - (a^3 + 9b^2)t^4, \\ 9B(a,b,s,t) =& 9bs^6 - 12a^2s^5t - 45abs^4t^2 - 90b^2s^3t^3 + 15a^2bs^2t^4 \\ & - 2a(2a^3 + 9b^2)st^5 - 3b(a^3 + 6b^2)t^6, \end{aligned}$$

Now one thing you notice straight away about these equations is that they **change** when one replaces a, b by $a\lambda^4, b\lambda^6$, namely:

$$A(\lambda^4 a, \lambda^6 b, s, t) = A(a, b, \lambda s, \lambda^3 t)$$

and the same equation holds for B. That is, the parametrization of $\mathbf{P}^1$ changes, and so the family is not literally projective space over this stack. Of course, if

$$\Delta(a, b) = 16(-4a^3 - 27b^2),$$

then

$$\Delta(a\lambda^4, b\lambda^6) = \lambda^{12}\Delta(a, b),$$

where Δ trivializes ω^{12}. In order to remove the ambiguity, one can then define

$$A^*(a, b, s, t) = A\left(a, b, \frac{s}{\Delta^{1/12}}, \frac{t}{\Delta^{3/12}}\right)$$

and similarly with B^*, then the equation is well defined, at least after addressing the issue of taking 12th roots correctly. This suggests that after pulling back to the space where you adjoin $\Delta^{1/12}$ you get projective space, but that the original space is not projective space at all but maybe something like the projective bundle

$$\mathrm{Proj}(\mathcal{O}_X \oplus \omega^2) = \mathrm{Proj}(\omega \oplus \omega^3)$$

where ω is the usual line bundle which has order 12 in the Picard group of $\mathcal{A}_1$.

Something very similar happens for the equations for families of fixed three torsion over $\mathcal{M}_2^w$, the moduli stack of genus two curves with a fixed Weierstrass point. In this case, the base looks like

$$y^2 = x^5 + ax^3 + bx^2 + cx + d$$

or $\mathbf{P}(4, 6, 8, 10)$ minus $\Delta = 0$. (You need to be a little bit more careful at the prime 5.) Here the corresponding identity for A, B, C, D is

$$A(\lambda^4 a, \lambda^6 b, \lambda^8 c, \lambda^{10} d, s, t, u, v) = A(a, b, c, d, \lambda s, \lambda^7 t, \lambda^{13} u, \lambda^{19} v)$$

and

$$\Delta(\lambda^4 a, \lambda^6 b, \lambda^8 c, \lambda^{10} d) = \lambda^{40}\Delta(a, b, c, d).$$

So now one wants to trivialize the family by taking the cover with various roots of Δ, including $\Delta^{1/20}$. Except now I don't really know what the Picard group of $\mathcal{M}_2^w$ is. Somehow I first assumed that the Picard group would be the same as that of the corresponding moduli space of abelian surfaces $\mathcal{A}_2^w$, and since Δ seems to give a trivialization of some power of the determinant bundle it should be related to torsion in $H_1(\Gamma, \mathbf{Z})$ for the corresponding congruence subgroup Γ of $\mathrm{Sp}_4(\mathbf{Z})$. But because of the congruence subgroup property, presumably $H_1(\mathrm{Sp}_4(\mathbf{Z}), \mathbf{Z})$ is equal to $\mathbf{Z}/2\mathbf{Z}$, and that's not going to change by taking the map to $S_6 = \mathrm{PSp}_4(\mathbf{F}_2)$ and taking the pre-image of S_5. But it is pure folly to imagine the Picard group of $\mathcal{M}_2^w$ and $\mathcal{A}_2^w$ coincide. The latter contains an extra divisor, the Humbert divisor, consisting of direct sums of elliptic curves. Moreover, (I guess) the Siegel modular form corresponding to Δ is probably very close to the Igusa form, which vanishes not only at the cusp but also along the Humbert divisor. So the line bundle ω on $\mathcal{A}_2$ has infinite order even though its pullback to $\mathcal{M}_2$ does not because Δ itself is giving a trivialization of some power of ω. So it is indeed plausible that abelianization of the corresponding (index five subgroup of) the $g = 2$ Torelli group has 20-torsion. One way to try to compute this is to explicitly compute the abelianization of the corresponding cover of the mapping class group (I guess there are explicit presentations?). So the first question is can someone confirm that $\mathrm{Pic}(\mathcal{M}_2^w)$ does indeed have 20-torsion? If only there was someone in my department who could *prime* me on the properties of mapping class groups ... Actually, Andrew Putman is probably the obvious person to ask. The second problem is confirm that the family explicitly computed in the last post does indeed coincide with $\mathrm{Proj}(\mathcal{O}_X \oplus \omega^6 \oplus \omega^{12} \oplus \omega^{18})$.

I confess my efforts to do a literature search in this case have not been very thorough. In my mind I somehow thought that the Picard group of the stack $\mathcal{M}_g$ (for $g \geq 2$) was $\mathbf{Z}$, but that is transparently false, at least for $g = 2$. I got as far as doing a google search for Picard groups of moduli stacks and found a few pages of notes written by Daniel Litt. So I naturally zoomed in to Daniel Litt's office hours once after he advertised them on twitter ... but I soon realized that it would take too long to explain and he had better things to do like explaining modular forms to his students ... so here it is now in blog form!

Comment 125.1 (Najmuddin Fakhruddin)**.** $g = 2$ is special, for $g > 2$ the Picard group of $\mathcal{M}_g$ is indeed $\mathbf{Z}$ (at least over $\mathbf{C}$) (see, e.g., Albarello–Cornalba Picard groups of the moduli spaces of curves (see [AC87]). The paper of Arsie–Vistoli Stacks of cyclic covers of projective spaces (see [AV04]) contains a computation of the Picard group of the moduli stack of hyperelliptic curves from which it follows that $\mathrm{Pic}(\mathcal{M}_2) = \mathbf{Z}/10\mathbf{Z}$.

My response: Excellent! So that surely means that ω is the generator 10 and that Δ as a Siegel modular form has weight 10. So now I guess I am implicitly suggesting that ω admits a square root over $\mathcal{M}_2^w$. Naïvely I would have even guess that there is an $\mathcal{O}(1)$ on $\mathbf{P}(4, 6, 8, 10) \smallsetminus (\Delta = 0)$ which had order 40 and may even generate $\mathrm{Pic}(\mathcal{M}_2^w)$ but stacky weighted projective spaces confuse me.

Najmuddin Fakhruddin replies: I don't see any conceptual way of showing that a square root of ω (which is indeed a generator of $\mathrm{Pic}(\mathcal{M}_2)$) exists on $\mathcal{M}_2^w$, but it should be easy to compute the Picard group of $\mathbf{P}(4, 6, 8, 10)\backslash(\Delta = 0)$. In general,

if X is say a smooth variety and G is an algebraic group acting on X, then the Picard group of the quotient stack $[X/G]$ can be computed as follows:

Let V be any linear representation of G on which the action is sufficiently free–to compute the Picard group it suffices to assume that the codimension of the non-free locus is at least two–and let U be the locus on which the action is free. Then $\mathrm{Pic}([X/G]) = \mathrm{Pic}((X \times U)/G)$ where the action is the diagonal action. (Of course, the point here is that $(X \times U)/G$ is a smooth variety.)

If $G = \mathbf{G}_m$, then one can just take $V = \mathbf{A}^2$ with the action $\lambda \cdot (x, y) = (\lambda x, \lambda y)$. In the case of weighted projective space one may even replace $X = \mathbf{A}^n - 0$ by $\mathbf{A}^n$ when computing $\mathrm{Pic}(X \times U/\mathbf{G}_m)$ so the quotient is an affine bundle over $\mathbf{P}^1$. In your example, I think for any irreducible Δ (with the given homogeneity properties) one can then see that the Picard group is indeed $\mathbf{Z}/(40)$.

126. Picard groups of moduli stacks update

Wed, 27 May 2020

A tiny update on § 125. I was chatting with Benson and realized that I may as well ask him directly for a presentation of the mapping class group of a genus two surface. Perhaps unsurprisingly, it can be found in his book with Dan Margalit (see page 122 of their book [FM12] which might be downloadable from a Russian website) and is given as follows:

$$G \simeq \langle a_1, a_2, a_3, a_4, a_5 |\ [a_i, a_j] \text{ for } |i-j| > 1, a_i a_{i+1} a_i = a_{i+1} a_i a_{i+1},$$

$$(a_1 a_2 a_3)^4 = a_5^2, [(a_5 a_4 a_3 a_2 a_1 a_1 a_2 a_3 a_4 a_5), a_1], (a_5 a_4 a_3 a_2 a_1 a_1 a_2 a_3 a_4 a_5)^2, \rangle.$$

The next task is to find the representation

$$G \to \mathrm{Sp}_4(\mathbf{Z}) \to \mathrm{Sp}_4(\mathbf{F}_2) \simeq S_6$$

and then take the index 6 preimage $\Gamma \subset G$ of the $S_5 \subset S_6$ corresponding to fixing a Weierstrass point. Note there are two conjugacy classes of S_5, the correct one is the one whose restriction to A_5 still acts absolutely irreducibly on $(\mathbf{F}_2)^4$. Then one can use Reidemeister–Schreier to compute a presentation of Γ and then compute $H_1(\Gamma, \mathbf{Z})$. This is all good in theory, and Farb-Margalit does have a chapter on the symplectic representation, but actually having to read the book in detail to extract the precise symplectic representation sounded like too much work, especially since all of this is ultimately just for a two sentence comment in a paper that might be removed for space reasons anyway. So instead I just fired up magma with the representation G and asked it to find *all* index six subgroups. It turns out that there are only two of them (up to conjugation), which must come exactly from the two subgroups of $S_5 \subset S_6$. The abelianization of one is $\mathbf{Z}/10\mathbf{Z} \simeq G^{\mathrm{ab}}$, but the other group is $\Gamma = \langle a_1, a_2, a_3, a_4 \rangle$, and one finds that $H_1(\Gamma, \mathbf{Z}) \simeq \mathbf{Z}/40\mathbf{Z}$. Hence this is (in light of the previous discussion) the correct subgroup, and this (unsurprisingly although not entirely independently) confirms the analysis of Najmuddin Fakhruddin in the comments. Now I suspect that if you think a little harder than I am prepared to do (or if you just know a little bit more than me), you might be able to see directly from the definition of the a_i that a_1, a_2, a_3, a_4 fix a Weierstrass point; if you are such a person please make a comment!

127. Families of Hilbert modular forms of partial weight one.

Wed, 03 Jun 2020

Today I would like to talk about a beautiful new theorem of my student Eric Stubley. The first version of Eric's result assumed (unknown) cases of the general Ramanujan conjecture for Hilbert modular forms, and relied on a beautiful idea due to Hida. The final argument, however, is unconditional, and goes beyond Hida's ideas in a way (I hope) that he would be delighted to see.

Suppose that F is a real quadratic field in which $p = vw$ splits. If f is a Hilbert modular form of (paritious) weight $(1, 2k+1)$ and level prime to p, then the corresponding Galois representation (really only defined up to twist):

$$\rho_f : G_F \to \mathrm{GL}_2(\overline{\mathbf{Q}}_p)$$

has the property that, for exactly one of the places $v|p$, the restriction $\rho_f|_{G_v}$ is unramified. Forms of partial weight one are slippery objects — one can construct such forms which are CM, but the existence of any such form which is not CM was open until an example was found by my students Richard Moy and Joel Specter see § 8, 73. They behave in many ways like tempered cohomological automorphic forms for groups without discrete series, more specifically Bianchi modular forms or cohomological forms for $\mathrm{GL}(3)/\mathbf{Q}$. In each of these cases, the invariant l_0 as considered in Calegari–Geraghty (see for example [Cal20, §2.8]) is equal to 1. Following work of Ash–Stevens and Calegari–Mazur, one might consider whether or not f deforms into a family of classical forms. For example, the form f will be ordinary at v, and so it lives in a Hida family $\mathcal{H}$ over $\Lambda = \mathbf{Z}_p[[\mathcal{O}_v^\times(p)]] \simeq \mathbf{Z}_p[[T]]$ where we keep the weight and level at w fixed and consider (nearly) ordinary forms at v. The specialization of this family to *regular* paritious weights will give a space of classical Hilbert modular forms. What can one say about the other specializations in partial weight one?

Theorem 127.1 (Stubley)**.** *Only finitely many partial weight one specializations of the one variable v-adic Hida family $\mathcal{H}$ associated to f are both classical and not CM.*

This gives a completely general rigidity result for all partial weight one Hilbert modular forms in the split case. Over the past decade or so, the prevailing philosophy is that the only algebraic automorphic forms which are not exceedingly rare are either those coming from automorphic forms which are discrete series at infinity, or come from such forms on lower rank groups by functoriality. In this setting, this predicts that non-CM forms of partial weight one should be rare. It might even be plausible to conjecture that, up to twisting, there are only finitely many such forms of fixed tame level. However, such conjectures are completely open, and Stubley's result is one of the first general theorems which points in that direction. (Stronger results for very specific F and p and tame level were obtained by Richard Moy and are discussed in some of the links above.)

One way to think about this theorem is in terms of the Galois representation associated to $\mathcal{H}$. Assume for convenience of exposition that the family is free of rank one over Λ. The Galois representation ρ_f extends to a family:

$$\rho : G_F \to \mathrm{GL}_2(\mathbf{Z}_p[[T]])$$

where $\Lambda = \mathbf{Z}_p[[T]]$ represents weight space, so $T = 0$ corresponds to the original specialization, and $T = \zeta - 1$ for a p-power root of unity ζ corresponds to

a specialization to partial weight one with non-trivial level structure at v. These representations are all nearly ordinary at v. Is it possible that they could be *split* locally at v for infinitely many specializations to partial weight one? Since a non-zero Iwasawa function has only finitely many zeros, this would actually force the local representation to split for *all* T. Moreover, it should imply (and does in many cases) that the specializations $T = \zeta - 1$ are all classical by modularity lifting theorems. Thus, by Stubley's theorem, this can only happen when the family ρ is CM. In particular, Stubley's result implies a theorem (assuming some Taylor–Wiles hypothesis) that a family of Galois representations which is (say) nearly ordinary at w of fixed weight and level and nearly ordinary at v is locally split at v if and only if it is CM.

Experts should recognize the similarity between the Galois theoretic version of Stubley's theorem and the work of Ghate–Vatsal [CV19], who prove that an ordinary family over $\mathbf{Q}$ cannot be locally split unless it is CM. The main ingredient in their proof is the fact that there are only finitely many weight one forms of fixed tame level (up to twist) which are not CM, since these correspond either to A_4, S_4, A_5 extensions of $\mathbf{Q}$ unramified outside a fixed set of primes, which are clearly finite, or real multiplication forms, whose finiteness comes down to the finiteness of the ray class group of conductor $N\mathfrak{p}^\infty$ for a split prime $\mathfrak{p}$ in a real quadratic field. However, the analogous statement for partial weight one forms is completely open as mentioned above, so Stubley's theorem requires a quite different argument.

Before discussing the proof, we first need to discuss a result of Hida (see this paper) (see [Hid16]) about fields of definition of ordinary forms in families. Consider an ordinary family over $\mathbf{Q}$, and consider specializations in some fixed weight, amounting (with some normalization) to specializing T to $\zeta - 1$ for a pth power root of unity. The coefficient field will automatically contain $\mathbf{Q}(\zeta)$. Suppose that for any prime q, the degrees $[\mathbf{Q}(a_q, \zeta) : \mathbf{Q}(\zeta)]$ are bounded for infinitely many specializations. Then Hida proves the family has to be a CM family. Let α_q be one of the corresponding Frobenius eigenvalues. Hida's key insight is to note that α_q is a Weil number, and that Weil numbers over extensions of $\mathbf{Q}(\zeta)$ of uniformly bounded degree are extremely restricted, and in particular given an infinite collection of such numbers then infinitely many of them have to be of the form $\alpha\zeta$ for a fixed α. Using a rigidity lemma fashioned for this very purpose, he then deduces that α_q in the Iwasawa algebra more or less has to equal $\alpha(1+T)^s$ for some $s \in \mathbf{Z}_p$, and this puts enough restrictions on a_q for him to be able to deduce the family is CM.

Stubley's first idea is to use Hida's result in the context of partial weight one forms. The key fact that is different in partial weight one is that when $a_v \neq 0$, the form f is automatically ordinary at v, and hence the $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}(\zeta))$ conjugates of f will still be ordinary at v! This is completely false in regular weights. However, in partial weight one, the only possible (finite) slope of any form at a split prime is 0. As a consequence, the boundedness assumption of Hida's theorem is always going to be satisfied, because all of the conjugates have to lie on one of the finitely many Hida families which all have bounded rank over Λ.

There is, however, a problem. Hida's argument crucially uses the fact that α_q is a Weil number, which uses the Ramanujan conjecture for forms of regular weight. The Ramanujan conjecture is completely open for partial weight one forms, since we have no idea how to prove they occur motivically (nor prove modularity of their symmetric powers). This is where Stubley's second idea comes in. Instead of the

Ramanujan conjecture, one does have standard bounds on the coefficients a_q. This is not enough to deduce that α_q has the form $\alpha\zeta$ for some fixed α. Instead, Stubley shows that it *does* allow one to show that the trace of a_q (together with the trace of any if its powers) to $\mathbf{Q}(\zeta)$ (which has uniformly bounded degree) can be written as a finite sum of roots of unity where the number of terms does not depend on ζ. Again for convenience of exposition and to avoid circumlocutions with traces, let us suppose that the rank of the Hida algebra is one and so $\mathbf{Q}(\zeta, f) = \mathbf{Q}(\zeta)$. Then Eric shows that infinitely many of the a_q satisfy:

$$a_q = \alpha_1\zeta_1 + \alpha_2\zeta_2 + \ldots + \alpha_N\zeta_N$$

for varying p-power roots of unity ζ_i, but where α_i and N are fixed. Then Stubley proves a new rigidity theorem in this context (not unrelated to results of Serban) showing that one must have an equality

$$a_q = \alpha_1(1+T)^{s_1} + \alpha_2(1+T)^{s_2} + \ldots + \alpha_N(1+T)^{s_n}$$

over the Iwasawa algebra. This is probably enough to show the family has to be CM using ideas similar to Hida, but even that is not necessary — by using this formula for specializations in *regular* weight one deduces that the α_i are in $\overline{\mathbf{Q}}$, and then applying Hida's theorem in this fixed regular weight one deduces that the family is CM.

Stubley's theorem is the first result that gives general theoretical evidence towards the conjecture (if one is so bold to make such a conjecture) that there are only finitely many non-CM partial weight one forms of fixed tame level up to twist. It also shows that certain v-ordinary deformations of a non-CM partial weight one form f will not be classical. But there is also a second possible way to deform f, namely, to vary the weight at $w|p$ instead (or as well). For example, if the form f was also ordinary at $w|p$, one could look at the ordinary at w Hida family. One might also conjecture that this family only contains finitely many non-CM points, but this is still open. (Boxer has raised this question.) I think this is an interesting but very hard question!

Notes 127.2. One version of Stubley's results can be found here, [Stu21], although unfortunately I am not sure he plans to submit this to a journal having left mathematics.

128. Chidambaram on Galois representations (not) associated to abelian varieties

Tue, 13 Oct 2020

Today's post is about a new paper [Chi24] by my student Shiva Chidambaram. Suppose that $A/\mathbf{Q}$ is a principally polarized abelian variety of dimension g and p is a prime. The Galois representation on the p-torsion points $A[p]$ gives rise to a Galois representation:

$$\rho : G_{\mathbf{Q}} \to \mathrm{GSp}_{2g}(\mathbf{F}_p)$$

with the property that the similitude character coincides with the mod-p cyclotomic character. A natural question to ask is whether the converse holds. Namely, given such a representation as above with the constraint on the similitude character, does it necessarily come from an abelian variety (principally polarized or not)?

When $g = 1$, the answer is that all such representations come from elliptic curves when $p \leq 5$, but that for $p \geq 7$ there exist representations for any p which do not. For $p \leq 5$, more is true: the twisted modular curves $X(\rho)$ all are isomorphic to $\mathbf{P}^1$. When $p \geq 7$, the curves $X(\rho)$ are of general type, so one might expect a "random" such example to have no rational points. Dieulefait was the first person to find explicit representations (for any such p) which do not come from elliptic curves (and there is a similar result in my paper here) (see [Cal06]). Both of these arguments exploit the Hasse bound. Namely, if

$$\rho : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{F}_p)$$

is unramified at $l \neq p \geq 5$ and ρ comes from $E/\mathbf{Q}$, then E must have either good or multiplicative reduction at l. But this puts a constraint on the possible trace of Frobenius at the prime l. For $l = 2$, for example, this leads to explicit examples of non-elliptic mod-p representations for $p \geq 11$. The case $p = 7$, however, requires a different argument. More generally, while the Hasse argument does generalize to larger g, it only works when p is large compared to g. On the other hand, the Siegel modular varieties $\mathcal{A}_g(p)$ of principal level p are rational over $\mathbf{C}$ for only very few values of g and p. Indeed, they are rational only for

$$(g, p) = (1, 2), (1, 3), (1, 5), (2, 2), (2, 3), (3, 2)$$

whereas $\mathcal{A}_g(p)$ turns out to be of general type for all other such pairs. When (g, p) *is* on this list, then, as discussed in §123, 124, the twists $\mathcal{A}_g(\rho)$ can all be shown to be *unirational* over $\mathbf{Q}$ and so any such representation ρ does indeed come from infinitely many (principally polarized) abelian varieties.

Thus one is left to consider all the remaining pairs. This is exactly the question resolved by Shiva:

Theorem 128.1 (Chidambaram [Chi24])**.** *Suppose that (g, p) is not one of the six pairs above such that $\mathcal{A}_g(p)/\mathbf{C}$ is rational. Then there exists a representation:*

$$\rho : G_{\mathbf{Q}} \to \mathrm{GSp}_{2g}(\mathbf{F}_p)$$

with cyclotomic similitude character which does not come from an abelian variety over $\mathbf{Q}$.

Shiva's argument does not use the Weil bound. Instead, the starting point for his argument is based on the following idea. Start by assuming that ρ comes from an abelian variety A. Suppose also that ρ is ramified at $v \neq p$ and the image of the inertia group at v contains an element of order n for some $(n, p) = 1$. Using this, one deduces (using independence of p arguments) that

$$|\mathrm{Sp}_{2g}(\mathbf{F}_l)| = l^{g^2} \prod_{m=1}^{g} l^{2m} - 1$$

is divisible by n for all large enough primes l, and hence divides the greatest common divisor K_g of all these orders. This is actually a very restrictive condition on n. For example, using Dirichlet's theorem, the number K_g is only divisible by primes at most $2g + 1$. But now if the order of the group $\mathrm{Sp}_{2g}(\mathbf{F}_p)$ for any *particular* p is divisible by a prime power n with n not dividing K_g, then one can hope to construct a mod-p Galois representation whose inertial image at some prime v has order divisible by this n, and this representation cannot come from an abelian variety over $\mathbf{Q}$.

The good news is that one can show that (most) symplectic groups have orders divisible by large primes using Zsigmondy's theorem. Combined with a few extra tricks and calculations for some boundary cases, the groups $\mathrm{Sp}_{2g}(\mathbf{F}_p)$ contain elements of "forbidden" orders *exactly* when one is not in the case of the six exceptional pairs (g, p). Note that Zsigmondy's theorem already arises in the literature in this context in order to understand prime factors of the (corresponding) simple groups.

So now one would be "done" if one could (for example) solve the inverse Galois problem for $\mathrm{GSp}_{2g}(\mathbf{F}_p)$ with local conditions. The inverse Galois problem *is* solved for these groups, but only because there is an obvious source of such representations coming from abelian varieties. Of course, these are precisely the representations Shiva wants to avoid.

Instead Shiva looks for *solvable* groups inside $\mathrm{GSp}_{2g}(\mathbf{F}_p)$ containing elements of order n for suitable large prime powers n. Note that the obvious thing would simply be to take the *cyclic* group generated by the element of the corresponding order. The problem is that there is no way to turn the corresponding representation into a Galois representation whose similitude character is cyclotomic. The groups Shiva actually uses are constructed as follows. Start by finding prime powers $n|p^m + 1$ for some $m \leq g$, then embed the non-split Cartan subgroup of $\mathrm{SL}_2(\mathbf{F}_{p^m})$ into $\mathrm{GSp}_{2g}(\mathbf{F}_p)$, and then consider the normalizer of this image. One finds a particularly nice metabelian subgroup whose similitude character surjects onto $\mathbf{F}_p^\times$. Shiva then has to prove the existence of a number field whose Galois group is this metabelian extension with the desired ramification properties at some auxiliary prime v but also crucially satisfying the cyclotomic similitude character condition. This translates into a (typically) non-split embedding problem — such problems can be quite subtle! Shiva solves it by a nice trick where he relates the obstruction to a similar one which can be shown to vanish using methods related to the proof of the Grunwald-Wang Theorem. Very nice! In retrospect, the case of $g = 1$ and $p = 7$ in my original paper is a special example of Shiva's argument, except it falls into one of the "easy" cases where the relevant metabelian extension actually is a split extension over the cyclotomic field. In general, this only happens when the the maximum power of 2 dividing g is strictly smaller than the maximum power of 2 dividing $p - 1$ which is automatic when g is odd. (The case when $p = 2$ is easier because the cyclotomic similitude character condition disappears!)

129. Hire my students!

Wed, 18 Nov 2020

(**Excised:** some exhortations to hire my graduating students.) Here's a result from Noah's thesis which I haven't discussed before:

Let N be prime, and let $\mathbf{T}$ denote the $\mathbf{Z}_2$-Hecke algebra generated by T_l for l prime to 2, and let $\widetilde{\mathbf{T}}$ denote the Hecke algebra where T_2 is also included. These Hecke algebras are famously not the same in general. For example, when $N = 23$, the space of cusp forms is 2-dimensional and has a pair of conjugate cusp forms as follows:

$$q - \frac{\sqrt{5}+1}{2}q^2 + \sqrt{5}q^3 + \frac{\sqrt{5}-1}{2}q^4 - (1+\sqrt{5})q^5 + \dots$$

So $\mathbf{T} = \mathbf{Z}[\sqrt{5}]$ whereas $\widetilde{\mathbf{T}} = \mathbf{Z}\left[\frac{\sqrt{5}+1}{2}\right]$. Noah gives a formula for the index:

Theorem 129.1 (Noah Taylor [Tay21])**.** *Let N be prime. Then the index $[\widetilde{\mathbf{T}} : \mathbf{T}]$ is given by the order of the space*

$$S_1(\Gamma_0(N), \mathbf{F}_2)$$

of Katz modular forms of weight one and level $\Gamma_0(N)$.

In particular, the index at level 23 is coming from the fact that there is a classical weight one form of this level. From this one sees that the index is non-trivial for all primes $N \equiv 3 \bmod 4$ except for $N = 3, 7, 11, 19, 43, 67$ and 163. For primes $N \equiv 1 \bmod 4$, on the other hand, I might guess that there would be a positive density of primes for which either the index was trivial or non-trivial. The question more or less hinges on the expected number of $\mathbf{SL}_2(\mathbf{F}_{2^n})$ representations of $\mathbf{Q}$ (with $n \geq 2$) which become unramified at all finite places over $\mathbf{Q}(\sqrt{N})$.

130. Ramanujan machine redux

Thu, 11 Feb 2021

I had no intention to discuss the Ramanujan Machine again, but over the past few days there has been a flurry of (attempted) trollish comments on that post, so after taking a brief look at the latest version, I thought I would offer you my updates. (I promise for the last time.)

Probably the nicest thing I have to say about the updated paper is that it is better than the original. My complaints about the tone of the paper remain the same, but I don't think it is necessary for me to revisit them here.

Concerning the intellectual merit, I think it is worth making the following remarks. First, I am only address the contributions to mathematics, Second, what counts as a *new* conjecture is not really as obvious as it sounds. Since continued fractions are somewhat recherché, it might be more helpful to give an analogy with infinite series. Suppose I claimed it was a new result that

$$2G = \sum_{n=0}^{\infty} a_n = 1 + \frac{1}{2} + \frac{5}{36} + \frac{5}{72} + \frac{269}{3600} - \frac{1219}{705600} + \dots$$

where for $n \geq 4$ one has

$$2n^2 a_n = n^2 a_{n-1} - 2(n-2)^2 a_{n-2} + (n-2)^2 a_{n-3}.$$

How can you evaluate this claim? Quite probably this is the first time this result has been written down, and you will not find it anywhere in the literature. But it turns out that

$$\left(\sum_{n=0}^{\infty} \frac{x^n}{2^n}\right) \times \left(\sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)^2}\right) = \sum_{n=0}^{\infty} a_n x^n$$

and letting $x = 1$ recovers the identity above and immediately explains how to prove it. To a mathematician, it is clear that the proof explains not only why the originally identity is true, but also why it is not at all interesting. It arises as more or less a formal manipulation of a definition, with a few minor things thrown in like the sum of a geometric series and facts about which functions satisfy certain

types of ordinary differential equations. The point is that the identities produced by the Ramanujan Machine have all been of this type. That is, upon further scrutiny, they have not yet revealed any new mathematical insights, even if any particular example, depending on what you know, may be more or less tricky to compute.

What then about the *improved* irrationality measures for the Catalan constant? I think that is a polite way of describing a failed attempt to prove that Catalan's constant was irrational. It's something that would be only marginally publishable in a mathematics journal even with a proof. Results about the irrationality measure *in the range where they are irrational* have genuine implications about the arithmetic of the relevant numbers, but these results do not.

What then about the new continued fractions developed over the last year — maybe these are now deeper? Here you have to remember that continued fractions, especially of the kind considered in this paper, are more or less equivalent to questions about certain types of ordinary differential equations and their related periods. (But importantly, not conversely: most of these interesting ODEs have nothing to do with continued fractions since they are associated with recurrences of length greater than two.) For your sake, dear reader, I voluntarily chose to give up an hour or two of my life and took a closer look at one of their "new conjectures." I deliberately chose one that they specifically highlighted in their paper, namely:

$$\frac{2}{-1+2G} = 3 + 0 \times 7 - \frac{6 \times 1^3}{3 + 1 \times 10 - \frac{8\times 2^3}{3+2\times 13 - \frac{10\times 3^3}{\cdots}}}$$

Where G here is Catalan's constant $L(2,\chi_4)$. As you might find unsurprising, once you start to unravel what is going on you find that, just as in the example above, the mystery of these numbers goes away. This example can be generalized in a number of ways without much change to the argument. Let $p_0 = 1$ and $q_0 = 0$, and otherwise let

$$\frac{p_n}{q_n} = \frac{3}{1}, \frac{33}{13}, \frac{765}{313}, \frac{30105}{12453}, \frac{1790775}{743403}, \dots$$

denote the (non-reduced) partial fraction convergents. If

$$P(z) = \sum \frac{4^n p_n z^n}{n!^2} = 1 + 12z + 132z^2 + \dots \quad Q(z) = \sum \frac{4^n q_n z^n}{n!^2} = 4z + 52z^2 + \dots$$

Then, completely formally, $DP(z) = 0$ where

$$D = z(8z-1)(4z-1)\frac{d^2}{dz^2} + (160z^2 - 40z + 1)\frac{d}{dz} + 12(8z-1)$$

and $DQ(z) = 4$. If K and E denote the standard elliptic functions, one observes that $P(z)$ is nothing but the hypergeometric function

$${}_2F_1\left[\begin{matrix}3/2 & 1/2 \\ 1 \end{matrix}; 16z(1-4z)\right] = \frac{2E(16z(1-4z))}{\pi(1-8z)^2}$$

But now one is more or less done! The argument is easily finished with a little help from mathematica. Another solution to $DF(z) = 0$ is of course

$$R(z) = \frac{2E((1-8z)^2) - 2K((1-8z)^2)}{(1-8z)^2} = \log(z) + 2 + \dots$$

and knowing both homogenous solutions allows one to write $Q(z) = u(z)P(z) + v(z)R(z)$ and then easily compute that

$$\lim_{n\to\infty} \frac{p_n}{q_n} = \lim_{z\to 1/8} \frac{P(z)}{Q(z)} = \frac{2}{-1+2G}.$$

as desired. For those playing at home, note that a convenient choice of $u(z)$ and $v(z)$ can be given by

$$v(z) = \int \frac{E(16z(1-4z))}{\pi} = 4z - 8z^2 + \dots$$

$$u(z) = \frac{-1+2G}{2} + \frac{\pi(1-8z)}{2}\left({}_3F_2\left[\begin{matrix}-1/2 & 1/2 & 1/2\\ & 1 & 3/2\end{matrix}; (1-8z)^2\right] - {}_3F_2\left[\begin{matrix}1/2 & 1/2 & 1/2\\ & 1 & 3/2\end{matrix}; (1-8z)^2\right]\right)$$
$$= -4z\log(z) - 4z + \dots$$

Comment 130.1 (Pupshaw)**.** "Here you have to remember that continued fractions, especially of the kind considered in this paper, are more or less equivalent to questions about certain types of ordinary differential equations and their related periods." — I'd love to understand this statement, where might I start?

Comment 130.2 (Persiflage)**.** Suppose you have a Picard–Fuchs equation over $\mathbf{P}^1 \smallsetminus \{0, 1, \infty\}$ with rational coefficients (so an ODE with coefficients in $\mathbf{Q}(z)$. You can expand a basis of solutions around 0 as power series (with perhaps some logarithm terms) but with rational coefficients. The coefficients of the holomorphic solutions will satisfy recurrence relations of the form

$$A_0(n)u_n - A_1(n)u_{n-1} - \dots - A_k(n)U_{n-k} = 0$$

for certain polynomials $A_i(n)$. Given two solutions $P(z)$ and $Q(z)$, say, then since 1 is a singular point, typically what will happen is that $P(z) - \alpha Q(z)$ will have better convergence properties for some α which will imply that the ratio of the coefficients of P and Q converge to α. So how to determine α? Well, you can also find a basis of solutions with rational coefficients in power series expanded around the point $z = 1$. To determine α, you want to write the rational basis around $z = 0$ in terms of the rational basis around $z = 1$, and for a Picard–Fuchs equation you will exactly see the periods arising in this matrix (in particular periods of the degenerate motives in the Picard–Fuchs equations above the singular points).

What is the relation with continued fractions? Well, given any recurrence relation of length 2, i.e. $u_n = R(n)u_{n-1} + S(n)u_{n-2}$, then for any two such sequences p_n and q_n you can write down a generalized continued fraction out of $R(n)$ and $S(n)$ with partial convergents p_n/q_n. Of course, you can do this even if the original ODE is not of geometric type, but then it is a little less obvious what the change of basis matrix will be, although if you know enough about the solutions to the ODE then this might not be a problem.

Comment 130.3 (Will Sawin)**.** And if I understand correctly, you can reverse this process — i.e. given a polynomial continued fraction, find the recurrence satisfied by its convergents, and write down the associated ODE? Then the next step in constructing a proof like the one you have is relating that ODE to ODE's satisfied by classical special functions (e..g hypergeometrics) and calculating some special values of these functions?

Comment 130.4 (Persiflage)**.** There are of course many wrinkles! One particularly fascinating one is that for Picard–Fuchs equations you can have non-singular points where the Mumford–Tate group of the corresponding Motive changes, and this has implications for the special values of the corresponding solutions to the ODE. The classical manifestation of this is of course the hypergeometric function associated to the Legrendre curve and then specialized at CM points. (This coincidentally comes full circle back to some hypergeometric identities observed by Ramanujan.) For generalized hypergeometric functions this degeneration also happens but more sporadically, because the codimension of the special locus is > 1. But it still happens! See, for example, the paper of Dembélé, Panchiskin, Voight, and Zudilin here. (see [DPVZ22]).

131. Test Your Intuition: p-adic local Langlands edition

Tue, 09 Mar 2021

Taking a page from Gil Kalai, here is a question to test your intuition about 2-dimensional crystalline deformation rings.

Fix a representation:

$$\rho : G_{\mathbf{Q}_p} \to \mathrm{GL}_2(\overline{\mathbf{F}}_p)$$

after twisting, let me assume that this representation has a crystalline lift of weight $[0, k]$ for some $1 \leq k \leq p$. Let R denote the universal framed local deformation ring with fixed determinant. Now consider positive integers $n \equiv k \bmod p-1$, and let R_n denote the Kisin crystalline deformation ring also with fixed determinant. Global considerations suggest that for $n \equiv m \equiv k \bmod p-1$ and $n \geq m$, there should be a surjection $R_n/p \to R_m/p$, and quite possibly one even knows this to be true. Global considerations also suggest that any representation can be seen in high enough weight, which leads to the following problem:

Question 131.1. How large does n have to be to see the entire tangent space of the unrestricted local deformation ring R? That is, how large does n have to be for the map $R/(p, \mathfrak{m}^2) \to R_n/(p, \mathfrak{m}^2)$ to be an isomorphism? Naturally, one can also ask the same question with $\mathfrak{m}^2$ replaced by $\mathfrak{m}^k$ for any $k \geq 2$.

The first question came up in a discussion with my student Chengyang. I made a guess, and then we proceeded (during our meeting) to do a test computation on magma, where my prediction utterly failed, but in retrospect my computation itself may have been dodgy so now I'm doubly confused.

Matt remarked that this question is not entirely unrelated in spirit to the Breuil–Mezard conjecture. Instead of counting multiplicities of geometric cycles, one is measuring the Hilbert-Samuel function and its "convergence" to that of the free module. Also, if you know everything about $\mathrm{GL}_2(\mathbf{Q}_p)$ and 2-dimensional Galois representations then you should be able to answer this question too.

Of course I could have re-done the initial computation for this blog post, but I think at least some readers are happier when I ask questions for which I don't know the answer ...

Notes 131.2. This question was taken up by Chengyang in her thesis, see § 152. The (still conjectural) answer is surprising to me — I would have guessed that

with $\mathfrak{m}^k$ the answer would be of order $O(p^k)$, but it turns out (conjecturally, and asyptotically) to have order $O(k^2)$, according to Chengyang's conjectures.

There was also some interaction between me and Vytas Paškūnas (somewhat at cross purposes) in the comments about this question (related to whether one took local deformation rings to include T_p or not. One point Vytas highlighted was (with the standard way of defining the local Kisin deformation rings) whether the map $R_{n+p-1}/p \to R_n$ was surjective. This is still open, although Chengyang's conjectures certainly seem to strongly suggest this is true.

132. Potential automorphy for GL(n)

Thu, 29 Apr 2021

Fresh on the arXiv, a nice new paper (see [Qia23]) by Lie Qian proving potential automorphy results for ordinary Galois representations

$$\rho : G_F \to \mathrm{GL}_n(\mathbf{Q}_p)$$

of regular weight $[0, 1, \dots, n-1]$ for arbitrary CM fields F. The key step in light of the 10-author paper is to construct suitable auxiliary compatible families of Galois representations for which:

(1) The mod-p representation coincides with the one coming from ρ,
(2) The compatible family can itself be shown to be potentially automorphic.

The main result then follows by an application of the p-q switch. Something similar was done by Harris–Shepherd-Barron–Taylor [HSBT10] in the self-dual case. They ultimately found the motives inside the Dwork family. Perhaps surprisingly, Qian also finds his motives in the same Dwork family, except now taken from a part of the cohomology which is not self-dual!

This result doesn't *quite* have immediate implications for the potential modularity of compatible families: If you take a (generically irreducible) compatible family with Hodge–Tate weights $[0, 1, \dots, n-1]$ then one certainly expects (with some assumption on the monodromy group) that the representations are generically ordinary, but this is a notorious open problem even in the analogous case of modular forms of high weight. One way to try to avoid this would be by proving analogous results for non-ordinary representations. But then you run into genuine difficulties trying to find such arbitrary residual representations inside the Dwork family over extensions unramified at p. This difficulty also arises in the self-dual situation, and the ultimate fix in [BLGGT14] was to bypass such questions by applying Khare–Wintenberger lifting style results. However, such lifting results can't immediately be adapted to the $l_0 > 0$ situation under discussion here.

On the other hand, I guess one should be OK for very small n: If M is (say) a rank three motive over $\mathbf{Q}$ with HT weights $[0, 1, 2]$, determinant ε^3, and coefficients in some CM quadratic field E (you have to allow coefficients since otherwise the motive is automatically self-dual, (see § 4), then one is probably in good shape. For example, the characteristic polynomials of Frobenius are Weil numbers α, β, γ of absolute value p and will have (as noted in the blog post linked to in the previous sentence) the shape

$$X^3 - a_p X^2 + \overline{a_p} p X + p^3,$$

and now for primes p which split in E, the corresponding v-adic representation will be ordinary for at least one of the $v|p$ unless a_p is divisible by p, which by purity forces

$$a_p \in \{-3p, -2p, -p, 0, p, 2p, 3p\}.$$

From the usual arguments, one sees that there is at least one ordinary v for almost all split primes p. The rest of the Taylor–Wiles hypotheses should also be generically satisfied assuming the monodromy of M is GL(3), potential modularity in any other case surely being more or less easy to handle directly. Hence Qian thus proves such motives are potentially automorphic. A funny thing about this game is that actually finding examples of non-self dual motives is very difficult, but in this case, van Geemen and Top studied (see [vGT94]) a family of such motives S_t occurring inside H^2 of the surface

$$z^2 = xy(x^2 - 1)(y^2 - 1)(x^2 - y^2 + txy)$$

for varying t (they note that this family was first considered by Ash and Grayson. Also apologies for changing the notation slightly from the paper, but I prefer to denote the parameter of the base by t). They then compare their particular motive when $t = 2$ to an explicit non-self dual form for GL(3)/$\mathbf{Q}$ of level 128. I'm sure by this time (after [HLTT16] and [Sch15b]) someone has verified using the Faltings–Serre method that S_2 is automorphic, but now by Qian's result we know that the S_t are potentially automorphic for all t.

Notes 132.1. The details are carried out in this paper [Mia24]. In particular, [Mia24, Thm 1.1] says that S_t is potentially automorphic.

133. Divisors near $\sqrt{n}$

Tue, 08 Jun 2021

Analytic Number Theory Alert! An even more idle question than normal (that's because it comes from twitter). Alex Kontorovich noted with pleasure the following pictorial representation of the integers from a Veritasium youtube video, where prime numbers are represented by $1 \times n$ rectangles and all other numbers represented as $a \times b$ rectangles (of area n) for some $a > 1$.

This leads to the natural followup questions.

Question 133.1. How much horizontal space does it take to graph the first X integers this way if one either:

(1) Plots the integers n as $a \times b$ with $a \leq b$ as big as possible?
(2) Plot the integers n as $a \times b$ with $a = 1$ if n is prime, and otherwise with a as small as possible, that is, the smallest divisor of n greater than 1?

(From the graph, it actually appears that the second algorithm is actually used.)

In both cases, there is a trivial upper bound $\ll X^{3/2}$. On the other hand, simply by considering products of primes in the interval $[X^{1/2}/C, X^{1/2}]$ for some constant $C > 1$ you get at least a constant times $(X^{1/2}/\log X)^2$ integers less than X with $a \gg X^{1/2}$, and hence a lower bound (in both cases) of $\gg X^{3/2}/(\log X)^2$. But neither of these bounds are presumably best possible. What then are the precise asymptotics? This seems like the type of question Kevin Ford might be able to answer. Actually, this might be a question that Kevin Ford *already* knows how to

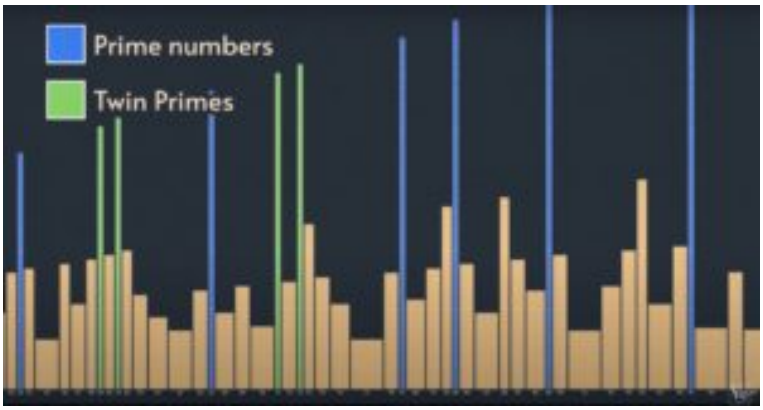

FIGURE 12. Number boxes

answer. I summon his spirit from the whispers of the internet to come and answer this for me. But if that doesn't work, anyone else should feel free to give an answer or make a guess.

Friend of the blog Boaty McBoatface emails me to say: I think the second one is quite easy. I think you just want to compute

$$\sum_{p<X^{1/2}} pF(X/p,p)$$

where $F(y,p)$ is the number of integers $\leq y$ with all prime factors $\geq p$. (This has a name, the Buchstab function). Here the X/p should be $\lfloor X/p \rfloor$ but this is of little consequence. Using the trivial bound $F(y,p) \leq y$ shows that essentially all the contribution is from $p > X^{1/2-\varepsilon}$, and in this range a number $\leq X/p$ has all its prime factors $\geq p$ if and only if it is in fact a prime $\geq p$. So in fact you want to compute

$$\sum_{p\leq X^{1/2}} p\pi(X/p).$$

There are various ways to do this more or less carefully, but by splitting into ranges $cX^{1/2} < p < (c+1/N)X^{1/2}$, summing over $c = 0, 1, 2, \ldots, N-1$ and then letting $N \to \infty$ I think one gets

$$\frac{4X^{3/2}}{(\log X)^2}\int_0^1 (1-c^2)dc \sim \frac{8}{3}\cdot\frac{X^{3/2}}{(\log X)^2}.$$

The first question is definitely much harder and, as you guess, feels pretty close to the kind of stuff Ford and Tenenbaum do in their work.

Comment 133.2 (John Voight)**.** Carl Pomerance points to this paper.

Comment 133.3 (Persiflage)**.** Indeed the Erdös' multiplication problem (how many distinct integers can you form by multiplying two integers less than N) seems closely related, and the work of Tenenbaum also seemed relevant (though I didn't know he had worked on the *exact* problem, giving upper bounds

$$O(X^{3/2}/(\log X)^c(\log\log X)^{1/2})$$

and lower bounds

$$O(X^{3/2}/(\log X)^{c+\varepsilon}).$$

But since then Kevin Ford made significant advances on this circle of problems (see here, [For08]), and now perhaps by these methods one can give the precise growth rate, or at least up to some constant terms. For example, maybe the upper bound above is actually correct up to a constant here. (Note I am lazy and haven't tried to work out if it follows easily from the linked paper above . . .)

134. 59281

Thu, 19 Aug 2021

The target audience of this blog (especially the mathematics) is usually professional mathematicians in the Langlands program. I do sometimes, however, have posts suitable for a broader mathematical audience. Very rarely though do I have anything (possibly) interesting to say to a popular audience. In my recent talk in the Number Theory Web Seminar, I gave a talk about some math that I've discussed with Soundararajan (and which will possibly be written up at some day) about the "average" digit of $1/p$ in its decimal expansion, in particular, discussing the distribution of primes for which the average digit of $1/p$ is less than, equal to, or greater than 4.5 respectively. An easy argument using Cebotarev shows that the density of primes for which the average is exactly 4.5 is 2/3. More subtle, however, is that there are *more* primes for which the average is less than 4.5 than greater than 4.5, but still the (upper and lower) density of primes for which the average is greater than 4.5 is still positive, assuming GRH (the actual percentages of primes with digit average less than, equal to, and greater than 4.5 are approximately 28%, 67%, and 5% respectively).

I think the talk went well, and one reason I suspect is that it was self-contained. Moreover, quite a lot of the setup was completely elementary, although certainly it did move towards deeper topics (Kummer's Conjecture and work of Patterson and Heath-Brown on equidistribution of Gauss sums, and work of Granville–Soundararajan on the distribution of L-values), it was a result that could more or less be appreciated by an undergraduate.

I decided that this was the time — if ever — that I should make a video post. I decided to make a "numberphile" style video — complete with brown paper and a title consisting of a single number — by taking my talk and significantly scaling back the mathematical content. My first attempt was, to put it mildly, a bit of a disaster. First of all, the aspects of making a video that I know nothing about (lighting, audio, glare, video, editing) were unsurprisingly a complete mess and a distraction from the actual mathematics. Second, my resident expert felt that it was still a bit too long, a bit too much like a recording of some lecture, and lacking a hook. So I cut down the script and made a second even more elementary version. This version (unfortunately) no longer has me writing on physical brown paper, but it might at least reach a bare minimum audio/video quality.

What's Special about 59,281?

Just in case you want to skip the video and skip straight to the challenge problem, here it is:

Conjecture 134.1. *Let $p \neq 2, 5$ be prime, and let $C(p)$ denote the average of the digits of $1/p$ in its decimal expansion. (Since the digits repeat this makes sense.) Then the maximum of $C(p)$ for all primes is achieved by $p = 59281$, with:*

$$C(59281) = \frac{486}{95} = 5.11\ldots$$

$$\frac{1}{59281} = 0.\overline{000016868811254870869249843963495892444459438943337662994}$$
$$\overline{888750189774126617297279060744589328 79}$$

A rough heuristic why this should be true: if the period of p is sufficiently large, then, if the digits are sufficiently random, the probability that the average deviates that much from 4.5 becomes exponentially small. Since there are not that many primes with small period, this leads to the heuristic that all but finitely many primes should have $C(p)$ very close to 4.5. Moreover, it suggests finding them as factors of $10^n - 1$ for small values of n. (59281 is a factor of $10^{95} - 1$.) Making the above idea more precise suggests that it is highly unlikely to find a counterexample with period more than 400 or so. Now pari/gp can't factor most numbers of this form even for small n, but there is a second competing heuristic. If p is too *large* and still has small period, then because $1/p$ starts out with a bunch of zeros, this suppresses the digit average. So any big prime factors of $10^n - 1$ that pari/gp doesn't find probably won't be counterexamples anyway. Note this secondary effect also explains why $C(p)$ *can* be significantly less than 4.5 — if $p = (10^q - 1)/9$ is prime, for example, then $C(p) = 9/q$. Since one expects infinitely many primes of this form ($q = 2, 19, 23, 317, 1031, \ldots$) one expects that $C(p)$ can be arbitrarily small.

That said, I certainly have not done any significant computation on this question — possibly `pari/gp` is not finding 10 digit factors of $10^n - 1$ for odd $n < 400$ — it was just an idle question I added to the end of my talk for fun. Hence:

(1) I offer a beer to the person who finds the first counterexample.
(2) I offer a bottle of fine Australian wine to the first person who proves the result. Proofs assuming GRH, for example, are certainly acceptable.

Probably the first thing to try (in order to look for a counter-example) would be to test all primes $p < 10^{10}$ (say) which are factors of $10^n - 1$ for some odd $n < 1000$ or so.

Comment 134.2 (Persiflage)**.** In binary, the corresponding prime could well be $p = 4721$ with period 295, where the string of length 295 has 160 ones and 135 zeros for a corresponding average

$$C_2(4721) = \frac{160}{295} = \frac{32}{59} = 0.542\ldots$$

Notes 134.3. Update from the youtube link: Matthew Bolan has carried out the computation I suggested above, using in addition information about the factorization of $10^n - 1$ for small n given at the Cunningham Project (Jonathan Webster told me about this link). The current records for the primes p with the six highest values of $C(p) = A(1/p)$ are given in the following table. (I had already found the four smallest of these primes in my initial search.) After this computation, it looks like my beer is pretty safe!

Prime	$C(p) = A(1/p)$	Period
59281	5.115789474	95
307627	4.898734177	79
9269802917	4.866028708	209
53	4.846153846	13
173	4.813953488	43
561470969	4.803108808	193

135. Polymath proposal: 4-folds of Mumford's type

Tue, 24 Aug 2021

Let A/K be an abelian variety of dimension g over a number field. If $g \not\equiv 0 \bmod 4$ and $\mathrm{End}(A/\mathbf{C}) = \mathbf{Z}$, then Serre proved that the Galois representations associated to A have open image in $\mathrm{GSp}_{2g}(\mathbf{Z}_p)$. The result is not true, however, when $g = 4$, as first noted by Mumford (in this paper) (see [Mum69]).

The goal of this polymath project is to find an "explicit" example of such a Mumford 4-fold over $\mathbf{Q}$. There are a number of things I have in mind for what "explicit" might mean (this is, after all, supposed to be a polymath project so I'm not supposed to know how to do everything). But here is one way: associated to A is a compatible family of Galois representations

$$\rho_p : G_{\mathbf{Q}} \to \mathrm{GSp}_8(\mathbf{Z}_p)$$

such that, for some integer N, the Galois representations ρ_p are unramified outside Np, and for all other primes q the characteristic polynomial of $\rho_p(\mathrm{Frob}_q)$ is equal to

$$Q_q(T) \in \mathbf{Z}[T]$$

for some polynomial which does not depend on p. Then for example one could hope to give a list of the polynomials $Q_q(T)$ for a collection of primes q.

Here is the strategy to find such Galois representations. We start by choosing a totally real cubic field, which for reasons to possibly be explained later should perhaps be $F = \mathbf{Q}(\zeta_7)^+$. (One reason: it is the Galois cubic field of smallest possible discriminant.)

Step I: Find a Hilbert modular form over F of weight $(1, 1, 2)$ with coefficients in F.

The idea here will be to follow the strategy employed by [MS15] (following Schaeffer, [Sch15a]) to compute a Hilbert modular form of weight $(1, 3)$ over the field $\mathbf{Q}(\sqrt{5})$. Namely, Let W denote the space of Hilbert modular forms of weight $(2, 2, 3)$ of some fixed level. Now divide by some suitable Eisenstein series of weight $(1, 1, 1)$ to get a space V of meromorphic forms of weight $(1, 1, 2)$. This will contain the (possibly zero) space U of holomorphic forms of weight $(1, 1, 2)$. The holomorphic forms will be preserved under the action of Hecke operators whereas V in general will not be. Hence one can start computing the intersection of V with its Hecke translates, which will also contain U. Either you eventually get zero, or you (most likely) end up with an eigenform which you can hope to prove is holomorphic by proving its square is holomorphic.

Some Issues: The way that Moy–Specter compute the (analogue) of W is to use Dembélé's programs to compute the Hecke eigensystems of that weight, and

then use the fact that q-expansions are determined by the Hecke eigenvalues for Hilbert modular forms (suitably interpreted, one has to compute spaces of old forms of lower level etc.). The same idea should certainly work, but note that we are working here in non-paritious weight (that is, not all weights are congruent modulo 2). My memory is that the current programs on the contrary assume that the weight is paritious. This would have to be fixed! Perhaps this is an opportunity for someone to code up Dembélé's algorithms in sage?

Step II: Suppose one finds such a form π. Note that I am also insisting that the coefficient field be as small as possible, namely the field F itself. Even though π is of non-paritious weight, there are still associated Galois representations (Some relevant references are this paper of Patrikis (see [Pat19]) and also this paper of Dembélé, Loeffler, and Pacetti) [DLP19]. More precisely, there are nice projective Galois representations, and these lift to actual representations, but they will not be Hodge–Tate; rather, up to twist (making the determinant have finite order, for example), they will have Hodge–Tate weights $[0,0]$, $[0,0]$, and $[-1/2,1/2]$. But now consider the tensor induction (twisted by a half!) of this representation from G_F to $G_{\mathbf{Q}}$, that is, for $\sigma \in \mathrm{Gal}(F/\mathbf{Q})$, the representations

$$\varrho := \rho(\pi) \otimes \rho^{\sigma}(\pi) \otimes \rho^{\sigma^2}(\pi)(1/2)$$

Now these representations will be crystalline with Hodge–Tate weights

$$[0,0,0,0,1,1,1,1].$$

Moreover, they will be symplectic, have cyclotomic similitude character, and (this is where the assumption on the coefficients of π comes in) will also have Frobenius traces in $\mathbf{Q}$. OK, I literally have not checked any of those statements at all, but it kind of feels like it has to be true so that's what I'm going with. The point of insisting that the coefficients of π was just F is to make the coefficients of this new system in $\mathbf{Q}$. But this means (at least conjecturally) that these Galois representations have to come exactly from an abelian variety of Mumford's type, because the Galois representations tell you that the Mumford–Tate group has Lie algebra $(\mathfrak{sl}_2)^3$.

Step III: Find this family in a different way. One issue with the construction above is that the Galois representations are not obviously motivic (or even satisfy purity!), so they certainly don't provably come from an abelian variety. But it might be easier to find the actual variety once one knows its exact level. I'm not quite sure what I mean by "find" here — it's an open question as to whether these Mumford 4-folds are Jacobians so I'm not entirely sure what one should be looking for.

Step IV: Bonus: prove that these 4-folds have L-functions with meromorphic continuations (at least for H^1 but it's worth checking the other degrees as well) using triple product L-functions.

Some Further Remarks: There are a number of relevant papers by Rutger Noot that one should be aware of (An particularly relevant example: this one) (see [Noo06]). There are restrictions on the possible level structures that can arise for Hilbert modular forms of this weight (in particular, they can't be Steinberg at some place), so make sure not to waste time computing at those levels. This is related to the fact that the corresponding Shimura variety is compact. The actual associated Shimura variety is isomorphic to $\mathbf{P}^1$ over the complex numbers; there's some discussion in section 5.4 of Elkies' paper (see [Elk08]). These Shimura curves

naturally have models over the reflex field, which is F in this case, but actually they can sometimes be defined over even smaller fields, such as $\mathbf{Q}$. Now I confess I am confused by a number of points, in increasing order:

(1) What is the exact relationship between the model of this Shimura curve over $\mathbf{Q}$ and the moduli problem? This is an issue both with understanding the moduli problem but also (because of the stackiness issues) differences between fields of moduli and fields of definition.
(2) Does this Shimura curve have points over $\mathbf{R}$? I think so. If I understand Shimura's paper here, I think the answer is yes. (see [Shi75]).
(3) Does this Shimura curve have points over $\mathbf{Q}$? I think so! Assuming it has points over $\mathbf{R}$ you only need to check all other finite primes, and the one that is most worrying is $p = 7$ but you don't really even need to check that one either if the others all work.
(4) Assuming it is $\mathbf{P}^1_{\mathbf{Q}}$, does that help at all? At the very least it provides succor that lots of A should exist over $\mathbf{Q}$, but it's not so clear how to go from a point to an equation. (Consider the easier case of Shimura curves corresponding to fake elliptic curves, for example.) Given a complex point, can one at least reconstruct some complex invariants of A such as its periods? Probably understanding this Shimura curve and its relationship with the moduli problem (over different fields) as concretely as possible would be a "second track" in this problem. (Presumably an advantage of a polymath project is that you can attack it from several angles at once.)

Notes 135.1. Some of these steps are active work in progress of my student Abhijit Mudigonda — stay tuned and ask me more before working on this!

136. Schur–Siegel–Smyth–Serre–Smith

Thu, 25 Nov 2021

If α is an algebraic number, the normalized trace of α is defined to be

$$T(\alpha) := \frac{\mathrm{Tr}(\alpha)}{[\mathbf{Q}(\alpha) : \mathbf{Q}]}.$$

If α is an algebraic integer that is totally positive, then the normalized trace is at least one. This follows from the AM-GM inequality, since the normalized trace is at least the nth root of the norm, and the norm of a non-zero integer is at least one. But it turns out that one can do better, as long as one excludes the special case $\alpha = 1$. One reason you might suspect this to be true is as follows. The AM-GM inequality is strict only when all the terms are equal. Hence the normalized trace will be close to one only when many of the conjugates of α are themselves close together. But the conjugates of algebraic integers have a tendency to repel one another since the product of their differences (the discriminant is also a non-zero integer.) In an Annals paper from 1945 [Sie45], Siegel (building on a previous inequality of Schur) proved the following:

Theorem 136.1 (Siegel)**.** *There are only finitely many algebraic integers with* $T(\alpha) < \lambda$ *for* $\lambda = 1.7336105\ldots$

Siegel was also able to find that the only such integers with normalized trace at most $3/2$ are 1 and $(3 \pm \sqrt{5})/2 = \phi^{\pm 2}$ for the golden ratio ϕ (We will also prove this below). On the other hand (generalizing these examples), one has

$$T(((\zeta_p + \zeta_p^{-1})^2) = 2\left(1 - \frac{1}{p-1}\right),$$

and hence the optimal value of λ is at most 2. Sometime later, Smyth had a very nice idea to extend the result of Siegel. (An early paper with these ideas can be found here.) Consider a collection of polynomials $P_i(x)$ with integral coefficients, and suppose that

$$Q(x) = -\lambda + x \ - \sum a_i \log |P_i(x)| \geq 0$$

for all real positive x where $Q(x)$ is well-defined, and where the coefficients a_i are also real and non-negative. Now take the sum of $Q(x)$ as x ranges over all conjugates of α. The key point is that the sum of $\log |P_i(\sigma\alpha)|$ is log of the absolute value of the norm of $P_i(\alpha)$. Assuming that α is not a root of this polynomial, it follows that the norm is at least one, and so the log of the norm is non-negative, and so the contribution to the sum (since $-a_i$ is negative) is zero or negative. On the other hand, after we divide by the degree, the sum of λ is just λ and the sum of $\sigma\alpha$ is the normalized trace. Hence one deduces that $T(\alpha) \geq \lambda$ unless α is actually a root of the polynomial $P_i(x)$. So the strategy is to first find a bunch of polynomials with small normalized traces, and then to see if one can construct for a constant λ as close to 2 as possible some function $Q(x)$ which is always positive. One can make this very explicit. Suppose that

$$Q(x) = -\lambda + x - \frac{43}{50} \cdot \log |x| - \frac{18}{25} \cdot \log |x-1| - \frac{7}{50} \cdot \log |x-2|,$$

Problem 136.2 (Calculus Exercise)**.** Show that, with $\lambda = 1.488753\ldots$, that $Q(x) \geq 0$ for all x where it is defined. Deduce that the only totally real algebraic integer with $T(\alpha) \leq \lambda$ is $\alpha = 1$.

The graph is as follows:

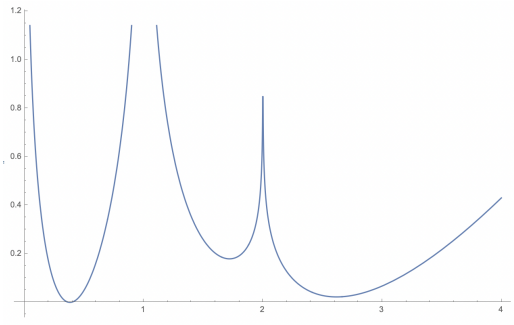

FIGURE 13. The function $Q(x)$ is positive

One can improve this by increasing λ and modifying the coefficients slightly, but note that we can't possibly modify this with the given polynomials to get $\lambda >$

3/2, because $T(\phi^2) = 3/2$. Somewhat surprisingly, we *can* massage the coefficients reprove the theorem of Siegel and push this bound to 3/2. Namely, take

$$Q(x) = -\frac{3}{2} + x - a\log|x| - (2a-1)\log|x-1| - (1-a)\log|x-2|,$$

and note that the derivative satisfies

$$Q'(x)x(x-1)(x-2) = (x^2 - 3x + 1)(x - 2a),$$

Hence the minimum occurs at either $x = 2a$ or at the conjugates of ϕ^2 where ϕ is the golden ratio. Since $\phi^2 - 1 = \phi$ and $\phi^2 - 2 = \phi^{-1}$, one finds that

$$Q(\phi^2) = -\frac{3}{2} + \phi^2 + (2 - 5a)\log\phi,$$

and so choosing a so that this vanishes when, we get

$$a = \frac{2}{5} + \frac{1}{2\sqrt{5}\log\phi} = 0.864674\ldots$$

and then we find that $Q(x) \geq 0$ for all x where it is defined with equality at ϕ^2 and ϕ^{-2}. So this reproves Siegel's theorem by elementary calculus. Of course we can strictly improve upon this result by including the polynomial $x^2 - 3x + 1$, for example, replacing $Q(x)$ by

$$P(x) = Q(x) - \frac{1}{15}\cdot\log|x^2 - 3x + 1| + \left(\frac{3}{2} - \lambda\right)$$

where $\lambda = 1.5444\ldots$ is now strictly greater than 3/2. By choosing enough polynomials and optimizing the coefficients by hook or crook, Smyth beat Siegel's value of λ (even with an explicit list of exceptions), although he did not push λ all the way to 2. This left open the following problem: is 2 the first limit point? That is, does Siegel's theorem hold for any $\lambda < 2$? This was already asked by Siegel and it became known as the Schur–Siegel–Smyth problem.

Some point later, Serre made a very interesting observation about Smyth's argument. (Serre's original remarks were in some letter which was hard to track down, but a more recent exposition of these ideas is contained in this Bourbaki seminar, see [Ser19].) He more or less proved that Smyth's method could **never** prove that 2 was the first limit point. Serre basically observed that there existed a measure μ on the positive real line (compactly supported) such that

$$\int \log|P(x)|d\mu \geq 0$$

for every polynomial $P(x)$ with integer coefficients, and yet with

$$\int x d\mu = \lambda < 2$$

for some $\lambda \sim 1.89\ldots$. Since Smyth's method only used the positivity of these integrals as an ingredient, this means the optimal inequality one could obtain by these methods is bounded above by Serre's λ. On the other hand, Serre's result certainly doesn't imply that the first limit point of normalized traces of totally positive algebraic integers is less than 2. A polynomial with roots chosen uniformly from μ will have normalized trace close to λ, but it is not at all clear that one can deform the polynomial to have integral coefficients and still have roots that are all positive and real.

I for one felt that Serre's construction pointed to a limitation of Smyth's method. Take the example of $Q(x)$ we considered above. We were able to prove the result for $\lambda = 3/2$ by virtue of the fact that $Q(x) = 0$ at these points. But that required the fact that the three quantities:

$$\phi^2, \phi^2 - 1 = \phi, \phi^2 - 2 = \phi^{-1}$$

were all units and so of norm one. The more and more polynomials one inputs into Smyth's method, the inequalities are optimal only when $P_i(\alpha)$ is a **unit** for all the polynomials P_i. But maybe there are arithmetic reasons why non-Chebychev polynomials (suitably shifted and normalized) must be far from being a unit when evaluated at $(\zeta + \zeta^{-1})^2$ for a root of unity ζ.

However, it turns out my intuition was completely wrong! Alex Smith has just proved (see [Smi24]) that, for a measure μ on (say) a compact subset of $\mathbf{R}$ with countably many components and capacitance greater than one, that if Serre's (necessary) inequality

$$\int \log|Q(x)| d\mu \geq 0$$

holds for every integer polynomial $Q(x)$, then you **can** indeed find a sequence of polynomials with integer coefficients whose associated atomic measure is weakly converging to μ. In particular, this shows that Serre's example actually proves the maximal λ in the Schur-Siegel-Smyth problem is strictly less than 2, and indeed is probably equal to something around 1.81 or so. Remarkable! I generally feel that my number theory intuition is pretty good, so I am always really excited when I am proved wrong, and this result is no exception.

Problem 136.3 (Exercise for the reader)**.** One minor consequence of Smith's argument is that for any constant $\varepsilon > 0$, there exist non-Chebyshev polynomials $P(x) \in \mathbf{Z}[x]$ such that, for primes p say and primitive roots of unity ζ, one has

$$\log \left| N_{\mathbf{Q}(\zeta)/\mathbf{Q}} P(\zeta + \zeta^{-1}) \right| < \varepsilon [\mathbf{Q}(\zeta_p) : \mathbf{Q}]$$

for all sufficiently large primes p. Here by non-Chebyshev I mean to rule out "trivial" examples that one should think of as coming from circular units, for example with $P(\zeta + \zeta^{-1}) = \zeta^k + \zeta^{-k}$ for some fixed k. Is there any other immediate construction of such polynomials?

For that matter, what are the best known bounds for the (normalized) norm of an element in $\mathbf{Z}(\zeta)$ which is not equal to 1, and ruling out bounds of elements in the group generated by units and Galois conjugates of $1 - \zeta$? I guess one expects the class number h^+ of the totally real subfield field to be quite small, perhaps even 1 infinitely often. Then, assuming GRH, there should exist primes which split completely of order some bounded power of $\log |\Delta_K|$, which gives an element of very small norm (bounded by some power of $[\mathbf{Q}(\zeta) : \mathbf{Q}]$). However, this both uses many conjectures and doesn't come from a fixed polynomial. In the opposite direction, the most trivial example is to take the element 2 which has normalized norm 2, but I wonder if there is an easy improvement on that bound. There is an entire circle of questions here that seems interesting but may well have easy answers.

Comment 136.4 (Noah Snyder)**.** Ooh, this is the calculus argument in Lemma 5 of our joint paper [CMS11]! Really cool to see the bigger picture that it's part of. My memory is playing tricks on me, I distinctly remember (and our emails seem to confirm it) that I'd come up with this argument myself (while on a train along

the Hudson on Christmas Eve). But we cite the Smyth paper elsewhere so surely it must somehow have come from there indirectly?

Comment 136.5 (Persiflage)**.** I think you indeed did come up with an idea more or less equivalent to this independently (although not in the '80s!), but I don't think we fully realized it. By the time of Zoey Guo's thesis paper [CG18b], there is a more direct reference to Smyth's work.

Comment 136.6 (Noah Snyder)**.** Ah great, yes that turn of events is spelled out nicely in Guo's paper. Glad my memory isn't totally gone, though disappointed that now all my "original" number theory arguments turned out to be known, because my other number theory argument here (see [Sny00]) is in some lectures of Osterle from the late 80s.

Notes 136.7. Alex's paper [Smi24] is actually one of only two papers I ever solicited for the Annals during my time as an associate editor (so far). The other was by Lue Pan, but by then he had already submitted it [Pan22].

137. ArXiv × 3

Fri, 04 Mar 2022

Three recent arXiv preprints this week caught my interest and seemed worth mentioning here.

The first is a paper by Oscar Randal-Williams (see [RW22]) which considers (among other things) the cohomology of congruence subgroups of $\mathrm{SL}_N(\mathbf{Z})$ in the stable range. (See Note 111.2). This is definitely something I have talked on the blog about a number of times, including § 64 and § 111. To recall; Matthew Emerton and I proved that the completed cohomology groups

$$\widetilde{H}^d(\mathbf{F}_p) = \lim H^d(\mathrm{SL}_N(\mathbf{Z}, p^n), \mathbf{F}_p)$$

are independent of N for N sufficiently large with respect to d, and are moreover finite vector spaces with a trivial action of $G = \mathrm{SL}_N(\mathbf{Z}_p)$. I later explained moreover how these groups are the cohomology groups of the homotopy fibre of the map from $\mathrm{SK}(\mathbf{Z}; \mathbf{Z}_p)$ to $\mathrm{SK}(\mathbf{Z}_p; \mathbf{Z}_p)$. But now the Quillen–Lichtenbaum conjecture shows (thanks to Blumberg and Mandell, see [BM20]) how the homotopy groups of these spaces are identified with Galois cohomology groups, which allows one to compute the maps between homotopy groups and understand (at the very least) the cohomology groups in degrees less than p. Since one has a Hochschild–Serre spectral sequence

$$E_2^{i,j} = H^i(G(p), \widetilde{H}^j(\mathbf{F}_p)) \Rightarrow H^{i+j}(\mathrm{SL}(\mathbf{Z}, p), \mathbf{F}_p),$$

this allows one to compute the cohomology of $\mathrm{SL}(\mathbf{Z}, p)$ over $\mathbf{F}_p$ in low degree by analyzing this spectral sequence. I later came to suspect that for regular primes p this spectral sequence degenerated immediately at least in degrees less than p or so, which would allow one to compute the cohomology groups in degree d explicitly for all large regular p. Actually the prediction was slightly stronger: in the range of cohomology degrees at most d one only had to avoid a finite set of primes (those dividing B_{2k} for small k together with the set of primes p which divided the finitely many zeta values $\zeta_p(3), \zeta_p(5), \ldots \zeta_p(2k+1)$ also for small k). Oscar not only proves this but goes one step further, by showing that it degenerates in small degrees

for *any* prime p, even as a $\mathrm{SL}(\mathbf{F}_p)$-module. This implies, for example, that, with $H^1(G(p), \mathbf{F}_p) = M$ being more or less the adjoint representation, that

$$H^4(\mathrm{SL}_N(\mathbf{Z}, p), \mathbf{F}_p) = \mathbf{F}_p \oplus \wedge^2 M \oplus \wedge^4 M$$

for $p > 5$ if and only if p does not divide the p-adic zeta function $\zeta_p(3)$, and

$$H^4(\mathrm{SL}_N(\mathbf{Z}, p), \mathbf{F}_p) = \mathbf{F}_p \oplus \mathbf{F}_p \oplus \wedge^2 M \oplus \wedge^4 M$$

otherwise. Note this condition implies that p is irregular but is much more restrictive. But it does actually happen! The only known primes with this property are $p = 16843$ and $p = 2124679$.

Part of my original interest in this problem came from Benson Farb and Tom Church — they noted that these groups should be stable in the weaker sense that they should be "independent of N" more or less exactly in the sense that there is a uniform description as above (proved later by Andrew Putman, see [Put15]), but this left open the question of what the groups actually were. Of course my feeling is that the completed cohomology groups are more "fundamental" and the cohomology at finite level is really just a frothy mix of unwinding what happens in the limit, but one has to admit that this new result is pretty satisfying.

The second is a paper by Will Sawin and Melanie Wood (see [SW24]). I remember 20 years ago or so being one of three BPs at Harvard asked to give a small presentation to the Harvard "Friends of Math" (Will Hearst and the gang), along with William Stein and Nathan Dunfield. One memory was that my talk was a chalk talk and theirs were both involved much snazzier technology. But I also remember that Nathan talked about his very nice paper with Bill Thurston (see [DT06]) on random 3-manifolds. In Melanie and Will's new paper, they beautifully exploit many of the recent progress on "random groups" (much of it due to the authors themselves) to show that the profinite completion of a random 3-manifold (in the sense of a random Heegaard splitting for larger and larger genus) itself has a limiting distribution.

Here is just one immediate corollary of their results which ties into previous problems considered both by Nathan and me and also Nigel Boston and Jordan Ellenberg. (Actually I say corollary, but I am just guessing that this should easily be a corollary without actually doing any of the computation so any error here is due to me!)

Corollary 137.1 (Expected)**.** *For a fixed prime $p > 2$ and a "random" 3-manifold M, there is a positive probability that:*

1. *There is a surjection: $\pi_1(M) \to \mathrm{SL}_2(\mathbf{Z}_p)$,*
2. *The corresponding tower of covers M_n coming from congruence subgroups all have trivial first Betti number.*

The point of course being that (as in Boston–Ellenberg [BE06]) one can deduce this from the more restrictive condition that the kernel N of the map

$$\pi_1(M) \to \mathrm{SL}_2(\mathbf{F}_p)$$

has $N/N^p = (\mathbf{F}_p)^3$ and no larger, and hence it can be phrased as the pro-finite completion of $\pi_1(M)$ surjecting onto one pro-finite group but not some other finite group. (Here $N/N^p = (\mathbf{F}_p)^3$ can I think be weakened to $N/N^p[N, N] = (\mathbf{F}_p)^3$ by an argument of Simon Marshall). I guess another way of saying this is that the

pro-p completion of the cover N can be described explicitly as the p-congruence subgroup of $\mathrm{SL}_2(\mathbf{Z}_p)$. Of course, this work also raises the very natural question:

Question 137.2. What is the distribution of $\widehat{\pi_1(M)}$ on arithmetic 3-manifolds? What about congruence arithmetic 3-manifolds?

The main point of course is that the existence of Hecke operators imposes a lot of extra structure, which one certainly expects (and can be numerically observed) changes the distribution of any given finite group occurring. Here I think the sensible question is to ask for a conjecture rather than a theorem, of course! (Maybe the first sensible question is actually to give a good conjecture for the distribution of the *abelianization* of these groups...)

The last paper is this one by Peter Kravchuk, Dalimil Mazáč, and Sridip Pal [KMP24], which I am even less qualified to talk about, which gives remarkable upper bounds for the smallest Laplacian eigenvalue of a (closed) hyperbolic orbifold of fixed genus. For example, when $g = 2$, they give the bound $\lambda_1 < 3.8388976481$, which is not too shabby given that there is an example with $\lambda_1 = 3.83888725\ldots$! The paper has a number of other gems, including more or less identifying the complete spectrum of all λ_1 as comprising a set of isolated points combined with the entire interval $[0, \alpha]$ for some $\alpha = 15.8\ldots$.

138. A random curve over $\mathbf{Q}$

Sat, 02 Apr 2022

Let $X/\mathbf{Q}$ be a smooth projective curve. I would like to be able to say that the motive M associated to X "generally" determines X. That is, I would like to say it in a talk without feeling like I'm telling too much of a fib. But is this true? There are two issues. Recall that, by the Torelli Theorem, the Jacobian together with a principle polarization determines $X/\mathbf{C}$. So there are two things to worry about:

(1) Knowing M only recovers the Jacobian up to isogeny, and you can certainly have two different curves with isogenous Jacobians, even isomorphic Jacobians with different polarizations.
(2) Knowing $X/\mathbf{C}$ does not determine $X/\mathbf{Q}$.

To overcome the second issue, it is sufficient and necessary to assume that $\mathrm{Aut}_{\mathbf{C}}(X)$ is trivial. Let me ignore the first point, since I both assume it generically doesn't happen but since I can't even address the second point yet I haven't thought about it yet.

Perhaps this is obvious to a geometer, but I don't see why a "random" curve $X/\mathbf{Q}$ doesn't have automorphisms. My model of a random curve is to take, for example, an embedding of M_g into projective space and then count points by the ambient height function and see what ratio of points has trivial automorphisms. (Presumably any other counting function like Faltings height or whatever will more or less be the same.) Certainly a generic $\mathbf{C}$-point of M_g has no automorphisms (at least for $g > 2$), but since M_g is of general type for large enough g I don't whether one can find enough rational points which are generic!

Probably the most natural way to answer this is to give a positive answer to the following question:

Question 138.1. Does M_g contain a subvariety X which is unirational over $\mathbf{Q}$ and has dimension strictly greater than the hyperelliptic locus?

Or, to put it more naturally, can you just explicitly write down enough generic curves which don't have any automorphisms to see that they dominate any point count?

Comment 138.2 (Felipe Voloch)**.** Trigonal curves. See the comments (especially the one from Jason Starr) to this answer of mine: this answer.

Comment 138.3 (Will Sawin)**.** I don't think anyone can rule out that there is a unirational subvariety of dimension $3g-4$ for g arbitrarily large, for example. Even if that was known, that wouldn't answer your question about heights, since a lower-dimensional subvariety can have a larger count of points with bounded height than a higher-dimensional one. I don't think there's much hope to formulate the claim for anything but a specific family — plane curves, n-gonal curves for n from 2 to 5, complete intersections in higher-dimensional space ...

There's a general philosophy related to the Bombieri–Lang and Manin conjectures – I don't remember exactly which combination of conjectures one could assume to draw this as a conclusion but probably there is one — that most of the rational points on a general type variety should lie on its Fano subvarieties, whichever ones have the most Manin-predicted points, if it has any Fano subvarieties at all.

The point being that all but finitely many rational points should be explained by varieties that have a lot of rational points (Fano, abelian varieties, Calabi-Yaus) but Fano varieties have vastly more points than the others (except Calabi-Yaus which themselves contain Fanos). For example, any rational curve contained in any variety has vastly more rational points than any abelian subvariety, no matter the ample height function you choose.

Thus, Manin-type predictions could come from looking for Fano subvarieties and trying to find the ones that have the most predicted rational points. I don't remember my idea above but I think it had to do with assuming that if there are many ration points of M_g with automorphisms then there must be a Fano variety parameterizing rational points of M_g with automorphisms and then showing that this maps to a Fano variety parameterizing points of $M_{g'}$ for $g' < g$ and trying to compare the numbers of rational points.

Comment 138.4 (Andrew Sutherland)**.** How about this example, the genus 3 curves

$$y^2 + (x^4 + x^3 + x^2 + 1)y = x^7 - 8x^5 - 4x^4 + 18x^3 - 3x^2 - 16x + 8$$

and

$$x^3z + x^2yz + x^2z^2 + xy^3 - xy^2z + y^4 - y^3z - yz^3$$

are both generic in the sense that the have no extra endomorphisms, but one is hyperelliptic and the other is not.

While I have not tried to run a Faltings-Serre computation to prove it, I'm morally certain these two genus 3 curves have the same L-function (their Frobenius traces agree out to 2^{28} and the conductor is only 8233), hence they have isogenous Jacobians. But these two curves really are different in meaningful sense.

139. What would Deuring do?

Wed, 13 Apr 2022

This is an incredibly lazy post, but why not!

Matt is running a seminar this quarter on the Weil conjectures. It came up that one possible way to prove the Weil conjectures for elliptic curves over finite fields is to lift them to CM elliptic curves using Deuring's theorem. But after some discussions we couldn't quite work out whether this was circular or not.

Certainly if you can lift to a CM elliptic curve and lift Frobenius to an endomorphism ϕ of the lift you get Weil immediately; the degree of ϕ is p which implies the norm of ϕ is p, but for imaginary quadratic fields the norm coincides with the absolute value. But how did Deuring prove his theorem?

The most obvious way to lift an (ordinary, say) elliptic curve $E/\mathbf{F}_p$ to characteristic zero is to note that, by the Weil conjectures, the order $\mathcal{O} = \mathbf{Z}[\phi]$ generated by Frobenius lies inside an imaginary quadratic field K (this is equivalent to the Weil conjectures), and so one can consider $\mathbf{C}/\mathbf{Z}[\phi]$. To make things simple, if the order is maximal, then this is defined over the Hilbert class field H of K, and since p splits principally in K (since ϕ has norm p) it follows that p splits principally in H as well by class field theory, and so the CM elliptic curve is also defined over $\mathbf{Z}_p$ and gives a lift. Of course, this argument uses the Weil conjectures! Without that, the ring $\mathcal{O}$ lives inside a real quadratic field and it's not clear what one can do.

One approach is to prove the existence of the canonical lift, which automatically will have extra endomorphisms and thus be CM since it lives in characteristic zero. This doesn't depend on the Weil conjectures. But the canonical lift is a construction I associate more with Serre–Tate than with Deuring. But it's certainly possible that Deuring's argument was via the canonical lift.

Some might say that the easy way to solve this is simply to look in one of Deuring's papers. But instead I will try to call upon my readers (possibly either number theorists who speak German or Brian Conrad) to save me the work and tell reveal all in the comments!

Comment 139.1. Dick Gross] Frank, You may still need to brush up on your German, but I believe that the first proof of the Riemann hypothesis (Artin's conjecture) for elliptic curves over finite fields via lifting to characteristic zero is due to Hasse [Has36]. Surely you are not too lazy to have a look!

What Deuring did was to refine these arguments to describe all possible endomorphism rings of elliptic curves in char $p \geq 0$. They are: $\mathbf{Z}$ (which cannot occur over finite fields), an order of conductor prime to p in an imaginary quadratic field where p splits, a maximal order in the quaternion algebra over $\mathbf{Q}$ ramified at p and ∞.

140. Murphy's law for Galois deformation rings and Ozaki's theorem

Sat, 23 Apr 2022

A theorem of Ozaki from 2011 [Oza11], perhaps not as widely known as expected, says the following:

Theorem 140.1. [Oza11] *Let p be prime, and let G be a finite p-group. Then there exists a number field F and an extension H/F such that:*

(1) H/F *is the maximal pro-*p *extension of* F *which is everywhere unramified.*
(2) $\mathrm{Gal}(H/F) = G$.

Since any non-trivial p-group G has a non-trivial center, it can be written as a central extension of a smaller p-group G' by $\mathbf{Z}/p\mathbf{Z}$, and thus the proof is (as one might imagine) by induction. But the structure of the argument is quite tricky, and it's a little hard to absorb all the ideas at once.

This post is to report on two extensions of Ozaki's result. The first is a new preprint by Hajir, Maire, and Ramakrishna [HMR24] which gives both a simplification and also an extension of Ozaki's result (the extension being that one has more explicit control over the degree of F).

But this post will actually be about a somewhat different generalization due to my student Andreea Iorga. Let me give her result now:

Theorem 140.2. *Iorga Let* Φ *be a finite group of order prime to* p*, and let* G *be a finite* p*-group with an action of* Φ*. Assume there exists an extension* L/K *with Galois group* G *such that:*

(1) L/K *is Galois with Galois group* Φ*,*
(2) L *contains* ζ_p*,*
(3) L *has trivial* p*-class group.*

Then there exists number fields $H/F/E$ *such that:*

(1) H/F *is the maximal pro-*p *extension of* F *which is everywhere unramified.*
(2) $\mathrm{Gal}(H/F) = G$.
(3) $\mathrm{Gal}(H/E) = G \ltimes \Phi$*, where the semi-direct action is the given action of* Φ *on* G.

When Φ is trivial one recovers Ozaki's theorem in the case when $\mathfrak{p}$ is a regular prime. In fact, Ozaki's first proof also has a similar hypothesis. Most likely Iorga's argument extends to the more general case where one does not need to assume that $\zeta_p \in E$. (Of course, in order not to accidentally solve the inverse Galois problem, the other two conditions on L and K will be necessary! One nice consequence (and a motivating example) of Iorga's theorem is as follows. Consider an absolutely irreducible residual representation:

$$\overline{\rho} : G_K \to \mathrm{GL}_2(\mathbf{F}_p)$$

to a finite field. What possible rings R can occur as deformation rings of $\overline{\rho}$? In this setting, let R denote the deformation ring of everywhere unramified representations. Let's also assume that the image (to be absolutely concrete) has image which has order prime to p, say projectively $\Phi = A_4$ or S_4. The Fontaine–Mazur conjecture predicts that the only $\overline{\mathbf{Q}_p}$-points will have finite image, and thus correspond to the natural lift (assuming the characteristic of k is $p > 5$. An argument with class groups then implies that one should expect $R[1/p] = \mathbf{Q}_p$, or equivalently that R is a ring admitting a map

$$R \to \mathbf{Z}_p$$

with finite (as a set) kernel I. A consequence of Iorga's theorem is the following:

Theorem 140.3. *Iorga Let* R *be any local ring admitting a surjection to* $\mathbf{Z}_5$ *with finite kernel. Then* R *is a universal everywhere unramified deformation ring.*

The key idea to reduce this theorem to the previous one is a follows. Suppose that the image of $\overline{\rho}$ is $\widetilde{\Phi}$. Since this has order prime to p, it lifts to a representation

$\widetilde{\Phi} \subset \mathrm{GL}_2(\mathbf{Z}_p)$. Then denote Γ denote the inverse image of this group inside $\mathrm{GL}_2(R)$, so it lives inside an exact (split) sequence:

$$1 \to 1 + M_2(I) \to \Gamma \to \widetilde{\Phi} \to 1$$

The group Γ admits a natural residual representation via $\overline{\rho}$, and clearly Γ admits a deformation to $\mathrm{GL}_2(R)$ by construction. The point is that one can show that this R is the universal deformation ring, and hence providing one has extensions $H/F/E$ with $\mathrm{Gal}(H/E) = \Gamma$ and H/F the maximal everywhere unramified pro-p extension of F (using the previous theorem) one is in good shape. (There is a trick to reduce the problem in this case to Φ in order to make the "base case" easier, since one has fields $F/\mathbf{Q}$ and $\widetilde{F}/\mathbf{Q}$ with $\mathrm{Gal}(F/\mathbf{Q}) = \Phi$ and $\mathrm{Gal}(\widetilde{F}/\mathbf{Q}) = \widetilde{\Phi}$ and if $\Phi = A_4$ and $p = 5$ then proving that $F(\zeta_5)$ of degree 96 has class number prime to 5 is easier than the same clam for $\widetilde{F}(\zeta_5)$ of degree some multiple of 192.

One way to view this result is as an example of "Murphy's Law" for moduli spaces. Here Ravi Vakil proves [Vak06] in a different setting that all possible singularities occur inside deformation spaces. The analog is to say that all possible local rings (subject to some obvious constraints) occur as Galois deformation rings. Another natural class of rings one might expect to arise in this way (still considering everywhere unramified deformation rings) is all *finite* rings. Of course for such rings one would have to consider residual representations whose images have order divisible by p, requiring a further modification of the theorems of Ozaki and Iorga. In a different direction, one can ask what happens for deformation conditions with other local conditions at p. Here are two natural such questions:

Problem 140.4. Let $(R, \mathbf{m})$ be any complete local Noetherian ring with finite residue field which is finite over $W(k)$. Then does R occur as the finite flat deformation ring of some absolutely irreducible residual representation?

Problem 140.5. Let $(R, \mathbf{m})$ be any complete local Noetherian ring with finite residue field over $W(k)$, and assume that:

(1) R is a complete intersection, namely that there is a presentation:
$$R \simeq W(k)[[x_1, \ldots, x_d]]/(f_1, \ldots, f_r)$$
where $d \geq r$.
(2) $p \in R$ is a regular element.

Then does R occur as the universal deformation ring (with fixed determinant) of some absolutely irreducible residual representation?

I would guess the first problem has a positive answer but I'm honestly not even sure about the second one!

Comment 140.6 (Persiflage)**.** There are some countability issues with Problem 140.5.

141. Joël Bellaïche

Tue, 14 Jun 2022

Very sad to hear that Joël Bellaïche has just died. He got his PhD at the same time as me, and I first got to know him during the Durham conference in 2004 and later at the eigenvarieties semester at Harvard (was that in 2005 or 2006?).

Joël was an original mathematician, and his papers (many written with Gaëtan Chenevier) contain many really good ideas. As a postdoc, I was totally immersed in thinking about Galois deformations of reducible representations when the paper lisseté de la courbe de Hecke de GL_2 aux points Eisenstein critiques [BC06] appeared on the arXiv. In that paper, they study the *ideal of reducibility* for certain Galois deformation rings (or pseudo-deformation rings). By studying the ring-theoretic properties of this ideal, they proved the Eigencurve was smooth at the evil Eisenstein points. It clarified immediately a number of the phenomena I had been thinking about, but it was also simply the "right" way to think about these things. I also learnt from Joël at Durham the problem of proving the non-vanishing of p-adic zeta values like $\zeta_p(3) \neq 0$, which remains 18 years later one of my favourite problems.

Another really beautiful idea was the approach by Joël and Gaëtan to Bloch-Kato type conjectures (including the Selmer group part of the Birch–Swinnerton-Dyer conjecture) via the geometry of eigenvarieties (including those associated to $U(3)$). This is of course related to the ideal of reducibility. Their joint asterisque paper Families of Galois representations and Selmer groups [BC09] is a very nice read on this topic, as are Joël's notes for the Clay summer school as well as his recent book on Eigenvarieties (see [Bel21]).

In more recent times, Joël had been exploring ideas in some interesting directions, including his intriguing work on self-correspondences on curves [Bel23]. What was consistent about his research was that his primary motivation always seemed to be rooted in coming to an original understanding of interesting math rather than simply making incremental improvements on work of others.

Last but not least, one should not forget his sense of humor with a decidedly irreverent streak. This is probably best appreciated with a beer or a glass of wine in a summer evening in Luminy, but to take a quote from on of Joël's own papers:

> Let p be a prime number that, we shall assume, splits in E. We shall also assume that $p \neq 13$. I don't think this is really useful, but who knows?

My thoughts are with his family.

Comment 141.1 (Clémentine Fauré-Bellaïche)**.** I am Joël's wife and I came upon this blog post by chance — merci beaucoup, it deeply touched me.

Comment 141.2 (Danièle Kuzmanovic)**.** Extrêmement peinée par le décès de Joël ! Je l'ai connu enfant, il venait jouer chez nous à la maison. Notre fils, Djordje Kuzmanovic, en avait fait son meilleur ami lors de la rentrée scolaire au collège Montaigne de Paris (1984). Nos pensées affectueuses vont à ses parents et à son épouse. Danièle et Dejan Kuzmanovic.

Comment 141.3 (Régis Lebrun)**.** J'apprends le décès de Joël avec une infinie tristesse. Il était mon meilleur ami lors de l'année passée en TC1 au lycée Louis le Grand en 1990/1991 et dans les années qui ont suivi. Nous avons partagé des moments inoubliables ensemble dans les gorges du Verdon, puis les concours et des parcours professionnels différents nous ont fait nous perdre de vue. Je voudrais exprimer mes sincères condoléances à ses parents et son épouse. Cette disparition prématurée enlève une personne exceptionnelle à ses proches et à la communauté. Régis Lebrun

142. Locally induced representations

Thu, 16 Jun 2022

Today is a post about work of my student Chengyang Bao [Bao22]. Recall that Lehmer's conjecture asks whether $\tau(p) \neq 0$ for all primes p, where

$$\Delta = q \prod_{n=1}^{\infty} (1-q^n)^{24} = \sum \tau(n) q^n$$

is Ramanujan's modular form. You might recall that Naser Talebizadeh Sardari and I (in [CTS21], see also § 120) studied a "vertical" version of Lehmer's conjecture where instead of fixing a modular form, we fixed a prime p and a tame level N and showed that there were only finitely many normalized eigenforms f of level N and even weight k with $a_p(f) = 0$ which were not CM. We exploited the fact that such forms give rise to Galois representations

$$\rho : G_{\mathbf{Q}} \to \mathrm{GL}_2(\overline{\mathbf{Q}}_p)$$

which are crystalline at p but also locally induced at p from the unique unramified quadratic extension $K/\mathbf{Q}_p$. As explained in § 122, it's hard to see this method being able to say much more to this (for example, to say anything about Lehmer's actual conjecture), since there do actually exist non-CM forms with $a_p(f) = 0$.

In practice, we don't even know in level $N = 1$ whether there exist infinitely many normalized eigenforms f with $a_p(f) \equiv 0 \bmod p$. As mentioned in § 69, one source of such representations comes from modular forms with exceptional image. For example, if $f = \Delta E_4$ is the normalized eigenform of weight 16, then as first observed by Serre and Swinnerton-Dyer, the mod-59 representation

$$\overline{\rho}_f : G_{\mathbf{Q}} \to \mathrm{GL}_2(\overline{\mathbf{F}}_p)$$

has projective image S_4 coming from the splitting field of $x^4 - x^3 - 7x^2 + 11x + 3$. But the local residual representation in this case is induced, which implies that $a_{59} \equiv 0 \bmod 59$. As explained in that post, standard conjectures about primes predict that there should be infinitely many S_4-representations unramified outside a single prime p giving rise to modular Galois representations which will then come from level one modular forms f with $a_p(f) \equiv 0 \bmod p$.

Chengyang's work concerns examples precisely of this sort. From my work with Naser, we can deduce that there are at most finitely many f of level one with $a_{59}(f) = 0$. Chengyang proves that there are no such forms. More precisely:

Theorem 142.1 (Bao [Bao22])**.** *Suppose that f is a modular form of level one, and suppose that $a_p(f) = 0$. Then all of the residual mod-p representations $\overline{\rho}$ associated to f have big image, that is, image containing $\mathrm{SL}_2(\mathbf{F}_p)$.*

In other words, none of the (presumably many) infinite examples of S_4 representations giving rise to f of level one with $a_p(f) \equiv 0 \bmod p$ can ever give an f with $a_p(f) = 0$.

Chengyang also proves some further results about the deformations of representations with exceptional image. For example, for the mod-59 representation above, the *only* deformations to characteristic zero unramified outside $p = 59$ which are locally induced are the representations which up to twist are the ones which up to twist coincide with the unique lift with finite image and order prime to $p = 59$.

In contrast, one might ask what happens for $p = 79$, the next case where there exists a form f of level one with $a_p(f) \equiv 0 \bmod p$. I suspect that in this case a (possibly quite complicated computation) should show that there should be at *most one* form with $a_p(f) = 0$, but that it might be quite difficult to prove using p-adic methods that there are no such forms. The problem will be that there will exist a deformation which will have infinite image and be locally induced, but now it will have generalized Hodge–Tate weights $[0, \kappa]$ for some p-adic number κ for which it will be very hard to show is not an integer. This is analogous to the family of Eisenstein series of level one with $p = 37$. One knows that the p-adic zeta function will have a unique zero, but it is very hard to probe the arithmetic nature of that zero and to rule out it occurring at some arithmetic weight. To put this slightly differently, is there an integer $k \geq 1$ such that

$$\zeta_{37}(-31 + 36k) = 0?$$

Presumably not, but this seems extremely difficult; the difficulty of course is that there will be a solution with $k \in \mathbf{Z}_{37}$.

143. The future is now; recap from Cetraro

Sun, 31 Jul 2022

I've just returned from the second Journal of Number Theory biennial conference in Italy. It's always nice to get a chance to see slices of number theory one wouldn't otherwise see at the conferences I usually go to (although this was the first conference of any kind I attended in person since 2019). Here is a brief and incomplete recap.

There were more talks that mentioned the Manin-Mumford conjecture and its various generalizations (particularly to uniform bounds in families) than I have ever previously attended in my life. There were probably equally many talks which mentioned Ax–Schanuel as well. It was nice to see these subjects and I learnt quite a lot from these talks.

143.1. **Modularity of elliptic curves over imaginary quadratic fields.** In my ICM talk, I claimed that the modularity of elliptic curves over the Gaussian integers is "within our grasp"; well, the future is now! James Newton talked about his work in progress with Ana Caraiani [CN23] where they prove modularity of all curves over imaginary quadratic fields F such that $\#X_0(15)(F) < \infty$, which includes $\mathbf{Q}(\sqrt{-1})$. One of the key tools in their proof is a suitable local-global compatibility statement for Galois representations coming from torsion classes in the crystalline setting where one is *not* in the Fontaine–Laffaille range (because of ramification in the base, for example). This was a situation where I had even been hesitant to make a precise conjecture. The problem is that the natural conjecture one might want to make is that the map of Hecke algebras factors locally through the Kisin deformation rings. But the construction of Kisin deformation rings as closures which are flat over $W(k)$ by default might make one worried whether it is the correct integral object for torsion representations. But Caraiani and Newton show that such concerns are unfounded, and the $W(k)$-flat deformation rings are indeed the correct objects. One key point of their argument is showing that the (possibly torsion) representations $\rho \oplus \rho^\vee$ (for suitable twists of ρ occur inside the cohomology

of the Shimura variety in such a way where (using some notion of ordinary for a parabolic other than the Borel) the local representations *in characteristic zero* are reducible and realize the required crystalline lifts of each factor. One remaining annoyance is that one would like to find points over twists $X(\overline{\rho}_E, \wedge)$ of the Klein quartic $X(7)$ over F corresponding to $E[7]$ which lie on solvable CM fields (in order to do a switch at the prime 7. You could (for example) start with the point E and the 15-isogenous curve E' and connect them via a line. This line will go through two further points defined over a quadratic extension H/F, but there is no reason to suppose a quadratic extension of an imaginary quadratic field will be CM. I had some idea related to a half-forgotten fact I learnt from John Cremona walking in the woods near Oberwolfach, but upon further consideration this half-forgotten fact was sufficiently ephemeral that it could not be reconstructed and didn't appear to correspond to any actual facts (See Note 143.6 below). I did learn from Tom Fisher the nice fact that the four curves 3-isogenous to E are collinear on the curve $X(\overline{\rho}_E), 3, \wedge)$ corresponding to the same mod-7 representation with the other choice of symplectic form (that is, $\wedge$ scaled by a quadratic non-residue).

143.2. **BSD for abelian surfaces.** This was my first chance seeing a talk on the work of Loeffler–Zerbes [LZ21] on BSD for abelian surfaces. The most difficult condition to verify in their theorem is that a certain (characteristic zero) deformation problem is unobstructed. It seems very plausible to me that one could numerically verify some interesting examples and get truly unconditional results on BSD for some autochthonous abelian surfaces, or at least autochthonous elliptic curves over imaginary quadratic fields. The idea is that to prove a certain ordinary (of some flavour) deformation problem is unobstructed it suffices to prove that it is unobstructed modulo p, which reduces to a computation with ray class groups in the splitting field of $\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GSp}_4(\mathbf{F}_p)$. This seems within the realm of practicality. Moreover, once one verifies the condition for A, one immediately deduces it for all the twists of A as well. It is important here to take p small, that is at most 3. Certainly if the mod-3 image is surjective the extension will be too large, but the case of a representation induced from an imaginary quadratic field seems completely manageable. The other possibility is to work with $p = 2$. Here I think one should work with $H^1(\mathbf{Q}, \mathrm{ad}_0(\overline{\rho}))$ where ad_0 is the quotient of the 11-dimensional adjoint representation by the diagonal (so slightly different from trace zero matrices). Here I'm imagining starting with a modular abelian surface which has good ordinary at 2 and whose mod-2 image is S_5. It might also be convenient if the local factor at 2 is congruent to $(1 + T + T^2)T^2$ mod 2 so that the local deformation problem has good integral properties. Anyone interested in computing such an example?

143.3. **The David Goss Prize.** During the conference the second "David Goss Prize" was awarded. This prize is for work done in the past two years in number theory by someone at most 35(!) and also by someone who has not (yet!) won any other major prizes. (I joked that it might be nice to have a prize for people 50 or older who have not won any prizes but there is something nice about no longer being eligible for any prize except those one has no hope of winning.) This year's winners were Ziyang Gao and Vesselin Dimitrov. The laudatio is here. Congratulations to both!

Talking of prizes, I can't quite work out whether there are far more prizes than there used to be, or whether I was simply oblivious about them when I was younger.

Comment 143.4 (David Hansen)**.** Do you know the smallest d for which $X_0(15)(\mathbf{Q}(\sqrt{-d}))$ is infinite? Glancing at [LMF24], it looks this set is finite for $d = 1, 2, 3$ and infinite for $d = 7$, but the info there doesn't seem to cover the cases $d = 5, 6 \ldots$ Anyway, it's a great theorem!

Comment 143.5 (Persiflage)**.** James mentioned it in his talk. I think $d = 6$? For full disclosure, I also believe that the theorem also includes the condition that $F \neq \mathbf{Q}(\sqrt{-55})$. I *assume* this is because one also has to rule out elliptic curves with other level structures at 3 and 5 which are too small, and this leads to a number of curves of genus $g \geq 2$, and one of those curves has an exceptional point over F. In principle, one can check "by hand" that this E is modular. However, if it has enormously large level, that could plausibly be an annoying computation. I did think that James missed a trick by not calling the talk "imagining Elliptic Curves over imaginary quadratic fields, but particularly not those involving the squareroot of -55."

Notes 143.6. Now that I can search through posts, the "half-forgotten fact" was discussed here § 102. It is not helpful here.

144. Potential modularity of K3 surfaces

Tue, 15 Nov 2022

This post is to report on results of my student Chao Gu who is graduating this (academic) year.

If A/F is an abelian surface, then one can associate to A a K3 surface X (the Kummer surface) by blowing up $A/[-1]$ at the 16 singular points (corresponding to 2-torsion points of A. If F is a totally real field, then one knows that A is (potentially) automorphic, and it follows that X is also (potentially) automorphic, which in particular implies the Hasse-Weil conjecture for X. It also proves that

$$\rho(X/F) = -\mathrm{ord}_{s=1} L(H^2(X/\overline{F}, \mathbf{Q}_p(1)), s),$$

where $H^2(X/\overline{F}, \mathbf{Q}_p(1))$ is the etale cohomology group considered as a Galois representation of G_F; this was conjectured by Tate in the same paper where he makes the "usual" Tate conjecture on algebraic cycles. Not all K3 surfaces arise in this way. For a start, if A has (geometric) Picard rank $\rho(A) \geq 1$, then X has geometric Picard rank $16 + \rho(A) \geq 17$. If the Picard rank of X is at least 19, then X also has to arise (at least in the category of Motives) as a Kummer surface, but more subtly this is not true in rank 17 and 18, where there are further obstructions relating to the structure of the transcendental lattice (as first observed by Morrison in this paper, see [Mor84]). What Chao does is prove the following:

Theorem 144.1 (Gu)**.** *Let $X/\mathbf{Q}$ be a K3 surface of Picard rank at least 17. Then X is potentially automorphic, and the Hasse-Weil conjecture holds for X.*

In the most interesting case of rank 17, the approach is to lift the compatible family of 5-dimensional orthogonal representations associated to the transcendental lattice to a compatible family of 4-dimensional symplectic representations which one hopes to prove is potentially automorphic. Finding (motivic) lifts of K3 surfaces is a well-studied problem, and a nice analysis of what happens arithmetically can be found in Patrikis' thesis. From the Kuga–Satake construction, one can certainly

reduce to considering certain abelian varieties. The question is then narrowing down the precise endomorphism structures of these varieties as well as their fields of definition. It turns out that for the problem of interest, there are more or less three types of abelian varieties one might want to consider beyond abelian surfaces over a totally real field F:

(1) Abelian varieties A/F of dimension $2d$, where A admits endomorphisms by an order in the ring of integers of a totally real field E of degree $[E : \mathbf{Q}] = d$.
(2) Abelian surfaces A/H over some Galois extension H/F where the conjugate of A by $\mathrm{Gal}(H/F)$ are all isogenous over H.
(3) Abelian 4-folds A/F with endomorphisms by an order in a quaternion algebra $D/\mathbf{Q}$.

More generally, one needs to consider the "cross-product" where several (or all) of these phenomena may occur at once. For those more familiar with the story of two-dimensional Galois representations over $\mathbf{Q}$, these three extensions correspond to replacing elliptic curves over $\mathbf{Q}$ by abelian varieties of GL_2-type, to $\mathbf{Q}$-curves, and to fake elliptic curves respectively. It turns out that the last case doesn't happen over totally real fields but the analog for abelian surfaces does, see § 98.

The optimal generalization of [BCGP21] to this setting would be to prove that all of these varieties are potentially modular. However, it turns out that there is an obstruction to proving this: namely, is not always possible to prove that these varieties have enough ordinary primes (one needs something slightly stronger, namely ordinary primes whose unit eigenvalues are distinct modulo p). This puts some restrictions on what can be proved unconditionally, but everything works as long as there are enough ordinary primes. Chao's proof requires a number of modifications from [BCGP], in particular to the Moret-Bailly part of the argument. In our original paper, we exploited the fact that we were working only with abelian surfaces which allowed us to use some tricks to simplify this step. In particular, the problem of finding an appropriate point on the desired moduli space over $\mathbf{Q}_p$ was made much simpler by virtue of the fact that the original abelian surface produced such a point. In Chao's generalization, however, this trick doesn't work, and one must use more subtle arguments using Serre–Tate theory. Fortunately, enough tricks are available concerning ordinary primes to settle the general case of K3 surfaces of (geometric) Picard rank at least 17 when they are defined over $\mathbf{Q}$. But note there do exist many such K3 surfaces (not related to abelian surfaces) over $\mathbf{Q}$ that one can write down explicitly; see the examples due to Nori discussed in [BCGP21, §9.4].

Note that this result is new even for Picard rank 18. For Picard rank 19 and 20, the (potential) modularity of any X/F for a totally real field F reduces to the corresponding problem for elliptic curves. The case of Picard rank 16 appears as hopeless as the case of generic genus three curves.

145. Check the arXiv regularly!

Sat, 18 Feb 2023

In § 136 I discussed a new result of Smith which addressed the following question: given a measure μ on $\mathbf{R}$ supported on some finite union of intervals Σ, under what conditions do there exist polynomials of arbitrarily large degree whose roots all lie in Σ whose distribution (in the limit) converge to μ?

A natural generalization is to replace Σ by a subset of $\mathbf{C}$ subject to certain natural constrains, including that μ is invariant under complex conjugation. I decided that this had a chance of being a good thesis problem and scheduled a meeting with one of my graduate students to discuss it. Our meeting was scheduled at 11AM. Then, around 9:30AM, I read my daily arXiv summary email and noticed the preprint (https://arxiv.org/abs/2302.02872) (see [OS23]) by Orloski and Sardari solving this exact problem! There are a number of other very natural questions along these lines of course, so this was certainly excellent timing. When I chatted with Naser over email about this, he mentioned he had become interested in this problem (in part) by reading my blog post!

There is of course a general danger of giving my students problems related to my blog posts, and indeed I have refrained from posting a number of times on possible thesis problems, but in this case everything turned out quite well.

Notes 145.1. My student Kapil Chandran is currently thinking about an adelic generalization of these theorems.

146. What the slopes are

Sat, 25 Feb 2023

Let f be a classical modular eigenform of weight k, for example, $f = \Delta$. The Ramanujan conjecture states that the Hecke eigenvalues a_p satisfy the bound

$$|a_p| \leq 2p^{(k-1)/2}.$$

A slightly fancier but cleaner way of saying this is as follows. Associated to f of weight k, level N prime to p and finite order Nebentypus character χ is a polynomial

$$X^2 - a_p X + p^{k-1}\chi(p).$$

(For Δ one has $\chi(p) = 1$ for all p, but in general it can be some other root of unity.) This is the characteristic polynomial of Frobenius on the ℓ-adic representation associated to f for $\ell \neq p$, or the characteristic polynomial of crystalline Frobenius for $\ell = p$. The Ramanujan conjecture is now that the Frobenius eigenvalues α_p and β_p satisfy

$$|\alpha_p| = |\beta_p| = p^{(k-1)/2}.$$

This conjecture was famously proved by Deligne. Having determined the complex valuation of these eigenvalues, one might ask about their ℓ-adic valuations as well. If $\ell \neq p$, then since the polynomial above has constant term prime to ℓ, then α_p and β_p are ℓ-adic units. So the remaining case is $p = \ell$.

When $p = \ell$, the p-adic valuation of the two roots α_p and β_p certainly satisfy $v(\alpha_p), v(\beta_p) \geq 0$ and $v_p(\alpha) + v_p(\beta) = k - 1$. Given f, we call these valuations the *slopes* of f. The first observation is that the slopes depend on more than just the weight k. For example, suppose that f is the weight 2 eigenform associated to a (modular) elliptic curve E with good reduction at p, then either E has ordinary reduction, in which case the slopes are 0 and 1, or E has supersingular reduction, and the slopes are both 1/2.

Instead of fixing f and varying p, one can fix both p and the tame level N and ask how the slopes vary as the weight changes. For example if $N = 1$ and $p = 2$,

then the first interesting case is when $k = 12$ and $f = \Delta$. In this case $\tau(2) = 24$, and the two slopes are 3 and $11 - 3 = 8$.

I've already tried to suggest by analogy to the Ramanujan conjecture why determining the p-adic valuations (the slopes) might naturally be an interesting question. But let me mention two other natural reasons. The first is that, from p-adic Hodge Theory, the crystalline eigenvalues α_p and β_p determine the restriction of the p-adic Galois representation associated to f to the local Galois group at p. The valuation of these slopes while containing less information than the eigenvalues themselves still tell you a lot about the p-adic representation, although determining exactly what is still a question of active and open interest. Secondly, as Coleman observed (following Hida in the case when one of the slopes is 0), one can deform the Galois representations associated to these finite slope forms into continuous p-adic families of Galois representations associated to cuspidal eigenforms, which leads to the story of the eigencurve of Coleman–Mazur and beyond.

Gouvêa was one of the first people to undertake a numerical study of the roots. In this paper where the slopes are (see [Gou01]), Gouvêa observed a number of interesting behavior of the slopes which were all somewhat mysterious. First, the slopes were almost all integers. This is not surprising when $a_p(f) \in \mathbf{Z}$, but in general $a_p(f)$ will be a random algebraic integer. The slopes also seemed to be distributed in a number of surprisings way. For example, although the slopes in weight k associated to f add up to $k - 1$, they tended to be concentrated in the intervals $[0, (k-1)/(p+1)]$ and $[p(k-1)/(p+1), k-1]$.

Kevin Buzzard went one step further and looked more closely at the slopes when $N = 1$ and $p = 2$. Ostensibly, according to Kevin, this was to test William Stein's latest magma code for bugs! Kevin found that the slopes in this case satisfied a much more regular pattern. As mentioned above, when $k = 12$, there is a single eigenform whose smallest slope is 3. When $k = 12 + 64$, there are six eigenforms whose smallest slopes are $3, 7, 13, 15, 17, 25$, and when $k = 12 + 2^{10}$, there are 86 forms whose slopes are

$$3, 7, 13, 15, 17, 25, 29, 31, 33, 37, 47, 49, 51, 57 \ldots$$

where all of these sequences are given explicitly by the 2-adic valuation of $2((3n)!/n!)^2$ (explained in our paper [BC05]). Note that the Gouvêa–Mazur conjecture says that these sequences should have some initial segment in common, but in fact the Gouvêa–Mazur conjecture only implies that the slopes in weights $16+2^6$ and $16+2^{10}$ should be the same up to slope 6, and in practice they agree ridiculously further than this. This was all very mysterious. Kevin found a general algorithm (see [Buz05]) which conjecturally computed by an inductive procedure all the slopes in all weights for a fixed tame level N. (In what I always regarded as a missed opportunity, he did not call the paper "What the slopes are").

In fact, Kevin's conjectural answer required an assumption on N and p which for $p > 2$ was equivalent to asking that the local residual representations associated to all low-weight forms are locally reducible. For $N = 1$, the first case for which this does not happen is $p = 59$, and this story is related to our counterexample to the original form of the Gouvêa–Mazur conjecture (see [BC04]).

Kevin Buzzard was a speaker at the Arizona Winter School in 2001, and for his project he outlined a special case of his conjecture corresponding to overconvergent modular forms of weight $k = 0$. Of course there are no classical modular cuspidal forms of weight $k = 0$, but by Coleman's theory there is a direct link between the

questions about classical forms in all weights and overconvergent forms of finite slope in all weights. Kevin's problem in its original formulation is described here.

I was a graduate student at the time, and although I wasn't actually assigned to Kevin's group, I did see his problem and had an idea which more or less amounted to the idea of using a slightly different explicit basis for this space than Kevin had considered and for which the U_2 operator has a much nicer explicit form. In fact this basis was related to computations I had done in the summer of 93-94 and shortly afterwards during my first year at Melbourne uni, using what Matthew Emerton and I had come to definitely know as the "f" function

$$f = q \prod_{n=1} (1+q^n)^{24}.$$

(When It came to writing my paper with Kevin, I still felt strongly enough to insist we call it by this letter.) Kevin and I put quite a bit of effort into proving his conjecture for more general k, still with $N = 1$ and $p = 2$, but only succeeded in a few cases, including $k = -12$, but also $k = -84$ which was somewhat randomly chosen as a case where our approach failed but only by a little bit which could be overcome. There are a few other cases where $X_0(p)$ has genus zero and one can do something similar, but otherwise very little progress was made on these conjectures in this situation.

There are a plethora of related conjectures which also came out (at least indirectly) from similar calculations. For example, Kevin's algorithm certainly always produces integers, so, in light of the p-adic theory, there is the natural question of whether a crystalline representation with Hodge–Tate weights $[0, k-1]$ for k even whose residual representation is reducible always has finite slope, Then there are questions of the exact relationship between the slope and the Galois representation, and so on.

Concerning the exact slopes themselves, perhaps the biggest advance over time were refinements of Kevin's conjecture. The Ghost Conjecture, formulated by Bergdall and Pollack [BP19], is some "master conjecture" of a combinatorial nature generalizing a number of previous conjectures concerning the slopes of classical overconvergent modular forms over the centre of weight space. (Although, perhaps amusingly, it's not clear that these conjectures do actually imply Buzzard's original conjecture.)

Now cut to the present day. In a recent preprint, Ruochuan Liu, Nha Xuan Truong, Liang Xiao, Bin Zhao [LTXZ23] have now proved all of these conjectures, at least up to some genericity hypotheses (excluding the case $N = 1$ and $p = 2$!). The authors certainly employ technologies that didn't exist 20 years ago (p-adic Langlands, for example), but that was not the only obstruction to previous progress: the paper contains a number of very original and clever ideas. Very amazingly and satisfyingly, it resolves a large number of the open problems discussed above, including Gouvêa's conjectures about the distribution of slopes, the integrality properties of slopes for locally reducible representations, and even a version of the Gouvêa–Mazur conjecture. It is also satisfying that the arguments use p-adic local Langlands, given that some of the initial computations of slopes served at least in part as inspiration for some aspects of this program. I myself have not really worked on this circle of problems for almost 20 years but I am still very happy to see these questions answered!

147. Deciphering Quanta

Thu, 23 Mar 2023

Sometimes it is claimed that Quanta articles are so watered down of mathematical content that they become meaningless. That presents a challenge: do I understand the quanta article on my own work? Let's consider New Proof Distinguishes Mysterious and Powerful 'Modular Forms'. I can confirm that I did not see this article in any form before it appeared. Overall I would say that it is faithful to the facts and I can interpret what everything means. I'm not quite sure why there is an Alex Kontorovich explainer about the Langlands Program in there but why not? I did, however, have to stop and wonder when I saw the following picture:

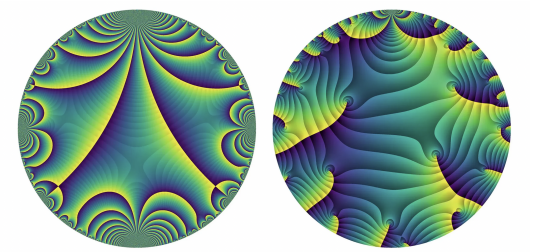

Figure 14. Congruence modular forms (left) have additional structure that noncongruence modular forms (right) lack

This appeared with the caption "Congruence modular forms (left) have additional structure that noncongruence modular forms (right) lack." OK, here is the challenge: can I work out what this picture actually represents?

In both pictures, there is clearly some color gradient between yellow and blue, and this has discontinuities along some regions we call L and R for left and right. These are also clearly pictures inside $\mathbf{H}^2$ in the Poincaré disc model. Here L looks at least superficially like a fundamental domain for some Fuchsian group. The rays of L going towards the point i on the boundary do look like geodesics in the hyperbolic metric (which are circular arcs meeting the boundary at right angles). There is a corresponding invariance for the figure by the parabolic element (with some cusp width) through i. Consider a conformal map from the upper half plane to this model which sends ∞ to i: Using the standard conformal map with the upper half plane:

$$\phi : \mathbf{H} \to D(0,1), \quad z \mapsto \frac{1+iz}{1+z}$$

From the picture, we see the images of the geodesics from the infinite cusp to $n\alpha$ for n odd. All parabolic elements are conjugate, but if we are to guess what Γ is by looking at the picture then working out α for this specific model is important. I tried to eyeball it for a while before printing it out and trying to compute the angle using continued fractions, which wasn't so accurate but gave $\tan\theta \sim 1/(1+1/3)$ with θ in the mid 30s. Then using some angle tools in keynote it came out to somewhere between 36 and 37 degrees, closer to the latter. So maybe it was exactly a 10th root of unity, which would make α live in a degree 4 field, or (much better!) maybe $\alpha = 1/2$ in which case the angle is $36.8698\ldots$, and then L (or rather $\phi(L)$

is invariant under exactly

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Maybe I should have guessed this in retrospect! Let's look at the geodesic from $i\infty$ to $1/2$, which in the picture goes from i to $4/5 - 3i/5$. There appears to be a point with 4-fold symmetry. That suggests invariance under some order four element corresponding to an elliptic point. But this is a little worrying, since $\mathrm{PSL}_2(\mathbf{Z})$ does not have any elements of order 4! Now there are some ways around this. For example, this could be the level set of a function which satisfies an extra invariance property under the normalizer of some congruence subgroup, and so does not itself come from an element of $\mathrm{SL}_2(\mathbf{Z})$ but $\mathrm{GL}_2(\mathbf{Q})^+$. For example, the Fricke involution on $X_0(N)$ is $\tau \mapsto -1/N\tau$ which corresponds to. a matrix in $\mathrm{GL}_2(\mathbf{Q})$ with determinant N, although that has order 2. Any element of order 4 has to have eigenvalues of the form α and αi, with $\alpha + i\alpha \in \mathbf{Z}$, and thus eigenvalues $1 + i$ and $1 - i$ up to rational multiples. That means the determinant should be 2 times a square. The unique element such element up to conjugacy is

$$S = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

which has order 4 and fixes i. If we want to fix a point on the geodesic $1/2 + it$, we can just make a hyperbolic scaling and then conjugate to get the matrix

$$S_t = \begin{pmatrix} 1 - 1/2t & t + 1/4t \\ -1/t & 1 + 1/2t \end{pmatrix}.$$

Here I started to run into a problem because it's hard to approximate t to any precision from the picture, although $t \sim 0.1$ numerically. There are a few natural integral matrices one can write down but it wasn't clear what was going on.

Next I turned to the picture around the cusp $-i$. This looks kind of weird because the biggest curves decidedly do not look like geodesics. But if we ignore this, we might decide that $-i$ is another cusp. It's also disturbing that the fundamental domain appears to contain enough of the boundary to suggest that Γ is a thin group, but let's ignore this as well. So what is the cusp width (without normalizations) at $-i$? Here numerically I simply get nonsense because if there really were geodesics around i translated by a fixed parabolic element and the first one was of the size indicated in the picture then there would be many fewer visible ones. As a typical example, one should expect a picture like this:

which doesn't look anything like the original picture. So I'm ready to give up now. What about the picture on the right? Well here I don't even know how to begin. There is not any obvious evidence of parabolic elements, which are not only present in congruence subgroups of $\mathrm{SL}_2(\mathbf{Z})$ but of any non-congruence subgroup as well. Perhaps there is a cusp at i with some large cusp width. There also seem to be singularities inside the disc. But that just suggests this might be a level set of a meromorphic form. But why choose a meromorphic form rather than one that is holomorphic away from the cusps? I wonder if that gives hints about the first picture; perhaps the transition from yellow to blue occurs when $\mathrm{Im}(f)$ goes from positive to negative, and where it might be 0 along some geodesic arc (For example: the Fourier coefficients are rational so the form is real along the purely imaginary axis) but the form can also have imaginary part 0 along some non-geodesic regions and that is what one is seeing around $-i$. That might allow one to reinterpret the

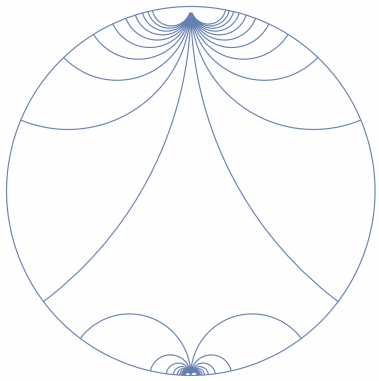

FIGURE 15. Geodesics through i with fixed cusp width

graphs as something related to $\mathrm{Im}(f)$. This suggests that the point on the left that appears to have degree four instead is related to a "local" symmetry coming from a vanishing point of the *derivative* of the modular function, but I'm just guessing now.

I should probably spend no more time on this, so instead I open up speculations to the comments!

Full Disclosure: The picture comes with an attribution to David Lowry-Duda but I did not try to follow that lead.

147.1. **Update.** So here is an new example. Like the graphs above, it plots the argument of a two modular forms on (different) finite index subgroups of $\mathbf{SL}_2(\mathbf{Z})$. I have normalized both pictures so that the cusp width at the cusp i is the same in both cases (under the normalizations above invariant under $\tau \to \tau + 3$; making the cusp width too small makes the behavior at $\tau = i\infty$ dominate the picture as the covolume goes up. There are dome differences with this example however:

(1) I have used (holomorphic) modular functions rather than modular forms.
(2) For these particular choices of functions, they are non-zero everywhere.
(3) I made the picture myself so it is not as pretty as the ones above.
(4) The functions I chose have the property that they are real precisely on geodesics, and thus the singularities here do form a tessellation.

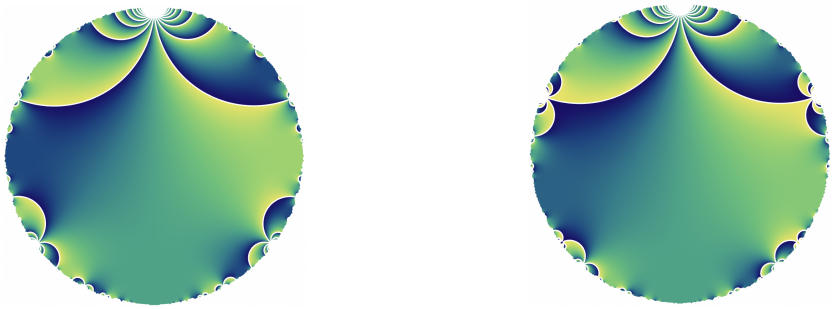

FIGURE 16. Which one is congruence?

Now one of the forms is defined on a congruence subgroup and has a Fourier series in $\mathbf{Z}[[q]]$, the other has a Fourier series in $\mathbf{Q}[[q]]$ but *not* in $\mathbf{Q} \otimes \mathbf{Z}[[q]]$ and is only defined on a non-congruence subgroup. But which is which? The pictures here are clearly much more similar than the pair of pictures above!

Comment 147.2 (Will Sawin)**.** In the spirit of your update let me mention what I thought upon seeing the image, before seeing Peter's link:

The color ranges from blue to yellow. If we think of the color spectrum from blue to yellow as an interval, the function taking a point to its color looks continuous, except for the lines of discontinuity where it switches from completely blue to completely yellow. In other words, this function is continuous after we glue both ends of the interval, forming a circle.

How do we get a continuous function to the circle from a modular form, or other complex-valued function? The only obvious thing is to take the argument. So the color in the picture must depict the argument of the modular form.

Now something I figured out after seeing the links, but feel like I could have figured out purely mathematically/visually: The argument does have another type of discontinuity, at the roots, where all colors collide around a single point. These can be distinguished from the potentially similar-looking points where the function is negative real and the derivative is zero because those points only have bright blue and yellow near that point. If the root has order 1, we will see a line of blue-yellow discontinuity ending at that point. The left picture only appears to have the brightly-colored kind of singularity, i.e. no roots, which could have been the clue to tell us it was Δ, while the right picture has clearly visible roots.

The roots in the right picture are significant because, while the argument is not invariant under the group of symmetries of the modular form (and unlike the absolute value can't even be normalized to become invariant), the set of roots certainly is, so examining the roots in the disc might give us a clue to the non-congruence group (but this is far beyond my visual-calculational abilities).

Comment 147.3 (David Lowry-Duda)**.** I can't say that this is "speculation" as I guess I hold all the cards, but I thought I would comment.

Will is exactly right about how the left picture is (essentially) the argument of the Δ function.

I first used those colors because the colormap works well for various types of insensitivity to certain colors. When Quanta asked me for images. But it's true that the sharp color discontinuity when across values when $\arg(z) = \pi$ suggests an important feature even where there isn't one.

The specific Cayley transform that I use to relate the disk to the upper halfplane is $\phi(z) = (1 - iz)/(z - i)$, which preserves the "apparent" vertical orientation of the imaginary axis. That is, the points $(-i, 0, i)$ in the disk correspond to the points $(0, i, \infty)$ in the upper halfplane.

The form on the right was computed using the work of Berghaus, Monien, and Radchenko here (see [BMR24]). It's one of the forms of weight 4 on the group with signature $(8, 0, 1, 2, 2)$ (using their notation and their data).

I also didn't see the article (or even really know much about what it was going to be about) when they asked me for an image. It just happens to be that when you google how to visualize modular forms, my little paper here comes up (see [LD21]).

148. Quadratic reciprocity

Tue, 02 May 2023

I accidentally proved quadratic reciprocity in class today, or at least three quarters of a proof. Can you finish it off? Here's the proof: start with a real quadratic field K, and the sequence

$$1 \to \mathcal{O}_K^\times \to K^\times \to K^\times/\mathcal{O}^\times \to 1$$

then take cohomology. If P_K is the group of principal ideals of K, then from Hilbert Theorem 90 you deduce that

$$P_K^G/P_{\mathbf{Q}} \simeq H^1(G, \mathcal{O}_K^\times).$$

If I_K is the group of all ideals of K, the left hand side is a subgroup of $I_K^G/P_{\mathbf{Q}}$ which is a product of groups of order two for each ramified prime. Now if $K = \mathbf{Q}(\sqrt{pq})$ with $p \equiv q \equiv 3 \bmod 4$, there is no unit of norm -1 because $(-1/p) = (-1/q) = -1$, so you deduce that the cohomology of the unit group has order 4 and so from order considerations that the prime ideals of norm p and q are principal. Write $\mathfrak{p} = (\alpha)$ and $\mathfrak{q} = (\beta)$. One has x and y with

$$x^2 - pqy^2 = N(\alpha)$$

Here $N(\alpha) = \pm p$. But the right hand side must be a square modulo q, and so $N(\alpha) = p$ if $(p/q) = +1$ and $-p$ if $(p/q) = -1$. Equivalently, $N(\alpha) = (p/q)p$, and similarly $N(\beta) = (q/p)q$. But $\mathfrak{pq} = (\sqrt{pq})$, so $\alpha\beta = \varepsilon\sqrt{pq}$ for some unit ε, and since all units in K have norm one, it follows that

$$pq(p/q)(q/p) = N(\alpha\beta) = N(\varepsilon)N(\sqrt{pq}) = -pq,$$

which is quadratic reciprocity! There is a similar argument for $p \equiv 1 \bmod 4$ and $q \equiv -1 \bmod 4$ (though now using facts about $(2/p)$ rather than $(-1/p)$). However, it is not so clear if one can prove the case $p \equiv q \equiv 1 \bmod 4$ using this argument. Is there one? Of course the challenge here is to "only" use Hilbert 90 and unique factorization of ideals, but never to make any arguments about how primes are splitting.

Notes 148.1. Actually, the "similar" argument doesn't quite work. So only one quarter of a proof!

149. Clozel 70, Part I

Sun, 24 Sep 2023

I recently returned home from a trip to Paris for Clozel's 70th birthday conference. Naturally I stayed in an airbnb downtown, and the RER B gods smiled on me with a hassle free commute for the entire week. Tekés was an interesting find, a fun (and surprisingly cheap) Israeli vegetarian restaurant right near where I was staying. But surely the food highlight of the week was the lunch spreads during the conference at Orsay — certainly the best conference food I've ever had! Great vegetarian food with amazing eggplant dishes, feta, figs, all the good stuff. (Rumor was it was chosen by Valentin Hernandez and paid for by Vincent Pilloni; I'm not sure if that's true but a great job all round.)

There were quite a number of interesting talks, although as mentioned before I don't like singling out because that sometimes seems like an implicit criticism of

the other talks. But a few thoughts spurred by some of the talks (which you can more or less guess if you wanted to), some of which were already raised by others in conversation during the conference:

149.1. **(Global) modules for Galois deformation rings.** As a result of Taylor–Wiles patching, one usually constructs a CM-module M_∞ defined over some local deformation ring R. Often quite a bit of mileage can be gained by exploiting the ring theoretic structure of R to conclude something about the module M_∞ and vice versa. Perhaps the ur-version of this argument is Diamond's argument showing (in some circumstances) that the formal smoothness of R implies that M_∞ is free. A more recent example is in the work of Jeff Manning where he exploits the geometry of some particular R (by relating to a more geometric situation where one can perhaps understand the Picard group) to restrict the possible M_∞ to a very small number of possibilities from which one can then get some mileage. But one question raised is the extent to which this one can always do this. As one considers more and more complicated R, is there some constraint on possible R which means that there are only going to ever exist a small number of faithful maximal rank one CM-modules M, or are there going to be situations where R is very complicated and one can't hope classify all such M, but only (for some mysterious situation) a very small number of them turn up in global situations. Note that globally there is often a few possibilities of the type of cohomology one patches, and even for GL(2) you can be in situations where you can force M to be free or self-dual by working in coherent cohomology or etale cohomology respectively and these modules are not always the same.

149.2. **The work of Arthur.** (Some) experts are at the point where they no longer expect Arthur to publish proofs of results he has claimed, leaving a huge gap in the literature. The summary seems to be that many very smart people are putting lots of effort into filling in some of these details, and that this seems to require new arguments. For example, it seems to be the case that one of Arthur's proposed inductive arguments will not (at least naively) work. The mathematical community should be immensely grateful to people working on this!

149.3. **Shimura Varieties.** I once joked that today's generation is more likely to learn about Galois representations associated to elliptic curves and modular curves before learning any class field theory. Well that generation has passed! We may be approaching a moment where people learn about Shimura varieties without ever thinking about modular curves, let alone getting close and personal with $X_0(11)$. (Note: I don't think that Richard knows the genus of $X_0(11)$ and that never stopped him, of course.) Some people can look at the abstract definition of a Shimura variety and then start proving things; I am certainly not one of those people. Fortunately there are still many interesting open questions even about classical modular forms.

149.4. **A result of Garland.** (at least) two talks reminded me of a vanishing result of Garland. Suppose that Γ is (for example) a lattice in $\mathrm{SL}_n(\mathbf{Q}_p)$. Then the cohomology (with coefficients in $\mathbf{Q}$) vanish in positive degrees below $(n-1)$. But I think that much more should be true, namely, that the cohomology should all have a "trivial" Hecke action in the expected ways, i.e. the completed cohomology groups should all be finite in this range, as they are for $\mathrm{SL}_n(\mathbf{Z})$ (more or less, let's not be precise about ranges and what exactly is known). It feels like conjectures of

this sort are not completely out of reach. Is this too optimistic? This is already an interesting problem in the case of $H^2(\Gamma, \mathbf{F}_p)$.

150. Clozel 70, Part II

Sat, 30 Sep 2023

Many years ago, Khare asked me (as I think he asked many others at the time) whether I believed their existed an irreducible motive M over $\mathbf{Z}$ (so good reduction everywhere) with Hodge–Tate weights $[0, 1, 2, \ldots, n-1]$ for any $n > 1$. (Here the Motive is allowed to have coefficients.) When $n = 2$, the answer is no. Assuming all conjectures, such an M must be modular associated to a cusp form of weight 2 and level one, but no such cuspform exists. But the answer is also no unconditionally (for any notion of motive), and this fact is intertwined with the (inductive) proof by Khare and Wintenberger of Serre's Conjecture. The hope might be that if no such motive existed for all n, it could serve as the inductive basis for a more general form of Serre's Conjecture.

My response at the time was that I guessed that no such motive existed for any n. I generally feel that my intuition is quite good in these matters, so it was surprising to learn some time later a convincing meta-argument that such motives should really exist. This idea, which I can't now remember whether I learned from Chenevier or Clozel, is related to trying to construct such forms which are in addition self-dual and so come from a classical group. In favourable situations, there exists a compact inner form on this group, so that computing these forms "reduces" to computing on certain finite sets. One such finite set turns out to be the set of positive definite lattices of discriminant one and dimension n. As is well-known, they only occur in dimensions a multiple of 8. For $n = 8$ there is just E_8, and for $n = 16$ there are two, and for $n = 24$ there turn out to be exactly 24, as classified by Niemeier, and which include the famous Leech lattice whose automorphism group is a central extension of the first sporadic group discovered Conway. Easier to compute is the weighted sum of such lattices by automorphisms; for $n = 24$, for example, this weighted sum is

$$\frac{1027637932586061520960267}{129477933340026851560636148613120000000}$$

which is very small, and of course is related to the fact that these lattices are quite symmetric. For $n = 32$, however, the weighed sum is bigger than 10^7, and so there are lots of lattices. You might then think that the existence of these lattices (even just E_8 when $n = 8$) implies the existence of automorphic forms which then should give rise to the desired automorphic forms on $\mathrm{GL}(n)$. But there are issues. One concerns the technical issue of transferring forms between groups which is of course a subtle problem. But there is another. A form which is cuspidal on some group need no longer be cuspidal after transferring to $\mathrm{GL}(n)$. So to see which forms are cuspidal you really need to do a computation. But these objects are of large complexity — already computing Hecke actions on supersingular points for $X_0(11)$ is a non-trivial exercise; here the objects involve lattices of enormously high dimension. Chenevier and his co-authors, (including Lannes, Renard, and Taïbi, [CR15, Che20, CT20]) have done a remarkable job understanding what is going on here. The most basic example of the type of theorem they prove is as follows. When $n = 16$ so there

are two lattices; one can try to compute the action of a Hecke operator T_p, and it turns out (see for example Theorem A here) that the answer involves Ramanujan's function $\tau(p)$. But this also tells you that the transfer to GL_{16} will have some explicit isobaric decomposition corresponding to twists of the modular form Δ, and in particular the associated π will clearly not be cuspidal.

At the same time, there are some automorphic arguments which show that cusp forms of level one (and cohomologically trivial weight) cannot exist. Here the idea goes back to the (automorphic) proof of lower bounds for discriminants of number fields by Stark and Odlyzko. The idea in that case to use the explicit formula for $\zeta_K(s)$ to construct an expression (and in particular the normalized version $\Lambda_K(s)$ which satisfies the functional equation and involves N^s where N is the level which is directly related to the discriminant of K) and ultimately arrive at some expression which is provably non-negative unless the root discriminant of K is larger than some explicit constant (minus some explicit $o(1)$ depending on the degree). Mestré generalized this argument [Mes86] to automorphic forms corresponding to other Motives, in particular proving that, assuming conjectures of Langlands type, that there did not exist any abelian varieties over $\mathbf{Z}$ of dimension at least one (which was proved unconditionally by Fontaine [Fon85]) but also that the conductor of such an abelian variety had to be at least 10^g. This was then later generalized by Fermigier (a student of Mestré) and then by Stephen Miller (Rutgers!) [Fer96, Mil02] to prove that there are no automorphic forms π for $\mathrm{GL}_n/\mathbf{Q}$ of level one which are cohomological for the trivial representation when $n < 27$. These are exactly the forms associated (conjecturally) to the motives of weight $[0, 1, \ldots, n-1]$.

Returning to the conference at Orsay: Chenevier gave a talk on understanding automorphic representations π of level one and low motivic weight, and once again raised the automorphic version of Khare's question. Now I have known about this question for a long time, but somehow being reminded of a problem can sometimes be the spark to help one think about the question again.

Correcting what was a past failing of my own intuition, I was very happy that George Boxer, Toby Gee, and I were able to come up with a very simple argument [BCG23a] to answer both questions; there does exist a compatible family of crystalline Galois representations with Hodge–Tate weights $[0, 1, 2, \ldots, n-1]$ for some n; for example one can take $n = 105$. Moreover, this compatible system is even automorphic and associated to a cuspidal π of level one and cohomological weight zero for GL_n. (With work, it is even "motivic" in the sense that the compatible system can be found inside an explicit algebraic variety over $\mathbf{Q}$, so it is in particular also pure.) Now while the argument is very simple, it must also be said that is uses some extremely hard theorems; for a start, it uses both the full modularity lifting results of [BLGGT14] (Barnet-Lamb, Gee, Geraghty, Taylor), following Clozel–Harris–Taylor and many others, *and* it uses the even more recent full symmetric power functoriality result [NT21a, NT21b] for classical modular forms by Newton and Thorne. (Since the paper is only nine pages and the proof only half of that, I won't explain it here.)

One would still like to prove, of course, that there are a huge number of self-dual forms for all sufficiently large n. And one can naturally ask what is the smallest such n, which we now know satisfies $27 \leq n \leq 105$. The expectation is certainly that n is probably close to around 32. It would be nice to know!

Of course, there is an endless list of other tricky problems one can pose of this form. For example, does there exist a regular motive (with coefficients) over $\mathbf{Z}$ with Hodge–Tate weights $[0, 1, \ldots, n-1]$ for some n which is *note* essentially self-dual?

Comment 150.1 (Will Sawin)**.** Discussing, I think, exactly this question, I believe I heard a long time ago the argument that if one counts cuspidal self-dual automorphic representations of GL_n with Hodge–Tate weights consecutive integers via the trace formula for the corresponding classical group, the main term coming from the adelic volume should grow so fast that all the other contributions, including the counts of Eisenstein series on GL_n arising from cusp forms on the classical group, should be dominated, and hence such representations should exist for all large enough n. Of course actually doing this kind of analysis with the trace formula in the large n limit seems essentially impossible, though the work of Deligne and Flicker does give a function field analogue.

I'm confused about one point in the Chenevier research report you link — why do the Galois representations considered there have a repeated Hodge–Tate weight, while the ones you are interested in do not?

Let me see if I see what is going on here. All forms on SO_{2n} arising from functions on even unimodular lattice have the same infinity type. Since the trivial representation is obviously among them, they all have the infinity-type of the trivial representation. Transferring them along any sort of Langlands functoriality will preserve these infinity-types, so you will get forms which have the infinity-type of the transfer of the trivial representation. However, the Arthur parameter of the trivial representation of SO_{2n} is the representation Sym^{2n-2} + trivial of SL_2 into the dual SO_{2n}, and transferring to GL_{2n} gets you a representation with Arthur parameter Sym^{2n-2} + trivial, i.e. not the trivial representation which would be Sym^{2n-1}.

. . . I was tempted to try to, rather than find an explicit eigenvalue of the set of rank 32 even unimodular lattices and check it gives a cusp form on GL_{32}, bound the number of possible Eisenstein series on GL_{32} that contribute to eigenforms on the set of rank 32 even unimodular lattices, but now I see why this is too hard: The worst case is perhaps the direct sum of Delta with a cusp form on GL_{30} that has exactly two gaps in its sequence of Hodge–Tate weights matching the weights of Delta, so we can sum them with Delta to get a problematic form on GL_{32}. Such forms on GL_{30} surely can't be ruled out by L-function methods as their weights are similarly spaced to forms that can't be ruled out, and counting them with the trace formula is clearly out of reach.

On the lower bounds side, I wonder if the results of Miller can be improved at all if one (a) assumes GRH and/or (b) restricts attention to self-dual representations. Given a lower bound of 27 and a potential upper bound of 32, even a tiny improvement would do a lot to close the gap.

With regards to the existence of modular forms of weight k and level 1 with reducible mod p reductions, I guess by Serre's conjecture this is equivalent to the existence of certain Galois representations into $\mathrm{GL}_2(\mathbf{F}_p)$. By a natural extension of Bhargava's heuristics for counting number fields to counting $\mathrm{GL}_2(\mathbf{F}_p)$ extensions with fixed $\mathrm{GL}_1(\mathbf{F}_p)$-part I get that the probability one should exist for each k is $1/(p-1)$ so the probability none exists for any k is $(1-1/(p-1))^{\phi(p-1)}$ which agrees closely with your heuristic. (The local factor at each non-archimedean prime is one so one only needs the local factor at infinity, which is the proportion of elements of

$\mathrm{GL}_2(\mathbf{F}_p)$ with determinant -1 that have order 2 and could be complex conjugation, which is $1/(p-1)$ for centralizer reasons.)

Comment 150.2 (Persiflage)**.** Concerning GRH: In Chenevier, the use of GRH does make a difference but only a very slight one in this range (a more general result for $m \leq 23$ can be improved to $m \leq 24$ but no more). So my guess that for this specific problem GRH probably wouldn't move the bar much, quite possibly even not at all.

151. Magma instability

Fri, 13 Oct 2023

I had occasion to return to some magma scripts I wrote in 2012. I the script used a number of pre-computed auxiliary files with computations, and was a little complicated, but didn't use anything particularly complicated. So I was really surprised to run them in 2023 and find that they no longer worked. That is, they compiled, but the results they gave were different (and also incompatible with the truth). It was quite confusing to understand what has gone wrong, but eventually I traced it to the following. Early on in the file one has (having defined t as a variable using code that's easy to write but which is somehow causing issues with wordpress):

```
F := NumberField(t^2 - 5);
ZF:=Integers(F);
```

So far, so good. But later on, the script called upon elements of F of the form `ZF![a,b]`. But it turns out that if `x:=ZF[0,1]` that

$$x^2 + x = 1$$

in 2012, but

$$x^2 - x = 1$$

in 2023; that is, x was replaced by $-x$, which is not an automorphism. I have no idea how or why that changed, but it certainly broke everything and took several days to fix. Annoying!

Comment 151.1 (Nathan Dunfield)**.** I get $x^2 - x = 1$ in an old copy of Magma 2.21-8. As the first version of Magma 2.21 came out in 2014, your scripts might have broken not long after you wrote them.

My guess is the cause is a change in the algorithm that computes an integral basis for a number field, which has the side-effect of giving a different answer for this one. I suspect the change was introduced in 2.21, where the release notes include:

> There is a new implementation of size reduction of an element by multiplication by units of the maximal order. This size reduction is now done in all situations where one would expect it to be done, for example in choosing a generator for a principal ideal and so on.

given that the two answers do differ by -1.

152. The horizontal Breuil–Mezard conjecture

Thu, 19 Oct 2023

Today I wanted to talk about Chengyang Bao's thesis. Fix a local mod-p representation, say

$$\overline{\rho} : G_{\mathbf{Q}_p} \to \mathrm{GL}_2(\mathbf{F}_p)$$

given on inertia by $\omega_2 \oplus \omega_2^p$. Associated to this residual representation is a Kisin deformation ring R corresponding to fixed determinant crystalline lifts of weights $[0, k-1]$, for some fixed positive integer $k \equiv 2 \bmod (p-1)$. The special fibres R/p of these rings have dimension one, and so, if one denotes their maximal ideal by $\mathfrak{m}$), then the Hilbert series

$$H_k(x) = \sum \dim(\mathfrak{m}^k/\mathfrak{m}^{k+1})X^k$$

has the form

$$H_k(x) = \frac{P_k(x)}{1-x}$$

where $P_k(x)$ is a polynomial. The Hilbert-Samuel multiplicities of these rings are given by the Breuil–Mezard conjecture (also proved by Kisin). These numbers are explicitly given by the values $P_k(1)$. It seems quite surprising that understanding the seemingly simple number $P_k(1)$ is so intimately linked to the proof of the Fontaine–Mazur conjecture. At the same time, we know very little about these rings R (or their special fibres) when k is large.

In the example above, the unrestricted (fixed determinant) local deformation ring R^{loc} is is formally smooth of dimension three over $\mathbf{Z}_p$. Although the rings R/p only have dimension one, one expects that for larger and larger k they start to "fill out" the unrestricted deformation ring. It is natural to wonder: how fast does this happen?

More explicitly, the Hilbert-Samuel series of the special fibre of the unrestricted deformation ring with fixed determinant is

$$\frac{1}{(1-x)^3} = 1 + 3x + 6x^2 + 10x^3 + \dots$$

So one can ask: what weight does one have to go to see all three dimensions of the tangent space? How far does one have to go to see all of $R^{\mathrm{loc}}/(p, \mathfrak{m}^n)$?

This was the thesis problem of Chengyang Bao, which grew out of (in part) questions arising during her work [Bao22]. In this particular case, it seems that one has to go to weight $k = p^2 + 1$ to see the entire tangent space. Actually, an even more basic question is whether there exists surjective map

$$R_{k+p-1}/p \to R_k/p,$$

this seems very tricky and is still open (but Chengyang's work strongly suggests that it is true).

One of the difficulties in this project is that close to nothing was known about the rings R for k anything larger than $k = 2p$ or so (although there has certainly been quite a bit of work understanding the link between a_p and the residual representation, including Buzzard–Gee [BG13], Sandra Rozensztajn [Roz18, Roz20], and many others using p-adic Langlands for larger k).

Chengyang's approach was, perhaps surprisingly, to use global methods. The basic summary of the Taylor–Wiles method as formulated by Kisin is that via

patching one finds that a patched Hecke ring may be identified with a power series ring over R. By reverse engineering this, if one finds a residual representation with sufficiently nice global properties, one can use explicit Taylor–Wiles primes to get arbitrarily close approximations to the Kisin deformation ring R. One of the tricks here is to be able to do this in a way where one can work efficiently after fixing a residual representation and then increasing the weight.

By doing these computations, Chengyang generated lots of explicit data about these rings R from which one can start making conjectures. I said before how the Breuil–Mezard conjecture amounts to predicting the value of $P_k(x)$ at $x = 1$. Chengyang has, at least in this particular case, been able to formulate an exact conjectural answer for the *entire* polynomial $P_k(x)$. As a consequence, one can read off from this the answer of how large a weight one has to go to see all the directions in $R^{\mathrm{loc}}/\mathfrak{m}^{\mathfrak{n}}$. I mentioned before that for $n = 2$ the answer is $p^2 + 1$. and my guess was that the answer in general might be of order p^n. But somehow the conjectural answer (up to constants which depend on p) turns out to be of order $O(n^2)$, which is honestly completely different from anything that I would have guessed. I think of this conjecture as a new "horizontal Breuil–Mezard conjecture." But really, it's only *half* a conjecture; the hope is that one can understand and interpret Chengyang's conjecture on the $\mathrm{GL}_2(\mathbf{Q}_p)$-side, and working this out is an exciting problem.

At the same time, there are lots of other things one can start to guess from looking at these explicit rings. Chengyang has a precise conjecture which says when the rings R in this setting are complete intersections or Gorenstein, and it also seems that they are always Cohen–Macaulay.

Even though we "know" p-adic Langlands for $\mathrm{GL}_2(\mathbf{Q}_p)$ much better than in any other situation, there seems to be a real opportunity here to tease out many more precise and explicit conjectures from Chengyang's work, and really to discover new phenomena which have hitherto never been noticed because computations of these rings has been so limited. (Another basic question: how many components does the generic fibre of R have in terms of k?).

153. Unramified Fontaine–Mazur for representations coming from abelian varieties

Thu, 09 May 2024

Mark Kisin gave a talk at the number theory seminar last week where the following problem arose:

Let W be the Galois representation associated to the Tate module of an abelian variety A over a number field, and suppose that $W = U \otimes V$. Now suppose that the Galois action on U is unramified at all primes above p. Can you prove that the Galois action on U has finite image?

Of course this is a special case of the unramified Fontaine–Mazur conjecture. But here the representation U literally "comes from an abelian variety" although as a tensor factor rather than a direct factor. At first sight it seems like it should be much easier than the actual Fontaine–Mazur conjecture if you just find the right trick, but I don't see how to do it! Here at least is a very special case.

Lemma 153.1. *Suppose that A/K has ordinary reduction at a set of primes of density one, and that U is a representation which is unramified at all primes dividing*

p of odd dimension which occurs as a tensor factor of $W = H^1(A) = U \otimes V$. Then, after some finite extension of K, U contains a copy of the trivial representation.

Proof. One may as well assume by induction that the action of the Galois group on U is absolutely irreducible of odd dimension d and remains so for every finite extension (otherwise decompose it into such pieces and take one of odd dimension).

Now choose a prime v of K. Let α_i be the eigenvalues of Frobenius at v on U, and let β_j be the corresponding eigenvalues on V. We know that $\alpha_i\beta_j$ are algebraic numbers which are Weil numbers of norm $N(v)$. The ratios of any two roots thus are also algebraic numbers with absolute value 1 at all real places, and so α_i/α_1 has this property.

Let's suppose that the ratios α_i/α_1 are actually roots of unity for a set v of density one. Since W is be defined over a fixed finite extension $E = \mathbf{Q}_p$, the degrees of these ratios has uniformly bounded order over E, and the the orders of these roots of unity also have uniformly bounded order. But then (projectively) only finitely many characteristic polynomials will arise from Frobeneii for a set of density one, , which would imply that U has finite projective image, from which it easily follows that U becomes trivial over a finite extension (remember the determinant is unramified so of finite image). Hence it suffices to show that the α_i/α_1 are all algebraic integers and then use Kronecker's theorem.

For finite places not dividing $N(v)$ this is clear because the valuations of the $\alpha_i\beta_j$ are all trivial and so are their ratios. For finite places dividing $N(v)$ now suppose in addition that A is ordinary. Fix a place above v. If the α_i/α_1 have valuation given by a_i, and β_j/β_1 have valuation b_i, it follows that the quantities $a_i + b_j$ take on precisely two values, zero and either 1 or -1, and they take on each of these values exactly half the time. But then either a_i is constant and thus (considering $i = 1$) equal to 0, or the b_j are all zero, and then half the a_i are zero and half are 1 or -1. But that's clearly only possible if U has odd dimension. So done! □

I suspect the case that $\dim(U) = 2$, even with an ordinary hypothesis, is probably quite hard. But I would be happy to be mistaken.

(I did avoid mentioning Pink's paper [Pin98] in part because in Kisin's talk he *used* the Mumford–Tate conjecture as an ingredient to avoid having to address this Fontaine–Mazur question. Pink also proves some very nice results "at almost all primes".)

154. A talk on my new work with Vesselin Dimitrov and Yunqing Tang on irrationality

Sun, 16 Jun 2024

Here is a video of my talk from the recent 70th birthday conference of Peter Sarnak. During a talk one always forgets to say certain things, so I realized that my blog could be a good place to give some extra context on points I missed. There are three things off the top that I can add before rewatching the talk. The first is that I made a typo in one of my collaborator's name (oops!). The second is that I didn't mention the work of Bost–Charles [BC22], whose influence on our work is clear. Indeed the $m = 0$ version of the holonomy theorem (version III) in this talk is a theorem in their monograph. The third is that my presentation of known irrationality results for *explicit* zeta values makes sense in the context of

framing of my talk, but it's good to note that the irrationality results of Rivoal, Ball-Rivoal, and Zudilin [Riv00, BR01, Zud01] (for example, at least (**edit:** one) of $\zeta(5), \zeta(7), \zeta(9), \zeta(11)$ is irrational) in a closely related direction are amazing theorems. There's probably more to say, and I might add some extra comments if I watch the video again).

The talk

154.1. **Some incidental remarks concerning history I thought about when preparing my talk.** I know from popular accounts [vdP79] that Apéry's result came as a complete surprise. Similarly, the result of Gelfand–Schneider was a complete shock as well. (Hilbert was reputed to say that he didn't think this problem would be solved within his lifetime.) Now these two theorems are "recent enough" so that the memory of their resolution is still within the collective consciousness of mathematicians. In the first case, I still know a bunch of people (Henri Cohen and Frits Beukers) who were actually at Apéry's infamous lecture. But what about Lindemann's proof that π is transcendental? I have no sense as to what was the reaction at the time, in part due to my lack of historical knowledge but also to the lack (as far as I can see) of easily available informal discussions about contemporary mathematics from the 19th century (I assume that personal letters would be the best source). The best (?) I could find was the following (quoted from here):

> In fact his [Lindemann's] proof is based on the proof that e is transcendental together with the fact that $e^{i\pi} = -1$. Many historians of science regret that Hermite, despite doing most of the hard work, failed to make the final step to prove the result concerning which would have brought him fame outside the world of mathematics. This fame was instead heaped on Lindemann but many feel that he was a mathematician clearly inferior to Hermite who, by good luck, stumbled on a famous result.

First, this seems pretty brutal towards Lindemann (to be fair, the continuation of the text does give some more grudging praise of Lindemann). Second, which historians are being referred to here? This seems far too judgmental for the historians I have ever spoken to in real life. If this text is at all accurate, it seems to suggest that Lindemann's result was lauded but perhaps not considered *surprising* to his contemporaries? I feel that this is recent enough that one should be able to get a fuller idea of what was going on at the time.

Going back in time further, I also wonder what Lambert's contemporaries thought of his proof (in the 1760s) that π was irrational. When I was giving a public talk on π in Sydney I looked up Lambert's paper. The introduction is quite amusing, with the following remark that suggests a modern way of thinking not much different to how I think about things today:

> Démontrer que le diametre du cercle n'est point à sa circonférence comme un nombre enteir à nombre entrier, c'est là une chose, dont les géometres ne seront gueres sorpris. On connoit les nombres de Ludolph, les rapports trouvés par Archimede, par Metius, etc. de même qu'un grand nombre de suites infinies, qui toures se rapportent à la quadrature du cercle. Et si la somme de ces suites est unq quantité rationelle, on doit assez naturellement conclure, qu'elle sera ou un nombre entier, ou one fraction très simple. Car, s'il y

> falloit une fraction fort composée, quoi raison y auroit-il, pourquoi plutôt relle que telle autre quelconque?

(Or in translation, errors some combination of mine and google translate):

> We prove that the ratio of the diameter of the circle to its circumference is not rational; something that geometers will hardly be surprised by. We know the number π of Ludolph, and expressions for this number found by Archimedes, by Metius, etc. in terms of a large number of infinite series of rational numbers, which all relate to the squaring of the circle. If the sum of these sequences was a rational quantity, we must quite naturally conclude that it will be either a whole number, or a very simple fraction. For, if a very complicated fraction were necessary, what reason would there be to be equal to such a number rather than any other real (irrational) number?

I guess Occam dates back to the 14th century!

Comment 154.2 (Anonymous)**.** Michel Waldschimdt writes here that all the main ideas for showing the transcendence of π were present in Hermite's 1873 memoir, but doesn't explicitly quote any contemporary reaction de Hermite's work.

This text of Emile Picard [Pic01] on the work of Hermite doesn't either.

Incidentally, in a footnote of a text on the correspondence of Lebesgue, there is mention of the tragic loss all of Hermite's correspondence in a fire of a storage unit ... So all handwritten reactions are probably lost.

Comment 154.3 (Vesselin Dimitrov)**.** It seems that Hermite's own opinion is captured by a letter to Borchardt from 1873 (from right-after Hermite's memoir on the exponential function, where he proved the transcendence of e and started the modern theory of functional rational approximation). The English translation could be something like this:

> I will not venture in search of a demonstration of the transcendence of the number π. Let others try to pull it off. No one would be happier than me in their success. But, believe me, my dear friend, it will not fail to cost them some effort.

In a nice recent popular (and historically well-researched) book "Tales of Impossibility" (David Richeson, Princeton University Press, 2019), there is the following discussion where the "historian" cited is none other than the famed mathematician Hans Freudenthal, but which completely contrasts with how Hermite viewed these things:

> In 1873, the same year that Hermite proved the transcendence of e, a German mathematics student named Ferdinand von Lindemann earned his PhD from Felix Klein in Erlangen. Shortly afterwards, he headed abroad to visit mathematicians in England and France. It was during his stop in Paris that Lindemann met Hermite and was able to discuss with him his methods.
>
> Nine years later, in 1882, Lindemann proved that π is transcendental. Some mathematicians have expressed disappointment that Hermite, who had devised the key ideas for what would eventually be Lindemann's proof, did not prove it. Although Lindemann had a good career, producing 60 PhD students, he was not seen as

> Hermite's equal. But Hermite was generous, inspiring, and shared his knowledge far and wide through his correspondences. Freudenthal pointed out this was often to Hermite's detriment, because it allowed others to achieve the important results — such as Lindemann's proof that it was impossible to square the circle — that were "rightfully" Hermite's.

and then follows Freudenthal's strange and scathing quote in the "Dictionary of Scientific Biography":

> Freudenthal: In a sense, [the proof of the transcendence of e] is paradigmatic of all of Hermite's discoveries. By a slight adaptation of Hermite's proof, Felix Lindemann, in 1882, obtained the much more exciting transcendence of π. Thus, Lindemann, a mediocre mathematician, became even more famous than Hermite for a discovery for which Hermite had laid all the groundwork and that he had come within a gnat's eye of making.

With all the hindsight, it is all too easy to make strong statements!

155. SL_n versus GL_n

Thu, 18 Jul 2024

I recently wrote a paper (with Toby Gee and George Boxer [BCG23a]) on constructing regular algebraic automorphic representations π of (cohomological) weight zero and level one, and therefore also cuspidal cohomology classes in the cohomology of $\mathrm{GL}_n(\mathbf{Z})$ for some values of n.

There was one slightly subtle point which we had to address concerning the relation between the cohomology of $\mathrm{SL}_n(\mathbf{Z})$ and $\mathrm{GL}_n(\mathbf{Z})$, or at least the relationship between the parts of cohomology which come from cuspidal modular forms. I have observed this issue turn up in some different contexts, and that is what I wanted to talk about today. The main message is that from the perspective of the Langlands program, the cohomology of $\mathrm{GL}_n(\mathcal{O}_F)$ is more fundamental than tbe cohomology of $\mathrm{SL}_n(\mathcal{O}_F)$. When $F = \mathbf{Q}$, these groups are "more or less" the same (more on that below), but the differences are more pronounced and significant when $F \neq \mathbf{Q}$. But let's start by talking about the case of classical modular forms, where there is already something a little bit interesting to say. A regular algebraic automorphic representation π for $\mathrm{GL}(2)/\mathbf{Q}$ of level one corresponds to a cuspidal modular eigenform of weight $k \geq 2$ and level one. We know that cuspidal modular forms of weight $k \geq 2$ and level one contribute via Eichler–Shimura to the Betti cohomology groups of the modular curve. As an orbifold, the modular curve can be realized as $\mathbf{H}/\Gamma$ where now $\Gamma = \mathrm{SL}_2(\mathbf{Z})$ rather than $\mathrm{GL}_2(\mathbf{Z})$. In this situation at least, we understand quite well what is happening. These eigenforms give rise to a two-dimensional space inside H^1 of the modular curve, and thus inside $H^1(\Gamma)$, and we understand what the "extra" action of the element

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

is; namely under the Eichler–Shimura isomorphism, it corresponds to the action of complex conjugation (so from the perspective of the Hodge filtration, it takes the holomorphic forms to the antiholomorphic forms and vice-versa). It acts on the

relevant piece of cohomology with trace zero. Note that this no longer holds on non-cuspidal cohomology, for example H^0 is one dimensional in both cases. Of course in cohomological weight zero (which corresponds to weight $k = 2$), there turn out to be no such forms, but the point is that the vanishing of the cuspidal cohomology for $\mathrm{GL}_2(\mathbf{Z})$ is equivalent to the same statement for $\mathrm{SL}_2(\mathbf{Z})$. (Something similar is also true in higher weight as well when there really do exist such forms.)

For larger n there is a similar equivalence; but now the behavior depends on the parity of n. For n odd, the cohomology of $\mathrm{GL}_n(\mathbf{Z})$ and $\mathrm{SL}_n(\mathbf{Z})$ is (rationally) the same because $\mathrm{GL}_n(\mathbf{Z}) \simeq \mathrm{SL}_n(\mathbf{Z}) \times \mathbf{Z}/2\mathbf{Z}$ (then use the Künneth formula). But for n even, a level one weight zero π gives rise to *two copies* of the exterior algebra

$$\bigwedge^{*} \mathbf{C}^{\ell_0}$$

in degrees $[q_0, \ldots, q_0+\ell_0]$, with $\ell_0 = (n-2)/2$, and the action of the "extra" element acts freely on these two copies. All this comes down to the differences in the real representation theory of $\mathrm{GL}_n(\mathbf{R})$ and $\mathrm{GL}_n(\mathbf{R})^+$ which is discussed briefly in the paper but which I won't talk about here.

But what happens for general number fields F? There's a confusion which I have seen in various places even for $n = 2$ about whether one should be considering the cohomology of $\mathrm{SL}_n(\mathcal{O}_F)$ or $\mathrm{GL}_n(\mathcal{O}_F)$. Of course it depends on what exactly one wants to do. But at least if one is interested in computing automorphic representations conjecturally associated to motives which have level one, one should really be considering the cohomology of $\mathrm{GL}_2(\mathcal{O}_F)$. This confusion comes with good pedigree — It turns up in the Serre–Tate correspondence! Tate mentions (October 15, 1969, [Col15a, p.382]) a colloquium by Swan who "disappointed everybody" by computing that $H_1(\mathrm{SL}_2(\mathbf{Z}[\sqrt{-14}]), \mathbf{Z})$ has rank three, compared to the lower bound (coming from the boundary tori) of two. (Note: Tate notes in a later letter [Dec 15] it should be $\sqrt{-10}$, not $\sqrt{-14}$.) Serre responds [Col15a, p.384] on November 15 that he doesn't find this at all surprising, and in fact:

> (via la théorie de Weil cela signifait qu'il existe de courbes elliptiques sur le corps en question qui n'ont pas de multiplication complexe — on n'en doute pas). En fait, vu Weil, il s'impose d'essayer de construire une courbe elliptique sur $\mathbf{Q}(\sqrt{-56})$ ayant bonne réduction partout;

Now I confess that when I first read this quote I interpreted it as a misapprehension on Serre's part, because (since this is SL_2 not GL_2) there need not exist any such elliptic curve. But looking it up again now, I started to have my doubts, and Serre was perhaps more circumspect than I assumed. Indeed chatgpt tells me:

> The phrase "il s'impose d'essayer" in French does not have the same strict sense of necessity as "it is necessary" in English. A more nuanced translation could be "it is imperative to try" or "it is important to try" It suggests a strong recommendation or importance, rather than an absolute necessity.

(Possibly Colmez can confirm this; AI has rendered his go playing superfluous but not yet his skills interpreting for anglophones the nuances of Serre's correspondence.) That's also consistent with how Serre continues:

> je connais trop mal la théorie de Weil pour être sûr que ça doit exister; mais il vaut la peine d'essayer

Later (note the remark on $d = -56$ versus $d = -40$ above), Serre says:

> C'est bien $\mathbf{Q}(\sqrt{-40})$ le corps où Mennicke a trouve que le rang de SL_2 rendu abélien est nombre de classes. Mais il a un corps encore plus beau: $\mathbf{Q}(\sqrt{-109})$ où le GL_2 rendu abélien est infini (c'est une propriété plus forte S1 que la précédente). Ici aussi, on a envie de chercher des courbes elliptiques à bonne réduction.

Perhaps worth adding the modern footnote as well:

> «via la théorie de Weil cela signifiait que...» je m'avançais beaucoup en disant ça (I was talking through my hat).

Of course, 45 years later things have been clarified, at least conjecturally. (We still have no general way to produce motives from cohomology, even for Hilbert modular forms of parallel weight 2.) One perspective which I think is helpful (at least to those who care more about Galois representations) is thinking about the differences between the Galois representations associated to automorphic forms on SL_n versus GL_n. Given a π for the former (say cuspidal algebraic of weight zero and level one), you should think about this as giving a compatible family of *projective* representations:

$$\rho(\pi) : G_F \to \mathrm{PGL}_n(\overline{\mathbf{Q}}_p)$$

which are absolutely irreducible and crystalline of the expected weights and unramified outside $v|p$. Now in this situation,one knows (following for example Patrikis ([Pat19]) that there exists for any such ρ a lift to a genuine representation of G_F which is crystalline at $v|p$ of the right weight for all $v|p$ – this generally requires some parity condition on the weight but we are assuming that here. What is not automatic, however, is that this lift has level $N = 1$ any more; that is, the image of inertia at other primes v may be non-trivial (though of course the image lies in the center). Here there is something special which happens only for $F = \mathbf{Q}$; as observed by Tate, you can *globalize* these local characters and then twist to eliminate all the auxiliary ramification. (This argument is explained by Serre in his 1975 Durham paper which is always impossible to find online; it is used to show that a complex projective representation can be lifted to an Artin representation ramified at the same set of primes.) For other fields, even if the class number is trivial, you get global obstructions coming (via class field theory) from the unit group. (Even for imaginary quadratic fields, where the unit group is not very big, this is still an issue, and the general problem can only be avoided for fields for which the unit group has order 2 and which have a real place, which is quite a restrictive condition when you think about it.) The direct automorphic argument is ultimately quite similar, but there are some traps waiting for the unwary (related to Grunwald–Wang); see the discussion in [LS19].

So for example, it is true that as F ranges over all imaginary quadratic fields, one has

$$H^1_{\mathrm{cusp}}(\mathrm{SL}_2(\mathcal{O}_F), \mathbf{C}) \neq 0$$

for all but finitely many F. But the analogue for $\mathrm{GL}_2(\mathcal{O}_F)$ is not only unknown, but we certainly have:

Conjecture 155.1. *There are infinitely many imaginary quadratic fields F with*

$$H^1_{\mathrm{cusp}}(\mathrm{GL}_2(\mathcal{O}_F), \mathbf{C}) = 0.$$

By the way, from the perspective of Galois representations, one can see why the group above should be non-zero in the case of $\mathrm{SL}_2(\mathcal{O}_F)$. Let $F = \mathbf{Q}(\sqrt{-D})$. All we need to find are modular forms π of weight two with the property that, locally at primes $p|D$, the corresponding Weil-(Deligne) representation on restriction to inertia becomes trivial after restriction to $\mathbf{Q}_p(\sqrt{-D})$ *up to twist*. One easy way to achieve this is to take ramified principal series $\mathrm{PS}(1, \chi)$ for some (local) ramified quadratic character χ. The problem is this leads (globally) to a sign difficulty; if F has prime discriminant, then globally you would want the weight of π to be two and the Nebentypus character to be the quadratic character of conductor Δ_F which is odd, which is a problem. (Sometimes it is not; if $F = \mathbf{Q}(\sqrt{-p})$ and $p \equiv 1 \bmod 4$ then you can take the real character of conductor p, but if $p \equiv -1 \bmod 4$ this doesn't work.) But instead of principal series, one can take certain supercuspidal representations: Assume that $F_p/\mathbf{Q}_p$ is a ramified quadratic extension. Then if χ is a totally ramified character of $F_p^\times$ of order 2^m where $2^m \| p-1$, then the base change of this supercuspidal representation will be unramified up to twist, but the original representation will not be unramified up to twist. It's now easy to construct such forms (and even compute how many of them there are), and see there are plenty of them when the discriminant of Δ_K gets large (one has to avoid CM forms over K which can become non-cuspidal but these are easy to bound.) It's also easy to see that while these base changes are unramified at every place up to a local twist they are not in general unramified everywhere up to a global twist.

The forms one finds in this way by base change are invariant under complex conjugation (now acting on the group), and there is another "geometric" way to show they exist which was originally done by Rohlfs [Roh85], who I believe was the first person to prove the non-vanishing claim above. (In fact, this is one way to start proving base change in this situation.)

When it comes to general number fields, one certainly expects (by functoriality!) that $H^*_{\mathrm{cusp}}(\mathrm{GL}_n(\mathcal{O}_F), \mathbf{C})$ should be non-zero for $n = 79$ say and every number field F, but this is hopeless for almost all fields. Using our arguments (and Newton-Thorne for totally real fields!) One certainly can prove it for many totally real and CM fields (some ramification conditions are required for the arguments to work) using the exact same argument. Of course, when for such fields there exists a cuspidal Hilbert modular form of weight two and level one then you can just used Newton–Thorne directly (see [NT22])! For general fields, as usual, the problem of understanding automorphic forms eludes us.

Curiously enough, while writing this post, there appeared a very recent preprint by Darshan and Raghuram here (see [DR24]) which constructs, for example, cuspidal cohomology classes for GL_n/F of (for example) cohomological weight zero for any number field F which is Galois over a totally real field F^+ of some deep enough level). Clozel [Clo87] did something similar when n is even by automorphic induction, but already for $n = 3$ this no longer works. Assuming all conjectures, the simplest way to construct such forms for $F = \mathbf{Q}$ or any totally real field is to take symmetric squares of Hilbert modular forms (these more or less constitute all the self-dual forms). It seems to me that the forms found by Darshan and Raghuram must be some shadow of these forms over the largest totally real subfield F^+ of F and so one is seeing a hint of non-cyclic base change here which is intriguing! I hope to return to this later when I understand it better.

References

[ABC56] N. C. Ankeny, R. Brauer, and S. Chowla, *A note on the class-numbers of algebraic number fields*, Amer. J. Math. **78** (1956), 51–61. 168

[AC87] Enrico Arbarello and Maurizio Cornalba, *The Picard groups of the moduli spaces of curves*, Topology **26** (1987), no. 2, 153–171. 253

[AC14] Patrick B. Allen and Frank Calegari, *Finiteness of unramified deformation rings*, Algebra Number Theory **8** (2014), no. 9, 2263–2272. 33, 133, 240

[ACC+23] Patrick B. Allen, Frank Calegari, Ana Caraiani, Toby Gee, David Helm, Bao V. Le Hung, James Newton, Peter Scholze, Richard Taylor, and Jack A. Thorne, *Potential automorphy over CM fields*, Ann. of Math. (2) **197** (2023), no. 3, 897–1113. 34, 57, 79, 180, 182, 183, 194, 202, 210, 220

[ADP02] Avner Ash, Darrin Doud, and David Pollack, *Galois representations with conjectural connections to arithmetic cohomology*, Duke Math. J. **112** (2002), no. 3, 521–579. 38

[Ago13] Ian Agol, *The virtual Haken conjecture*, Doc. Math. **18** (2013), 1045–1087, With an appendix by Agol, Daniel Groves, and Jason Manning. 16, 173, 231

[AKT23] Patrick B. Allen, Chandrashekhar Khare, and Jack A. Thorne, *Modularity of* $\mathrm{GL}_2(\mathbf{F}_p)$*-representations over CM fields*, Camb. J. Math. **11** (2023), no. 1, 1–158. 88

[AMP23] Narjess Afzaly, Scott Morrison, and David Penneys, *The classification of subfactors with index at most* $5\frac{1}{4}$, Mem. Amer. Math. Soc. **284** (2023), no. 1405, v+81. 142

[ANT20] Patrick B. Allen, James Newton, and Jack A. Thorne, *Automorphy lifting for residually reducible* l*-adic Galois representations, II*, Compos. Math. **156** (2020), no. 11, 2399–2422. 235

[AP08] Avner Ash and David Pollack, *Everywhere unramified automorphic cohomology for* $\mathrm{SL}_3(\mathbf{Z})$, Int. J. Number Theory **4** (2008), no. 4, 663–675. 154

[AT22] Christos Anastassiades and Jack A. Thorne, *Raising the level of automorphic representations of* GL_{2n} *of unitary type*, J. Inst. Math. Jussieu **21** (2022), no. 4, 1421–1444. 235

[AV04] Alessandro Arsie and Angelo Vistoli, *Stacks of cyclic covers of projective spaces*, Compos. Math. **140** (2004), no. 3, 647–666. 253

[AVA17] Dan Abramovich and Anthony Várilly-Alvarado, *Level structures on abelian varieties and Vojta's conjecture*, Compos. Math. **153** (2017), no. 2, 373–394, With an appendix by Keerthi Madapusi Pera. 208

[Bao22] Chengyang Bao, *Locally induced galois representations with exceptional residual images*, 2022. 283, 301

[Bas63] Hyman Bass, *On the ubiquity of Gorenstein rings*, Math. Z. **82** (1963), 8–28. 75

[BC04] Kevin Buzzard and Frank Calegari, *A counterexample to the Gouvêa-Mazur conjecture*, C. R. Math. Acad. Sci. Paris **338** (2004), no. 10, 751–753. 289

[BC05] ———, *Slopes of overconvergent 2-adic modular forms*, Compos. Math. **141** (2005), no. 3, 591–604. 289

[BC06] J. Bellaïche and G. Chenevier, *Lissité de la courbe de Hecke de* GL_2 *aux points Eisenstein critiques*, J. Inst. Math. Jussieu **5** (2006), no. 2, 333–349. 282

[BC09] Joël Bellaïche and Gaëtan Chenevier, *Families of Galois representations and Selmer groups*, Astérisque (2009), no. 324, xii+314. 282

[BC22] Jean-Benoît Bost and François Charles, *Projective and formal-analytic arithmetic surfaces*, 2022, p. 189. 303

[BCDT01] Christophe Breuil, Brian Conrad, Fred Diamond, and Richard Taylor, *On the modularity of elliptic curves over* $\mathbf{Q}$*: wild 3-adic exercises*, J. Amer. Math. Soc. **14** (2001), no. 4, 843–939 (electronic). 185

[BCG23a] George Boxer, Frank Calegari, and Toby Gee, *Cuspidal cohomology classes for* $\mathrm{GL}_n(\mathbf{Z})$, 2023, submitted. 298, 306

[BCG+23b] George Boxer, Frank Calegari, Toby Gee, James Newton, and Jack Thorne, *The Ramanujan and Sato–Tate conjectures for Bianchi modular forms*, arXiv 2309.15880, 2023. 189

[BCGP21] George Boxer, Frank Calegari, Toby Gee, and Vincent Pilloni, *Abelian surfaces over totally real fields are potentially modular*, Publ. Math. Inst. Hautes Études Sci. **134** (2021), 153–501. 192, 193, 204, 210, 213, 246, 248, 249, 287

[BCGP24] ———, *Modularity theorems for abelian surfaces*, 2024, submitted. 196

[BD05] M. Bertolini and H. Darmon, *Iwasawa's main conjecture for elliptic curves over anticyclotomic $\mathbf{Z}_p$-extensions*, Ann. of Math. (2) **162** (2005), no. 1, 1–64. 236

[BE06] Nigel Boston and Jordan S. Ellenberg, *Pro-p groups and towers of rational homology spheres*, Geom. Topol. **10** (2006), 331–334 (electronic). 154, 276

[Bel21] Joël Bellaïche, *The eigenbook—eigenvarieties, families of Galois representations, p-adic L-functions*, Pathways in Mathematics, Birkhäuser/Springer, Cham, [2021] ©2021. 282

[Bel23] ———, *On self-correspondences on curves*, Algebra Number Theory **17** (2023), no. 11, 1867–1899. 282

[BG99] Jerzy Browkin and Herbert Gangl, *Tame and wild kernels of quadratic imaginary number fields*, Math. Comp. **68** (1999), no. 225, 291–305. 113, 114

[BG04a] Karim Belabas and Herbert Gangl, *Generators and relations for $K_2\mathcal{O}_F$*, K-Theory **31** (2004), no. 3, 195–231. 110, 112

[BG04b] Martin R. Bridson and Fritz J. Grunewald, *Grothendieck's problems concerning profinite completions and representations of groups*, Ann. of Math. (2) **160** (2004), no. 1, 359–373. 74

[BG13] Kevin Buzzard and Toby Gee, *Explicit reduction modulo p of certain 2-dimensional crystalline representations, II*, Bull. Lond. Math. Soc. **45** (2013), no. 4, 779–788. 301

[BG16] ———, *Slopes of modular forms*, Families of automorphic forms and the trace formula, Simons Symp., Springer, [Cham], 2016, pp. 93–109. 147

[BGW17] Manjul Bhargava, Benedict H. Gross, and Xiaoheng Wang, *A positive proportion of locally soluble hyperelliptic curves over $\mathbf{Q}$ have no point over any odd degree extension*, J. Amer. Math. Soc. **30** (2017), no. 2, 451–493, With an appendix by Tim Dokchitser and Vladimir Dokchitser. 57

[BK05] Kevin Buzzard and L. J. P. Kilford, *The 2-adic eigencurve at the boundary of weight space*, Compos. Math. **141** (2005), no. 3, 605–619. 219

[BK14] Armand Brumer and Kenneth Kramer, *Paramodular abelian varieties of odd conductor*, Trans. Amer. Math. Soc. **366** (2014), no. 5, 2463–2516. 204

[BK20] Tobias Berger and Krzysztof Klosin, *Deformations of Saito-Kurokawa type and the paramodular conjecture*, Amer. J. Math. **142** (2020), no. 6, 1821–1875, With and appendix by Chris Poor, Jerry Shurman, and David S. Yuen. 194

[BL05] Yuri F. Bilu and Florian Luca, *Divisibility of class numbers: enumerative approach*, J. Reine Angew. Math. **578** (2005), 79–91. 89

[BL17] Kevin Buzzard and Alan Lauder, *A computation of modular forms of weight one and small level*, Ann. Math. Qué. **41** (2017), no. 2, 213–219. 159

[BLGGT14] Thomas Barnet-Lamb, Toby Gee, David Geraghty, and Richard Taylor, *Potential automorphy and change of weight*, Ann. of Math. (2) **179** (2014), no. 2, 501–609. 5, 8, 32, 122, 188, 189, 210, 264, 298

[BM20] Andrew J. Blumberg and Michael A. Mandell, *K-theoretic Tate-Poitou duality and the fiber of the cyclotomic trace*, Invent. Math. **221** (2020), no. 2, 397–419. 275

[BMR24] David Berghaus, Hartmut Monien, and Danylo Radchenko, *On the computation of modular forms on noncongruence subgroups*, Math. Comp. **93** (2024), no. 347, 1399–1425. 294

[BN18] Nils Bruin and Brett Nasserden, *Arithmetic aspects of the Burkhardt quartic threefold*, J. Lond. Math. Soc. (2) **98** (2018), no. 3, 536–556. 246

[Bor74] Armand Borel, *Stable real cohomology of arithmetic groups*, Ann. Sci. École Norm. Sup. (4) **7** (1974), 235–272 (1975). 136

[BP19] John Bergdall and Robert Pollack, *Slopes of modular forms and the ghost conjecture*, Int. Math. Res. Not. IMRN (2019), no. 4, 1125–1144. 290

[BPP+19] Armand Brumer, Ariel Pacetti, Cris Poor, Gonzalo Tornaría, John Voight, and David S. Yuen, *On the paramodularity of typical abelian surfaces*, Algebra Number Theory **13** (2019), no. 5, 1145–1195. 194

[BR01] Keith Ball and Tanguy Rivoal, *Irrationalité d'une infinité de valeurs de la fonction zêta aux entiers impairs*, Invent. Math. **146** (2001), no. 1, 193–207. 304

[BS21] J. Brüdern and K. Soundararajan, *Common divisors of totients of polynomial sequences*, Math. Z. **299** (2021), no. 1-2, 527–542. 162, 230

[Buz05] Kevin Buzzard, *Questions about slopes of modular forms*, Astérisque (2005), no. 298, 1–15, Automorphic forms. I. 289

[BW00] Armand Borel and Nolan R. Wallach, *Continuous cohomology, discrete subgroups, and representations of reductive groups*, second ed., Mathematical Surveys and Monographs, vol. 67, American Mathematical Society, Providence, RI, 2000. 106

[Cal] Frank Calegari, *Semistable modularity lifting over imaginary quadratic fields*, unpublished. 202

[Cal05] ———, *Irrationality of certain p-adic periods for small p*, Int. Math. Res. Not. (2005), no. 20, 1235–1249. 28

[Cal06] ———, *Mod p representations on elliptic curves*, Pacific J. Math. **225** (2006), no. 1, 1–11. 258

[Cal12] ———, *Even Galois Representations and the Fontaine-Mazur conjecture II*, J. Amer. Math. Soc. **25** (2012), no. 2, 533–554. 216, 234

[Cal15] ———, *The stable homology of congruence subgroups*, Geom. Topol. **19** (2015), no. 6, 3149–3191. 45, 55, 135

[Cal18] ———, *Non-minimal modularity lifting in weight one*, J. Reine Angew. Math. **740** (2018), 41–62. 91

[Cal20] ———, *Motives and L-functions*, Current developments in mathematics 2018, Int. Press, Somerville, MA, [2020] ©2020, pp. 57–123. 255

[Cas69] J. W. S. Cassels, *On a conjecture of R. M. Robinson about sums of roots of unity*, J. Reine Angew. Math. **238** (1969), 112–131. 100

[CC22] Frank Calegari and Shiva Chidambaram, *Rationality of twists of the Siegel modular variety of genus 2 and level 3*, Proc. Amer. Math. Soc. **150** (2022), no. 5, 1975–1984. 248

[CCR20] Frank Calegari, Shiva Chidambaram, and David P. Roberts, *Abelian surfaces with fixed 3-torsion*, ANTS XIV—Proceedings of the Fourteenth Algorithmic Number Theory Symposium, Open Book Ser., vol. 4, Math. Sci. Publ., Berkeley, CA, 2020, pp. 91–108. 249, 251

[CD06] Frank Calegari and Nathan M. Dunfield, *Automorphic forms and rational homology 3-spheres*, Geom. Topol. **10** (2006), 295–329 (electronic). 154

[CDT99] Brian Conrad, Fred Diamond, and Richard Taylor, *Modularity of certain potentially Barsotti-Tate Galois representations*, J. Amer. Math. Soc. **12** (1999), no. 2, 521–567. 185

[CE05] Frank Calegari and Matthew Emerton, *On the ramification of Hecke algebras at Eisenstein primes*, Invent. Math. **160** (2005), no. 1, 97–144. 230

[CE09] ———, *Bounds for multiplicities of unitary representations of cohomological type in spaces of cusp forms*, Ann. of Math. (2) **170** (2009), no. 3, 1437–1446. 40

[CE12] ———, *Completed cohomology—a survey*, Non-abelian fundamental groups and Iwasawa theory, London Math. Soc. Lecture Note Ser., vol. 393, Cambridge Univ. Press, Cambridge, 2012, pp. 239–257. 39, 40

[CEG+16] Ana Caraiani, Matthew Emerton, Toby Gee, David Geraghty, Vytautas Paškūnas, and Sug Woo Shin, *Patching and the p-adic local Langlands correspondence*, Camb. J. Math. **4** (2016), no. 2, 197–287. 94

[CG18a] Frank Calegari and David Geraghty, *Modularity lifting beyond the Taylor–Wiles method*, Inventiones mathematicae **211** (2018), no. 1, 297–433. 49, 68, 77, 89, 107, 135, 170

[CG18b] Frank Calegari and Zoey Guo, *Abelian spiders and real cyclotomic integers*, Trans. Amer. Math. Soc. **370** (2018), no. 9, 6515–6533. 139, 141, 142, 275

[CG20] Frank Calegari and David Geraghty, *Minimal modularity lifting for nonregular symplectic representations*, Duke Math. J. **169** (2020), no. 5, 801–896, With an appendix by Calegari, Geraghty and Michael Harris. 194

[CGZ23] Frank Calegari, Stavros Garoufalidis, and Don Zagier, *Bloch groups, algebraic K-theory, units, and Nahm's conjecture*, Ann. Sci. Éc. Norm. Supér. (4) **56** (2023), no. 2, 383–426. 118

[CH17] Frank Calegari and Zili Huang, *Counting Perron numbers by absolute value*, J. Lond. Math. Soc. (2) **96** (2017), no. 1, 181–200. 79, 84

[Che14] Gaëtan Chenevier, *The p-adic analytic space of pseudocharacters of a profinite group and pseudorepresentations over arbitrary rings*, Automorphic forms and Galois representations. Vol. 1, London Math. Soc. Lecture Note Ser., vol. 414, Cambridge Univ. Press, Cambridge, 2014, pp. 221–285. 38

[Che20] ———, *An automorphic generalization of the Hermite-Minkowski theorem*, Duke Math. J. **169** (2020), no. 6, 1039–1075. 297

[Chi24] Shiva Chidambaram, Mod-p *Galois representations not arising from abelian varieties*, J. Number Theory **259** (2024), 219–237. 257, 258

[CHT08] Laurent Clozel, Michael Harris, and Richard Taylor, *Automorphy for some l-adic lifts of automorphic mod l Galois representations*, Pub. Math. IHES **108** (2008), 1–181. 186

[CLH16] Ana Caraiani and Bao V. Le Hung, *On the image of complex conjugation in certain Galois representations*, Compos. Math. **152** (2016), no. 7, 1476–1488. 93

[Clo87] L. Clozel, *On the cuspidal cohomology of arithmetic subgroups of* SL($2n$) *and the first Betti number of arithmetic* 3*-manifolds*, Duke Math. J. **55** (1987), no. 2, 475–486. 309

[CM09] Frank Calegari and Barry Mazur, *Nearly ordinary Galois deformations over arbitrary number fields*, J. Inst. Math. Jussieu **8** (2009), no. 1, 99–177. 35, 72, 166

[CMS11] Frank Calegari, Scott Morrison, and Noah Snyder, *Cyclotomic integers, fusion categories, and subfactors*, Comm. Math. Phys. **303** (2011), no. 3, 845–896. 100, 274

[CN23] Ana Caraiani and James Newton, *On the modularity of elliptic curves over imaginary quadratic fields*, arXiv 2301.10509, 2023. 88, 184, 284

[Col93] Pierre Colmez, *Périodes des variétés abéliennes à multiplication complexe*, Ann. of Math. (2) **138** (1993), no. 3, 625–683. 130

[Col15a] *Correspondance Serre-Tate. Vol. I*, Documents Mathématiques (Paris) [Mathematical Documents (Paris)], vol. 13, Société Mathématique de France, Paris, 2015, Edited, and with notes and commentaries by Pierre Colmez and Jean-Pierre Serre. 171, 235, 307

[Col15b] *Correspondance Serre-Tate. Vol. II*, Documents Mathématiques (Paris) [Mathematical Documents (Paris)], vol. 14, Société Mathématique de France, Paris, 2015, Edited, and with notes and commentaries by Pierre Colmez and Jean-Pierre Serre. 171, 235

[Col17] Pierre Colmez, *Tate's work and the Serre-Tate correspondence*, Bull. Amer. Math. Soc. (N.S.) **54** (2017), no. 4, 559–573. 171

[CR15] Gaëtan Chenevier and David Renard, *Level one algebraic cusp forms of classical groups of small rank*, Mem. Amer. Math. Soc. **237** (2015), no. 1121, v+122. 297

[CS17] Ana Caraiani and Peter Scholze, *On the generic part of the cohomology of compact unitary Shimura varieties*, Ann. of Math. (2) **186** (2017), no. 3, 649–766. 187

[CS19] Frank Calegari and Joel Specter, *Pseudorepresentations of weight one are unramified*, Algebra Number Theory **13** (2019), no. 7, 1583–1596. 98

[CS24] Ana Caraiani and Peter Scholze, *On the generic part of the cohomology of non-compact unitary Shimura varieties*, Ann. of Math. (2) **199** (2024), no. 2, 483–590. 187, 194

[CT20] Gaëtan Chenevier and Olivier Taïbi, *Discrete series multiplicities for classical groups over* $\mathbf{Z}$ *and level 1 algebraic cusp forms*, Publ. Math. Inst. Hautes Études Sci. **131** (2020), 261–323. 297

[CTS21] Frank Calegari and Naser Talebizadeh Sardari, *Vanishing Fourier coefficients of Hecke eigenforms*, Math. Ann. **381** (2021), no. 3-4, 1197–1215. 153, 240, 283

[CV92] Robert F. Coleman and José Felipe Voloch, *Companion forms and Kodaira-Spencer theory*, Invent. Math. **110** (1992), no. 2, 263–281. 68, 108

[CV19] Frank Calegari and Akshay Venkatesh, *A torsion Jacquet-Langlands correspondence*, Astérisque (2019), no. 409, x+226. 45, 221, 225, 256

[Dem09] Lassina Dembélé, *A non-solvable Galois extension of* $\mathbf{Q}$ *ramified at 2 only*, C. R. Math. Acad. Sci. Paris **347** (2009), no. 3-4, 111–116. 202

[DGV11] Lassina Dembélé, Matthew Greenberg, and John Voight, *Nonsolvable number fields ramified only at 3 and 5*, Compos. Math. **147** (2011), no. 3, 716–734. 202

[Dim19] Vesselin Dimitrov, *A proof of the schinzel-zassenhaus conjecture on polynomials*, 2019. 238

[dJSBVdV90] A. J. de Jong, N. I. Shepherd-Barron, and A. Van de Ven, *On the Burkhardt quartic*, Math. Ann. **286** (1990), no. 1-3, 309–328. 246

[DL16] Hansheng Diao and Ruochuan Liu, *The eigencurve is proper*, Duke Math. J. **165** (2016), no. 7, 1381–1395. 148, 150

[DLP19] Lassina Dembélé, David Loeffler, and Ariel Pacetti, *Non-paritious Hilbert modular forms*, Math. Z. **292** (2019), no. 1-2, 361–385. 270

[Don80] Ron Donagi, *Group law on the intersection of two quadrics*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **7** (1980), no. 2, 217–239. 58

[Dou99] Darrin Doud, *S_4 and $\widetilde{S}_4$ extensions of $\mathbf{Q}$ ramified at only one prime*, J. Number Theory **75** (1999), no. 2, 185–197. 149

[DPSZ02] Amir Dembo, Bjorn Poonen, Qi-Man Shao, and Ofer Zeitouni, *Random polynomials having few or no real zeros*, J. Amer. Math. Soc. **15** (2002), no. 4, 857–892. 84

[DPVZ22] Lassina Dembélé, Alexei Panchishkin, John Voight, and Wadim Zudilin, *Special hypergeometric motives and their L-functions: Asai recognition*, Exp. Math. **31** (2022), no. 4, 1278–1290. 263

[DR24] Nasit Darshan and A. Raghuram, *Cuspidal cohomology for $gl(n)$ over a number field*, 2024. 309

[DS67] H. Davenport and A. Schinzel, *Diophantine approximation and sums of roots of unity*, Math. Ann. **169** (1967), 118–135. 102

[DT94] Fred Diamond and Richard Taylor, *Nonoptimal levels of mod l modular representations*, Invent. Math. **115** (1994), no. 3, 435–462. 163

[DT06] Nathan M. Dunfield and William P. Thurston, *Finite covers of random 3-manifolds*, Invent. Math. **166** (2006), no. 3, 457–521. 276

[DY23] Hansheng Diao and Zijian Yao, *The halo conjecture for gl2*, 2023. 150

[Edi92] Bas Edixhoven, *The weight in Serre's conjectures on modular forms*, Invent. Math. **109** (1992), no. 3, 563–594. 49, 108

[EG23] Matthew Emerton and Toby Gee, *Moduli stacks of étale (φ, Γ)-modules and the existence of crystalline lifts*, Annals of Mathematics Studies, vol. 215, Princeton University Press, Princeton, NJ, [2023] ©2023. 216

[Elk08] Noam D. Elkies, *Shimura curve computations via $K3$ surfaces of Néron-Severi rank at least 19*, Algorithmic number theory, Lecture Notes in Comput. Sci., vol. 5011, Springer, Berlin, 2008, pp. 196–211. 270

[Fer96] Stéfane Fermigier, *Annulation de la cohomologie cuspidale de sous-groupes de congruence de* $\mathrm{GL}_n(\mathbf{Z})$, Math. Ann. **306** (1996), no. 2, 247–256. 298

[FI10] John Friedlander and Henryk Iwaniec, *Opera de cribro*, American Mathematical Society Colloquium Publications, vol. 57, American Mathematical Society, Providence, RI, 2010. 161, 162

[FJ13] Étienne Fouvry and Florent Jouve, *A positive density of fundamental discriminants with large regulator*, Pacific J. Math. **262** (2013), no. 1, 81–107. 92

[FKP22] Najmuddin Fakhruddin, Chandrashekhar Khare, and Stefan Patrikis, *Lifting and automorphy of reducible* $\mathrm{mod}\, p$ *Galois representations over global fields*, Invent. Math. **228** (2022), no. 1, 415–492. 216

[FKRS12] Francesc Fité, Kiran S. Kedlaya, Víctor Rotger, and Andrew V. Sutherland, *Sato-Tate distributions and Galois endomorphism modules in genus 2*, Compos. Math. **148** (2012), no. 5, 1390–1442. 210

[FM12] Benson Farb and Dan Margalit, *A primer on mapping class groups*, Princeton Mathematical Series, vol. 49, Princeton University Press, Princeton, NJ, 2012. 254

[Fon85] Jean-Marc Fontaine, *Il n'y a pas de variété abélienne sur* $\mathbf{Z}$, Invent. Math. **81** (1985), no. 3, 515–538. 6, 298

[For08] Kevin Ford, *The distribution of integers with a divisor in a given interval*, Ann. of Math. (2) **168** (2008), no. 2, 367–433. 267

[FW08] Peter J. Forrester and S. Ole Warnaar, *The importance of the Selberg integral*, Bull. Amer. Math. Soc. (N.S.) **45** (2008), no. 4, 489–534. 82

[GG23] Mathilde Gerbelli-Gauthier, *Limit multiplicity for unitary groups and the stable trace formula*, Algebra Number Theory **17** (2023), no. 12, 2181–2228. 17

[GHM09] Benedict H. Gross, Eriko Hironaka, and Curtis T. McMullen, *Cyclotomic factors of Coxeter polynomials*, J. Number Theory **129** (2009), no. 5, 1034–1043. 140

[Gou01] Fernando Q. Gouvêa, *Where the slopes are*, J. Ramanujan Math. Soc. **16** (2001), no. 1, 75–99. 289

[Gro90] Benedict H. Gross, *A tameness criterion for Galois representations associated to modular forms (mod p)*, Duke Math. J. **61** (1990), no. 2, 445–517. 68, 108

[GV04] Eknath Ghate and Vinayak Vatsal, *On the local behaviour of ordinary Λ-adic representations*, Ann. Inst. Fourier (Grenoble) **54** (2004), no. 7, 2143–2162 (2005). 241

[GZ85] Benedict H. Gross and Don B. Zagier, *On singular moduli*, J. Reine Angew. Math. **355** (1985), 191–220. 130

[Has36] Helmut Hasse, *Zur Theorie der abstrakten elliptischen Funktionenkörper III. Die Struktur des Meromorphismenrings. Die Riemannsche Vermutung*, J. Reine Angew. Math. **175** (1936), 193–208. 279

[Hel07] David Helm, *On maps between modular Jacobians and Jacobians of Shimura curves*, Israel J. Math. **160** (2007), 61–117. 223

[Hid16] Haruzo Hida, *Growth of Hecke fields along a p-adic analytic family of modular forms*, Families of automorphic forms and the trace formula, Simons Symp., Springer, [Cham], 2016, pp. 129–173. 256

[HJK+19] S. Holmin, N. Jones, P. Kurlberg, C. McLeman, and K. Petersen, *Missing class groups and class number statistics for imaginary quadratic fields*, Exp. Math. **28** (2019), no. 2, 233–254. 173

[HK03] Emmanuel Halberstadt and Alain Kraus, *Sur la courbe modulaire $X_E(7)$*, Experiment. Math. **12** (2003), no. 1, 27–40. 215

[HLTT16] Michael Harris, Kai-Wen Lan, Richard Taylor, and Jack Thorne, *On the rigid cohomology of certain Shimura varieties*, Res. Math. Sci. **3** (2016), 3:37. 20, 27, 33, 34, 56, 137, 138, 265

[HMR24] Farshid Hajir, Christian Maire, and Ravi Ramakrishna, *On Ozaki's theorem realizing prescribed p-groups as p-class tower groups*, Algebra Number Theory **18** (2024), no. 4, 771–786. 280

[HS02] Klaus Hulek and G. K. Sankaran, *The geometry of Siegel modular varieties*, Higher dimensional birational geometry (Kyoto, 1997), Adv. Stud. Pure Math., vol. 35, Math. Soc. Japan, Tokyo, 2002, pp. 89–156. 213

[HSBT10] Michael Harris, Nick Shepherd-Barron, and Richard Taylor, *A family of Calabi-Yau varieties and potential automorphy*, Ann. of Math. (2) **171** (2010), no. 2, 779–813. 188, 230, 264

[HW01] J. William Hoffman and Steven H. Weintraub, *The Siegel modular variety of degree two and level three*, Trans. Amer. Math. Soc. **353** (2001), no. 8, 3267–3305. 247

[Iim86] Kiyoaki Iimura, *On the l-rank of ideal class groups of certain number fields*, Acta Arith. **47** (1986), no. 2, 153–166. 180

[Jau81] Jean-François Jaulent, *Unités et classes dans les extensions métabéliennes de degré nl^s sur un corps de nombres algébriques*, Ann. Inst. Fourier (Grenoble) **31** (1981), no. 1, ix–x, 39–62. 180

[Joh17] Christian Johansson, *On the Sato-Tate conjecture for non-generic abelian surfaces*, Trans. Amer. Math. Soc. **369** (2017), no. 9, 6303–6325, With an appendix by Francesc Fité. 210

[Jon75] Antonia J. Jones, *Cyclic overlattices III, IV*, J. Number Theory **7** (1975), no. 3, 267–282; ibid. 7 (1975), no. 3, 283–292. 101, 102

[Kat73] Nicholas M. Katz, *p-adic properties of modular schemes and modular forms*, Modular functions of one variable, III (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972), Lecture Notes in Math., Vol. 350, Springer, Berlin-New York, 1973, pp. 69–190. 108

[Kat04] ———, *Larsen's alternative, moments, and the monodromy of Lefschetz pencils*, Contributions to automorphic forms, geometry, and number theory, Johns Hopkins Univ. Press, Baltimore, MD, 2004, pp. 521–560. 157

[KM19] Kiran S. Kedlaya and Anna Medvedovsky, *Mod-2 dihedral Galois representations of prime conductor*, Proceedings of the Thirteenth Algorithmic Number Theory

Symposium, Open Book Ser., vol. 2, Math. Sci. Publ., Berkeley, CA, 2019, pp. 325–342. 236

[KMP24] Petr Kravchuk, Dalimil Mazáč, and Sridip Pal, *Automorphic spectra and the conformal bootstrap*, Comm. Amer. Math. Soc. **4** (2024), 1–63. 277

[Kob16] Hirotomo Kobayashi, *Class numbers of pure quintic fields*, J. Number Theory **160** (2016), 463–477. 180

[Koh05] Winfried Kohnen, *A very simple proof of the q-product expansion of the* Δ*-function*, Ramanujan J. **10** (2005), no. 1, 71–73. 11

[KW09a] Chandrashekhar Khare and Jean-Pierre Wintenberger, *Serre's modularity conjecture. I*, Invent. Math. **178** (2009), no. 3, 485–504. 6

[KW09b] ———, *Serre's modularity conjecture. II*, Invent. Math. **178** (2009), no. 3, 505–586. 6

[Lam17] Youness Lamzouri, *The number of imaginary quadratic fields with prime discriminant and class number up to H*, Q. J. Math. **68** (2017), no. 4, 1379–1393. 173

[Lan16] Jaclyn Lang, *On the image of the Galois representation associated to a non-CM Hida family*, Algebra Number Theory **10** (2016), no. 1, 155–194. 157

[Laz65] Michel Lazard, *Groupes analytiques p-adiques*, Inst. Hautes Études Sci. Publ. Math. (1965), no. 26, 389–603. 19, 40

[LD21] David Lowry-Duda, *Visualizing modular forms*, Arithmetic geometry, number theory, and computation, Simons Symp., Springer, Cham, [2021] ©2021, pp. 537–557. 294

[Lec18] Emmanuel Lecouturier, *On the Galois structure of the class group of certain Kummer extensions*, J. Lond. Math. Soc. (2) **98** (2018), no. 1, 35–58. 177

[LMF24] The LMFDB Collaboration, *The L-functions and modular forms database*, https://www.lmfdb.org, 2024, [Online; accessed 1 June 2024]. 214, 245, 286

[LMY13] Sheng-Chi Liu, Riad Masri, and Matthew P. Young, *Subconvexity and equidistribution of Heegner points in the level aspect*, Compos. Math. **149** (2013), no. 7, 1150–1174. 30

[Loc94] P. Lockhart, *On the discriminant of a hyperelliptic curve*, Trans. Amer. Math. Soc. **342** (1994), no. 2, 729–752. 209

[Loe11] David Loeffler, *Density of classical points in eigenvarieties*, Math. Res. Lett. **18** (2011), no. 5, 983–990. 72

[LR11] Benjamin Lundell and Ravi Ramakrishna, *New parts of Hecke rings*, Math. Res. Lett. **18** (2011), no. 1, 59–73. 243

[LS76] Ronnie Lee and R. H. Szczarba, *The group* $K_3(Z)$ *is cyclic of order forty-eight*, Ann. of Math. (2) **104** (1976), no. 1, 31–60. 136

[LS19] Jean-Pierre Labesse and Joachim Schwermer, *Central morphisms and cuspidal automorphic representations*, J. Number Theory **205** (2019), 170–193. 308

[LT16] D. D. Long and Morwen B. Thistlethwaite, *Lenstra-Hurwitz cliques and the class number one problem*, J. Number Theory **162** (2016), 564–577. 129

[LTX+22] Yifeng Liu, Yichao Tian, Liang Xiao, Wei Zhang, and Xinwen Zhu, *On the Beilinson-Bloch-Kato conjecture for Rankin-Selberg motives*, Invent. Math. **228** (2022), no. 1, 107–375. 235

[LTXZ23] Ruochuan Liu, Nha Xuan Truong, Liang Xiao, and Bin Zhao, *Slopes of modular forms and geometry of eigencurves*, 2023. 290

[LW22] Jaclyn Lang and Preston Wake, *A modular construction of unramified p-extensions of* $\mathbf{Q}(N^{1/p})$, Proc. Amer. Math. Soc. Ser. B **9** (2022), 415–431. 178

[LWX17] Ruochuan Liu, Daqing Wan, and Liang Xiao, *The eigencurve over the boundary of weight space*, Duke Math. J. **166** (2017), no. 9, 1739–1787. 150, 215

[LZ21] David Loeffler and Sarah Livia Zerbes, *On the birch-swinnerton-dyer conjecture for modular abelian surfaces*, 2021. 285

[Man86] Yu. I. Manin, *Cubic forms*, second ed., North-Holland Mathematical Library, vol. 4, North-Holland Publishing Co., Amsterdam, 1986, Algebra, geometry, arithmetic, Translated from the Russian by M. Hazewinkel. 247

[Man21] Jeffrey Manning, *Patching and multiplicity* 2^k *for Shimura curves*, Algebra Number Theory **15** (2021), no. 2, 387–434. 223, 226

[Mar14] Simon Marshall, *Endoscopy and cohomology growth on* $U(3)$, Compos. Math. **150** (2014), no. 6, 903–910. 17

[Mat23] Kojiro Matsumoto, *On the potential automorphy and the local-global compatibility for the monodromy operators at $p \neq l$ over cm fields*, 2023. 189

[MB90] Laurent Moret-Bailly, *Extensions de corps globaux à ramification et groupe de Galois donnés*, C. R. Acad. Sci. Paris Sér. I Math. **311** (1990), no. 6, 273–276. 234

[Mes86] Jean-François Mestre, *Formules explicites et minorations de conducteurs de variétés algébriques*, Compositio Math. **58** (1986), no. 2, 209–232. 298

[Mia24] Konstantin Miagkov, *Potential automorphy of certain non self-dual 3-dimensional galois representations*, 2024. 265

[Mic04] P. Michel, *The subconvexity problem for Rankin-Selberg L-functions and equidistribution of Heegner points*, Ann. of Math. (2) **160** (2004), no. 1, 185–236. 31

[Mil02] Stephen D. Miller, *The highest lowest zero and other applications of positivity*, Duke Math. J. **112** (2002), no. 1, 83–116. 298

[Mil15a] John C. Miller, *Class numbers in cyclotomic $\mathbf{Z}_p$-extensions*, J. Number Theory **150** (2015), 47–73. 128

[Mil15b] ———, *Real cyclotomic fields of prime conductor and their class numbers*, Math. Comp. **84** (2015), no. 295, 2459–2469. 128

[Mir89] Rick Miranda, *The basic theory of elliptic surfaces*, Dottorato di Ricerca in Matematica. [Doctorate in Mathematical Research], ETS Editrice, Pisa, 1989. 144

[Moc21a] Shinichi Mochizuki, *Inter-universal Teichmüller theory I: Construction of Hodge theaters*, Publ. Res. Inst. Math. Sci. **57** (2021), no. 1-2, 3–207. 201

[Moc21b] ———, *Inter-universal Teichmüller theory II: Hodge-Arakelov-theoretic evaluation*, Publ. Res. Inst. Math. Sci. **57** (2021), no. 1-2, 209–401. 201

[Moc21c] ———, *Inter-universal Teichmüller theory III: Canonical splittings of the log-theta-lattice*, Publ. Res. Inst. Math. Sci. **57** (2021), no. 1-2, 403–626. 201

[Moc21d] ———, *Inter-universal Teichmüller theory IV: Log-volume computations and set-theoretic foundations*, Publ. Res. Inst. Math. Sci. **57** (2021), no. 1-2, 627–723. 201

[Mor84] David R. Morrison, *On K3 surfaces with large Picard number*, Invent. Math. **75** (1984), no. 1, 105–121. 286

[MS15] Richard A. Moy and Joel Specter, *There exist non-CM Hilbert modular forms of partial weight 1*, Int. Math. Res. Not. IMRN (2015), no. 24, 13047–13061. 14, 98, 244, 269

[Mum69] D. Mumford, *A note of Shimura's paper "Discontinuous groups and abelian varieties"*, Math. Ann. **181** (1969), 345–351. 269

[Nek18] Jan Nekovář, *Eichler-Shimura relations and semisimplicity of étale cohomology of quaternionic Shimura varieties*, Ann. Sci. Éc. Norm. Supér. (4) **51** (2018), no. 5, 1179–1252. 92

[Noo06] Rutger Noot, *Lifting systems of Galois representations associated to abelian varieties*, J. Ramanujan Math. Soc. **21** (2006), no. 4, 299–342. 270

[NT21a] James Newton and Jack A. Thorne, *Symmetric power functoriality for holomorphic modular forms*, Publ. Math. Inst. Hautes Études Sci. **134** (2021), 1–116. 235, 298

[NT21b] ———, *Symmetric power functoriality for holomorphic modular forms, II*, Publ. Math. Inst. Hautes Études Sci. **134** (2021), 117–152. 235, 298

[NT22] James Newton and Jack A. Thorne, *Symmetric power functoriality for hilbert modular forms*, 2022. 309

[NT23] James Newton and Jack A. Thorne, *Adjoint Selmer groups of automorphic Galois representations of unitary type*, J. Eur. Math. Soc. (JEMS) **25** (2023), no. 5, 1919–1967. 235

[OS23] Bryce Joseph Orloski and Naser Talebizadeh Sardari, *Limiting distributions of conjugate algebraic integers*, 2023. 288

[Oza11] Manabu Ozaki, *Construction of maximal unramified p-extensions with prescribed Galois groups*, Invent. Math. **183** (2011), no. 3, 649–680. 279

[Pan22] Lue Pan, *On locally analytic vectors of the completed cohomology of modular curves*, Forum Math. Pi **10** (2022), Paper No. e7, 82. 275

[Pat19] Stefan Patrikis, *Variations on a theorem of Tate*, Mem. Amer. Math. Soc. **258** (2019), no. 1238, viii+156. 270, 308

[Pic01] Émile Picard, *L'œuvre scientifique de Charles Hermite*, Ann. Sci. École Norm. Sup. (3) **18** (1901), 9–34. 305

[Pil20] V. Pilloni, *Higher coherent cohomology and p-adic modular forms of singular weights*, Duke Math. J. **169** (2020), no. 9, 1647–1807. 194

[Pin98] Richard Pink, *l-adic algebraic monodromy groups, cocharacters, and the Mumford-Tate conjecture*, J. Reine Angew. Math. **495** (1998), 187–237. 303

[PS18] Mihail Poplavskyi and Grégory Schehr, *Exact persistence exponent for the 2d-diffusion equation and related kac polynomials*, Phys. Rev. Lett. **121** (2018), 150601. 84

[PT15] Stefan Patrikis and Richard Taylor, *Automorphy and irreducibility of some l-adic representations*, Compos. Math. **151** (2015), no. 2, 207–229. 7

[Put15] Andrew Putman, *Stability in the homology of congruence subgroups*, Invent. Math. **202** (2015), no. 3, 987–1027. 276

[PY15] Cris Poor and David S. Yuen, *Paramodular cusp forms*, Math. Comp. **84** (2015), no. 293, 1401–1438. 204

[Qia23] Lie Qian, *Potential automorphy for GL_n*, Invent. Math. **231** (2023), no. 3, 1239–1275. 264

[Qui72] Daniel Quillen, *On the cohomology and K-theory of the general linear groups over a finite field*, Ann. of Math. (2) **96** (1972), 552–586. 136

[Raj04] C. S. Rajan, *Unique decomposition of tensor products of irreducible representations of simple algebraic groups*, Ann. of Math. (2) **160** (2004), no. 2, 683–704. 157

[Rib90] Kenneth A. Ribet, *Multiplicities of Galois representations in Jacobians of Shimura curves*, Festschrift in honor of I. I. Piatetski-Shapiro on the occasion of his sixtieth birthday, Part II (Ramat Aviv, 1989), Israel Math. Conf. Proc., vol. 3, Weizmann, Jerusalem, 1990, pp. 221–236. 221

[Riv00] Tanguy Rivoal, *La fonction zêta de Riemann prend une infinité de valeurs irrationnelles aux entiers impairs*, C. R. Acad. Sci. Paris Sér. I Math. **331** (2000), no. 4, 267–270. 304

[Rob65] Raphael M. Robinson, *Some conjectures about cyclotomic integers*, Math. Comp. **19** (1965), 210–217. 18

[Roh85] J. Rohlfs, *On the cuspidal cohomology of the Bianchi modular groups*, Math. Z. **188** (1985), no. 2, 253–269. 309

[Roz18] Sandra Rozensztajn, *An algorithm for computing the reduction of 2-dimensional crystalline representations of* $\mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$, Int. J. Number Theory **14** (2018), no. 7, 1857–1894. 301

[Roz20] ———, *On the locus of 2-dimensional crystalline representations with a given reduction modulo p*, Algebra Number Theory **14** (2020), no. 3, 643–700. 301

[RS95] K. Rubin and A. Silverberg, *Families of elliptic curves with constant mod p representations*, Elliptic curves, modular forms, & Fermat's last theorem (Hong Kong, 1993), Ser. Number Theory, I, Int. Press, Cambridge, MA, 1995, pp. 148–161. 245

[RW13] Frederick Robinson and Michael Wurtz, *On the magnitudes of some small cyclotomic integers*, Acta Arith. **160** (2013), no. 4, 317–332. 18, 19, 100

[RW22] Oscar Randal-Williams, *Stable cohomology of congruence subgroups*, 2022. 229, 275

[San14] Fabian Sander, *Hilbert-Samuel multiplicities of certain deformation rings*, Math. Res. Lett. **21** (2014), no. 3, 605–615. 76

[Sav04] David Savitt, *Modularity of some potentially Barsotti-Tate Galois representations*, Compos. Math. **140** (2004), no. 1, 31–63. 185

[SBT97] Nicholas I. Shepherd-Barron and Richard Taylor, mod 2 *and* mod 5 *icosahedral representations*, J. Amer. Math. Soc. **10** (1997), no. 2, 283–298. 195

[SC67] Atle Selberg and S. Chowla, *On Epstein's zeta-function*, J. Reine Angew. Math. **227** (1967), 86–110. 130, 131

[Sch03] René Schoof, *Class numbers of real cyclotomic fields of prime conductor*, Math. Comp. **72** (2003), no. 242, 913–937. 172

[Sch15a] George J. Schaeffer, *Hecke stability and weight 1 modular forms*, Math. Z. **281** (2015), no. 1-2, 159–191. 269

[Sch15b] Peter Scholze, *On torsion in the cohomology of locally symmetric varieties*, Ann. of Math. (2) **182** (2015), no. 3, 945–1066. 33, 37, 56, 156, 265

[Sch19] Ciaran Schembri, *Examples of genuine QM abelian surfaces which are modular*, Res. Number Theory **5** (2019), no. 1, Paper No. 11, 12. 206

[SD73] Henry Peter Francis Swinnerton-Dyer, *On l-adic representations and congruences for coefficients of modular forms*, Modular functions of one variable, III (Proc. Internat. Summer School, Univ. Antwerp, 1972), Springer, Berlin, 1973, pp. 1–55. Lecture Notes in Math., Vol. 350. 148

[Ser87] Jean-Pierre Serre, *Sur les représentations modulaires de degré* 2 *de* $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, Duke Math. J. **54** (1987), no. 1, 179–230. 5, 108

[Ser03] ———, *On a theorem of Jordan*, Bull. Amer. Math. Soc. (N.S.) **40** (2003), no. 4, 429–440 (electronic). 26

[Ser18] Vlad Serban, *An infinitesimal p-adic multiplicative Manin-Mumford conjecture*, J. Théor. Nombres Bordeaux **30** (2018), no. 2, 393–408. 72

[Ser19] Jean-Pierre Serre, *Distribution asymptotique des valeurs propres des endomorphismes de Frobenius [d'après Abel, Chebyshev, Robinson,...]*, Astérisque (2019), no. 414, Séminaire Bourbaki. Vol. 2017/2018. Exposés 1136–1150, Exp. No. 1146, 379–425. 273

[Ser22] Vlad Serban, *A finiteness result for p-adic families of Bianchi modular forms*, J. Number Theory **233** (2022), 405–431. 72, 230

[She89] James B. Shearer, *On the distribution of the maximum eigenvalue of graphs*, Linear Algebra Appl. **114/115** (1989), 17–20. 101

[Shi75] Goro Shimura, *On the real points of an arithmetic quotient of a bounded symmetric domain*, Math. Ann. **215** (1975), 135–164. 271

[Sie45] Carl Ludwig Siegel, *The trace of totally positive and real algebraic integers*, Ann. of Math. (2) **46** (1945), 302–312. 271

[Smi24] Alexander Smith, *Algebraic integers with conjugates in a prescribed distribution*, Ann. of Math. (2) **200** (2024), no. 1, 71–122. 274, 275

[Smy71] C. J. Smyth, *On the product of the conjugates outside the unit circle of an algebraic integer*, Bull. London Math. Soc. **3** (1971), 169–175. 238

[Smy84] ———, *The mean values of totally real algebraic integers*, Math. Comp. **42** (1984), no. 166, 663–681. 100

[Sno18] Andrew Snowden, *Singularities of ordinary deformation rings*, Math. Z. **288** (2018), no. 3-4, 759–781. 9, 109

[Sny00] Noah Snyder, *An alternate proof of Mason's theorem*, Elem. Math. **55** (2000), no. 3, 93–94. 275

[Sou79] C. Soulé, *K-théorie des anneaux d'entiers de corps de nombres et cohomologie étale*, Invent. Math. **55** (1979), no. 3, 251–295. 136

[Sou07] K. Soundararajan, *The number of imaginary quadratic fields with a given class number*, Hardy-Ramanujan J. **30** (2007), 13–18. 173

[Spe18] Joel Specter, *The crystalline period of a height one p-adic dynamical system*, Trans. Amer. Math. Soc. **370** (2018), no. 5, 3591–3608. 98

[SS19] Karl Schaefer and Eric Stubley, *Class groups of Kummer extensions via cup products in Galois cohomology*, Trans. Amer. Math. Soc. **372** (2019), no. 10, 6927–6980. 189

[Sta74] H. M. Stark, *Some effective cases of the Brauer-Siegel theorem*, Invent. Math. **23** (1974), 135–152. 119

[Stu21] Eric Stubley, *Classical forms of weight one in ordinary families*, 2021. 257

[SW24] Will Sawin and Melanie Matchett Wood, *Finite quotients of 3-manifold groups*, Invent. Math. **237** (2024), no. 1, 349–440. 276

[Tat76] John Tate, *Relations between K_2 and Galois cohomology*, Invent. Math. **36** (1976), 257–274. 136

[Tay12] Richard Taylor, *The image of complex conjugation in l-adic representations associated to automorphic forms*, Algebra Number Theory **6** (2012), no. 3, 405–435. 188

[Tay20] Noah Taylor, *Sato-Tate distributions on Abelian surfaces*, Trans. Amer. Math. Soc. **373** (2020), no. 5, 3541–3559. 210

[Tay21] Noah Taylor, *The index of $\mathbf{T}^{an}$ in $\mathbf{T}$*, 2021. 260

[Tay22] Noah Taylor, *On seven conjectures of Kedlaya and Medvedovsky*, J. Number Theory **231** (2022), 333–348. 236

[Tho12] Jack A. Thorne, *On the automorphy of l-adic Galois representations with small residual image*, J. Inst. Math. Jussieu **11** (2012), no. 4, 855–920, With an appendix by Robert Guralnick, Florian Herzig, Richard Taylor and Thorne. 32, 122

[TV16] David Treumann and Akshay Venkatesh, *Functoriality, Smith theory, and the Brauer homomorphism*, Ann. of Math. (2) **183** (2016), no. 1, 177–228. 220

[Vak06] Ravi Vakil, *Murphy's law in algebraic geometry: badly-behaved deformation spaces*, Invent. Math. **164** (2006), no. 3, 569–590. 281

[Č18] Kęstutis Česnavičius, *The Manin constant in the semistable case*, Compos. Math. **154** (2018), no. 9, 1889–1920. 159

[vdP79] Alfred van der Poorten, *A proof that Euler missed...Apéry's proof of the irrationality of* $\zeta(3)$, Math. Intelligencer **1** (1978/79), no. 4, 195–203, An informal report. 304

[Ven02] Otmar Venjakob, *On the structure theory of the Iwasawa algebra of a p-adic Lie group*, J. Eur. Math. Soc. (JEMS) **4** (2002), no. 3, 271–311. 19, 40

[vGT94] Bert van Geemen and Jaap Top, *A non-selfdual automorphic representation of* GL_3 *and a Galois representation*, Invent. Math. **117** (1994), no. 3, 391–401. 7, 265

[Von18] Jan Vonk, *Modular eigenforms at the boundary of weight space*, Res. Number Theory **4** (2018), no. 2, Paper No. 23, 13. 220

[Wag76] J. B. Wagoner, *Continuous cohomology and p-adic K-theory*, Algebraic K-theory (Proc. Conf., Northwestern Univ., Evanston, Ill., 1976), Lecture Notes in Math., Vol. 551, Springer, Berlin-New York, 1976, pp. 241–248. 136

[Wat04] Mark Watkins, *Class numbers of imaginary quadratic fields*, Math. Comp. **73** (2004), no. 246, 907–938. 172

[Wie14] Gabor Wiese, *On Galois representations of weight one*, Doc. Math. **19** (2014), 689–707. 108

[WWE20] Preston Wake and Carl Wang-Erickson, *The rank of Mazur's Eisenstein ideal*, Duke Math. J. **169** (2020), no. 1, 31–115. 176, 177

[Yoo19] Hwajong Yoo, *Non-optimal levels of a reducible* $\bmod \ell$ *modular representation*, Trans. Amer. Math. Soc. **371** (2019), no. 6, 3805–3830. 71

[Zud01] Wadim Zudilin, *One of the numbers* $\zeta(5)$, $\zeta(7)$, $\zeta(9)$, $\zeta(11)$ *is irrational*, Uspekhi Mat. Nauk **56** (2001), no. 4(340), 149–150. 304

[Zyw15] David Zywina, *The inverse Galois problem for* $\mathrm{PSL}_2(\mathbf{F}_p)$, Duke Math. J. **164** (2015), no. 12, 2253–2292. 25

[Zyw23] ———, *The inverse Galois problem for orthogonal groups*, Trans. Amer. Math. Soc. Ser. B **10** (2023), 1173–1211. 144

Email address: fcale@math.uchicago.edu

The University of Chicago, 5734 S University Ave, Chicago, IL 60637, USA