Lecture 3

Kevin Buzzard *

November 14, 2001

1 Maps between spaces of forms

Recall that we have been thinking of modular forms as rules defined on test objects $(E/R,\omega)$ or $(E/R,\omega,Y)$. Hence if one has a "natural" map F which sends a test object to another test object, then F induces a natural map between spaces of modular forms: if f is a modular form, then one can define a modular form F^*f as being the rule sending a test object T to $f(F(T))$. One needs to check that this rule satisfies the axioms, but if F is sufficiently natural then this kind of check should be very straightforward.[1] We use the upper star notation because one easily checks that the map on forms goes the other way to the map F. We now give some concrete examples of this phenomenon.

- If $(E/R,\omega,Y)$ is a ρ-overconvergent test object, then forgetting Y gives us a classical test object $(E/R,\omega)$. Hence if f is a classical modular form, one can define a ρ-overconvergent form as being the rule sending $(E/R,\omega,Y)$ to $f(E/R,\omega)$. This way we get a natural map from classical forms to ρ-overconvergent forms.
- if $r\in R_0$ and $\rho_1=r\rho_2$, and if $(E/R,\omega,Y)$ is a ρ_2-overconvergent test object, then $(E/R,\omega,rY)$ is a ρ_1-overconvergent test object, and hence r gives us a natural map from the space of ρ_1-overconvergent forms to the space of ρ_2-overconvergent forms.

One very interesting collection of maps between spaces of forms, namely the Hecke operators, do not quite fit into this framework, but are only a mild generalisation of it, as we shall now see.

2 Hecke operators

If $(E/R,\omega)$ is a classical test object, and l is a prime, then for a finite locally-free subgroup scheme $C\subset E$ of order l defined over R, we can form the quotient

*The author would like to thank David Whitehouse for supplying him with a copy of the notes for the lecture

[1] One could in fact define "natural" as meaning "such that this check works"!

curve E/C and we have a natural projection map $\pi : E \to (E/C)$, an isogeny of degree l. One can form the dual isogeny $\pi^\vee : (E/C) \to E$. It is often true that $(E/C, (\pi^\vee)^*\omega)$ is another classical test object—the only troublesome point is that the differential might vanish, but this will not happen if, for example, R is a $\mathbf{Z}[1/l]$-algebra, as then π and its dual will both be etale. Let us hence assume that R is a $\mathbf{Z}[1/l]$-algebra.

There could be, in general, more than one subgroup of E of order l, and hence we do not yet have a "natural" rule sending one test object to another, as in the previous section. However, we get around this difficulty by simply considering *all* subgroups at once! Before we make this rigorous, we recall some facts about the group scheme $(\mathbf{Z}/l\mathbf{Z})^2$, considered as an etale group scheme over $\mathbf{Z}[1/l]$. This group scheme is essentially just a copy of the abelian group $(\mathbf{Z}/l\mathbf{Z})^2$ over each point of $\mathrm{Spec}(\mathbf{Z}[1/l])$, and one can easily check that it has precisely $l+1$ locally free subgroups of order l, corresponding to our usual intuition from group theory. Let us label these subgroups $C_1, C_2, \ldots, C_{l+1}$. Note that if R is any $\mathbf{Z}[1/l]$-algebra then the base extensions C_i/R are still subgroups of $(\mathbf{Z}/l\mathbf{Z})^2/R$, and we shall refer to these groups as C_i for short.

Let f be a modular form of weight k, defined over a $\mathbf{Z}[1/l]$-algebra R_0. We will define a new modular form $T_l f$ as follows: If $(E/R, \omega)$ is a test object, then $E[l]$ will be locally isomorphic, in the etale topology, to $(\mathbf{Z}/l\mathbf{Z})^2$. More concretely, this implies that there will be a finite etale over-ring $R' \supset R$ such that over R', $E[l]$ becomes isomorphic to $(\mathbf{Z}/l\mathbf{Z})^2$. Choose such an isomorphism. Let $C_1, C_2, \ldots, C_{l+1}$ be the corresponding $l+1$ subgroups of $E[l] \cong (\mathbf{Z}/l\mathbf{Z})^2$, and define $T_l f(E/R, \omega) = l^{k-1} \sum_{i=1}^{l+1} f((E/C_i)/R', (\pi_i^\vee)^*\omega)$. Here π_i denotes the projection $E \to E/C_i$.

This definition has two subtle problems associated to it, one of which I was not in fact aware of before the Arizona Winter School, and I shall sketch how one gets around these problems. The first is that we chose an isomorphism $E[l] \cong (\mathbf{Z}/l\mathbf{Z})^2$ over R'. If $\mathrm{Spec}(R')$ is not connected then there is the issue that different isomorphisms will yield different choices of $C_1, \ldots, C_{l+1}$. So one has to check that different choices yield the same result. Fortunately, this is not difficult to do, because one can reduce to the case of a local ring, and the spectrum of a local ring is connected. We thank Bjorn Poonen for pointing out this subtlety, and Brian Conrad for explaining how to get around it.

The second problem is that we extended our base from R to R', and hence it looks like $T_l f(E/R, \omega)$ will be an element of R' rather than R. One can use a generalisation of Galois theory, or what the experts would call "a descent argument", to prove that $T_l f(E/R, \omega)$ is in fact in R.

The above discussion yields a map T_l from the space of classical weight k modular forms over $\mathbf{Z}[1/l]$ to itself, and also a map T_l from the space of p-adic modular forms to itself, as long as $l \neq p$. In the p-adic setting there is also a very important Hecke operator at p, but its definition is slightly more subtle and we shall come back to it later. As a brief summary of the problem, what will happen is that for an elliptic curve defined over a p-adic ring, it is frequently the case that not all subgroups of order p are the same—one of them is more "canonical" than the others. We can define a Hecke operator U_p by quotienting out by the p non-canonical subgroups of order p. We now make all this more precise.

3 A measure of supersingularity on an elliptic curve

For simplicity now, let K be a finite extension of $\mathbf{Q}_p$, Let $\mathcal{O}_K$ denote the integers of K. There is a valuation map $v : \mathcal{O}_K\backslash\{0\} \to \mathbf{Q}_{\geq 0}$, normalised so that $v(p) = 1$.

Let R denote $\mathcal{O}_K/p\mathcal{O}_K$. Note that R may well not be the residue field of K—in fact this is exactly the point: if K is highly ramified, then R will contain lots of nilpotent elements. The valuation map above induces $v : R\backslash\{0\} \to [0,1)\cap\mathbf{Q}$, with the property that $v(ur) = v(r)$ for all $u \in R^{\times}$.

Let E/K be an elliptic curve with good reduction. By definition of "good reduction", there is an elliptic curve $\mathcal{E}/\mathcal{O}_K$ with generic fibre E. Define $\overline{E}/R$ to be the base change of $\mathcal{E}$ to R. The R-module $H^0(\overline{E},\Omega^1_{\overline{E}/R})$ is projective of rank 1, and hence free of rank 1, over R. If ω is an R-basis for this module, then by definition, ω is a non-vanishing differential. Furthermore, such ω exist, and are unique up to multiplication by an element of $R^{\times}$.

If A denotes the Hasse invariant, then $A(\overline{E},\omega) \in R$ is an element which is either equal to 0, or has a valuation which is independent of choice of non-vanishing ω. Let us say that E is "very supersingular" if $A(\overline{E},\omega) = 0$, and that E is "not too supersingular" otherwise.

Assume that E is not too supersingular. Then $v(A(\overline{E},\omega))$ is independent of choice of ω, and is a rational in $[0,1)$. Define $v(E)$ to be this rational. By the definition of the Hasse invariant, $v(E) = 0$ iff E has good ordinary reduction. For completion, define $v(E) = 0$ if E has bad reduction.

This definition gives us another way of understanding the "Y" part of the definition of an overconvergent test object, in some simple cases: if R_0 is the integers in a finite extension of $\mathbf{Q}_p$, and $0 \neq \rho \in R_0$ with $0 \leq v(\rho) < 1$, then for a test object $(E/R_0,\omega,Y)$ we have $YE_{p-1}(E,\omega) = \rho$ and this implies that $v(E) \leq v(\rho)$. On the other hand, if $(E/R_0,\omega)$ is a classical test object, then Y will exist making $(E/R_0,\omega,Y)$ a ρ-overconvergent test object iff $v(E) \leq v(\rho)$, because if the inequality holds then one can define $Y = \rho/E_{p-1}(E,\omega)$.

More generally, if R is an arbitrary p-adically complete R_0-algebra, then a ρ-overconvergent test object defined over R can be thought of, loosely speaking, as a family of elliptic curves E all of which have $v(E) \leq v(\rho)$. In fact, this can be made more rigorous, as we are about to see.

4 The rigid-analytic viewpoint

This section is rather vague, because I did not want to get bogged down with the details of the foundations of rigid analysis. The reader is hence asked to take on board the fact that there is a good p-adic analogue of the theory of Riemann surfaces, namely the theory of rigid-analytic curves.

Let N be an integer prime to p. The modular curve $X_1(N)$ parameterises (generalised) elliptic curves equipped with a point of order N. If $\mathcal{E}$ is the universal elliptic curve over $X_1(N)$, then one can define a sheaf ω on $X_1(N)$ as being the pushforward of the differentials on $\mathcal{E}/X_1(N)$ (and being careful at cusps). This sheaf is locally free of rank 1, and one can think of a classical modular form as being a section of $\omega^{\otimes k}$.

The problem comes when one wants to start "throwing away" elliptic curves. For example, let us try and consider only the ordinary locus of $X_1(N)$, that is, let us consider $X_1(N)$ over, say, $\overline{\mathbf{Q}}_p$, and let us consider the locus of points $X_1(N)^{\mathrm{ord}}$ which correspond to curves with good ordinary, or multiplicative, reduction. This set contains infinitely many points, as does its complement. Hence there is no way that this set can possibly be the set of points of some kind of subvariety of $X_1(N)$, as any non-trivial closed subvariety of a curve is finite, and any non-trivial open subvariety has finite complement.

Fortunately, if one words over a complete base field like $\mathbf{Q}_p$ or $\mathbf{C}_p$, then $X_1(N)^{\mathrm{ord}}$ has the structure of a *rigid-analytic space*. What is happening here is that $X_1(N)^{\mathrm{ord}}$ is some kind of p-adic analogue of a Riemann surface. The theory of rigid analytic spaces is set up from scratch in the book "non-Archimedean analysis" by Bosch, Guentzer and Remmert, and in several other places, but the reader with less patience can find a summary of the theory in Peter Schneider's article in the 1996 Durham proceedings. Let us just think of these things as being p-adic analogues of Riemann surfaces, and let us use the theory of complex analytic geometry as a guide to what we can do. From this viewpoint, $X_1(N)^{\mathrm{ord}}$ is an open subvariety of $X_1(N)$, and it will inherit an analytic sheaf ω^{an} of rank 1. The theory of rigid spaces is precisely what one needs to give a good geometric feel to the theory of p-adic modular forms. For example, one can check that if K is a finite extension of $\mathbf{Q}_p$ with integers $\mathcal{O}_K$ then the global sections of $(\omega^{\mathrm{an}})^{\otimes k}$ over K are precisely the 1-overconvergent modular forms defined over $\mathcal{O}_K$.

More generally, if $0 \leq r < 1$ is rational, one can define $X_1(N)_{\geq r}$ as $X_1(N)$ with all points corresponding to elliptic curves E which are either much too supersingular, or have $v(E) > r$, removed. Although it is slightly dangerous to draw a picture of a p-adic Riemann surface, one can think of these objects as looking rather like classical Riemann surfaces with small discs removed. This is because the regions of $X_1(N)$ corresponding to elliptic curves with supersingular reduction are the preimages in the generic fibre of the supersingular points, and one can easily be convinced that the pre-image of a smooth point in the special fibre is a disc in the generic fibre (for example, consider the projective line over $\mathbf{C}_p$: the pre-image of the origin in the special fibre is $\{z \in \mathbf{C}_p : |z| < 1\}$).

One can easily analyse these so-called "supersingular discs". If one chooses a trivialisation of ω^{p-1} on each disc, then the form E_{p-1} gives a parameter on these discs, which can now be thought of as open discs with radius 1. The elliptic curves with $0 < v(E) < 1$ are the curves on the boundary of these discs, and v can be thought of as the valuation of the parameter. The closed disc of radius $1/p$ with centre the zero of E_{p-1} is the region consisting of elliptic curves which are "much too supersingular". The space $X_1(N)_{\geq r}$ corresponds to $X_1(N)$ with open discs radius p^{-r} removed.

Again, one has the powerful geometric definition of a ρ-overconvergent form of weight k over K, a finite extension of $\mathbf{Q}_p$: it is a section of $(\omega^{\mathrm{an}})^{\otimes k}$ on $X_1(N)_{\geq r}$, where $r = v(\rho)$. Note that if we stick to ρ with $v(\rho) < 1$ then the definition does not even depend on a choice of lifting of the Hasse invariant, and in particular this method gives us a means of avoiding the thorny problems associated with lifting the Hasse invariant to characteristic zero in many cases of interest—one simply has to lift the Hasse invariant on each supersingular disc, which is possible even if p is small.

5 Canonical subgroups and the U operator

We now come back to the Hecke operator at p. We start with a specific example which the author finds very illuminating, because it really shows a concrete example of the canonical subgroup of an elliptic curve.

If $a \in \overline{\mathbf{Q}}_2$ with $|a| \leq 1$ then define the elliptic curve

$$E_a : y^2 + y + axy = x^3 + x^2.$$

One can reduce this curve mod the prime above 2, and there are two cases: if $|a| = 1$ then E_a reduces to the curve $y^2 + y + \bar{a}xy = x^3 + x^2$, which is an ordinary elliptic curve over $\overline{\mathbf{F}}_2$. On the other hand, if $|a| < 1$ then E_a reduces to $y^2 + y = x^3 + x^2$, which is supersingular.

Let's put this curve into canonical form: define $Y = y + \frac{1}{2}(1 + ax)$ and the equation for E_a becomes $Y^2 = f(x)$, where

$$f(x) = x^3 + \left(\frac{a^2}{4} + 1\right) x^2 + \frac{a}{2}x + \frac{1}{4}.$$

The points of order 2 on E_a correspond to the roots of $f(x)$. What are the valuations of these roots? This is easy to establish via the theory of the Newton Polygon.

If $|a| = 1$ then the valuations of the coefficients of $f(x)$ are $0, -2, -1, -2$ respectively, and hence f has one root with valuation -2 and two roots with valuation 0. The root with valuation -2 is of course the one that reduces to the point at infinity in the reduction map, and indeed one expects exactly one non-zero point to have this property because the 2-torsion in generic fibre has order 4, and the 2-torsion in the special fibre has order only 2.

In the supersingular reduction case, the special fibre has no 2-torsion at all. However, if $|a| = 1 - \epsilon$ with ϵ small, then a similar Newton polygon argument shows that one of the roots of f has valuation $-2 + 2\epsilon$ and the other two have valuation $-\epsilon$. All three roots have negative valuation, as expected, but one sticks out like a sore thumb. This point generates the so-called "canonical subgroup" of E_a.

Finally, if $|a|$ is very small, then all three roots have valuation $-2/3$ and it is hard to distinguish between them in any canonical manner.

One should think of a as being a function on $X_0(1)$, with $|a| < 1$ on the supersingular locus and $|a| \geq 1$ on the ordinary locus. The area where $|a| = 1 - \epsilon$ with ϵ small then corresponds to the region near the boundary of the supersingular disc, and the example shows that for elliptic curves near the boundary of the disc, even though they have supersingular reduction, they still have a canonical subgroup of order 2.

This example is a special case of the following phenomenon (whose proof is just a long elaboration of what we have just seen above):

Theorem. *If K is a finite extension of $\mathbf{Q}_p$ and E/K is an elliptic curve with $v(E) < \frac{p}{p+1}$ then E has a canonical subgroup of order p. Furthermore, this canonical subgroup varies smoothly as E varies smoothly, and hence can be defined for a family of elliptic curves over a p-adic base.*