ESSENTIAL NUMBER THEORY

Modularity lifting theorems

Toby Gee

2022

vol. 1 no. 1

msp

Essential Number Theory
Vol. 1, No. 1, 2022
https://doi.org/10.2140/ent.2022.1.73

Modularity lifting theorems

Toby Gee

Updated lecture notes from 2013 Arizona winter school.

1.	Introduction	73
2.	Galois representations	74
3.	Galois deformations	85
4.	Modular and automorphic forms	97
5.	The Taylor–Wiles–Kisin method	106
6.	Relaxing the hypotheses	121
Acknowledgements		123
References		123

1. Introduction

The main aim of these notes is to explain modularity/automorphy lifting theorems for two-dimensional p-adic representations, using wherever possible arguments that go over to the (essentially conjugate self-dual) n-dimensional case. In particular, we use improvements on the original Taylor–Wiles method due to Diamond, Fujiwara and Kisin, and we explain (in the case $n = 2$) Taylor's arguments [2008] that avoid the use of Ihara's lemma. For the most part I ignore the issues which are local at p, focusing on representations which satisfy the Fontaine–Laffaille condition.

1.1. ***Notation.*** Much of this notation will also be introduced in the text, but I have tried to collect together various definitions here, for ease of reading. Throughout these notes, $p > 2$ is a prime greater than two. In the earlier stages of the notes, we discuss n-dimensional p-adic and mod p representations, before specialising to the case $n = 2$. When we do so, we assume that $p \nmid n$. (Of course, in the case $n = 2$, this follows from our assumption that $p > 2$.)

If M is a field, we let G_M denote its absolute Galois group $\mathrm{Gal}(\overline{M}/M)$, where $\overline{M}$ is some choice of separable closure of M. We write ε_p (or just ε) for the p-adic cyclotomic character. We fix an algebraic closure $\overline{\mathbb{Q}}$ of $\mathbb{Q}$, and regard all algebraic extensions of $\mathbb{Q}$ as subfields of $\overline{\mathbb{Q}}$. For each prime p we fix an algebraic closure $\overline{\mathbb{Q}}_p$

MSC2020: 11S37.
Keywords: automorphic forms, modularity lifting, Langlands program.

of $\mathbb{Q}_p$, and we fix an embedding $\overline{\mathbb{Q}} \hookrightarrow \overline{\mathbb{Q}}_p$. In this way, if v is a finite place of a number field F, we have a homomorphism $G_{F_v} \hookrightarrow G_F$. We also fix an embedding $\overline{\mathbb{Q}} \hookrightarrow \mathbb{C}$. If $L/\mathbb{Q}_p$ is algebraic, then we write $\mathcal{O}_L$ for the ring of integers of L, and $k(L)$ for its residue field.

We normalize the definition of Hodge–Tate weights so that all the Hodge–Tate weights of the p-adic cyclotomic character ε_p are -1.

If R is a local ring, we write $\mathfrak{m}_R$ for the maximal ideal of R.

We let ζ_p be a primitive p-th root of unity.

We use the terms "modularity lifting theorem" and "automorphy lifting theorem" more or less interchangeably.

2. Galois representations

Modularity lifting theorems prove that certain Galois representations are modular, in the sense that they come from modular forms. We begin in this first chapter by introducing Galois representations, and explaining some of their basic properties.

2.1. ***Basics of Galois representations (and structure of Galois groups).*** Let K'/K be a (not necessarily finite) normal and separable extension of fields. Then the Galois group $\mathrm{Gal}(K'/K)$ is the group

$$\{\sigma \in \mathrm{Aut}(K') : \sigma|_K = \mathrm{id}_K\}.$$

This has a natural topology, making it a compact Hausdorff totally disconnected topological group; equivalently, it is a profinite group. This can be expressed by the topological isomorphism

$$\mathrm{Gal}(K'/K) \cong \varprojlim_{\substack{K''/K \text{ finite normal} \\ K'' \subseteq K'}} \mathrm{Gal}(K''/K),$$

where the finite groups $\mathrm{Gal}(K''/K)$ have the discrete topology. Then Galois theory gives a bijective correspondence between intermediate fields $K' \supset K'' \supset K$ and closed subgroups $H \subset \mathrm{Gal}(K'/K)$, with K'' corresponding to $\mathrm{Gal}(K'/K'')$ and H corresponding to K^H; see, e.g., Section 1.6 of [Gruenberg 1967].

Fix a separable closure $\overline{K}$ of K, and write $G_K := \mathrm{Gal}(\overline{K}/K)$. Let L be a topological field; then a *Galois representation* is a continuous homomorphism $\rho : G_K \to \mathrm{GL}_n(L)$ for some n. The nature of these representations depends on the topology on L. For example, if L has the discrete topology, then the image of ρ is finite, and ρ factors through a finite Galois group $\mathrm{Gal}(K''/K)$.

Exercise 2.2. If $L = \mathbb{C}$ with the usual topology, then $\rho(G_K)$ is finite, and ρ is conjugate to a representation valued in $\mathrm{GL}_n(\overline{\mathbb{Q}})$.

On the other hand, if $L/\mathbb{Q}_p$ is a finite extension with the p-adic topology, then there can be examples with infinite image. The rest of these notes will be concerned with these p-adic representations. For example, if $p \neq \operatorname{char} K$, we have the p-adic cyclotomic character $\varepsilon_p : G_K \to \mathbb{Z}_p^\times$, which is uniquely determined by the requirement that if $\sigma \in G_K$ and $\zeta \in \overline{K}$ with $\zeta^{p^m} = 1$ for some n, then $\sigma(\zeta) = \zeta^{\varepsilon_p(\sigma) \pmod{p^m}}$. More interesting examples arise from geometry, as we explain in Section 2.21 below.

Fact 2.3. If $L/\mathbb{Q}_p$ is an algebraic extension, and $\rho : G_K \to \mathrm{GL}_n(L)$ is a continuous representation, then $\rho(G_K) \subseteq \mathrm{GL}_n(M)$ for some $L \supset M \supset \mathbb{Q}_p$ with $M/\mathbb{Q}_p$ finite.

Proof. This follows from the Baire category theorem; see, e.g., the proof of Corollary 5 of [Dickinson 2001b] for the details. □

Exercise 2.4. If $L/\mathbb{Q}_p$ is an algebraic extension, and $\rho : G_K \to \mathrm{GL}_n(L)$ is a continuous representation, then ρ is conjugate to a representation in $\mathrm{GL}_n(\mathcal{O}_L)$.

Any finite-dimensional Galois representation has a Jordan–Hölder sequence, and thus a well-defined semisimplification.

Fact 2.5. Two Galois representations $\rho, \rho' : G_K \to \mathrm{GL}_n(L)$ have isomorphic semisimplifications if and only if $\rho(g), \rho'(g)$ have the same characteristic polynomials for each $g \in G_K$. If $\operatorname{char} L = 0$ (or indeed if $\operatorname{char} L > n$), then this is equivalent to $\operatorname{tr} \rho(g) = \operatorname{tr} \rho'(g)$ for all $g \in G_K$.

Proof. This is a consequence of the Brauer–Nesbitt theorem, [Curtis and Reiner 1962, 30.16] □

As a corollary of the previous exercise and fact, we see that p-adic representations have well-defined semisimplified reductions modulo p. Indeed, given $\rho : G_K \to \mathrm{GL}_n(L)$ with $L/\mathbb{Q}_p$ algebraic, we may conjugate ρ to be valued in $\mathrm{GL}_n(\mathcal{O}_L)$, reduce modulo the maximal ideal and semisimplify to get a semisimple representation $\bar{\rho} : G_K \to \mathrm{GL}_n(k(L))$, whose characteristic polynomials are determined by those of ρ.

Remark 2.6. We really do have to semisimplify here; to see why, think about the reductions modulo p of the matrices $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 1 & p \\ 0 & 1 \end{smallmatrix}\right)$.

2.7. ***Local representations with p ≠ l: the monodromy theorem.*** In this section we will let $K/\mathbb{Q}_l$ be a finite extension, for some prime $l \neq p$. In order to study the representations of G_K, we firstly recall something of the structure of G_K itself; see, e.g., [Serre 1979] for further details. Let ϖ_K be a uniformizer of $\mathcal{O}_K$, let $k = k(K)$ denote the residue field of K, and let $\mathrm{val}_K : K^\times \twoheadrightarrow \mathbb{Z}$ be the ϖ_K-adic valuation. Let $|\cdot|_K := (\#k)^{-\mathrm{val}_K(\cdot)}$ be the corresponding norm. The action of G_K on K preserves

val_K, and thus induces an action on k, so that we have a homomorphism $G_K \to G_k$, and in fact a short exact sequence

$$0 \to I_K \to G_K \to G_k \to 0$$

defining the inertia subgroup I_K. We let $\mathrm{Frob}_K = \mathrm{Frob}_k \in G_k$ be the geometric Frobenius element, a topological generator of $G_k \cong \hat{\mathbb{Z}}$.

Then we define the Weil group W_K via the commutative diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & I_K & \longrightarrow & G_K & \longrightarrow & G_k & \longrightarrow & 0 \\ & & \Vert & & \uparrow & & \uparrow & & \\ 0 & \longrightarrow & I_K & \longrightarrow & W_K & \longrightarrow & \mathrm{Frob}_k^{\mathbb{Z}} & \longrightarrow & 0 \end{array}$$

so that W_K is the subgroup of G_K consisting of elements which map to an integral power of the Frobenius in G_k. The group W_K is a topological group, but its topology is not the subspace topology of G_K; rather, the topology is determined by decreeing that I_K is open, and has its usual topology.

Let $K^{\mathrm{ur}} = \overline{K}^{I_K}$ be the maximal unramified extension of K, and let $K^{\mathrm{tame}} = \bigcup_{(m,l)=1} K^{\mathrm{ur}}(\varpi_K^{1/m})$ be the maximal tamely ramified extension. Then the wild inertia subgroup $P_K := \mathrm{Gal}(\overline{K}/K^{\mathrm{tame}})$ is the unique Sylow pro-l subgroup of I_K. Let $\zeta = (\zeta_m)_{(m,l)=1}$ be a compatible system of primitive roots of unity (i.e., $\zeta_{ab}^a = \zeta_b$). Then we have a character

$$t_\zeta : I_K/P_K \xrightarrow{\sim} \prod_{p \neq l} \mathbb{Z}_p,$$

defined by

$$\frac{\sigma(\varpi_K^{1/m})}{\varpi_K^{1/m}} = \zeta_m^{(t_\zeta(\sigma) \pmod{m})}.$$

Exercise 2.8. Any other compatible system of roots of unity is of the form ζ^u for some $u \in \prod_{p \neq l} \mathbb{Z}_p^\times$, and we have $t_{\zeta^u} = u^{-1} t_\zeta$.

If $\sigma \in W_K$, then $t_\zeta(\sigma\tau\sigma^{-1}) = \varepsilon(\sigma) t_\zeta(\tau)$, where ε is the cyclotomic character. We let $t_{\zeta,p}$ be the composite of t_ζ and the projection to $\mathbb{Z}_p$.

Local class field theory is summarized in the following statement. (See, for example, [Tate 1979] for this and the other facts about class field theory recalled below.)

Theorem 2.9. *Let W_K^{ab} denote the group $W_K/\overline{[W_K, W_K]}$. Then there are unique isomorphisms $\mathrm{Art}_K : K^\times \xrightarrow{\sim} W_K^{\mathrm{ab}}$ such that*

(1) *if K'/K is a finite extension, then* $\mathrm{Art}_{K'} = \mathrm{Art}_K \circ N_{K'/K}$, *and*

(2) *we have a commutative square*

$$\begin{array}{ccc} K^\times & \xrightarrow{\mathrm{Art}_K} & W_K^{\mathrm{ab}} \\ \Big\downarrow{\scriptstyle \mathrm{val}_K} & & \Big\downarrow \\ \mathbb{Z} & \longrightarrow & \mathrm{Frob}_K^{\mathbb{Z}} \end{array}$$

where the bottom arrow is the isomorphism sending $a \mapsto \mathrm{Frob}_K^a$.

The continuous irreducible representations of the group W_K^{ab} are just the continuous characters of W_K, and local class field theory gives a simple description of them, as representations of $K^\times = \mathrm{GL}_1(K)$. The local Langlands correspondence for GL_n (see Section 4.1) is a kind of n-dimensional generalization of this, giving a description of certain representations of $\mathrm{GL}_n(K)$ in terms of the n-dimensional representations of W_K.

Definition 2.10. Let L be a field of characteristic 0. A *representation* of W_K over L is a representation (on a finite-dimensional L-vector space) which is continuous if L has the discrete topology (i.e., a representation with open kernel).

A *Weil–Deligne* representation of W_K on a finite-dimensional L-vector space V is a pair (r, N) consisting of a representation $r : W_K \to \mathrm{GL}(V)$, and an endomorphism $N \in \mathrm{End}(V)$ such that for all $\sigma \in W_K$,

$$r(\sigma)Nr(\sigma)^{-1} = (\#k)^{-v_K(\sigma)}N,$$

where $v_K : W_K \to \mathbb{Z}$ is determined by $\sigma|_{K^{\mathrm{ur}}} = \mathrm{Frob}_K^{v_K(\sigma)}$.

Remark 2.11. (1) Since I_K is compact and open in W_K, if r is a representation of W_K then $r(I_K)$ is finite.

(2) N is necessarily nilpotent.

Exercise 2.12. (1) Show that if (r, V) is a representation of W_K and $m \geq 1$ then the following defines a Weil–Deligne representation $\mathrm{Sp}_m(r)$ with underlying vector space V^m: we let W_K act via

$$r|\mathrm{Art}_K^{-1}|_K^{m-1} \oplus r|\mathrm{Art}_K^{-1}|_K^{m-2} \oplus \cdots \oplus r,$$

and let N induce an isomorphism from $r|\mathrm{Art}_K^{-1}|_K^{i-1}$ to $r|\mathrm{Art}_K^{-1}|_K^{i}$ for each $i < m-1$, and be 0 on $r|\mathrm{Art}_K^{-1}|_K^{m-1}$.

(2) Show that every Weil–Deligne representation (r, V) for which r is semisimple is isomorphic to a direct sum of representations $\mathrm{Sp}_{m_i}(r_i)$.

(3) Show that if (r, V, N) is a Weil–Deligne representation of W_K, and K'/K is a finite extension, then $(r|_{W_{K'}}, V, N)$ is a Weil–Deligne representation of $W_{K'}$.

(4) Suppose that r is a representation of W_K. Show that if $\sigma \in W_K$ then for some positive integer n, $r(\sigma^n)$ is in the center of $r(W_K)$.

(5) Assume further that $\sigma \notin I_K$. Show that for any $\tau \in W_K$ there exists $n \in \mathbb{Z}$ and $m > 0$ such that $r(\sigma^n) = r(\tau^m)$.

(6) Show that for a representation r of W_K, the following conditions are equivalent:
   (a) r is semisimple.
   (b) $r(\sigma)$ is semisimple for all $\sigma \in W_K$.
   (c) $r(\sigma)$ is semisimple for some $\sigma \notin I_K$.

(7) Let (r, N) be a Weil–Deligne representation of W_K. Set $\tilde{r}(\sigma) = r(\sigma)^{\mathrm{ss}}$, the semisimplification of $r(\sigma)$. Prove that $(\tilde{r}, N)$ is also a Weil–Deligne representation of W_K.

Definition 2.13. We say that a Weil–Deligne representation (r, N) is *Frobenius semisimple* if r is semisimple. With notation as in Exercise 2.12(7), we say that $(\tilde{r}, N)$ is the *Frobenius semisimplification* of (r, N).

Definition 2.14. If L is an algebraic extension of $\mathbb{Q}_p$, then we say that an element $A \in \mathrm{GL}_n(L)$ is *bounded* if it has determinant in $\mathcal{O}_L^\times$, and characteristic polynomial in $\mathcal{O}_L[X]$.

Exercise 2.15. A is bounded if and only if it stabilizes an $\mathcal{O}_L$-lattice in L^n.

Definition 2.16. Let L be an algebraic extension of $\mathbb{Q}_p$. Then we say that r is *bounded* if $r(\sigma)$ is bounded for all $\sigma \in W_K$.

Exercise 2.17. Show r is bounded if and only if $r(\sigma)$ is bounded for some $\sigma \notin I_K$.

The reason for all of these definitions is the following theorem, which in practice gives us a rather concrete classification of the p-adic representations of G_K.

Proposition 2.18 (Grothendieck's monodromy theorem). *Suppose that $l \neq p$, that $K/\mathbb{Q}_l$ is finite, and that V is a finite-dimensional L-vector space, with L an algebraic extension of $\mathbb{Q}_p$. Fix $\varphi \in W_K$ a lift of* Frob_K *and a compatible system $(\zeta_m)_{(m,l)=1}$ of primitive roots of unity. If $\rho : G_K \to \mathrm{GL}(V)$ is a continuous representation then there is a finite extension K'/K and a uniquely determined nilpotent $N \in \mathrm{End}(V)$ such that for all $\sigma \in I_{K'}$,*

$$\rho(\sigma) = \exp(N t_{\zeta,p}(\sigma)).$$

For all $\sigma \in W_K$, we have $\rho(\sigma) N \rho(\sigma)^{-1} = \#k^{-v_K(\sigma)} N$. In fact, we have an equivalence of categories $\mathrm{WD} = \mathrm{WD}_{\zeta,\varphi}$ *from the category of continuous representations of G_K on finite-dimensional L-vector spaces to the category of bounded Weil–Deligne representations on finite-dimensional L-vector spaces, taking*

$$\rho \mapsto (V, r, N), \quad r(\tau) := \rho(\tau) \exp(-t_{\zeta,p}(\varphi^{-v_K(\tau)}\tau) N).$$

The functors $\mathrm{WD}_{\zeta',\varphi'}$ *and* $\mathrm{WD}_{\zeta,\varphi}$ *are naturally isomorphic.*

Remark 2.19. Note that since N is nilpotent, the exponential here is just a polynomial — there are no convergence issues!

The proof is contained in the following exercise.

Exercise 2.20. (1) By Exercise 2.4 there is a G_K-stable $\mathcal{O}_L$-lattice $\Lambda \subset V$. Show that if $G_{K'}$ is the kernel of the induced map $G_K \to \mathrm{Aut}(\Lambda/p\Lambda)$, then K'/K is a finite extension, and $\rho(G_{K'})$ is pro-p. Show that $\rho|_{I_{K'}}$ factors through $t_{\zeta,p} : I_{K'} \to \mathbb{Z}_p$.

(2) Choose $\sigma \in I_{K'}$ such that $t_{\zeta,p}(\sigma)$ topologically generates $t_{\zeta,p}(I_{K'})$. By considering the action of conjugation by φ, show that the eigenvalues of $\rho(\sigma)$ are all p-power roots of unity. Hence show that one may make a further finite extension K''/K' such that the elements of $\rho(I_{K''})$ are all unipotent.

(3) Deduce the existence of a unique nilpotent $N \in \mathrm{End}(V)$ such that for all $\sigma \in I_{K''}$, $\rho(\sigma) = \exp(N t_{\zeta,p}(\sigma))$. [Hint: use the logarithm map (why are there no convergence issues?).]

(4) Complete the proof of the proposition, by showing that (r, N) is a Weil–Deligne representation. Where does the condition that r is bounded come in?

One significant advantage of Weil–Deligne representations over Galois representations is that there are no subtle topological issues: the topology on the Weil–Deligne representation is the discrete topology. This allows one to describe representations in a way that is "independent of L", and is necessary to make sense of the notion of a compatible system of Galois representations (or at least to make sense of it at places at which the Galois representation is ramified); see Definition 2.32 below.

2.21. ***Local representations with p = l: p-adic Hodge theory.*** The case $l = p$ is far more complicated than the case $l \neq p$, largely because wild inertia can act in a highly nontrivial fashion, so there is no simple analogue of Grothendieck's monodromy theorem. (There is still an analogue, though, it's just much harder to state and prove, and doesn't apply to all p-adic Galois representations.) The study of representations $G_K \to \mathrm{GL}_n(\overline{\mathbb{Q}}_p)$ with $K/\mathbb{Q}_p$ finite is part of what is called *p-adic Hodge theory*, a subject initially developed by Fontaine in the 1980s. For an introduction to the part of p-adic Hodge theory concerned with Galois representations, the reader could consult [Berger 2004]. There is a lot more to p-adic Hodge theory than the study of Galois representations, and an excellent overview of some recent developments in the more geometric part of the theory can be found in [Bhatt 2021]. We will content ourselves with some terminology, some definitions, and some remarks intended to give intuition and motivation.

Fix $K/\mathbb{Q}_p$ finite. In some sense, "most" p-adic Galois representations $G_K \to \mathrm{GL}_n(\overline{\mathbb{Q}}_p)$ will not be relevant for us, because they do not arise in geometry, or in the Galois representations associated to automorphic representations. Instead, there is a hierarchy of classes of representations

$$\{\text{crystalline}\} \subsetneq \{\text{semistable}\} \subsetneq \{\text{de Rham}\} \subsetneq \{\text{Hodge–Tate}\}.$$

For any of these classes X, we say that ρ is *potentially* X if there is a finite extension K'/K such that $\rho|_{G_{K'}}$ is X. A representation is potentially de Rham if and only if it is de Rham, and potentially Hodge–Tate if and only if it is Hodge–Tate; the corresponding statements for crystalline and semistable representations are false, as we will see concretely in the case $n = 1$ later on. The p-adic analogue of Grothendieck's monodromy theorem is the following deep theorem of Berger.

Theorem 2.22 (the p-adic monodromy theorem). *A representation is de Rham if and only if it is potentially semistable.*

The notion of a de Rham representation is designed to capture the representations arising in geometry; it does so by the following result of Tsuji (building on the work of many people).

Theorem 2.23. *If X/K is a smooth projective variety, then each $H^i_{\text{ét}}(X \times_K \overline{K}, \overline{\mathbb{Q}}_p)$ is a de Rham representation.*

Similarly, the definitions of crystalline and semistable are designed to capture the notions of good and semistable reduction, and one has (again as a consequence of Tsuji's work); see Section 2.5 of [Berger 2004].

Theorem 2.24. *If X/K is a smooth projective variety with good (respectively, semistable) reduction, then each $H^i_{\text{ét}}(X \times_K \overline{K}, \overline{\mathbb{Q}}_p)$ is a crystalline (respectively, semistable) representation.*

Thus the p-adic monodromy theorem can be thought of as a Galois-theoretic incarnation of Grothendieck's semistable reduction theorem.

The case that $n = 1$ is particularly simple, as we now explain. In this case, every semistable character is crystalline, and the de Rham characters are exactly the Hodge–Tate characters. In the case $K = \mathbb{Q}_p$, these are precisely the characters whose restrictions to inertia are of the form $\psi\varepsilon_p^m$ where ψ has finite order and $m \in \mathbb{Z}$, while the crystalline characters are those for which ψ is trivial. A similar description exists for general K, with ε_p^m replaced by a product of so-called *Lubin–Tate characters*.

Fact 2.25. A character $\chi : G_K \to \overline{\mathbb{Q}}_p^\times$ is de Rham if and only if there is an open subgroup U of $K^\times$ and an integer n_τ for each $\tau : K \hookrightarrow \overline{\mathbb{Q}}_p$ such that $(\chi \circ \mathrm{Art}_K)(\alpha) = \prod_\tau \tau(\alpha)^{-n_\tau}$ for each $\alpha \in U$, and it is crystalline if and only if we can take $U = \mathcal{O}_K^\times$. See Exercise 6.4.3 of [Brinon and Conrad 2009].

As soon as $n > 1$, there are noncrystalline semistable representations, and non-de Rham Hodge–Tate representations. A useful heuristic when comparing to the $l \neq$

p case is that crystalline representations correspond to unramified representations, semistable representations correspond to representations for which inertia acts unipotently, and de Rham representations correspond to all representations.

Suppose that $\rho : G_K \to \mathrm{GL}_n(\overline{\mathbb{Q}}_p)$ is a Hodge–Tate representation. Then for each $\tau : K \hookrightarrow \overline{\mathbb{Q}}_p$ there is a multiset of *τ-labeled Hodge–Tate weights* (defined for example in the notation section of [Barnet-Lamb et al. 2014], where they are called "Hodge–Tate numbers") $\mathrm{HT}_\tau(\rho)$ associated to ρ; this is a multiset of integers, and in the case of a de Rham character χ as above, $\mathrm{HT}_\tau(\chi) = n_\tau$. In particular, the p-adic cyclotomic character ε_p has all Hodge–Tate weights equal to -1. If K'/K is a finite extension, and $\tau' : K' \hookrightarrow \overline{\mathbb{Q}}_p$ extends $\tau : K \hookrightarrow \overline{\mathbb{Q}}_p$, then

$$\mathrm{HT}_{\tau'}(\rho|_{G_{K'}}) = \mathrm{HT}_\tau(\rho).$$

If furthermore ρ is potentially semistable (equivalently, de Rham) then a construction of Fontaine associates a Weil–Deligne representation $\mathrm{WD}(\rho) = (r, N)$ of W_K to ρ. If K'/K is a finite extension, then $\mathrm{WD}(\rho|_{G_{K'}}) = (r|_{W_{K'}}, N)$. It is known that ρ is semistable if and only if r is unramified, and that ρ is crystalline if and only if r is unramified and $N = 0$. Thus ρ is potentially crystalline if and only $N = 0$.

2.26. ***Number fields.*** We now consider the case that K is a number field (that is, a finite extension of $\mathbb{Q}$). If v is a finite place of K, we let K_v denote the completion of K at v. If K'/K is a finite Galois extension, then $\mathrm{Gal}(K'/K)$ transitively permutes the places of K' above v; if we choose one such place w, then we define the *decomposition group*

$$\mathrm{Gal}(K'/K)_w := \{\sigma \in \mathrm{Gal}(K'/K) \mid w\sigma = w\}.$$

Then we have a natural isomorphism $\mathrm{Gal}(K'/K)_w \xrightarrow{\sim} \mathrm{Gal}(K'_w/K_v)$, and since $\mathrm{Gal}(K'/K)_{w\sigma} = \sigma^{-1}\,\mathrm{Gal}(K'/K)_w\sigma$, we see that the definition extends to general algebraic extensions, and in particular we have an embedding $G_{K_v} \hookrightarrow G_K$ which is well-defined up to conjugacy (alternatively, up to a choice of embedding $\overline{K} \hookrightarrow \overline{K}_v$). (Note that you need to be slightly careful with taking completions in the case that K'/K is infinite, as then the extension K'_w/K_v need not be algebraic; we can for example define $\mathrm{Gal}(K'_w/K_v)$ to be the group of continuous automorphisms of K'_w which fix K_v pointwise.)

If K'/K is Galois and unramified at v, and w is a place of K' lying over v, then we define

$$\mathrm{Frob}_w := \mathrm{Frob}_{K_v} \in \mathrm{Gal}(K'_w/K_v) \xrightarrow{\sim} \mathrm{Gal}(K'/K)_w \hookrightarrow \mathrm{Gal}(K'/K).$$

We have $\mathrm{Frob}_{w\sigma} = \sigma^{-1}\,\mathrm{Frob}_w\,\sigma$, and thus a well-defined conjugacy class $[\mathrm{Frob}_v] = \{\mathrm{Frob}_w\}_{w\,|\,v}$ in $\mathrm{Gal}(K'/K)$.

Fact 2.27 (Chebotarev density theorem). If K'/K is a Galois extension which is unramified outside of a finite set S of places of K, then the union of the conjugacy classes $[\mathrm{Frob}_v]$, $v \notin S$ is dense in $\mathrm{Gal}(K'/K)$.

We briefly recall a statement of global class field theory. Let $\mathbb{A}_K$ denote the adeles of K, and write $K_\infty = \prod_{v\,|\,\infty} K_v$. Let $K^{\mathrm{ab}} = \overline{K}^{[G_K,G_K]}$ be the maximal abelian extension of K. Then there is a homomorphism $\mathrm{Art}_K : \mathbb{A}_K^\times/(K_\infty^\times)^\circ \to \mathrm{Gal}(K^{\mathrm{ab}}/K)$, defined in the following way: for each finite place v of K, the restriction of Art_K to $K_v^\times$ agrees with the local Artin maps Art_{K_v}, and similarly at the infinite places, it agrees with the obvious isomorphisms $\mathrm{Art}_{K_v} : K_v^\times/(K_v^\times)^\circ \xrightarrow{\sim} \mathrm{Gal}(\overline{K}_v/K_v)$. (In both cases, the symbol $^\circ$ refers to the connected component of the identity.) Then global class field theory states that Art_K induces an isomorphism

$$\mathrm{Art}_K : \mathbb{A}_K^\times/\overline{K^\times(K_\infty^\times)^\circ} \xrightarrow{\sim} \mathrm{Gal}(K^{\mathrm{ab}}/K).$$

The global Galois representations that we will care about are those that Fontaine and Mazur call *geometric*. Let $L/\mathbb{Q}_p$ be an algebraic extension.

Definition 2.28. If K is a number field, then a continuous representation $\rho : G_K \to \mathrm{GL}_n(L)$ is *geometric* if it is unramified outside of a finite set of places of K, and if for each place $v \mid p$, $\rho|_{G_{K_v}}$ is de Rham.

Remark 2.29. It is known that both conditions are necessary; that is, there are examples of representations which are unramified outside of a finite set of places of K but not de Rham at places lying over p, and examples of representations which are de Rham at all places lying over p, but are ramified at infinitely many places. (As we will see in Theorem 2.43, these examples require $n > 1$.)

In practice (and conjecturally always), geometric Galois representations arise as part of a *compatible system* of Galois representations. There are a number of different definitions of a compatible system in the literature, all of which are conjecturally equivalent (although proving the equivalence of the definitions is probably very hard). The following definition, taken from [Barnet-Lamb et al. 2014], is simultaneously a strong enough set of assumptions under which one can hope to employ automorphy lifting theorems to study a compatible system, and is weak enough that the conditions can be verified in interesting examples.

Definition 2.30. Suppose that K and M are number fields, that S is a finite set of places of K and that n is a positive integer. By a *weakly compatible system* of n-dimensional p-adic representations (for varying p) of G_K defined over M and unramified outside S we mean a family of continuous semisimple representations

$$r_\lambda : G_K \to \mathrm{GL}_n(\overline{M}_\lambda),$$

where λ runs over the finite places of M, with the following properties:

- If $v \notin S$ is a finite place of K, then for all λ not dividing the residue characteristic of v, the representation r_λ is unramified at v and the characteristic polynomial of $r_\lambda(\mathrm{Frob}_v)$ lies in $M[X]$ and is independent of λ.
- Each representation r_λ is de Rham at all places above the residue characteristic of λ, and in fact crystalline at any place $v \notin S$ which divides the residue characteristic of λ.
- For each embedding $\tau : K \hookrightarrow \overline{M}$ the τ-labeled Hodge–Tate weights of r_λ are independent of λ.

Remark 2.31. By the Chebotarev density theorem and the Brauer–Nesbitt theorem, each r_λ is determined by the characteristic polynomials of the $r_\lambda(\mathrm{Frob}_v)$ for $v \notin S$, and in particular the compatible system is determined by a single r_λ. Note that for a general element $\sigma \in G_K$, there will be no relationship between the characteristic polynomials of the $r_\lambda(\sigma)$ as λ varies (and they won't even lie in $M[X]$, so there will be no way of comparing them).

There are various other properties one could demand; for example, we have the following definition (again following [Barnet-Lamb et al. 2014], although we have slightly strengthened the definition made there by allowing λ to divide the residue characteristic of v).

Definition 2.32. We say that a weakly compatible system is *strictly compatible* if for each finite place v of K there is a Weil–Deligne representation WD_v of W_{K_v} over $\overline{M}$ such that for each finite place λ of M and every M-linear embedding $\varsigma : \overline{M} \hookrightarrow \overline{M}_\lambda$ we have $\varsigma\, \mathrm{WD}_v \cong \mathrm{WD}(r_\lambda|_{G_{K_v}})^{\text{F-ss}}$.

Conjecturally, every weakly compatible system is strictly compatible, and even satisfies further properties, such as purity; see, e.g., Section 5 of [Barnet-Lamb et al. 2014]. We also have the following consequence of the Fontaine–Mazur conjecture (Conjecture 2.38 below) and standard conjectures on the étale cohomology of algebraic varieties over number fields.

Conjecture 2.33. *Any semisimple geometric representation $G_K \to \mathrm{GL}_n(L)$ is part of a strictly compatible system of Galois representations.*

In practice, most progress on understanding these conjectures has been made by using automorphy lifting theorems to prove special cases of the following conjecture.

Conjecture 2.34. *Any weakly compatible system of Galois representations is strictly compatible, and is in addition automorphic, in the sense that there is an algebraic automorphic representation (in the sense of* [Clozel 1990]*) π of $\mathrm{GL}_n(\mathbb{A}_K)$ with the property that $\mathrm{WD}_v \cong \mathrm{rec}_{K_v}(\pi_v|\det|^{(1-n)/2})$ for each finite place v of K, where rec_{K_v} is the local Langlands correspondence as in Section 4.1 below.*

2.35. ***Sources of Galois representations.*** The main source (and conjecturally the only source) of compatible systems of Galois representations is the étale cohomology of algebraic varieties. We have the following result, a consequence of Theorem 2.23 and "independence of l" results in étale cohomology [Katz and Messing 1974].

Theorem 2.36. *Let K be a number field, and let X/K be a smooth projective variety. Then for any i, j, the $H^i_{\text{ét}}(X \times_K \overline{K}, \mathbb{Q}_p)^{\text{ss}}(j)$ (the (j) denoting a Tate twist) form a weakly compatible system (defined over $\mathbb{Q}$) as p varies.*

Remark 2.37. Conjecturally, it is a strictly compatible system, and there is no need to semisimplify the representations. Both of these properties are known if X is an abelian variety; see Section 2.4 of [Fontaine 1994].

Conjecture 2.38 (the Fontaine–Mazur conjecture [Fontaine and Mazur 1995]). *Any irreducible geometric representation $\rho : G_K \to \mathrm{GL}_n(\overline{\mathbb{Q}}_p)$ is (the extension of scalars to $\overline{\mathbb{Q}}_p$ of) a subquotient of a representation arising from étale cohomology as in Theorem 2.36.*

Remark 2.39. The *Fontaine–Mazur–Langlands conjecture* is a somewhat ill-defined conjecture, which is essentially the union of Conjectures 2.33 and 2.34, expressing the expectation that an irreducible geometric Galois representation is automorphic.

When $n = 1$, all of these conjectures are essentially known, as we will now explain. For $n > 1$, we know very little (although the situation when $K = \mathbb{Q}$ and $n = 2$ is pretty good), and the main results that are known are as a consequence of automorphy lifting theorems (as discussed in these notes) and of potential automorphy theorems (which are not discussed in these notes, but should be accessible given the material we develop here; for a nice introduction, see [Buzzard 2012]).

Definition 2.40. A *Grössencharacter* is a continuous character $\chi : \mathbb{A}_K^\times / K^\times \to \mathbb{C}^\times$. We say that χ is *algebraic* (or "type A_0") if for each $\tau : K \hookrightarrow \mathbb{C}$ there is an integer n_τ, such that for each $\alpha \in (K_\infty^\times)^\circ$, we have $\chi(\alpha) = \prod_\tau (\tau(\alpha))^{-n_\tau}$.

Definition 2.41. Let L be a field of characteristic zero such that for each embedding $\tau : K \hookrightarrow \bar{L}$, we have $\tau(K) \subseteq L$. Then an *algebraic character* $\chi_0 : \mathbb{A}_K^\times \to \bar{L}^\times$ is a character with open kernel such that for each $\tau : K \hookrightarrow L$ there is an integer n_τ with the property that for all $\alpha \in K^\times$, we have $\chi_0(\alpha) = \prod_\tau (\tau(\alpha))^{n_\tau}$.

Exercise 2.42. Show that if χ_0 is an algebraic character, then χ_0 takes values in some number field. [Hint: show that $\mathbb{A}_K^\times / (K^\times \ker \chi_0)$ is finite, and that $\chi_0(K^\times \ker \chi_0)$ is contained in a number field.]

Theorem 2.43. *Let E be a number field containing the normal closure of K. Fix embeddings $\imath_\infty : \overline{E} \hookrightarrow \mathbb{C}$, $\imath_p : \overline{E} \hookrightarrow \overline{\mathbb{Q}}_p$. Then the following are in natural bijection:*

(1) *Algebraic characters $\chi_0 : \mathbb{A}_K^\times \to \overline{E}^\times$.*

(2) *Algebraic Grössencharacters $\chi : \mathbb{A}_K^\times / K^\times \to \mathbb{C}^\times$.*

(3) *Continuous representations* $\rho : G_K \to \overline{\mathbb{Q}}_p^\times$ *which are de Rham at all* $v \mid p$.

(4) *Geometric representations* $\rho : G_K \to \overline{\mathbb{Q}}_p^\times$.

Exercise 2.44. Prove Theorem 2.43 as follows; see, e.g., Section 1 of [Fargues 2011] for more details. Firstly, use Fact 2.25, together with global class field theory, to show that (3) and (4) are equivalent. For the correspondence between (1) and (2), show that we can pair up χ_0 and χ by

$$\chi(\alpha) = \iota_\infty\Bigg(\chi_0(\alpha) \prod_{\tau : K \hookrightarrow \mathbb{C}} \tau(\alpha_\infty)^{-n_{\iota_\infty^{-1}\tau}}\Bigg).$$

For the correspondence between (1) and (3), show that we can pair up χ_0 and ρ by

$$(\rho \circ \mathrm{Art}_K)(\alpha) = \iota_p\Bigg(\chi_0(\alpha) \prod_{\tau : K \hookrightarrow \overline{\mathbb{Q}}_p} \tau(\alpha_p)^{-n_{\iota_p^{-1}\tau}}\Bigg).$$

3. Galois deformations

The "lifting" in "modularity lifting theorems" refers to deducing the modularity of a p-adic Galois representation from the modularity of its reduction modulo p; so we "lift" the modularity property from characteristic p to characteristic zero. In this section we consider the Galois-theoretic aspects of this lifting, which are usually known as "Galois deformation theory".

There are a number of good introductions to the material in this section, and for the most part we will simply give basic definitions and motivation, and refer elsewhere for proofs. In particular, [Mazur 1997] is a very nice introduction to Galois deformations (although slightly out of date, as it does not treat liftings/framed deformations), and [Böckle 2013] is a thorough modern treatment.

3.1. ***Generalities.*** Take $L/\mathbb{Q}_p$ finite with ring of integers $\mathcal{O} = \mathcal{O}_L$ and maximal ideal λ, and write $\mathbb{F} = \mathcal{O}/\lambda$. Let G be a profinite group which satisfies the following condition (Mazur's condition Φ_p): for each open subgroup Δ of G, then $\Delta/\langle[\Delta, \Delta], \Delta^p\rangle$ is finite. Equivalently (see, e.g., Exercise 1.8.1 of [Böckle 2013]), for each Δ the maximal pro-p quotient of Δ is topologically finitely generated. If G is topologically finitely generated, then Φ_p holds, but we will need to use the condition for some G (the global Galois groups $G_{K,S}$ defined below) which are not known to be topologically finitely generated.

In particular, using class field theory or Kummer theory, it can be checked that Φ_p holds if $G = G_K = \mathrm{Gal}(\overline{K}/K)$ for some prime l (possibly equal to p) and some finite extension $K/\mathbb{Q}_l$, or if $G = G_{K,S} = \mathrm{Gal}(K_S/K)$ where K is a number field, S is a finite set of finite places of K, and K_S/K is the maximal extension unramified outside of S and the infinite places; see, e.g., the proof of Theorem 2.41 of [Darmon et al. 1997].

Fix a continuous representation $\bar{\rho} : G \to \mathrm{GL}_n(\mathbb{F})$. Let $\mathcal{C}_{\mathcal{O}}$ be the category of complete local Noetherian $\mathcal{O}$-algebras with residue field $\mathbb{F}$, and consider the functor $\mathcal{C}_{\mathcal{O}} \to \underline{Sets}$ which sends A to the set of continuous representations $\rho : G \to \mathrm{GL}_n(A)$ such that $\rho \bmod \mathfrak{m}_A = \bar{\rho}$ (that is, to the set of *lifts* of $\bar{\rho}$ to A).

Lemma 3.2. *This functor is represented by a representation* $\rho^{\square} : G \to \mathrm{GL}_n(R^{\square}_{\bar{\rho}})$.

Proof. This is straightforward; see Proposition 1.3.1(a) of [Böckle 2013] for a closely related result (showing the prorepresentability of the functor restricted to Artinian algebras), or to [Dickinson 2001a] for a complete proof of a more general result. □

Definition 3.3. We say that $R^{\square}_{\bar{\rho}}$ is the *universal lifting ring* (or in Kisin's terminology, the *universal framed deformation ring*). We say that $\rho^{\square}$ is the *universal lifting* of $\bar{\rho}$.

If $\mathrm{End}_{\mathbb{F}[G]}\, \bar{\rho} = \mathbb{F}$ we will say that $\bar{\rho}$ is *Schur*. By Schur's lemma, if $\bar{\rho}$ is absolutely irreducible, then $\bar{\rho}$ is Schur. In this case, there is a very useful (and historically earlier) variant on the above construction.

Definition 3.4. Suppose that $\bar{\rho}$ is Schur. Then a *deformation* of $\bar{\rho}$ to $A \in \mathrm{ob}\, \mathcal{C}_{\mathcal{O}}$ is an equivalence class of liftings, where $\rho \sim \rho'$ if and only if $\rho' = a\rho a^{-1}$ for some $a \in \ker(\mathrm{GL}_n(A) \to \mathrm{GL}_n(\mathbb{F}))$ (or equivalently, for some $a \in \mathrm{GL}_n(A)$).

Lemma 3.5. *If $\bar{\rho}$ is Schur, then the functor $\mathcal{C}_{\mathcal{O}} \to \underline{Sets}$ sending A to the set of deformations of $\bar{\rho}$ to A is representable by some* $\rho^{\mathrm{univ}} : G \to \mathrm{GL}_n(R^{\mathrm{univ}}_{\bar{\rho}})$.

Proof. See Proposition 1.3.1(b) of [Böckle 2013], or Theorem 2.36 of [Darmon et al. 1997] for a more hands-on approach. □

Definition 3.6. We say that ρ^{univ} (or more properly, its equivalence class) is the *universal deformation* of $\bar{\rho}$, and $R^{\mathrm{univ}}_{\bar{\rho}}$ is the *universal deformation ring*.

Deformations are representations considered up to conjugation, so it is reasonable to hope that deformations can be studied by considering their traces. In the case that $\bar{\rho}$ is absolutely irreducible, universal deformations are determined by traces in the following rather strong sense. This result is essentially due to Carayol [1994].

Lemma 3.7. *Suppose that $\bar{\rho}$ is absolutely irreducible. Let R be an object of $\mathcal{C}_{\mathcal{O}}$, and $\rho : G \to \mathrm{GL}_n(R)$ a lifting of $\bar{\rho}$:*

(1) *If $a \in \mathrm{GL}_n(R)$ and $a\rho a^{-1} = \rho$ then $a \in R^{\times}$.*
(2) *If $\rho' : G \to \mathrm{GL}_n(R)$ is another continuous lifting of $\bar{\rho}$ and $\mathrm{tr}\, \rho = \mathrm{tr}\, \rho'$, then there is some $a \in \ker(\mathrm{GL}_n(R) \to \mathrm{GL}_n(\mathbb{F}))$ such that $\rho' = a\rho a^{-1}$.*
(3) *If $S \subseteq R$ is a closed subring with $S \in \mathrm{ob}\, \mathcal{C}_{\mathcal{O}}$ and $\mathfrak{m}_S = \mathfrak{m}_R \cap S$, and if $\mathrm{tr}\, \rho(G) \subseteq S$, then there is some $a \in \ker(\mathrm{GL}_n(R) \to \mathrm{GL}_n(\mathbb{F}))$ such that $a\rho a^{-1} : G \to \mathrm{GL}_n(S)$.*

Proof. See Lemmas 2.1.8 and 2.1.10 of [Clozel et al. 2008], or Theorem 2.2.1 of [Böckle 2013]. □

Exercise 3.8. Deduce from Lemma 3.7 that if $\bar{\rho}$ is absolutely irreducible, then $R_{\bar{\rho}}^{\mathrm{univ}}$ is topologically generated over $\mathcal{O}$ by the values $\operatorname{tr}\rho^{\mathrm{univ}}(g)$ as g runs over any dense subset of G.

Exercise 3.9. Show that if $\bar{\rho}$ is absolutely irreducible, then $R_{\bar{\rho}}^{\square}$ is isomorphic to a power series ring in (n^2-1) variables over $R_{\bar{\rho}}^{\mathrm{univ}}$. Hint: let ρ^{univ} be a choice of universal deformation, and consider the homomorphism

$$\rho^{\square}: G \to \mathrm{GL}_n(R_{\bar{\rho}}^{\mathrm{univ}}[[X_{i,j}]]_{i,j=1,\ldots,n}/(X_{1,1}))$$

given by $\rho^{\square} = (1_n + (X_{i,j}))\rho^{\mathrm{univ}}(1_n + (X_{i,j}))^{-1}$. Show that this is the universal lifting.

3.10. *Tangent spaces.* The tangent spaces of universal lifting and deformation rings have a natural interpretation in terms of liftings and deformations to the ring of dual numbers, $\mathbb{F}[\varepsilon]/(\varepsilon^2)$.

Exercise 3.11. Show that we have natural bijections between:

(1) $\mathrm{Hom}_{\mathbb{F}}(\mathfrak{m}_{R_{\bar{\rho}}^{\square}}/(\mathfrak{m}_{R_{\bar{\rho}}^{\square}}^2, \lambda), \mathbb{F})$.

(2) $\mathrm{Hom}_{\mathcal{O}}(R_{\bar{\rho}}^{\square}, \mathbb{F}[\varepsilon]/(\varepsilon^2))$.

(3) The set of liftings of $\bar{\rho}$ to $\mathbb{F}[\varepsilon]/(\varepsilon^2)$.

(4) The set of cocycles $Z^1(G, \operatorname{ad}\bar{\rho})$.

Show that if $\bar{\rho}$ is absolutely irreducible, then we also have a bijection between $\mathrm{Hom}_{\mathbb{F}}(\mathfrak{m}_{R_{\bar{\rho}}^{\mathrm{univ}}}/(\mathfrak{m}_{R_{\bar{\rho}}^{\mathrm{univ}}}^2, \lambda), \mathbb{F})$ and $H^1(G, \operatorname{ad}\bar{\rho})$.

Hint: given $f \in \mathrm{Hom}_{\mathbb{F}}(\mathfrak{m}_{R_{\bar{\rho}}^{\square}}/(\mathfrak{m}_{R_{\bar{\rho}}^{\square}}^2, \lambda), \mathbb{F})$, define an element of $\mathrm{Hom}_{\mathcal{O}}(R_{\bar{\rho}}^{\square}, \mathbb{F}[\varepsilon]/(\varepsilon^2))$ by sending $a + x$ to $a + f(x)\varepsilon$ whenever $a \in \mathcal{O}$ and $x \in \mathfrak{m}_{R_{\bar{\rho}}^{\square}}$. Given a cocycle $\phi \in Z^1(G, \operatorname{ad}\bar{\rho})$, define a lifting $\rho: G \to \mathrm{GL}_n(\mathbb{F}[\varepsilon]/(\varepsilon^2))$ by $\rho(g) := (1 + \phi(g)\varepsilon)\bar{\rho}(g)$.

Corollary 3.12. *We have*

$$\dim_{\mathbb{F}} \mathfrak{m}_{R_{\bar{\rho}}^{\square}}/(\mathfrak{m}_{R_{\bar{\rho}}^{\square}}^2, \lambda) = \dim_{\mathbb{F}} H^1(G, \operatorname{ad}\bar{\rho}) + n^2 - \dim_{\mathbb{F}} H^0(G, \operatorname{ad}\bar{\rho}).$$

Proof. This follows from the exact sequence

$$0 \to (\operatorname{ad}\bar{\rho})^G \to \operatorname{ad}\bar{\rho} \to Z^1(G, \operatorname{ad}\bar{\rho}) \to H^1(G, \operatorname{ad}\bar{\rho}) \to 0. \qquad \square$$

In particular, if $d := \dim_{\mathbb{F}} Z^1(G, \operatorname{ad}\bar{\rho})$, then we can choose a surjection $\phi: \mathcal{O}[[x_1, \ldots, x_d]] \twoheadrightarrow R_{\bar{\rho}}^{\square}$. Similarly, if $\bar{\rho}$ is absolutely irreducible, we can choose a surjection $\phi': \mathcal{O}[[x_1, \ldots, x_{d'}]] \twoheadrightarrow R_{\bar{\rho}}^{\mathrm{univ}}$, where $d' := \dim_{\mathbb{F}} H^1(G, \operatorname{ad}\bar{\rho})$.

Lemma 3.13. *If $J = \ker\phi$ or $J = \ker\phi'$, then there is an injection*

$$\mathrm{Hom}_{\mathbb{F}}(J/\mathfrak{m}J, \mathbb{F}) \hookrightarrow H^2(G, \operatorname{ad}\bar{\rho}),$$

where $\mathfrak{m}$ denotes the maximal ideal of $\mathcal{O}[[x_1, \ldots, x_d]]$ or $\mathcal{O}[[x_1, \ldots, x_{d'}]]$ respectively.

Proof. See the proof of Proposition 2 of [Mazur 1989]. □

Corollary 3.14. *If* $H^2(G, \mathrm{ad}\,\bar\rho) = 0$, *then* $R^{\square}_{\bar\rho}$ *is formally smooth of relative dimension* $\dim_{\mathbb{F}} Z^1(G, \mathrm{ad}\,\bar\rho)$ *over* $\mathcal{O}$.

In any case, the Krull dimension of $R^{\square}_{\bar\rho}$ *is at least*

$$1 + n^2 - \dim_{\mathbb{F}} H^0(G, \mathrm{ad}\,\bar\rho) + \dim_{\mathbb{F}} H^1(G, \mathrm{ad}\,\bar\rho) - \dim_{\mathbb{F}} H^2(G, \mathrm{ad}\,\bar\rho).$$

If $\bar\rho$ *is absolutely irreducible, then the Krull dimension of* $R^{\mathrm{univ}}_{\bar\rho}$ *is at least*

$$1 + \dim_{\mathbb{F}} H^1(G, \mathrm{ad}\,\bar\rho) - \dim_{\mathbb{F}} H^2(G, \mathrm{ad}\,\bar\rho).$$

3.15. ***Deformation conditions.*** In practice, we frequently want to impose additional conditions on the liftings and deformations we consider. For example, if we are trying to prove the Fontaine–Mazur conjecture, we would like to be able to restrict to global deformations which are geometric. There are various ways in which to impose extra conditions; we will use the formalism of *deformation problems* introduced in [Clozel et al. 2008].

Definition 3.16. By a *deformation problem* $\mathcal{D}$ we mean a collection of liftings (R, ρ) of $(\mathbb{F}, \bar\rho)$ (with R an object of $\mathcal{C}_{\mathcal{O}}$), satisfying the following properties:

- $(\mathbb{F}, \bar\rho) \in \mathcal{D}$.
- If $f : R \to S$ is a morphism in $\mathcal{C}_{\mathcal{O}}$ and $(R, \rho) \in \mathcal{D}$, then $(S, f \circ \rho) \in \mathcal{D}$.
- If $f : R \hookrightarrow S$ is an injective morphism in $\mathcal{C}_{\mathcal{O}}$ then $(R, \rho) \in \mathcal{D}$ if and only if $(S, f \circ \rho) \in \mathcal{D}$.
- Suppose that $R_1, R_2 \in \mathrm{ob}\,\mathcal{C}_{\mathcal{O}}$ and I_1, I_2 are closed ideals of R_1, R_2 respectively such that there is an isomorphism $f : R_1/I_1 \xrightarrow{\sim} R_2/I_2$. Suppose also that $(R_1, \rho_1), (R_2, \rho_2) \in \mathcal{D}$, and that $f(\rho_1 \bmod I_1) = \rho_2 \bmod I_2$. Then $(\{(a, b) \in R_1 \oplus R_2 : f(a \bmod I_1) = b \bmod I_2\}, \rho_1 \oplus \rho_2) \in \mathcal{D}$.
- If (R, ρ) is a lifting of $(\mathbb{F}, \bar\rho)$ and $I_1 \supset I_2 \supset \cdots$ is a sequence of ideals of R with $\cap_j I_j = 0$, and $(R/I_j, \rho \bmod I_j) \in \mathcal{D}$ for all j, then $(R, \rho) \in \mathcal{D}$.
- If $(R, \rho) \in \mathcal{D}$ and $a \in \ker(\mathrm{GL}_n(R) \to \mathrm{GL}_n(\mathbb{F}))$, then $(R, a\rho a^{-1}) \in \mathcal{D}$.

In practice, when we want to impose a condition on our deformations, it will be easy to see that it satisfies these requirements. (An exception is that these properties are hard to check for certain conditions arising in p-adic Hodge theory, but we won't need those conditions in these notes.)

The relationship of this definition to the universal lifting ring is as follows. Note that each element $a \in \ker(\mathrm{GL}_n(R^{\square}_{\bar\rho}) \to \mathrm{GL}_n(\mathbb{F}))$ acts on $R^{\square}_{\bar\rho}$, via the universal property and by sending $\rho^{\square}$ to $a^{-1}\rho^{\square}a$. [Warning: this *isn't* a group action, though!]

Lemma 3.17. *(1) If $\mathcal{D}$ is a deformation problem then there is a* $\ker(\mathrm{GL}_n(R^\square_{\bar\rho}) \to \mathrm{GL}_n(\mathbb{F}))$*-invariant ideal $I(\mathcal{D})$ of $R^\square_{\bar\rho}$ such that $(R, \rho) \in \mathcal{D}$ if and only if the map $R^\square_{\bar\rho} \to R$ induced by ρ factors through the quotient $R^\square_{\bar\rho}/I(\mathcal{D})$.*

(2) *Let $\tilde{L}(\mathcal{D}) \subseteq Z^1(G, \operatorname{ad}\bar\rho) \cong \operatorname{Hom}(\mathfrak{m}_{R^\square_{\bar\rho}}/(\lambda, \mathfrak{m}^2_{R^\square_{\bar\rho}}), \mathbb{F})$ denote the annihilator of the image of $I(\mathcal{D})$ in $\mathfrak{m}_{R^\square_{\bar\rho}}/(\lambda, \mathfrak{m}^2_{R^\square_{\bar\rho}})$. Then $\tilde{L}(\mathcal{D})$ is the preimage of some subspace $L(\mathcal{D}) \subseteq H^1(G, \operatorname{ad}\bar\rho)$.*

(3) *If I is a* $\ker(\mathrm{GL}_n(R^\square_{\bar\rho}) \to \mathrm{GL}_n(\mathbb{F}))$*-invariant ideal of $R^\square_{\bar\rho}$ with $\sqrt{I} = I$ and $I \neq \mathfrak{m}_{R^\square_{\bar\rho}}$, then*

$$\mathcal{D}(I) := \{(R, \rho) : R^\square_{\bar\rho} \to R \text{ factors through } R^\square_{\bar\rho}/I\}$$

is a deformation problem. Furthermore, we have $I(\mathcal{D}(I)) = I$ and $\mathcal{D}(I(\mathcal{D})) = \mathcal{D}$.

Proof. See Lemma 2.2.3 of [Clozel et al. 2008] and Lemma 3.2 of [Barnet-Lamb et al. 2011] (and for (2), use that $I(\mathcal{D})$ is $\ker(\mathrm{GL}_n(R^\square_{\bar\rho}) \to \mathrm{GL}_n(\mathbb{F}))$-invariant). □

3.18. ***Fixing determinants.*** For technical reasons, we will want to fix the determinants of our Galois representations; see Remark 5.12 of [Calegari and Geraghty 2018]. To this end, let $\chi : G \to \mathcal{O}^\times$ be a continuous homomorphism such that $\chi \bmod \lambda = \det\bar\rho$. Then it makes sense to ask that a lifting has determinant χ, and we can define a universal lifting ring $R^\square_{\bar\rho,\chi}$ for lifts with determinant χ, and when $\bar\rho$ is Schur, a universal fixed determinant deformation ring $R^{\mathrm{univ}}_{\bar\rho,\chi}$.

Exercise 3.19. Check that the material developed in the previous section goes over unchanged, except that $\operatorname{ad}\bar\rho$ needs to be replaced with $\operatorname{ad}^0\bar\rho := \{x \in \operatorname{ad}\bar\rho : \operatorname{tr} x = 0\}$.

Note that since we are assuming throughout that $p \nmid n$, $\operatorname{ad}^0\bar\rho$ is a direct summand of $\operatorname{ad}\bar\rho$ (as a G-representation).

3.20. ***Global deformations with local conditions.*** Now fix a finite set S, and for each $v \in S$, a profinite group G_v satisfying Φ_p, together with a continuous homomorphism $G_v \to G$, and a deformation problem $\mathcal{D}_v$ for $\bar\rho|_{G_v}$. [In applications, G will be a global Galois group, and the G_v will be decomposition groups at finite places.]

Also fix $\chi : G \to \mathcal{O}^\times$, a continuous homomorphism such that $\chi \bmod \lambda = \det\bar\rho$. Assume that $\bar\rho$ is absolutely irreducible, and fix some subset $T \subseteq S$.

Definition 3.21. Fix $A \in \operatorname{ob}\mathcal{C}_\mathcal{O}$. A *$T$-framed deformation* of $\bar\rho$ of *type* $\mathcal{S} := (S, \{\mathcal{D}_v\}_{v\in S}, \chi)$ to A is an equivalence class of tuples $(\rho, \{\alpha_v\}_{v\in T})$, where $\rho : G \to \mathrm{GL}_n(A)$ is a lift of $\bar\rho$ such that $\det\rho = \chi$ and $\rho|_{G_v} \in \mathcal{D}_v$ for all $v \in S$, and α_v is an element of $\ker(\mathrm{GL}_n(A) \to \mathrm{GL}_n(\mathbb{F}))$.

The equivalence relation is defined by decreeing that for each $\beta \in \ker(\mathrm{GL}_n(A) \to \mathrm{GL}_n(\mathbb{F}))$, we have $(\rho, \{\alpha_v\}_{v\in T}) \sim (\beta\rho\beta^{-1}, \{\beta\alpha_v\}_{v\in T})$.

The point of considering T-framed deformations is that it allows us to study absolutely irreducible representations $\bar{\rho}$ for which some of the $\bar{\rho}|_{G_v}$ are reducible, because if $(\rho, \{\alpha_v\}_{v\in T})$ is a T-framed deformation of type $\mathcal{S}$, then $\alpha_v^{-1}\rho|_{G_v}\alpha_v$ is a well-defined element of $\mathcal{D}_v$ (independent of the choice of representative of the equivalence class). The following lemma should be unsurprising.

Lemma 3.22. *The functor $\mathcal{C}_{\mathcal{O}} \to \underline{Sets}$ sending A to the set of T-framed deformations of $\bar{\rho}$ of type $\mathcal{S}$ is represented by a universal object $\rho^{\square_T} : G \to \mathrm{GL}_n(R_{\mathcal{S}}^{\square_T})$.*

Proof. See Proposition 2.2.9 of [Clozel et al. 2008]. □

If $T = \varnothing$ then we will write $R_{\mathcal{S}}^{\mathrm{univ}}$ for $R_{\mathcal{S}}^{\square_T}$.

3.23. ***Presenting global deformation rings over local lifting rings.*** Continue to use the notation of the previous subsection. Since $\alpha_v^{-1}\rho^{\square_T}|_{G_v}\alpha_v$ is a well-defined element of $\mathcal{D}_v$, we have a tautological homomorphism $R^{\square}_{\bar{\rho}|_{G_v},\chi}/I(\mathcal{D}_v) \to R_{\mathcal{S}}^{\square_T}$. Define

$$R_{\mathcal{S},T}^{\mathrm{loc}} := \widehat{\otimes}_{v\in T}(R^{\square}_{\bar{\rho}|_{G_v},\chi}/I(\mathcal{D}_v)).$$

Then we have a natural map $R_{\mathcal{S},T}^{\mathrm{loc}} \to R_{\mathcal{S}}^{\square_T}$.

We now generalize Corollary 3.14 by considering presentations of $R_{\mathcal{S}}^{\square_T}$ over $R_{\mathcal{S},T}^{\mathrm{loc}}$. In order to compute how many variables are needed to present $R_{\mathcal{S}}^{\square_T}$ over $R_{\mathcal{S},T}^{\mathrm{loc}}$, we must compute $\dim_{\mathbb{F}} \mathfrak{m}_{R_{\mathcal{S}}^{\square_T}}/(\mathfrak{m}^2_{R_{\mathcal{S}}^{\square_T}}, \mathfrak{m}_{R_{\mathcal{S},T}^{\mathrm{loc}}}, \lambda)$. Unsurprisingly, in order to compute this, we will compute a certain H^1.

We define a complex as follows. As usual, given a group G and an $\mathbb{F}[G]$-module M, we let $C^i(G, M)$ be the space of functions $G^i \to M$, and we let $\partial : C^i(G, M) \to C^{i+1}(G, M)$ be the usual coboundary map. We define a complex $C^i_{\mathcal{S},T,\mathrm{loc}}(G, \mathrm{ad}^0\,\bar{\rho})$ by

$$C^0_{\mathcal{S},T,\mathrm{loc}}(G, \mathrm{ad}^0\,\bar{\rho}) = \oplus_{v\in T} C^0(G_v, \mathrm{ad}\,\bar{\rho}) \oplus \oplus_{v\in S\setminus T} 0,$$
$$C^1_{\mathcal{S},T,\mathrm{loc}}(G, \mathrm{ad}^0\,\bar{\rho}) = \oplus_{v\in T} C^1(G_v, \mathrm{ad}^0\,\bar{\rho}) \oplus \oplus_{v\in S\setminus T} C^1(G_v, \mathrm{ad}^0\,\bar{\rho})/\tilde{L}(\mathcal{D}_v),$$

and for $i \geq 2$,

$$C^i_{\mathcal{S},T,\mathrm{loc}}(G, \mathrm{ad}^0\,\bar{\rho}) = \oplus_{v\in S} C^i(G_v, \mathrm{ad}^0\,\bar{\rho}).$$

Let $C^0_0(G, \mathrm{ad}^0\,\bar{\rho}) := C^0(G, \mathrm{ad}\,\bar{\rho})$, and set $C^i_0(G, \mathrm{ad}^0\,\bar{\rho}) = C^i(G, \mathrm{ad}^0\,\bar{\rho})$ for $i > 0$. Then we let $H^i_{\mathcal{S},T}(G, \mathrm{ad}^0\,\bar{\rho})$ denote the cohomology of the complex

$$C^i_{\mathcal{S},T}(G, \mathrm{ad}^0\,\bar{\rho}) := C^i_0(G, \mathrm{ad}^0\,\bar{\rho}) \oplus C^{i-1}_{\mathcal{S},T,\mathrm{loc}}(G, \mathrm{ad}^0\,\bar{\rho})$$

where the coboundary map is given by

$$(\phi, (\psi_v)) \mapsto (\partial\phi, (\phi|_{G_v} - \partial\psi_v)).$$

Then we have an exact sequence of complexes

$$0 \to C^{i-1}_{\mathcal{S},T,\mathrm{loc}}(G, \mathrm{ad}^0\,\bar{\rho}) \to C^i_{\mathcal{S},T}(G, \mathrm{ad}^0\,\bar{\rho}) \to C^i_0(G, \mathrm{ad}^0\,\bar{\rho}) \to 0,$$

and the corresponding long exact sequence in cohomology is

$$0 \to H^0_{\mathcal{S},T}(G,\mathrm{ad}^0\bar\rho) \to H^0(G,\mathrm{ad}\bar\rho) \to \oplus_{v\in T} H^0(G_v,\mathrm{ad}\bar\rho) \to H^1_{\mathcal{S},T}(G,\mathrm{ad}^0\bar\rho) \to H^1(G,\mathrm{ad}^0\bar\rho) \to \oplus_{v\in T} H^1(G_v,\mathrm{ad}^0\bar\rho)\oplus_{v\in S\setminus T} H^1(G_v,\mathrm{ad}^0\bar\rho)/L(\mathcal{D}_v) \to H^2_{\mathcal{S},T}(G,\mathrm{ad}^0\bar\rho) \to H^2(G,\mathrm{ad}^0\bar\rho) \to \oplus_{v\in S} H^2(G_v,\mathrm{ad}^0\bar\rho) \to H^3_{\mathcal{S},T}(G,\mathrm{ad}^0\bar\rho) \to \cdots$$

Taking Euler characteristics, we see that if we define the negative Euler characteristic χ by $\chi(G, \mathrm{ad}^0\bar\rho) = \sum_i (-1)^{i-1} \dim_{\mathbb{F}} H^i(G, \mathrm{ad}^0\bar\rho)$, we have

$$\begin{aligned}\chi_{\mathcal{S},T}(G, \mathrm{ad}^0\bar\rho) = -1 + \chi(G, \mathrm{ad}^0\bar\rho) - \sum_{v\in S} \chi(G_v, \mathrm{ad}^0\bar\rho) \\ + \sum_{v\in T} (\dim_{\mathbb{F}} H^0(G_v, \mathrm{ad}\,\bar\rho) - \dim_{\mathbb{F}} H^0(G_v, \mathrm{ad}^0\bar\rho)) \\ + \sum_{v\in S\setminus T} (\dim_{\mathbb{F}} L(\mathcal{D}_v) - \dim_{\mathbb{F}} H^0(G_v, \mathrm{ad}^0\bar\rho)).\end{aligned}$$

From now on for the rest of the notes, we specialize to the case that F is a number field, S is a finite set of finite places of F including all the places lying over p, and we set $G = G_{F,S}$, $G_v = G_{F_v}$ for $v \in S$. (Since $G = G_{F,S}$, note in particular that all deformations we are considering are unramified outside of S.) We then employ standard results on Galois cohomology that can be found in [Milne 2006]. In particular, we have $H^i(G_{F_v}, \mathrm{ad}\,\bar\rho) = 0$ if $i \ge 3$, and

$$H^i(G_{F,S}, \mathrm{ad}^0\bar\rho) \cong \oplus_{v \text{ real}} H^i(G_{F_v}, \mathrm{ad}^0\bar\rho) = 0$$

if $i \ge 3$ (the vanishing of the local cohomology groups follows as $p > 2$, so G_{F_v} has order coprime to that of $\mathrm{ad}^0\bar\rho$). Consequently, $H^i_{\mathcal{S},T}(G_{F,S}, \mathrm{ad}^0\bar\rho) = 0$ if $i > 3$.

We now employ the local and global Euler characteristic formulas. For simplicity, assume from now on that T contains all the places of S lying over p. The global formula gives

$$\chi(G_{F,S}, \mathrm{ad}^0\bar\rho) = -\sum_{v\,|\,\infty} \dim_{\mathbb{F}} H^0(G_{F_v}, \mathrm{ad}^0\bar\rho) + [F:\mathbb{Q}](n^2-1),$$

and the local formula gives

$$\sum_{v\in S} \chi(G_{F_v}, \mathrm{ad}^0\bar\rho) = \sum_{v\,|\,p} (n^2-1)[F_v:\mathbb{Q}_p] = (n^2-1)[F:\mathbb{Q}],$$

so that

$$\begin{aligned}&\chi_{\mathcal{S},T}(G_{F,S},\mathrm{ad}^0\bar{\rho})\\ &=-1+\#T-\sum_{v\mid\infty}\dim_{\mathbb{F}}H^0(G_{F_v},\mathrm{ad}^0\bar{\rho})+\sum_{v\in S\setminus T}(\dim_{\mathbb{F}}L(\mathcal{D}_v)-\dim_{\mathbb{F}}H^0(G_{F_v},\mathrm{ad}^0\bar{\rho})).\end{aligned}$$

Assume now that $\bar{\rho}$ is absolutely irreducible; then $H^0(G_{F,S},\mathrm{ad}\,\bar{\rho})=\mathbb{F}$, so $H^0_{\mathcal{S},T}(G_{F,S},\mathrm{ad}^0\bar{\rho})=\mathbb{F}$. To say something sensible about $H^1_{\mathcal{S},T}(G_{F,S},\mathrm{ad}^0\bar{\rho})$ we still need to control the $H^2_{\mathcal{S},T}$ and $H^3_{\mathcal{S},T}$. Firstly, the above long exact sequence gives us in particular the exact sequence

$$\begin{aligned}&H^1(G_{F,S},\mathrm{ad}^0\bar{\rho})\to\oplus_{v\in T}H^1(G_{F_v},\mathrm{ad}^0\bar{\rho})\oplus_{v\in S\setminus T}H^1(G_{F_v},\mathrm{ad}^0\bar{\rho})/L(\mathcal{D}_v)\\ &\to H^2_{\mathcal{S},T}(G_{F,S},\mathrm{ad}^0\bar{\rho})\to H^2(G_{F,S},\mathrm{ad}^0\bar{\rho})\longrightarrow\oplus_{v\in S}H^2(G_{F_v},\mathrm{ad}^0\bar{\rho})\\ &\to H^3_{\mathcal{S},T}(G_{F,S},\mathrm{ad}^0\bar{\rho})\longrightarrow 0.\end{aligned}$$

On the other hand, from the Poitou–Tate exact sequence [Milne 2006, Proposition 4.10, Chapter 1] we have an exact sequence

$$\begin{aligned}&H^1(G_{F,S},\mathrm{ad}^0\bar{\rho})\longrightarrow\oplus_{v\in S}H^1(G_{F_v},\mathrm{ad}^0\bar{\rho})\longrightarrow H^1(G_{F,S},(\mathrm{ad}^0\bar{\rho})^\vee(1))^\vee\\ &\to H^2(G_{F,S},\mathrm{ad}^0\bar{\rho})\longrightarrow\oplus_{v\in S}H^2(G_{F_v},\mathrm{ad}^0\bar{\rho})\longrightarrow H^0(G_{F,S},(\mathrm{ad}^0\bar{\rho})^\vee(1))^\vee\longrightarrow 0.\end{aligned}$$

Note that $\mathrm{ad}^0\bar{\rho}$ is self-dual under the trace pairing, so we can and do identify $(\mathrm{ad}^0\bar{\rho})^\vee(1)$ and $(\mathrm{ad}^0\bar{\rho})(1)$. If we let $L(\mathcal{D}_v)^\perp\subseteq H^1(G_{F_v},(\mathrm{ad}^0\bar{\rho})(1))$ denote the annihilator of $L(\mathcal{D}_v)$ under the pairing coming from Tate local duality, and we define

$$\begin{aligned}&H^1_{\mathcal{S},T}(G_{F,S},(\mathrm{ad}^0\bar{\rho})(1))\\ &\qquad:=\ker(H^1(G_{F,S},(\mathrm{ad}^0\bar{\rho})(1))\to\oplus_{v\in S\setminus T}(H^1(G_{F_v},(\mathrm{ad}^0\bar{\rho})(1))/L(\mathcal{D}_v)^\perp)),\end{aligned}$$

then we deduce that we have an exact sequence

$$\begin{aligned}&H^1(G_{F,S},\mathrm{ad}^0\bar{\rho})\to\oplus_{v\in T}H^1(G_{F_v},\mathrm{ad}^0\bar{\rho})\oplus_{v\in S\setminus T}H^1(G_{F_v},\mathrm{ad}^0\bar{\rho})/L(\mathcal{D}_v)\\ &\to H^1_{\mathcal{S},T}(G_{F,S},\mathrm{ad}^0\bar{\rho}(1))^\vee\to H^2(G_{F,S},\mathrm{ad}^0\bar{\rho})\longrightarrow\oplus_{v\in S}H^2(G_{F_v},\mathrm{ad}^0\bar{\rho})\\ &\to H^0(G_{F,S},\mathrm{ad}^0\bar{\rho}(1))^\vee\longrightarrow 0,\end{aligned}$$

and comparing with the diagram above shows that

$$\begin{aligned}H^3_{\mathcal{S},T}(G_{F,S},\mathrm{ad}^0\bar{\rho})&\cong H^0(G_{F,S},\mathrm{ad}^0\bar{\rho}(1))^\vee,\\ H^2_{\mathcal{S},T}(G_{F,S},\mathrm{ad}^0\bar{\rho})&\cong H^1_{\mathcal{S},T}(G_{F,S},\mathrm{ad}^0\bar{\rho}(1))^\vee.\end{aligned}$$

Combining all of this, we see that

$$\begin{aligned}\dim_{\mathbb{F}} H^1_{\mathcal{S},T}(G_{F,S}, \operatorname{ad}^0 \bar\rho) = \#T - \sum_{v \mid \infty} \dim_{\mathbb{F}} H^0(G_{F_v}, \operatorname{ad}^0 \bar\rho) \\ + \sum_{v \in S \setminus T} (\dim_{\mathbb{F}} L(\mathcal{D}_v) - \dim_{\mathbb{F}} H^0(G_{F_v}, \operatorname{ad}^0 \bar\rho)) \\ + \dim_{\mathbb{F}} H^1_{\mathcal{S}^\perp,T}(G_{F,S}, \operatorname{ad}^0 \bar\rho(1)) \\ - \dim_{\mathbb{F}} H^0(G_{F,S}, \operatorname{ad}^0 \bar\rho(1)).\end{aligned}$$

Now, similar arguments to those we used above give us the following result; see Section 2.2 of [Clozel et al. 2008].

Proposition 3.24. (1) *There is a canonical isomorphism*

$$\operatorname{Hom}(\mathfrak{m}_{R_{\mathcal{S}}^{\square_T}}/(\mathfrak{m}^2_{R_{\mathcal{S}}^{\square_T}}, \mathfrak{m}_{R^{\mathrm{loc}}_{\mathcal{S},T}}, \lambda), \mathbb{F}) \cong H^1_{\mathcal{S},T}(G_{F,S}, \operatorname{ad}^0 \bar\rho).$$

(2) $R_{\mathcal{S}}^{\square_T}$ *is the quotient of a power series ring in* $\dim_{\mathbb{F}} H^1_{\mathcal{S}^\perp,T}(G_{F,S}, \operatorname{ad}^0 \bar\rho)$ *variables over* $R^{\mathrm{loc}}_{\mathcal{S},T}$.

(3) *The Krull dimension of* $R_{\mathcal{S}}^{\mathrm{univ}}$ *is at least*

$$\begin{aligned}1 + \sum_{v \in S} (\operatorname{Krull}\dim(R^{\square}_{\bar\rho|_{G_{F_v}},\chi}/I(\mathcal{D}_v)) - n^2) \\ - \sum_{v \mid \infty} \dim_{\mathbb{F}} H^0(G_{F_v}, \operatorname{ad}^0 \bar\rho) - \dim_{\mathbb{F}} H^0(G_{F,S}, \operatorname{ad}^0 \bar\rho(1)).\end{aligned}$$

3.25. ***Finiteness of maps between global deformation rings.*** Suppose that F'/F is a finite extension of number fields, and that S' is the set of places of F' lying over S. Assume that $\bar\rho|_{G_{F',S'}}$ is absolutely irreducible. Then restricting the universal deformation ρ^{univ} of $\bar\rho$ to $G_{F',S'}$ gives a ring homomorphism $R^{\mathrm{univ}}_{\bar\rho|_{G_{F',S'}}} \to R^{\mathrm{univ}}_{\bar\rho}$. The following very useful fact is due to Khare and Wintenberger.

Proposition 3.26. *The ring* $R^{\mathrm{univ}}_{\bar\rho}$ *is a finitely generated* $R^{\mathrm{univ}}_{\bar\rho|_{G_{F',S'}}}$*-module.*

Proof. See, e.g., Lemma 1.2.3 of [Barnet-Lamb et al. 2014]. □

3.27. ***Local deformation rings with l = p.*** For proving modularity lifting theorems, we typically need to consider local deformation rings when $l = p$ which capture certain properties in p-adic Hodge theory (for example being crystalline with fixed Hodge–Tate weights). These deformation rings are one of the most difficult and interesting parts of the subject; for example, a detailed computation of deformation rings with $l = p = 3$ was at the heart of the eventual proof of the Taniyama–Shimura–Weil conjecture.

For the most part, the relevant deformation rings when $l = p$ are still not well understood; we don't have a concrete description of the rings in most cases, or even basic information such as the number of irreducible components of the generic fiber. In these notes, we will ignore all of these difficulties, and work only with the "Fontaine–Laffaille" case, where the deformation rings are formally smooth. This is already enough to have important applications.

Assume that $K/\mathbb{Q}_p$ is a finite unramified extension, and assume that L is chosen large enough to contain the images of all embeddings $K \hookrightarrow \overline{\mathbb{Q}}_p$. For each $\sigma : K \hookrightarrow L$, let H_σ be a set of n distinct integers, such that the difference between the maximal and minimal elements of H_σ is less than or equal to $p-2$.

Theorem 3.28. *There is a unique reduced, p-torsion free quotient $R^{\square}_{\bar{\rho},\chi,\mathrm{cr},\{H_\sigma\}}$ of $R^{\square}_{\bar{\rho},\chi}$ with the property that a continuous homomorphism $\psi : R^{\square}_{\bar{\rho},\chi} \to \overline{\mathbb{Q}}_p$ factors through $R^{\square}_{\bar{\rho},\chi,\mathrm{cr},\{H_\sigma\}}$ if and only if $\psi \circ \rho^{\square}$ is crystalline, and for each $\sigma : K \hookrightarrow L$, we have* $\mathrm{HT}_\sigma(\psi \circ \rho^{\square}) = H_\sigma$.

Furthermore it has Krull dimension given by

$$\dim R^{\square}_{\bar{\rho},\chi,\mathrm{cr},\{H_\sigma\}} = n^2 + [K : \mathbb{Q}_p]\tfrac{1}{2}n(n-1),$$

and in fact $R^{\square}_{\bar{\rho},\chi,\mathrm{cr},\{H_\sigma\}}$ is formally smooth over $\mathcal{O}$, i.e., it is isomorphic to a power series ring in $n^2 - 1 + [K : \mathbb{Q}_p]\frac{1}{2}n(n-1)$ variables over $\mathcal{O}$.

In fact, if we remove the assertion of formal smoothness, Theorem 3.28 still holds without the assumption that $K/\mathbb{Q}_p$ is unramified, and without any assumption on the difference between the maximal and minimal elements of the H_σ, but in this case it is a much harder theorem of Kisin [2008]. In any case, the formal smoothness will be important for us.

Theorem 3.28 is essentially a consequence of Fontaine–Laffaille theory [Fontaine and Laffaille 1982], which is a form of integral p-adic Hodge theory; it classifies the Galois-stable lattices in crystalline representations, under the assumptions we've made above. The first proof of Theorem 3.28 was essentially in Ramakrishna's thesis [1993], and the general result is the content of Section 2.4 of [Clozel et al. 2008].

3.29. ***Local deformation rings with $p \neq l$.*** In contrast to the situation when $l = p$, we will need to consider several deformation problems when $l \neq p$. We will restrict ourselves to the two-dimensional case. Let $K/\mathbb{Q}_l$ be a finite extension, with $l \neq p$, and fix $n = 2$. As we saw in Section 2.7, there is essentially an incompatibility between the wild inertia subgroup of G_K and the p-adic topology on $\mathrm{GL}_2(\mathcal{O})$, which makes it possible to explicitly describe the p-adic representations of G_K, and consequently the corresponding universal deformation rings. This was done in varying degrees of generality over a long period of time; in particular, in the general n-dimensional case we highlight Section 2.4.4 of [Clozel et al. 2008] and [Choi 2009], and in the 2-dimensional setting [Pilloni 2008] and [Shotton 2016]. In fact [Shotton 2016] gives a complete description of the deformation rings for a fixed inertial type.

We will content ourselves with recalling some of the basic structural results, and with giving a sketch of how the results are proved in one particular case; see Exercise 3.34 below.

3.30. ***Deformations of fixed type.*** Recall from Proposition 2.18 that given a representation $\rho : G_K \to \mathrm{GL}_2(\overline{\mathbb{Q}}_p)$ there is a Weil–Deligne representation $\mathrm{WD}(\rho)$ associated to ρ. If $\mathrm{WD} = (r, N)$ is a Weil–Deligne representation, then we write $\mathrm{WD}\,|_{I_K}$ for $(r|_{I_K}, N)$, and call it an *inertial* WD-*type*.

Fix $\bar{\rho} : G_K \to \mathrm{GL}_2(\mathbb{F})$. Then (assuming as usual that L is sufficiently large) we have the following general result on $R^{\square}_{\bar{\rho},\chi}$; see, e.g., Theorem 3.3.1 of [Böckle 2013].

Theorem 3.31. *$R^{\square}_{\bar{\rho},\chi}$ is equidimensional of Krull dimension* 4, *and the generic fiber $R^{\square}_{\bar{\rho},\chi}[1/p]$ has Krull dimension* 3. *Furthermore*:

(a) *The function which takes a $\overline{\mathbb{Q}}_p$-point $x : R^{\square}_{\bar{\rho},\chi}[1/p] \to \overline{\mathbb{Q}}_p$ to* (*the isomorphism class of*) $\mathrm{WD}(x \circ \rho^{\square})|_{I_K}$ (*forgetting N*) *is constant on the irreducible components of $R^{\square}_{\bar{\rho},\chi}[1/p]$.*

(b) *The irreducible components of $R^{\square}_{\bar{\rho},\chi}[1/p]$ are all regular*, *and there are only finitely many of them.*

In light of Theorem 3.31, we make the following definition. Let τ be an inertial WD-type. Then there is a unique reduced, p-torsion free quotient $R^{\square}_{\bar{\rho},\chi,\tau}$ of $R^{\square}_{\bar{\rho},\chi}$ with the property that a continuous homomorphism $\psi : R^{\square}_{\bar{\rho},\chi} \to \overline{\mathbb{Q}}_p$ factors through $R^{\square}_{\bar{\rho},\chi,\tau}$ if and only if $\psi \circ \rho^{\square}$ has inertial Weil–Deligne type τ. (Of course, for all but finitely many τ, we will just have $R^{\square}_{\bar{\rho},\chi,\tau} = 0$.) By Theorem 3.31 we see that if $R^{\square}_{\bar{\rho},\chi,\tau}$ is nonzero then it has Krull dimension 4.

3.32. ***Taylor–Wiles deformations.*** As the name suggests, the deformations that we consider in this subsection will be of crucial importance for the Taylor–Wiles–Kisin method. Assume that $\bar{\rho}$ is unramified, that $\bar{\rho}(\mathrm{Frob}_K)$ has distinct eigenvalues, and that $\#k \equiv 1 \pmod{p}$. Suppose also that χ is unramified.

Lemma 3.33. *Suppose that* $(\#k - 1)$ *is exactly divisible by p^m. Then* $R^{\square}_{\bar{\rho},\chi} \cong \mathcal{O}[[x, y, B, u]]/((1+u)^{p^m} - 1)$. *Furthermore*, *if $\varphi \in G_K$ is a lift of* Frob_K, *then $\rho^{\square}(\varphi)$ is conjugate to a diagonal matrix.*

Exercise 3.34. Prove this lemma as follows. Note firstly that $\rho^{\square}(P_K) = \{1\}$, because $\bar{\rho}(P_K) = \{1\}$, so $\rho^{\square}(P_K)$ is a pro-l-subgroup of the pro-p-group $\ker(\mathrm{GL}_2(R^{\square}_{\bar{\rho},\chi}) \to \mathrm{GL}_2(\mathbb{F}))$.

Let φ be a fixed lift of Frob_K to G_K/P_K, and σ a topological generator of I_K/P_K, which as in Section 2.7 we can choose so that $\varphi^{-1}\sigma\varphi = \sigma^{\#k}$. Write $\bar{\rho}(\varphi) = \left(\begin{smallmatrix} \bar{\alpha} & 0 \\ 0 & \bar{\beta} \end{smallmatrix}\right)$, and fix lifts $\alpha, \beta \in \mathcal{O}$ of $\bar{\alpha}, \bar{\beta}$.

Then we will show that we can take

$$\rho^{\square}(\varphi) = \begin{pmatrix} 1 & y \\ x & 1 \end{pmatrix}^{-1} \begin{pmatrix} \alpha + B & 0 \\ 0 & \chi(\varphi)/(\alpha + B) \end{pmatrix} \begin{pmatrix} 1 & y \\ x & 1 \end{pmatrix},$$

$$\rho^{\square}(\sigma) = \begin{pmatrix} 1 & y \\ x & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1+u & 0 \\ 0 & (1+u)^{-1} \end{pmatrix} \begin{pmatrix} 1 & y \\ x & 1 \end{pmatrix}.$$

(1) Let $\rho : G_K \to \mathrm{GL}_2(A)$ be a lift of $\bar{\rho}$. By Hensel's lemma, there are $a, b \in \mathfrak{m}_A$ such that $\rho(\varphi)$ has characteristic polynomial $(X-(\alpha+a))(X-(\beta+b))$. Show that there are $x, y \in \mathfrak{m}_A$ such that

$$\rho(\varphi)\begin{pmatrix} 1 \\ x \end{pmatrix} = (\alpha + a)\begin{pmatrix} 1 \\ x \end{pmatrix} \quad \text{and} \quad \rho(\varphi)\begin{pmatrix} y \\ 1 \end{pmatrix} = (\beta + b)\begin{pmatrix} y \\ 1 \end{pmatrix}$$

(2) Since $\bar{\rho}$ is unramified, $\bar{\rho}(\sigma) = 1$, so we may write

$$\begin{pmatrix} 1 & y \\ x & 1 \end{pmatrix}^{-1} \rho(\sigma) \begin{pmatrix} 1 & y \\ x & 1 \end{pmatrix} = \begin{pmatrix} 1+u & v \\ w & 1+z \end{pmatrix}$$

with $u, v, w, z \in \mathfrak{m}_A$. Use the commutation relation between $\rho(\varphi)$ and $\rho(\sigma)$ to show that $v = w = 0$.

(3) Use the fact that χ is unramified to show that $1 + z = (1+u)^{-1}$.

(4) Show that $(1+u)^{\#k} = 1 + u$, and deduce that $(1+u)^{\#k-1} = 1$.

(5) Deduce that $(1+u)^{p^m} = 1$.

(6) Complete the proof of the lemma.

3.35. ***Taylor's "Ihara avoidance" deformations.*** The following deformation rings are crucial to Taylor's arguments [2008] which avoid the use of Ihara's lemma in proving automorphy lifting theorems. When $n = 2$ these arguments are not logically necessary, but they are crucial to all applications of automorphy lifting theorems when $n > 2$. They are used in order to compare Galois representations with differing ramification at places not dividing p.

Continue to let $K/\mathbb{Q}_l$ be a finite extension, and assume that $\bar{\rho}$ is the trivial 2-dimensional representation, that $\#k \equiv 1 \pmod{p}$, that χ is unramified, and that $\bar{\chi}$ is trivial. Again, we see that $\rho^{\square}(P_K)$ is trivial, so that $\rho^{\square}$ is determined by the two matrices $\rho^{\square}(\sigma)$ and $\rho^{\square}(\varphi)$, as in Exercise 3.34. A similar analysis then yields the following facts. (For the proof of the analogous results in the n-dimensional case, see Section 3 of [Taylor 2008].)

Definition 3.36. (1) Let $\mathcal{P}_{\mathrm{ur}}$ be the minimal ideal of $R^{\square}_{\bar{\rho},\chi}$ modulo which $\rho^{\square}(\sigma) = 1_2$.

(2) For any root of unity ζ which is trivial modulo λ, we let $\mathcal{P}_{\zeta}$ be the minimal ideal of $R^{\square}_{\bar{\rho},\chi}$ modulo which $\rho^{\square}(\sigma)$ has characteristic polynomial $(X-\zeta)(X-\zeta^{-1})$.

(3) Let $\mathcal{P}_{\mathrm{m}}$ be the minimal ideal of $R^{\square}_{\bar{\rho},\chi}$ modulo which $\rho^{\square}(\sigma)$ has characteristic polynomial $(X-1)^2$, and $\#k(\operatorname{tr}\rho^{\square}(\varphi))^2 = (1+\#k)^2 \det \rho^{\square}(\varphi)$.

[The motivation for the definition of $\mathcal{P}_{\mathrm{m}}$ is that we are attempting to describe the unipotent liftings, and if you assume that $\rho^{\square}(\sigma) = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, this is the relation forced on $\rho^{\square}(\varphi)$.]

Proposition 3.37. *The minimal primes of $R^{\square}_{\bar{\rho},\chi}$ are precisely $\sqrt{\mathcal{P}_{\mathrm{ur}}}$, $\sqrt{\mathcal{P}_{\mathrm{m}}}$, and the $\sqrt{\mathcal{P}_{\zeta}}$ for $\zeta \neq 1$. We have $\sqrt{\mathcal{P}_1} = \sqrt{\mathcal{P}_{\mathrm{ur}}} \cap \sqrt{\mathcal{P}_{\mathrm{m}}}$.*

Write $R^{\square}_{\bar{\rho},\chi,1}$, $R^{\square}_{\bar{\rho},\chi,\zeta}$, $R^{\square}_{\bar{\rho},\chi,\mathrm{ur}}$, $R^{\square}_{\bar{\rho},\chi,\mathrm{m}}$ for the corresponding quotients of $R^{\square}_{\bar{\rho},\chi}$.

Theorem 3.38. *We have $R^{\square}_{\bar{\rho},\chi,1}/\lambda = R^{\square}_{\bar{\rho},\chi,\zeta}/\lambda$. Furthermore*:

(1) *If $\zeta \neq 1$ then $R^{\square}_{\bar{\rho},\chi,\zeta}[1/p]$ is geometrically irreducible of dimension* 3.

(2) *$R^{\square}_{\bar{\rho},\chi,\mathrm{ur}}$ is formally smooth over $\mathcal{O}$ (and thus geometrically irreducible) of relative dimension* 3.

(3) *$R^{\square}_{\bar{\rho},\chi,\mathrm{m}}[1/p]$ is geometrically irreducible of dimension* 3.

(4) *Both*

$$\operatorname{Spec} R^{\square}_{\bar{\rho},\chi,1} = \operatorname{Spec} R^{\square}_{\bar{\rho},\chi,\mathrm{ur}} \cup \operatorname{Spec} R^{\square}_{\bar{\rho},\chi,\mathrm{m}}$$

and

$$\operatorname{Spec} R^{\square}_{\bar{\rho},\chi,1}/\lambda = \operatorname{Spec} R^{\square}_{\bar{\rho},\chi,\mathrm{ur}}/\lambda \cup \operatorname{Spec} R^{\square}_{\bar{\rho},\chi,\mathrm{m}}/\lambda$$

are unions of two irreducible components, and have relative dimension 3.

Proof. See Proposition 3.1 of [Taylor 2008] for an n-dimensional version of this result. In the 2-dimensional case it can be proved by explicitly computing equations for the lifting rings; see [Shotton 2016]. □

4. Modular and automorphic forms, and the Langlands correspondence

We now turn to the automorphic side of the Langlands correspondence, and define the spaces of modular forms to which our modularity lifting theorems pertain.

4.1. ***The local Langlands correspondence (and the Jacquet–Langlands correspondence).*** Weil–Deligne representations are the objects on the "Galois" side of the local Langlands correspondence. We now describe the objects on the "automorphic" side. These will be representations (π, V) of $\mathrm{GL}_n(K)$ on (usually infinite-dimensional) $\mathbb{C}$-vector spaces, where as above $K/\mathbb{Q}_l$ is a finite extension for some prime l.

Definition 4.2. We say that (π, V) is *smooth* if for any vector $v \in V$, the stabilizer of v in $\mathrm{GL}_n(K)$ is open. We say that (π, V) is *admissible* if it is smooth, and for any compact open subgroup $U \subset \mathrm{GL}_n(K)$, V^U is finite-dimensional.

For example, a smooth one-dimensional representation of $K^\times$ is the same thing as a continuous character (for the discrete topology on $\mathbb{C}$).

Fact 4.3. (1) If π is smooth and irreducible then it is admissible.

(2) Schur's lemma holds for admissible smooth representations, and in particular if π is smooth, admissible and irreducible then it has a central character $\chi_\pi : K^\times \to \mathbb{C}^\times$.

In general these representations are classified in terms of the (super)cuspidal representations. We won't need the details of this classification, and accordingly we won't define the cuspidal representations; see, for example, Chapter IV of [Bushnell and Henniart 2006].

Let B be the subgroup of $\mathrm{GL}_2(K)$ consisting of upper-triangular matrices. Define $\delta : B \to K^\times$ by

$$\delta\left(\begin{pmatrix} a & * \\ 0 & d \end{pmatrix}\right) = ad^{-1}.$$

Given two continuous characters $\chi_1, \chi_2 : K^\times \to \mathbb{C}^\times$, we may view $\chi_1 \otimes \chi_2$ as a representation of B by

$$\chi_1 \otimes \chi_2 : \begin{pmatrix} a & * \\ 0 & d \end{pmatrix} \mapsto \chi_1(a)\chi_2(d).$$

Then we define a representation $\chi_1 \times \chi_2$ of $\mathrm{GL}_2(K)$ by *normalized induction*:

$$\begin{aligned} \chi_1 \times \chi_2 &= \text{n-Ind}_B^{\mathrm{GL}_2(K)}(\chi_1 \otimes \chi_2) \\ &:= \{\varphi : \mathrm{GL}_2(K) \to \mathbb{C} \mid \varphi(hg) = (\chi_1 \otimes \chi_2)(h)|\delta(h)|_K^{1/2}\varphi(g) \\ &\qquad\qquad \text{for all } h \in B,\ g \in \mathrm{GL}_2(K)\} \end{aligned}$$

where $\mathrm{GL}_2(K)$ acts by $(g\varphi)(g') = \varphi(g'g)$, and we only allow smooth φ, i.e., functions for which there is an open subgroup U of $\mathrm{GL}_2(K)$ such that $\varphi(gu) = \varphi(g)$ for all $g \in \mathrm{GL}_2(K)$, $u \in U$.

The representation $\chi_1 \times \chi_2$ has length at most 2, but is not always irreducible. It is always the case that $\chi_1 \times \chi_2$ and $\chi_2 \times \chi_1$ have the same Jordan-Hölder factors. If $\chi_1 \times \chi_2$ is irreducible then we say that it is a *principal series* representation.

Fact 4.4. (1) $\chi_1 \times \chi_2$ is irreducible unless $\chi_1/\chi_2 = |\cdot|_K^{\pm 1}$.

(2) $\chi \times \chi|\cdot|_K$ has a one-dimensional irreducible subrepresentation, and the corresponding quotient is irreducible. We denote this quotient by $\mathrm{Sp}_2(\chi)$.

We will let $\chi_1 \boxplus \chi_2$ denote $\chi_1 \times \chi_2$ unless $\chi_1/\chi_2 = |\cdot|_K^{\pm 1}$, and we let

$$\chi \boxplus \chi|\cdot|_K = \chi|\cdot|_K \boxplus \chi = (\chi|\cdot|_K^{1/2}) \circ \det.$$

(While this notation may seem excessive, we remark that a similar construction is possible for n-dimensional representations, which is where the notation comes from.) These representations, and the $\mathrm{Sp}_2(\chi)$, are all the noncuspidal irreducible admissible representations of $\mathrm{GL}_2(K)$. We say that an irreducible smooth representation π of $\mathrm{GL}_2(K)$ is *discrete series* if it is of the form $\mathrm{Sp}_2(\chi)$ or is cuspidal.

The local Langlands correspondence provides a unique family of bijections rec_K from the set of isomorphism classes of irreducible smooth representations of $\mathrm{GL}_n(K)$ to the set of isomorphism classes of n-dimensional Frobenius semisimple Weil–Deligne representations of W_K over $\mathbb{C}$, satisfying a list of properties. In order to be uniquely determined, one needs to formulate the correspondence for all n at once, and the properties are expressed in terms of L- and ε-factors, neither of which we have defined. Accordingly, we will not make a complete statement of the local Langlands correspondence, but will rather state the properties of the correspondence that we will need to use. (Again, the reader could look at the book [Bushnell and Henniart 2006] for these properties, and many others.) It is also possible to define the correspondence in global terms, as we will see later, and indeed at present the only proof of the correspondence is global.

Fact 4.5. We now list some properties of rec_K for $n = 1, 2$:

(1) If $n = 1$ then $\mathrm{rec}_K(\pi) = \pi \circ \mathrm{Art}_K^{-1}$.

(2) If χ is a smooth character, $\mathrm{rec}_K(\pi \otimes (\chi \circ \det)) = \mathrm{rec}_K(\pi) \otimes \mathrm{rec}_K(\chi)$.

(3) $\mathrm{rec}_K(\mathrm{Sp}_2(\chi)) = \mathrm{Sp}_2(\mathrm{rec}_K(\chi))$; see Exercise 2.12 for this notation.

(4) $\mathrm{rec}_K(\chi_1 \boxplus \chi_2) = \mathrm{rec}_K(\chi_1) \oplus \mathrm{rec}_K(\chi_2)$.

(5) If $n = 2$, then $\mathrm{rec}_K(\pi)$ is unramified (i.e., $N = 0$ and the restriction to I_K is trivial) if and only if $\pi = \chi_1 \boxplus \chi_2$ with χ_1, χ_2 both unramified characters (i.e., trivial on $\mathcal{O}_K^\times$). These conditions are equivalent to $\pi^{\mathrm{GL}_2(\mathcal{O}_K)} \neq 0$, in which case it is one-dimensional.

(6) π is discrete series if and only if $\mathrm{rec}_K(\pi)$ is indecomposable, and cuspidal if and only if $\mathrm{rec}_K(\pi)$ is irreducible.

4.6. ***Hecke operators.*** Consider the set of compactly supported $\mathbb{C}$-valued functions on $\mathrm{GL}_2(\mathcal{O}_K)\backslash \mathrm{GL}_2(K)/\mathrm{GL}_2(\mathcal{O}_K)$. Concretely, these are functions which vanish outside of a finite number of double cosets $\mathrm{GL}_2(\mathcal{O}_K)g\,\mathrm{GL}_2(\mathcal{O}_K)$. The set of such functions is in fact a ring, with the multiplication being given by convolution. To be precise, we fix μ the (left and right) Haar measure on $\mathrm{GL}_2(K)$ such that $\mu(\mathrm{GL}_2(\mathcal{O}_K)) = 1$, and we define

$$(\varphi_1 * \varphi_2)(x) = \int_{\mathrm{GL}_2(K)} \varphi_1(g)\varphi_2(g^{-1}x)\,d\mu_g.$$

Of course, this integral is really just a finite sum. One can check without too much difficulty that the ring $\mathcal{H}$ of these Hecke operators is just $\mathbb{C}[T, S^{\pm 1}]$, where T is the characteristic function of

$$\mathrm{GL}_2(\mathcal{O}_K)\begin{pmatrix}\varpi_K & 0\\ 0 & 1\end{pmatrix}\mathrm{GL}_2(\mathcal{O}_K)$$

and S is the characteristic function of

$$\mathrm{GL}_2(\mathcal{O}_K)\begin{pmatrix}\varpi_K & 0\\ 0 & \varpi_K\end{pmatrix}\mathrm{GL}_2(\mathcal{O}_K).$$

The algebra $\mathcal{H}$ acts on an irreducible admissible $\mathrm{GL}_2(K)$-representation π. Given $\varphi \in \mathcal{H}$, we obtain a linear map $\pi(\varphi) : \pi \to \pi^{\mathrm{GL}_2(\mathcal{O}_K)}$, by

$$\pi(\varphi)(v) = \int_{\mathrm{GL}_2(K)} \varphi(g)\pi(g)v\,d\mu_g.$$

In particular, if π is unramified then $\pi(\varphi)$ acts via a scalar on the one-dimensional $\mathbb{C}$-vector space $\pi^{\mathrm{GL}_2(\mathcal{O}_K)}$. We will now compute this scalar explicitly.

Exercise 4.7. (1) Show that we have decompositions

$$\mathrm{GL}_2(\mathcal{O}_K)\begin{pmatrix}\varpi_K & 0\\ 0 & \varpi_K\end{pmatrix}\mathrm{GL}_2(\mathcal{O}_K) = \begin{pmatrix}\varpi_K & 0\\ 0 & \varpi_K\end{pmatrix}\mathrm{GL}_2(\mathcal{O}_K),$$

and

$$\mathrm{GL}_2(\mathcal{O}_K)\begin{pmatrix}\varpi_K & 0\\ 0 & 1\end{pmatrix}\mathrm{GL}_2(\mathcal{O}_K)$$
$$= \Bigg(\coprod_{\alpha\in\mathcal{O}_K \pmod{\varpi_K}} \begin{pmatrix}\varpi_K & \alpha\\ 0 & 1\end{pmatrix}\mathrm{GL}_2(\mathcal{O}_K)\Bigg) \coprod \begin{pmatrix}1 & 0\\ 0 & \varpi_K\end{pmatrix}\mathrm{GL}_2(\mathcal{O}_K).$$

(2) Suppose that $\pi = (\chi|\cdot|^{1/2}) \circ \det$ with χ unramified. Show that $\pi^{\mathrm{GL}_2(\mathcal{O}_K)} = \pi$, and that S acts via $\chi(\varpi_K)^2(\#k)^{-1}$, and that T acts via $(\#k^{1/2} + \#k^{-1/2})\chi(\varpi_K)$.

(3) Suppose that χ_1, χ_2 are unramified characters and that $\chi_1 \neq \chi_2|\cdot|_K^{\pm 1}$. Let $\pi = \chi_1 \boxplus \chi_2$. Using the Iwasawa decomposition $\mathrm{GL}_2(K) = B(K)\,\mathrm{GL}_2(\mathcal{O}_K)$, check that $\pi^{\mathrm{GL}_2(\mathcal{O}_K)}$ is one-dimensional, and is spanned by a function φ_0 with $\varphi_0\left(\left(\begin{smallmatrix}a & b\\ 0 & d\end{smallmatrix}\right)\right) = \chi_1(a)\chi_2(d)|a/d|^{1/2}$. Show that S acts on $\pi^{\mathrm{GL}_2(\mathcal{O}_K)}$ via $(\chi_1\chi_2)(\varpi_K)$, and that T acts via $\#k^{1/2}(\chi_1(\varpi_K) + \chi_2(\varpi_K))$.

4.8. ***Modular forms and automorphic forms on quaternion algebras.*** Let F be a totally real field, and let D/F be a quaternion algebra with center F, i.e., a central simple F-algebra of dimension 4. Letting $S(D)$ be the set of places v of F at which D is ramified, i.e., for which $D \otimes_F F_v$ is a division algebra (equivalently, is not isomorphic to $M_2(F_v)$), it is known that $S(D)$ classifies D up to isomorphism, and that $S(D)$ can be any finite set of places of F of even cardinality (so for example $S(D)$ is empty if and only if $D = M_2(F)$). We will now define some spaces of automorphic forms on $D^\times$.

For each $v \mid \infty$ fix $k_v \geq 2$ and $\eta_v \in \mathbb{Z}$ such that $k_v + 2\eta_v - 1 = w$ is independent of v. These will be the weights of our modular forms. Let G_D be the algebraic group over $\mathbb{Q}$ such that for any $\mathbb{Q}$-algebra R, $G_D(R) = (D \otimes_{\mathbb{Q}} R)^\times$. For each place $v \mid \infty$ of F, we define a subgroup U_v of $(D \otimes_F F_v)^\times$ as follows: if $v \in S(D)$ we let $U_v = (D \otimes_F F_v)^\times \cong \mathbb{H}^\times$ (where $\mathbb{H}$ denotes the Hamilton quaternions), and if $v \notin S(D)$, so that $(D \otimes_F F_v)^\times \cong \mathrm{GL}_2(\mathbb{R})$, we take $U_v = \mathbb{R}^\times \mathrm{SO}(2)$. If $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{R})$ and $z \in \mathbb{C} - \mathbb{R}$, we let $j(\gamma, z) = cz + d$. One checks easily that $j(\gamma\delta, z) = j(\gamma, \delta z) j(\delta, z)$.

We now define a representation (τ_v, W_v) of U_v over $\mathbb{C}$ for each $v \mid \infty$. If $v \in S(D)$, we have $U_v \hookrightarrow \mathrm{GL}_2(\bar{F}_v) \cong \mathrm{GL}_2(\mathbb{C})$ which acts on $\mathbb{C}^2$, and we let (τ_v, W_v) be the representation

$$(\mathrm{Sym}^{k_v - 2} \mathbb{C}^2) \otimes (\wedge^2 \mathbb{C}^2)^{\eta_v}.$$

If $v \notin S(D)$, then we have $U_v \cong \mathbb{R}^\times \mathrm{SO}(2)$, and we take $W_v = \mathbb{C}$, with

$$\tau_v(\gamma) = j(\gamma, i)^{k_v} (\det \gamma)^{\eta_v - 1}.$$

We write $U_\infty = \prod_{v \mid \infty} U_v$, $W_\infty = \otimes_{v \mid \infty} W_v$, $\tau_\infty = \otimes_{v \mid \infty} \tau_v$. Let $\mathbb{A} = \mathbb{A}_{\mathbb{Q}}$ be the adeles of $\mathbb{Q}$, and let $\mathbb{A}^\infty$ be the finite adeles. We then define $S_{D,k,\eta}$ (where k, η reflect the dependence on the integers k_v, η_v) to be the space of functions $\varphi : G_D(\mathbb{Q}) \backslash G_D(\mathbb{A}) \to W_\infty$ which satisfy:

(1) $\varphi(g u_\infty) = \tau_\infty(u_\infty)^{-1} \varphi(g)$ for all $u_\infty \in U_\infty$ and $g \in G_D(\mathbb{A})$.

(2) There is a nonempty open subset $U^\infty \subset G_D(\mathbb{A}^\infty)$ such that $\varphi(gu) = \varphi(g)$ for all $u \in U^\infty$, $g \in G_D(\mathbb{A})$.

(3) Let S_∞ denote the infinite places of F. If $g \in G_D(\mathbb{A}^\infty)$ then the function

$$(\mathbb{C} - \mathbb{R})^{S_\infty - S(D)} \to W_\infty$$

defined by

$$h_\infty(i, \ldots, i) \mapsto \tau_\infty(h_\infty) \phi(g h_\infty)$$

is holomorphic. [Note that this function is well-defined by the first condition, as U_∞ is the stabilizer of $(i, \ldots, i)$.]

(4) If $S(D) = \varnothing$ then for all $g \in G_D(\mathbb{A}) = \mathrm{GL}_2(\mathbb{A}_F)$, we have

$$\int_{F \backslash \mathbb{A}_F} \varphi\left(\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} g \right) dx = 0.$$

If in addition we have $F = \mathbb{Q}$, then we furthermore demand that for all $g \in G_D(\mathbb{A}^\infty)$, $h_\infty \in \mathrm{GL}_2(\mathbb{R})^+$ the function $\varphi(gh_\infty)|\mathrm{Im}(h_\infty i)|^{k/2}$ is bounded on $\mathbb{C} - \mathbb{R}$.

There is a natural action of $G_D(\mathbb{A}^\infty)$ on $S_{D,k,\eta}$ by right-translation, i.e., $(g\varphi)(x) := \varphi(xg)$.

Exercise 4.9. While this definition may at first sight appear rather mysterious, it is just a generalization of the familiar spaces of cuspidal modular forms. For example, take $F = \mathbb{Q}$, $S(D) = \varnothing$, $k_\infty = k$, and $\eta_\infty = 0$. Define

$$U_1(N) = \left\{ g \in \mathrm{GL}_2(\hat{\mathbb{Z}}) \,\middle|\, g \equiv \begin{pmatrix} * & * \\ 0 & 1 \end{pmatrix} \pmod{N} \right\}.$$

(1) Let $\mathrm{GL}_2(\mathbb{Q})^+$ be the subgroup of $\mathrm{GL}_2(\mathbb{Q})$ consisting of matrices with positive determinant. Show that the intersection of $\mathrm{GL}_2(\mathbb{Q})^+$ and $U_1(N)$ inside $\mathrm{GL}_2(\mathbb{A}^\infty)$ is $\Gamma_1(N)$, the matrices in $\mathrm{SL}_2(\mathbb{Z})$ congruent to $\left(\begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix}\right) \pmod{N}$. [Hint: what is $\hat{\mathbb{Z}}^\times \cap \mathbb{Q}^\times$?]

(2) Use the facts that $\mathrm{GL}_2(\mathbb{A}) = \mathrm{GL}_2(\mathbb{Q}) U_1(N) \mathrm{GL}_2(\mathbb{R})^+$ [which follows from strong approximation for SL_2 and the fact that $\det U_1(N) = \hat{\mathbb{Z}}^\times$] and that $\mathbb{A}^\times = \mathbb{Q}^\times \hat{\mathbb{Z}}^\times \mathbb{R}^\times_{>0}$ to show that $S^{U_1(N)}_{D,k,0}$ can naturally be identified with a space of functions

$$\varphi : \Gamma_1(N) \backslash \mathrm{GL}_2(\mathbb{R})^+ \to \mathbb{C}$$

satisfying

$$\varphi(gu_\infty) = j(u_\infty, i)^{-k} \varphi(g)$$

for all $g \in \mathrm{GL}_2(\mathbb{R})^+$, $u_\infty \in \mathbb{R}^\times_{>0} \mathrm{SO}(2)$.

(3) Show that the stabilizer of i in $\mathrm{GL}_2(\mathbb{R})^+$ is $\mathbb{R}^\times_{>0} \mathrm{SO}(2)$. Hence deduce a natural isomorphism between $S^{U_1(N)}_{D,k,0}$ and $S_k(\Gamma_1(N))$, which takes a function φ as above to the function $(gi \mapsto j(g,i)^k \varphi(g))$, $g \in \mathrm{GL}_2(\mathbb{R})^+$.

The case that $S_\infty \subset S(D)$ is particularly simple; then if $U \subset G_D(\mathbb{A}^\infty)$ is an open subgroup, then $S^U_{D,2,0}$ is just the set of $\mathbb{C}$-valued functions on

$$G_D(\mathbb{Q}) \backslash G_D(\mathbb{A}) / G_D(\mathbb{R}) U,$$

which is a finite set. When proving modularity lifting theorems, we will be able to reduce to the case that $S_\infty \subset S(D)$; when this condition holds, we say that D is a *definite* quaternion algebra.

We will now examine the action of Hecke operators on these spaces. Choose an $\mathcal{O}_F$-order $\mathcal{O}_D \subset D$ (that is, an $\mathcal{O}_F$-subalgebra of D which is finitely generated as a $\mathbb{Z}$-module and for which $\mathcal{O}_D \otimes_{\mathcal{O}_F} F \xrightarrow{\sim} D$). For example, if $D = M_2(F)$, one may take $\mathcal{O}_D = M_2(\mathcal{O}_F)$.

For all but finite many finite places v of F we can choose an isomorphism $D_v \cong M_2(F_v)$ such that this isomorphism induces an isomorphism $\mathcal{O}_D \otimes_{\mathcal{O}_F} \mathcal{O}_{F_v} \xrightarrow{\sim} M_2(\mathcal{O}_{F_v})$. Then $G_D(\mathbb{A}^\infty)$ is the subset of elements $g = (g_v) \in \prod_{v \nmid \infty} G_D(F_v)$ such that $g_v \in \mathrm{GL}_2(\mathcal{O}_{F_v})$ for almost all v.

We now wish to describe certain irreducible representations of $G_D(\mathbb{A}^\infty)$ in terms of irreducible representations of the $GL_2(F_v)$. More generally, we have the following construction. Let I be an indexing set and for each $i \in I$, let V_i be a $\mathbb{C}$-vector space. Suppose that we are given $0 \neq e_i \in V_i$ for almost all i (that is, all but finitely many i). Then we define the *restricted tensor product*

$$\otimes'_{\{e_i\}} V_i := \varinjlim_{J \subseteq I} \otimes_{i \in J} V_i,$$

where the colimit is over the finite subsets $J \subseteq I$ containing all the places for which e_i is not defined, and where the transition maps for the colimit are given by "tensoring with the e_i". It can be checked that $\otimes'_{\{e_i\}} V_i \cong \otimes'_{\{f_i\}} V_i$ if for almost all i, e_i and f_i span the same line.

Definition 4.10. We call a representation (π, V) of $G_D(\mathbb{A}^\infty)$ *admissible* if

(1) for any $x \in V$, the stabilizer of x is open, and

(2) for any $U \subset G_D(\mathbb{A}^\infty)$ an open subgroup, $\dim_{\mathbb{C}} V^U < \infty$.

Fact 4.11 [Flath 1979]. If π_v is an irreducible smooth (so admissible) representation of $(D \otimes_F F_v)^\times$ with $\pi_v^{\mathrm{GL}_2(\mathcal{O}_{F_v})} \neq 0$ for almost all v, then $\otimes' \pi_v := \otimes'_{\{\pi_v^{\mathrm{GL}_2(\mathcal{O}_{F_v})}\}} \pi_v$ is an irreducible admissible smooth representation of $G_D(\mathbb{A}^\infty)$, and any irreducible admissible smooth representation of $G_D(\mathbb{A}^\infty)$ arises in this way for unique π_v.

We have a global Hecke algebra, which decomposes as a restricted tensor product of the local Hecke algebras in the following way. For each finite place v of F we choose $U_v \subset (D \otimes_F F_v)^\times$ a compact open subgroup, such that $U_v = \mathrm{GL}_2(\mathcal{O}_{F_v})$ for almost all v. Let μ_v be a Haar measure on $(D \otimes_F F_v)^\times$, chosen such that for almost all v we have $\mu_v(\mathrm{GL}_2(\mathcal{O}_{F_v})) = 1$. Then there is a unique Haar measure μ on $G_D(\mathbb{A}^\infty)$ such that for any U_v as above, if we set $U = \prod_v U_v \subset G_D(\mathbb{A}^\infty)$, then $\mu(U) = \prod_v \mu_v(U_v)$. Then there is a decomposition

$$\mathcal{C}_c(U \backslash G_D(\mathbb{A}^\infty)/U)\mu \cong \otimes'_{\{1_{U_v}\mu_v\}} \mathcal{C}_c(U_v \backslash (D \otimes_F F_v)^\times / U_v)\mu_v,$$

and the actions of these Hecke algebras are compatible with the decomposition $\pi = \otimes' \pi_v$. For the following fact, see Lemma 1.3 of [Taylor 2006].

Fact 4.12. $S_{D,k,\eta}$ is a semisimple admissible representation of $G_D(\mathbb{A}^\infty)$.

Definition 4.13. The irreducible constituents of $S_{D,k,\eta}$ are called the *cuspidal automorphic representations* of $G_D(\mathbb{A}^\infty)$ of weight (k, η).

Remark 4.14. Note that these automorphic representations do not include Maass forms or weight one modular forms; they are the class of *regular algebraic* or *cohomological* cuspidal automorphic representations.

For the following facts, the reader could consult [Gelbart 1975].

Fact 4.15 (strong multiplicity one (and multiplicity one) for GL_2). Suppose that $S(D)=\varnothing$. Then every irreducible constituent of $S_{D,k,\eta}$ has multiplicity one. In fact if π (respectively π') is a cuspidal automorphic representation of weight (k,η) (respectively (k',η')) such that $\pi_v\cong\pi'_v$ for almost all v then $k=k'$, $\eta=\eta'$, and $\pi=\pi'$.

Fact 4.16 (the theory of newforms). Suppose that $S(D)=\varnothing$. If $\mathfrak{n}$ is an ideal of $\mathcal{O}_F$, write

$$U_1(\mathfrak{n})=\left\{g\in\mathrm{GL}_2(\hat{\mathcal{O}}_F)\;\middle|\;g\equiv\begin{pmatrix}* & *\\ 0 & 1\end{pmatrix}\pmod{\mathfrak{n}}\right\}.$$

If π is a cuspidal automorphic representation of $G_D(\mathbb{A}^\infty)$ then there is a unique ideal $\mathfrak{n}$ such that $\pi^{U_1(\mathfrak{n})}$ is one-dimensional, and $\pi^{U_1(\mathfrak{m})}\neq 0$ if and only if $\mathfrak{n}\mid\mathfrak{m}$. We call $\mathfrak{n}$ the *conductor* (or sometimes the *level*) of π.

Analogous to the theory of admissible representations of $\mathrm{GL}_2(K)$, $K/\mathbb{Q}_p$ finite that we sketched above, there is a theory of admissible representations of $M^\times$, M a nonsplit quaternion algebra over K. Since $M^\times/K^\times$ is compact, any irreducible smooth representation of $M^\times$ is finite-dimensional. There is a bijection JL, the *local Jacquet–Langlands correspondence*, from the irreducible smooth representations of $M^\times$ to the discrete series representations of $\mathrm{GL}_2(K)$, determined by a character identity.

Fact 4.17 (the global Jacquet–Langlands correspondence). We have the following facts about $G_D(\mathbb{A}^\infty)$:

(1) The only finite-dimensional cuspidal automorphic representations of $G_D(\mathbb{A}^\infty)$ are 1-dimensional representations which factor through the reduced norm; these only exist if $D\neq M_2(F)$.

(2) There is a bijection JL from the infinite-dimensional cuspidal automorphic representations of $G_D(\mathbb{A}^\infty)$ of weight (k,η) to the cuspidal automorphic representations of $\mathrm{GL}_2(\mathbb{A}_F^\infty)$ of weight (k,η) which are discrete series for all finite places $v\in S(D)$. Furthermore if $v\notin S(D)$ then $\mathrm{JL}(\pi)_v=\pi_v$, and if $v\in S(D)$ then $\mathrm{JL}(\pi)_v=\mathrm{JL}(\pi_v)$.

Remark 4.18. We will use the global Jacquet–Langlands correspondence together with base change (see below) to reduce ourselves to considering the case that $S(D)=S_\infty$ when proving automorphy lifting theorems.

4.19. ***Galois representations associated to automorphic representations.***

Fact 4.20 (the existence of Galois representations associated to regular algebraic cuspidal automorphic representations). Let π be a regular algebraic cuspidal automorphic representation of $\mathrm{GL}_2(\mathbb{A}_F^\infty)$ of weight (k,η). Then there is a CM field L_π which contains the eigenvalues of T_v and S_v on $\pi_v^{\mathrm{GL}_2(\mathcal{O}_{F_v})}$ for each finite place v at which π_v is unramified. Furthermore, for each finite place λ of L_π there is a continuous irreducible Galois representation

$$r_\lambda(\pi): G_F \to \mathrm{GL}_2(\bar{L}_{\pi,\lambda})$$

such that:

(1) If π_v is unramified and v does not divide the residue characteristic of λ, then $r_\lambda(\pi)|_{G_{F_v}}$ is unramified, and the characteristic polynomial of Frob_v is $X^2 - t_vX + (\#k(v))s_v$, where t_v and s_v are the eigenvalues of T_v and S_v respectively on $\pi_v^{\mathrm{GL}_2(\mathcal{O}_{F_v})}$, and $k(v)$ is the residue field of F_v. [Note that by the Chebotarev density theorem, this already characterizes $r_\lambda(\pi)$ up to isomorphism.]

(2) More generally, for all finite places v not dividing the residue characteristic of λ, $\mathrm{WD}(r_\lambda(\pi)|_{G_{F_v}})^{F-\mathrm{ss}} \cong \mathrm{rec}_{F_v}(\pi_v \otimes |\det|^{-1/2})$.

(3) If v divides the residue characteristic of λ then $r_\lambda(\pi)|_{G_{F_v}}$ is de Rham with τ-Hodge–Tate weights $\eta_\tau, \eta_\tau + k_\tau - 1$, where $\tau: F \hookrightarrow \bar{L}_\pi \subset \mathbb{C}$ is an embedding lying over v. If π_v is unramified then $r_\lambda(\pi)|_{G_{F_v}}$ is crystalline.

(4) If c_v is a complex conjugation, then $\det r_\lambda(\pi)(c_v) = -1$.

Remark 4.21. The representations $r_\lambda(\pi)$ in fact form a strictly compatible system; see Section 5 of [Barnet-Lamb et al. 2014] for a discussion of this in a more general context.

Remark 4.22. Using the Jacquet–Langlands correspondence, we get Galois representations for the infinite-dimensional cuspidal automorphic representations of $G_D(\mathbb{A}^\infty)$ for any D. In fact, the proof actually uses the Jacquet–Langlands correspondence; in most cases, you can transfer to a D for which $S(D)$ contains all but one infinite place, and the Galois representations are then realized in the étale cohomology of the associated Shimura curve. The remaining Galois representations are constructed from these ones via congruences.

Fact 4.23 (cyclic base change). Let E/F be a cyclic extension of totally real fields of prime degree. Let $\mathrm{Gal}(E/F) = \langle\sigma\rangle$ and let $\mathrm{Gal}(E/F)^\vee = \langle\delta_{E/F}\rangle$ (here $\mathrm{Gal}(E/F)^\vee$ is the dual abelian group of $\mathrm{Gal}(E/F)$). Let π be a cuspidal automorphic representation of $\mathrm{GL}_2(\mathbb{A}_F^\infty)$ of weight (k,η). Then there is a cuspidal automorphic representation $\mathrm{BC}_{E/F}(\pi)$ of $\mathrm{GL}_2(\mathbb{A}_E^\infty)$ of weight $(\mathrm{BC}_{E/F}(k), \mathrm{BC}_{E/F}(\eta))$ such that:

(1) For all finite places v of E, $\mathrm{rec}_{E_v}(\mathrm{BC}_{E/F}(\pi)_v) = (\mathrm{rec}_{F_{v|F}}(\pi_{v|F}))|_{W_{E_v}}$. In particular, $r_\lambda(BC_{E/F}(\pi)) \cong r_\lambda(\pi)|_{G_E}$.

(2) $\mathrm{BC}_{E/F}(k)_v = k_{v|_F}$, $\mathrm{BC}_{E/F}(\eta)_v = \eta_{v|_F}$.

(3) $\mathrm{BC}_{E/F}(\pi) \cong \mathrm{BC}_{E/F}(\pi')$ if and only if $\pi \cong \pi' \otimes (\delta^i_{E/F} \circ \mathrm{Art}_F \circ \det)$ for some i.

(4) A cuspidal automorphic representation π of $\mathrm{GL}_2(\mathbb{A}_E^\infty)$ is in the image of $\mathrm{BC}_{E/F}$ if and only if $\pi \circ \sigma \cong \pi$.

Definition 4.24. We say that $r : G_F \to \mathrm{GL}_2(\overline{\mathbb{Q}}_p)$ is *modular* (of weight (k, η)) if it is isomorphic to $r_\lambda(\pi)$ for some cuspidal automorphic representation π (of weight (k, η)) and some place λ of L_π lying over p.

Proposition 4.25. *Suppose that $r : G_F \to \mathrm{GL}_2(\overline{\mathbb{Q}}_p)$ is a continuous representation, and that E/F is a finite solvable Galois extension of totally real fields. Then $r|_{G_E}$ is modular if and only if r is modular.*

Exercise 4.26. Prove the above proposition as follows:

(1) Use induction to reduce to the case that E/F is cyclic of prime degree.

(2) Suppose that $r|_{G_E}$ is modular, say $r|_{G_E} \cong r_\lambda(\pi)$. Use strong multiplicity one to show that $\pi \circ \sigma \cong \pi$. Deduce that there is an automorphic representation π' such that $\mathrm{BC}_{E/F}(\pi') = \pi$.

(3) Use Schur's lemma to deduce that there is a character χ of G_F such that $r \cong r_\lambda(\pi') \otimes \chi$. Conclude that r is modular.

We can make use of this result to make considerable simplifications in our proofs of modularity lifting theorems. It is frequently employed in conjunction with the following fact from class field theory.

Fact 4.27 [Taylor 2003, Lemma 2.2]. Let K be a number field, and let S be a finite set of places of K. For each $v \in S$, let L_v be a finite Galois extension of K_v. Then there is a finite solvable Galois extension M/K such that for each place w of M above a place $v \in S$ there is an isomorphism $L_v \cong M_w$ of K_v-algebras.

Note that we are allowed to have infinite places in S, so that if K is totally real we may choose to make L totally real by an appropriate choice of the L_v.

5. The Taylor–Wiles–Kisin method

In this section we prove our modularity lifting theorem, using the Taylor–Wiles–Kisin patching method. Very roughly, the idea of this method is to patch together spaces of modular forms of varying levels, allowing more and more ramification at places away from p, in such a way as to "smooth out" the singularities of global deformation rings, reducing the problem to one about local deformation rings. This patching procedure is (at least on first acquaintance) somewhat strange, as it involves making many noncanonical choices to identify spaces of modular forms with level structures at different primes.

5.1. Our aim now is to prove the following theorem. Let $p > 3$ be a prime, and let $L/\mathbb{Q}_p$ be a finite extension with ring of integers $\mathcal{O}$, maximal ideal λ, and residue field $\mathbb{F} = \mathcal{O}/\lambda$. Let F be a totally real number field, and assume that L is sufficiently large that L contains the images of all embeddings $F \hookrightarrow \bar{L}$.

Theorem 5.2. *Let $\rho, \rho_0 : G_F \to \mathrm{GL}_2(\mathcal{O})$ be two continuous representations, such that $\bar{\rho} = \rho \pmod{\lambda} = \rho_0 \pmod{\lambda}$. Assume that ρ_0 is modular, that ρ is geometric, and that $p > 3$. Assume further that the following properties hold:*

(1) *For all $\sigma : F \hookrightarrow L$, $\mathrm{HT}_\sigma(\rho) = \mathrm{HT}_\sigma(\rho_0)$, and contains two distinct elements.*

(2)
- *For all $v \mid p$, $\rho|_{G_{F_v}}$ and $\rho_0|_{G_{F_v}}$ are crystalline.*
- *p is unramified in F.*
- *For all $\sigma : F \hookrightarrow L$, the elements of $\mathrm{HT}_\sigma(\rho)$ differ by at most $p-2$.*

(3) $\operatorname{Im} \bar{\rho} \supseteq \mathrm{SL}_2(\mathbb{F}_p)$.

Then ρ is modular.

5.3. ***The integral theory of automorphic forms.*** In order to prove Theorem 5.2, we will need to study congruences between automorphic forms. This is easier to do if we work with automorphic forms on $G_D(\mathbb{A}^\infty)$, where $S(D) = S_\infty$. In order to do this, assume that $[F : \mathbb{Q}]$ is even. (We will reduce to this case by base change.) Then such a D exists, and we have $G_D(\mathbb{A}^\infty) \cong \mathrm{GL}_2(\mathbb{A}_F^\infty)$, and $(D \otimes_\mathbb{Q} \mathbb{R})^\times/(F \otimes_\mathbb{Q} \mathbb{R})^\times$ is compact.

Fix an isomorphism $\imath : \bar{L} \xrightarrow{\sim} \mathbb{C}$, and some $k \in \mathbb{Z}_{\geq 2}^{\mathrm{Hom}(F,\mathbb{C})}$, $\eta \in \mathbb{Z}^{\mathrm{Hom}(F,\mathbb{C})}$ with $w := k_\tau + 2\eta_\tau - 1$ independent of τ. Let $U = \prod_v U_v \subset \mathrm{GL}_2(\mathbb{A}_F^\infty)$ be a compact open subgroup, and let S be a finite set of finite places of F, not containing any of the places lying over p, with the property that if $v \notin S$, then $U_v = \mathrm{GL}_2(\mathcal{O}_{F_v})$.

Let $U_S := \prod_{v \in S} U_v$, write $U = U_S U^S$, let $\psi : U_S \to \mathcal{O}^\times$ be a continuous homomorphism (which implies that it has open kernel), and let $\chi_0 : \mathbb{A}_F^\times/F^\times \to \mathbb{C}^\times$ be an algebraic Grössencharacter with the properties that

- χ_0 is unramified outside S,
- for each place $v \mid \infty$, $\chi_0|_{(F_v^\times)^\circ}(x) = x^{1-w}$, and
- $\chi_0|_{\left(\prod_{v \in S} F_v^\times\right) \cap U_S} = \imath \circ \psi^{-1}$.

As in Theorem 2.43, this gives us a character

$$\chi_{0,\imath} : \mathbb{A}_F^\times/\overline{F^\times (F_\infty^\times)^\circ} \to \bar{L}^\times,$$

$$x \mapsto \Big(\prod_{\tau : F \hookrightarrow L} \tau(x_p)^{1-w}\Big)\imath^{-1}\Big(\prod_{\tau : F \hookrightarrow \mathbb{C}} \tau(x_\infty)\Big)^{w-1} \chi_0(x).$$

Our spaces of (p-adic) algebraic automorphic forms will be defined in a similar way to the more classical spaces defined in Section 4.8, but with the role of the infinite places being played by the places lying over p. Accordingly, we define coefficient systems in the following way. Assume that L is sufficiently large that it contains the image of $\chi_{0,\iota}$.

Let $\Lambda = \Lambda_{k,\eta,\iota} = \otimes_{\tau:F\hookrightarrow\mathbb{C}} \mathrm{Sym}^{k_\tau-2}(\mathcal{O}^2)\otimes(\wedge^2\mathcal{O}^2)^{\otimes\eta_\tau}$, and let $\mathrm{GL}_2(\mathcal{O}_{F,p}) := \prod_{v\mid p}\mathrm{GL}_2(\mathcal{O}_{F_v})$ act on Λ via $\iota^{-1}\tau$ on the τ-factor. In particular, $\Lambda\otimes_{\mathcal{O},\iota}\mathbb{C}\cong \otimes_{\tau:F\hookrightarrow\mathbb{C}}\mathrm{Sym}^{k_\tau-2}(\mathbb{C}^2)\otimes(\wedge^2\mathbb{C}^2)^{\otimes\eta_\tau}$, which has an obvious action of $\mathrm{GL}_2(F_\infty)$, and the two actions of $\mathrm{GL}_2(\mathcal{O}_{F,(p)})$ (via its embeddings into $\mathrm{GL}_2(\mathcal{O}_{F,p})$ and $\mathrm{GL}_2(F_\infty)$) are compatible.

Let A be a finite $\mathcal{O}$-module. Since D is fixed, we drop it from the notation from now on. We define $S(U,A) = S_{k,\eta,\iota,\psi,\chi_0}(U,A)$ to be the space of functions

$$\phi : D^\times\backslash\,\mathrm{GL}_2(\mathbb{A}_F^\infty)\to\Lambda\otimes_{\mathcal{O}}A$$

such that for all $g\in\mathrm{GL}_2(\mathbb{A}_F^\infty)$, $u\in U$, $z\in(\mathbb{A}_F^\infty)^\times$, we have

$$\phi(guz) = \chi_{0,\iota}(z)\psi(u_S)^{-1}u_p^{-1}\phi(g).$$

Since $D^\times\backslash\,\mathrm{GL}_2(\mathbb{A}_F^\infty)/U(\mathbb{A}_F^\infty)^\times$ is finite, we see in particular that $S(U,\mathcal{O})$ is a finite free $\mathcal{O}$-module. It has a Hecke action in the obvious way: let $\tilde{\mathbb{T}} := \mathcal{O}[T_v, S_v : v\nmid p, v\notin S]$, let ϖ_v be a uniformizer of F_v, and let T_v, S_v act via the usual double coset operators corresponding to $\left(\begin{smallmatrix}\varpi_v & 0\\ 0 & 1\end{smallmatrix}\right)$, $\left(\begin{smallmatrix}\varpi_v & 0\\ 0 & \varpi_v\end{smallmatrix}\right)$. Let $\mathbb{T}_U$ be the image of $\tilde{\mathbb{T}}$ in $\mathrm{End}_{\mathcal{O}}(S(U,\mathcal{O}))$, so that $\mathbb{T}_U$ is a commutative $\mathcal{O}$-algebra which acts faithfully on $S(U,\mathcal{O})$, and is finite free as an $\mathcal{O}$-module.

As in [Taylor 2006, Lemma 1.3], to which we refer for more details, there is an isomorphism

$$S(U,\mathcal{O})\otimes_{\mathcal{O},\iota}\mathbb{C}\xrightarrow{\sim}\mathrm{Hom}_{U_S}(\mathbb{C}(\psi^{-1}), S_{k,\eta}^{U^S,\chi_0}),$$

with the map being

$$\phi\mapsto(g\mapsto g_\infty^{-1}\iota(g_p\phi(g^\infty))),$$

where g_p acts on $\Lambda\otimes_{\mathcal{O},\iota}\mathbb{C}$ via the obvious extension of the action of $\mathrm{GL}_2(\mathcal{O}_{F,(p)})$ defined above, and the target of the isomorphism is the elements $\phi'\in S_{k,\eta}$ with $z\phi' = \chi_0(z)\phi'$ for all $z\in(\mathbb{A}_F^\infty)^\times$, $u\phi' = \psi(u_S)^{-1}\phi'$ for all $u\in U$. This isomorphism is compatible with the actions of $\tilde{\mathbb{T}}$ on each side. The target is isomorphic to

$$\oplus_\pi\,\mathrm{Hom}_{U_S}(\mathbb{C}(\psi^{-1}),\pi_S)\otimes\otimes'_{v\notin S}\pi_v^{\mathrm{GL}_2(\mathcal{O}_{F_v})},$$

where the sum is over the cuspidal automorphic representations π of $G_D(\mathbb{A}^\infty)$ of weight (k,η), which have central character χ_0 and are unramified outside of S (so that in particular, for $v\notin S$, $\pi_v^{\mathrm{GL}_2(\mathcal{O}_{F_v})}$ is a one-dimensional $\mathbb{C}$-vector space).

By strong multiplicity one, this means that we have an isomorphism

$$\mathbb{T}_U \otimes_{\mathcal{O},\iota} \mathbb{C} \cong \prod_{\pi \text{ as above, with } \mathrm{Hom}_{U_S}(\mathbb{C}(\psi^{-1}),\pi_S)\neq 0} \mathbb{C}$$

sending T_v, S_v to their eigenvalues on $\pi_v^{\mathrm{GL}_2(\mathcal{O}_{F_v})}$. (Note in particular that this shows that $\mathbb{T}_U$ is reduced.) This shows that there is a bijection between ι-linear ring homomorphisms $\theta : \mathbb{T}_U \to \mathbb{C}$ and the set of π as above, where π corresponds to the character taking T_v, S_v to their corresponding eigenvalues.

Each π has a corresponding Galois representation. Taking the product of these representations, we obtain a representation

$$\rho^{\mathrm{mod}} : G_F \to \prod_{\pi} \mathrm{GL}_2(\bar{L}) = \mathrm{GL}_2(\mathbb{T}_U \otimes_{\mathcal{O}} \bar{L}),$$

which is characterized by the properties that it is unramified outside of $S \cup \{v \mid p\}$, and for any $v \notin S$, $v \nmid p$, we have $\operatorname{tr} \rho^{\mathrm{mod}}(\mathrm{Frob}_v) = T_v$, $\det \rho^{\mathrm{mod}}(\mathrm{Frob}_v) = \#k(v) S_v$.

Let $\mathfrak{m}$ be a maximal ideal of $\mathbb{T}_U$. Then if $\mathfrak{p} \subsetneq \mathfrak{m}$ is a minimal prime, then there is an injection $\theta : \mathbb{T}_U/\mathfrak{p} \hookrightarrow \bar{L}$, which corresponds to some π as above. (This follows from the going-up and going-down theorems, and the fact that $\mathbb{T}_U$ is finitely generated and free over $\mathcal{O}$.) The semisimple mod p Galois representation corresponding to π can be conjugated to give a representation $\bar{\rho}_{\mathfrak{m}} : G_F \to \mathrm{GL}_2(\mathbb{T}_U/\mathfrak{m})$ (because the trace and determinant are valued in $\mathbb{T}_U/\mathfrak{m}$, which is a finite field, and thus has trivial Brauer group, so the Schur index is trivial). This is well defined (up to isomorphism) independently of the choice of $\mathfrak{p}$ and θ (by the Chebotarev density theorem).

Since $\mathbb{T}_U$ is finite over the complete local ring $\mathcal{O}$, it is semilocal, and we can write $\mathbb{T}_U = \prod_{\mathfrak{m}} \mathbb{T}_{U,\mathfrak{m}}$. Suppose now that $\bar{\rho}_{\mathfrak{m}}$ is absolutely irreducible. Then we have the representation

$$\rho_{\mathfrak{m}}^{\mathrm{mod}} : G_F \to \mathrm{GL}_2(\mathbb{T}_{U,\mathfrak{m}} \otimes_{\mathcal{O}} \bar{L}) = \prod_{\pi} \mathrm{GL}_2(\bar{L}),$$

where the product is over the π as above with $\bar{\rho}_{\pi,\iota} \cong \bar{\rho}_{\mathfrak{m}}$. Each representation to $\mathrm{GL}_2(\bar{L})$ can be conjugated to lie in $\mathrm{GL}_2(\mathcal{O}_{\bar{L}})$, and after further conjugation (so that the residual representations are equal to $\bar{\rho}_{\mathfrak{m}}$, rather than just conjugate to it), the image of $\rho_{\mathfrak{m}}^{\mathrm{mod}}$ lies in the subring of $\prod_{\pi} \mathrm{GL}_2(\mathcal{O}_{\bar{L}})$ consisting of elements whose image modulo the maximal ideal of $\mathcal{O}_{\bar{L}}$ lie in $\mathbb{T}_U/\mathfrak{m}$. We can then apply Lemma 3.7 to see that $\rho_{\mathfrak{m}}^{\mathrm{mod}}$ can be conjugated to lie in $\mathrm{GL}_2(\mathbb{T}_{U,\mathfrak{m}})$. We will write $\rho_{\mathfrak{m}}^{\mathrm{mod}} : G_F \to \mathrm{GL}_2(\mathbb{T}_{U,\mathfrak{m}})$ for the resulting representation from now on.

We will sometimes want to consider Hecke operators at places in S. To this end, let $T \subseteq S$ satisfy $\psi|_{U_T} = 1$, and choose $g_v \in \mathrm{GL}_2(F_v)$ for each $v \in T$. Set $W_v = [U_v g_v U_v]$, and define $\mathbb{T}_U \subseteq \mathbb{T}'_U \subseteq \mathrm{End}_{\mathcal{O}}(S(U, \mathcal{O}))$ by adjoining the W_v for

$v \in T$. This is again commutative, and finite and flat over $\mathcal{O}$. However, it need not be reduced; indeed, we have

$$\mathbb{T}'_U \otimes_{\mathcal{O},\iota} \mathbb{C} \cong \oplus_\pi \otimes_{v \in T} \{ \text{ subalgebra of } \operatorname{End}_{\mathbb{C}}(\pi_v^{U_v}) \text{ generated by } W_v\},$$

so that there is a bijection between ι-linear homomorphisms $\mathbb{T}'_U \to \mathbb{C}$ and tuples $(\pi, \{\alpha_v\}_{v\in T})$, where α_v is an eigenvalue of W_v on $\pi_v^{U_v}$. (Note that we will not explicitly use the notation $\mathbb{T}'_U$ again for a Hecke algebra, but that for example the Hecke algebras $\mathbb{T}_{U_Q}$ used in the patching argument below, which incorporate Hecke operators at the places in Q, are an example of this construction.)

We can write

$$\mathrm{GL}_2(\mathbb{A}_F^\infty) = \coprod_{i\in I} D^\times g_i U (\mathbb{A}_F^\infty)^\times$$

for some finite indexing set I, and so we have an injection $S(U, A) \hookrightarrow \oplus_{i\in I}(\Lambda \otimes_{\mathcal{O}} A)$, by sending $\phi \mapsto (\phi(g_i))$. To determine the image, we need to consider when we can have $g_i = \delta g_i u z$ for $\delta \in D^\times$, $z \in (\mathbb{A}_F^\infty)^\times$, $u \in U$ (because then $\phi(g_i) = \phi(\delta g_i u z) = \chi_{0,\iota}(z)\psi(u_S)^{-1}u_p^{-1}\phi(g_i)$). We see in this way that we obtain an isomorphism

$$S(U, A) \xrightarrow{\sim} \oplus_{i\in I}(\Lambda \otimes A)^{(U(\mathbb{A}_F^\infty)^\times \cap g_i^{-1} D^\times g_i)/F^\times}.$$

We need to have some control on these finite groups

$$G_i := (U(\mathbb{A}_F^\infty)^\times \cap g_i^{-1} D^\times g_i)/F^\times.$$

(Note that they are finite, because $D^\times$ is discrete in $G_D(\mathbb{A}^\infty)$.) Since we have assumed that $p > 3$ and p is unramified in F, we see that $[F(\zeta_p) : F] > 2$. Then we claim that G_i has order prime to p. To see this, note that if $g_i^{-1}\delta g_i$ is in this group, with $\delta \in D^\times$, then $\delta^2/\det\delta \in D^\times \cap g_i U g_i^{-1}(\det U)$, the intersection of a discrete set and a compact set, so $\delta^2/\det\delta$ has finite order, i.e., is a root of unity. However any element of D generates an extension of F of degree at most 2, so by the assumption that $[F(\zeta_p) : F] > 2$, it must be a root of unity of degree prime to p, and there is some $p \nmid N$ with $\delta^{2N} \in F^\times$, so that $g_i^{-1}\delta g_i$ has order prime to p, as required.

Proposition 5.4. (1) *We have* $S(U, \mathcal{O}) \otimes_{\mathcal{O}} A \xrightarrow{\sim} S(U, A)$.

(2) *If* V *is an open normal subgroup of* U *with* $\#(U/V)$ *a power of* p*, then* $S(V, \mathcal{O})$ *is a free* $\mathcal{O}[U/V(U \cap (\mathbb{A}_F^\infty)^\times)]$*-module.*

Proof. (1) This is immediate from the isomorphism $S(U, A) \xrightarrow{\sim} \oplus_{i\in I}(\Lambda \otimes A)^{G_i}$, because the fact that the G_i have order prime to p means that $(\Lambda \otimes A)^{G_i} = (\Lambda)^{G_i} \otimes A$.

(2) Write $U = \coprod_{j\in J} u_j V(U \cap (\mathbb{A}_F^\infty)^\times)$. We claim that we have $\mathrm{GL}_2(\mathbb{A}_F^\infty) = \coprod_{i\in I, j\in J} D^\times g_i u_j V(\mathbb{A}_F^\infty)^\times$, from which the result is immediate. To see this, we need to show that if $g_i u_j = \delta g_{i'} u_{j'} v z$ then $i = i'$ and $j = j'$.

That $i = i'$ is immediate from the definition of I, so we have $u_{j'}vu_j^{-1}z = g_i^{-1}\delta^{-1}g_i$. As above, there is some positive integer N coprime to p such that $\delta^N \in F^\times$, thus $(u_{j'}vu_j^{-1})^N \in (\mathbb{A}_F^\infty)^\times$. Since V is normal in U, we can write $(u_{j'}vu_j^{-1})^N = (u_{j'}u_j^{-1})^N v'$ for some $v' \in V$, so that $(u_{j'}u_j^{-1})^N \in V(U \cap (\mathbb{A}_F^\infty)^\times)$. Since $\#(U/V)$ is a power of p, we see that in fact $u_{j'}u_j^{-1} \in V(U \cap (\mathbb{A}_F^\infty)^\times)$, so that $j = j'$ by the definition of J. □

5.5. *Base change.* We begin the proof of Theorem 5.2 by using base change to reduce to a special case. By Facts 4.23 and 4.27, we can replace F by a solvable totally real extension which is unramified at all primes above p, and assume that:

- $[F : \mathbb{Q}]$ is even.
- $\bar{\rho}$ is unramified outside p.
- For all places $v \nmid p$, both $\rho(I_{F_v})$ and $\rho_0(I_{F_v})$ are unipotent (possibly trivial).
- If ρ or ρ_0 are ramified at some place $v \nmid p$, then $\bar{\rho}|_{G_{F_v}}$ is trivial, and $\#k(v) \equiv 1 \pmod{p}$.
- $\det \rho = \det \rho_0$. [To see that we can assume this, note that the assumption that ρ, ρ_0 are crystalline with the same Hodge–Tate weights for all places dividing p implies that $\det \rho / \det \rho_0$ is unramified at all places dividing p. Since we have already assumed that $\rho(I_{F_v})$ and $\rho_0(I_{F_v})$ are unipotent for all places $v \nmid p$, we see that the character $\det \rho / \det \rho_0$ is unramified at all places, and thus has finite order. Since it is residually trivial, it has p-power order, and is thus trivial on all complex conjugations; so the extension cut out by its kernel is a finite, abelian, totally real extension which is unramified at all places dividing p.]

We will assume from now on that all of these conditions hold. Write χ for $\det \rho = \det \rho_0$; then we have $\chi \varepsilon_p = \chi_{0,\iota}$ for some algebraic Grössencharacter χ_0.

From now on, we will assume without further comment that the coefficient field L is sufficiently large, in the sense that L contains a primitive p-th root of unity, and for all $g \in G_F$, $\mathbb{F}$ contains the eigenvalues of $\bar{\rho}(g)$.

5.6. *Patching.* Having used base change to impose the additional conditions of the previous section, we are now in a position to begin the main patching argument.

We let D/F be a quaternion algebra ramified at exactly the infinite places (which exists by our assumption that $[F : \mathbb{Q}]$ is even). By the Jacquet–Langlands correspondence, we can and will work with automorphic representations of $G_D(\mathbb{A}^\infty)$ from now on.

Let T_p be the set of places of F lying over p, let T_r be the set of places not lying over p at which ρ or ρ_0 is ramified, and let $T = T_p \coprod T_r$. If $v \in T_r$, write σ_v for a choice of topological generator of I_{F_v}/P_{F_v}. By our assumptions above, if $v \in T_r$ then $\bar{\rho}|_{G_{F_v}}$ is trivial, $\rho|_{I_{F_v}}$, $\rho_0|_{I_{F_v}}$ are unipotent, and $\#k(v) \equiv 1 \pmod{p}$.

The patching argument will involve the consideration of various finite sets Q of auxiliary finite places. We will always assume that if $v \in Q$, then

- $v \notin T$,
- $\#k(v) \equiv 1 \pmod p$, and
- $\bar{\rho}(\mathrm{Frob}_v)$ has distinct eigenvalues, which we denote $\bar{\alpha}_v$ and $\bar{\beta}_v$.

For each set Q of places satisfying these conditions, we define deformation problems $\mathcal{S}_Q = (T \cup Q, \{\mathcal{D}_v\}, \chi)$ and $\mathcal{S}'_Q = (T \cup Q, \{\mathcal{D}'_v\}, \chi)$ as follows. (The reason for considering both problems is that the objects without a prime are the ones that we ultimately wish to study, but the objects with a prime have the advantage that the ring $(R^{\mathrm{loc},'})^{\mathrm{red}}$ defined below is irreducible. We will exploit this irreducibility, and the fact that the two deformation problems agree modulo p.) Let ζ be a fixed primitive p-th root of unity in L:

- If $v \in T_p$, then $\mathcal{D}_v = \mathcal{D}'_v$ is chosen so that $R^{\square}_{\bar{\rho}|_{G_{F_v}},\chi}/I(\mathcal{D}_v) = R^{\square}_{\bar{\rho}|_{G_{F_v}},\chi,\mathrm{cr},\{\mathrm{HT}_\sigma(\rho)\}}$.
- If $v \in Q$, then $\mathcal{D}_v = \mathcal{D}'_v$ consists of all lifts of $\bar{\rho}|_{G_{F_v}}$ with determinant χ.
- If $v \in T_r$, then $\mathcal{D}_v$ consists of all lifts of $\bar{\rho}|_{G_{F_v}}$ with $\mathrm{char}_{\rho(\sigma_v)}(X) = (X-1)^2$, while $\mathcal{D}'_v$ consists of all lifts with $\mathrm{char}_{\rho(\sigma_v)}(X) = (X-\zeta)(X-\zeta^{-1})$.

(In particular, the difference between $\mathcal{S}_Q$ and $\mathcal{S}_\varnothing$ is that we have allowed our deformations to ramify at places in Q.) We write

$$R^{\mathrm{loc}} = \widehat{\otimes}_{v\in T,\mathcal{O}} R^{\square}_{\bar{\rho}|_{G_{F_v}},\chi}/I(\mathcal{D}_v), \qquad R^{\mathrm{loc},'} = \widehat{\otimes}_{v\in T,\mathcal{O}} R^{\square}_{\bar{\rho}|_{G_{F_v}},\chi}/I(\mathcal{D}'_v).$$

Then $R^{\mathrm{loc}}/\lambda = R^{\mathrm{loc},'}/\lambda$, because $\zeta \equiv 1 \pmod \lambda$. In addition, we see from Theorems 3.28 and 3.38 that

- $(R^{\mathrm{loc},'})^{\mathrm{red}}$ is irreducible, $\mathcal{O}$-flat, and has Krull dimension $1+3\#T+[F:\mathbb{Q}]$,
- $(R^{\mathrm{loc}})^{\mathrm{red}}$ is $\mathcal{O}$-flat, equidimensional of Krull dimension $1+3\#T+[F:\mathbb{Q}]$, and reduction modulo λ gives a bijection between the irreducible components of $\mathrm{Spec}\, R^{\mathrm{loc}}$ and those of $\mathrm{Spec}\, R^{\mathrm{loc}}/\lambda$.

We have the global analogues $R^{\mathrm{univ}}_Q := R^{\mathrm{univ}}_{\bar{\rho},\mathcal{S}_Q}$, $R^{\mathrm{univ},'}_Q := R^{\mathrm{univ}}_{\bar{\rho},\mathcal{S}'_Q}$, $R^{\square}_Q := R^{\square_T}_{\bar{\rho},\mathcal{S}_Q}$, $R^{\square,'}_Q := R^{\square_T}_{\bar{\rho},\mathcal{S}'_Q}$, and we have $R^{\mathrm{univ}}_Q/\lambda = R^{\mathrm{univ},'}_Q/\lambda$, $R^{\square}_Q/\lambda = R^{\square,'}_Q/\lambda$. There are obvious natural maps $R^{\mathrm{loc}} \to R^{\square}_Q$, $R^{\mathrm{loc},'} \to R^{\square,'}_Q$, and these maps agree after reduction $\bmod \lambda$.

We can and do fix representatives ρ^{univ}_Q, $\rho^{\mathrm{univ},'}_Q$ for the universal deformations of $\bar{\rho}$ over R^{univ}_Q, $R^{\mathrm{univ},'}_Q$ respectively, which are compatible with the choices of $\rho^{\mathrm{univ}}_\varnothing$, $\rho^{\mathrm{univ},'}_\varnothing$, and so that the induced surjections

$$R^{\mathrm{univ}}_Q \twoheadrightarrow R^{\mathrm{univ}}_\varnothing, \;\; R^{\mathrm{univ},'}_Q \twoheadrightarrow R^{\mathrm{univ},'}_\varnothing$$

are identified modulo λ.

Fix a place $v_0 \in T$, and set $\mathcal{J} := \mathcal{O}[[X_{v,i,j}]]_{v\in T, i,j=1,2}/(X_{v_0,1,1})$. Let $\mathfrak{a}$ be the ideal of $\mathcal{J}$ generated by the $X_{v,i,j}$. Then our choice of ρ_Q^{univ} gives an identification $R_Q^{\square} \xrightarrow{\sim} R_Q^{\mathrm{univ}} \widehat{\otimes}_{\mathcal{O}} \mathcal{J}$, corresponding to the universal T-framed deformation $(\rho_Q^{\mathrm{univ}}, \{1 + (X_{v,i,j})\}_{v\in T})$.

Now, by Exercise 3.34, for each place $v \in Q$ we have an isomorphism $\rho_Q^{\mathrm{univ}}|_{G_{F_v}} \cong \chi_\alpha \oplus \chi_\beta$, where $\chi_\alpha, \chi_\beta : G_{F_v} \to (R_Q^{\mathrm{univ}})^\times$, where $(\chi_\alpha \bmod \mathfrak{m}_{R_Q^{\mathrm{univ}}})(\mathrm{Frob}_v) = \bar{\alpha}_v$, $(\chi_\beta \bmod \mathfrak{m}_{R_Q^{\mathrm{univ}}})(\mathrm{Frob}_v) = \bar{\beta}_v$.

Let Δ_v be the maximal p-power quotient of $k(v)^\times$ (which we sometimes regard as a subgroup of $k(v)^\times$). Then $\chi_\alpha|_{I_{F_v}}$ factors through the composite

$$I_{F_v} \twoheadrightarrow I_{F_v}/P_{F_v} \twoheadrightarrow k(v)^\times \twoheadrightarrow \Delta_v,$$

and if we write $\Delta_Q = \prod_{v\in Q} \Delta_v$, $\left(\prod \chi_\alpha\right) : \Delta_Q \to (R_Q^{\mathrm{univ}})^\times$, then we see that $(R_Q^{\mathrm{univ}})_{\Delta_Q} = R_\varnothing^{\mathrm{univ}}$.

The isomorphism $R_Q^{\square} \xrightarrow{\sim} R_Q^{\mathrm{univ}} \widehat{\otimes}_{\mathcal{O}} \mathcal{J}$ and the homomorphism $\Delta_Q \to (R_Q^{\mathrm{univ}})^\times$ together give a homomorphism $\mathcal{J}[\Delta_Q] \to R_Q^{\square}$. In the same way, we have a homomorphism $\mathcal{J}[\Delta_Q] \to R_Q^{\square,'}$, and again these agree modulo λ. If we write $\mathfrak{a}_Q := \langle \mathfrak{a}, \delta - 1\rangle_{\delta\in\Delta_Q} \lhd \mathcal{J}[\Delta_Q]$, then we see that $R_Q^{\square}/\mathfrak{a}_Q = R_\varnothing^{\mathrm{univ}}$, and that $R_Q^{\square,'}/\mathfrak{a}_Q = R_\varnothing^{\mathrm{univ},'}$, and again these agree modulo λ.

We now examine the spaces of modular forms that we will patch. We have our fixed isomorphism $\imath : \bar{L} \xrightarrow{\sim} \mathbb{C}$, and an algebraic Grössencharacter χ_0 such that $\chi\varepsilon_p = \chi_{0,\imath}$. Define k, η by $\mathrm{HT}_\tau(\rho_0) = \{\eta_{\imath\tau}, \eta_{\imath\tau} + k_{\imath\tau} - 1\}$. We define compact open subgroups $U_Q = \prod U_{Q,v}$, where

- $U_{Q,v} = \mathrm{GL}_2(\mathcal{O}_{F_v})$ if $v \notin Q \cup T_r$,
- $U_{Q,v} = U_0(v) = \left\{\left(\begin{smallmatrix} * & * \\ 0 & * \end{smallmatrix}\right) \pmod v\right\}$ if $v \in T_r$, and
- $U_{Q,v} = \left\{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in U_0(v) \mid a/d \pmod v \in k(v)^\times \mapsto 1 \in \Delta_v\right\}$ if $v \in Q$.

We let $\psi : \prod_{v\in Q\cup T_r} U_{Q,v} \to \mathcal{O}^\times$ be the trivial character. Similarly, we set $U_Q' = U_Q$, and we define $\psi' : \prod_{v\in Q\cup T_r} U_{Q,v} \to \mathcal{O}^\times$ in the following way. For each $v \in T_r$, we have a homomorphism $U_{Q,v} \to k(v)^\times$ given by sending $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ to $a/d \pmod v$, and we compose these characters with the characters $k(v)^\times \to \mathcal{O}^\times$ sending the image of σ_v to ζ, where σ_v is a generator of I_{F_v}/P_{F_v}. We let ψ' be trivial at the places in Q.

We obtain spaces of modular forms $S(U_Q, \mathcal{O})$, $S(U_Q', \mathcal{O})$ and corresponding Hecke algebras $\mathbb{T}_{U_Q}$, $\mathbb{T}_{U_Q'}$, generated by the Hecke operators T_v, S_v with $v \notin T \cup Q$, together with Hecke operators U_{ϖ_v} for $v \in Q$ (depending on a chosen uniformizer ϖ_v) defined by

$$U_{\varpi_v} = \left[U_{Q,v} \begin{pmatrix} \varpi_v & 0 \\ 0 & 1 \end{pmatrix} U_{Q,v} \right].$$

Note that $\psi = \psi' \pmod \lambda$, so we have $S(U_\varnothing, \mathcal{O})/\lambda = S(U_\varnothing', \mathcal{O})/\lambda$. We let $\mathfrak{m}_\varnothing \lhd \mathbb{T}_{U_\varnothing}$ be the ideal generated by λ and the $\operatorname{tr} \bar{\rho}(\mathrm{Frob}_v) - T_v$, $\det \bar{\rho}(\mathrm{Frob}_v) - \# k(v) S_v$,

$v \notin T$. This is a maximal ideal of $\mathbb{T}_{U_\varnothing}$, because it is the kernel of the homomorphism $\mathbb{T}_{U_\varnothing} \to \mathcal{O} \twoheadrightarrow \mathbb{F}$, where the map $\mathbb{T}_{U_\varnothing} \to \mathcal{O}$ is the one coming from the automorphicity of ρ_0, sending $T_v \mapsto \operatorname{tr}\rho_0(\mathrm{Frob}_v)$, $S_v \mapsto \#k(v)^{-1}\det\rho_0(\mathrm{Frob}_v)$.

Write $\mathbb{T}_\varnothing := \mathbb{T}_{U_\varnothing,\mathfrak{m}_\varnothing}$. We have a lifting $\rho^{\mathrm{mod}} : G_F \to \mathrm{GL}_2(\mathbb{T}_\varnothing)$ of type $\mathcal{S}_\varnothing$, so by the universal property of $R_\varnothing^{\mathrm{univ}}$, we have a surjection $R_\varnothing^{\mathrm{univ}} \twoheadrightarrow \mathbb{T}_\varnothing$ (it is surjective because local-global compatibility shows that the Hecke operators generating $\mathbb{T}_\varnothing$ are in the image). Similarly, we have a surjection $R_\varnothing^{\mathrm{univ},'} \twoheadrightarrow \mathbb{T}'_\varnothing := \mathbb{T}_{U'_\varnothing,\mathfrak{m}_\varnothing}$. Set $S_\varnothing := S(U_\varnothing, \mathcal{O})_{\mathfrak{m}_\varnothing}$, $S'_\varnothing := S(U'_\varnothing, \mathcal{O})_{\mathfrak{m}_\varnothing}$. Then the identification $R_\varnothing^{\mathrm{univ}}/\lambda \cong R_\varnothing^{\mathrm{univ},'}/\lambda$ is compatible with $S_\varnothing/\lambda = S'_\varnothing/\lambda$.

Lemma 5.7. *If* $\mathrm{Supp}_{R_\varnothing^{\mathrm{univ}}}(S_\varnothing) = \mathrm{Spec}\, R_\varnothing^{\mathrm{univ}}$, *then ρ is modular.*

Proof. Suppose that $\mathrm{Supp}_{R_\varnothing^{\mathrm{univ}}}(S_\varnothing) = \mathrm{Spec}\, R_\varnothing^{\mathrm{univ}}$. Since $S_\varnothing$ is a faithful $\mathbb{T}_\varnothing$-module by definition, we see that $\ker(R_\varnothing^{\mathrm{univ}} \to \mathbb{T}_\varnothing)$ is nilpotent, so that $(R_\varnothing^{\mathrm{univ}})^{\mathrm{red}} \xrightarrow{\sim} \mathbb{T}_\varnothing$. Then ρ corresponds to some homomorphism $R_\varnothing^{\mathrm{univ}} \to \mathcal{O}$, and thus to a homomorphism $\mathbb{T}_\varnothing \to \mathcal{O}$, and the composite of this homomorphism with $\imath : \mathcal{O} \hookrightarrow \mathbb{C}$ corresponds to a cuspidal automorphic representation π of $G_D(\mathbb{A}^\infty)$ of weight (k, η), which by construction has the property that $\rho \cong \rho_{\pi,\imath}$, as required. □

To show that $\mathrm{Supp}_{R_\varnothing^{\mathrm{univ}}}(S_\varnothing) = \mathrm{Spec}\, R_\varnothing^{\mathrm{univ}}$, we will study the above constructions as Q varies. Let $\mathfrak{m}_Q \lhd \mathbb{T}_{U_Q}$ be the maximal ideal generated by λ, the $\operatorname{tr}\bar\rho(\mathrm{Frob}_v) - T_v$ and $\det\bar\rho(\mathrm{Frob}_v) - \#k(v)S_v$ for $v \notin T \cup Q$, and the $U_{\varpi_v} - \bar\alpha_v$ for $v \in Q$.

Write $S_Q = S_{U_Q} := S(U_Q, \mathcal{O})_{\mathfrak{m}_Q}$ and $\mathbb{T}_Q := (\mathbb{T}_{U_Q})_{\mathfrak{m}_Q}$. We have a homomorphism $\Delta_Q \to \mathrm{End}(S_Q)$, given by sending $\delta \in \Delta_v$ to $\left(\begin{smallmatrix}\delta & 0\\ 0 & 1\end{smallmatrix}\right) \in U_0(v)$. We also have another homomorphism $\Delta_Q \to \mathrm{End}(S_Q)$, given by the composite

$$\Delta_Q \to R_Q^{\mathrm{univ}} \twoheadrightarrow \mathbb{T}_Q \to \mathrm{End}(S_Q).$$

Let $U_{Q,0} := \prod_{v \notin Q} U_{Q,v} \prod_{v \in Q} U_0(v)$. Then U_Q is a normal subgroup of $U_{Q,0}$, and $U_{Q,0}/U_Q = \Delta_Q$.

We now examine the consequences of local-global compatibility at the places in Q.

Proposition 5.8. (1) *The two homomorphisms $\Delta_Q \to \mathrm{End}(S_Q)$ (the other one coming via R_Q^{univ}) are equal.*

(2) *S_Q is finite free over $\mathcal{O}[\Delta_Q]$.*

Proof. A homomorphism $\theta : \mathbb{T}_Q \to \bar{L} \xrightarrow{\sim} \mathbb{C}$ corresponds to a cuspidal automorphic representation π, and for each $v \in Q$ the image α_v of U_{ϖ_v} is such that α_v is an eigenvalue of U_{ϖ_v} on $\pi_v^{U_{Q,v}}$.

It can be checked that since $\pi_v^{U_{Q,v}} \neq 0$, π_v is necessarily a subquotient of $\chi_1 \times \chi_2$ for some tamely ramified characters $\chi_1, \chi_2 : F_v^\times \to \mathbb{C}^\times$. Then one checks explicitly that

$$(\chi_1 \times \chi_2)^{U_{Q,v}} \cong \mathbb{C}\phi_1 \oplus \mathbb{C}\phi_w,$$

where $w = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, $\phi_1(1) = \phi_w(w) = 1$, and $\operatorname{Supp}\phi_1 = B(F_v)U_{Q,v}$, $\operatorname{Supp}\phi_w = B(F_v)wU_{Q,v}$.

Further explicit calculation shows that

$$U_{\varpi_v}\phi_1 = \#k(v)^{1/2}\chi_1(\varpi_v)\phi_1 + X\phi_w$$

for some X, which is 0 if χ_1/χ_2 is ramified, and

$$U_{\varpi_v}\phi_w = \#k(v)^{1/2}\chi_2(\varpi_v)\phi_w.$$

By local-global compatibility $\iota^{-1}(\#k(v)^{1/2}\chi_1(\varpi_v))$ and $\iota^{-1}(\#k(v)^{1/2}\chi_2(\varpi_v))$ are the eigenvalues of $\rho_{\pi,\iota}(\mathrm{Frob}_v)$, so one of them is a lift of $\bar{\alpha}_v$, and one is a lift of $\bar{\beta}_v$. As a consequence, we see that $\chi_1/\chi_2 \neq |\cdot|^{\pm 1}$ (as if this equality held, we would have $\bar{\alpha}_v/\bar{\beta}_v \equiv \#k(v)^{\pm 1} \equiv 1 \pmod{\lambda}$, contradicting our assumption that $\bar{\alpha}_v \neq \bar{\beta}_v$). Consequently we have $\pi_v = \chi_1 \times \chi_2 \cong \chi_2 \times \chi_1$, so that without loss of generality we have $\bar{\chi}_1(\varpi_v) = \bar{\beta}_v$, $\bar{\chi}_2(\varpi_v) = \bar{\alpha}_v$.

It is also easily checked that

$$\begin{pmatrix} \delta & 0 \\ 0 & 1 \end{pmatrix}\phi_1 = \chi_1(\delta)\phi_1, \quad \begin{pmatrix} \delta & 0 \\ 0 & 1 \end{pmatrix}\phi_w = \chi_2(\delta)\phi_w.$$

We see that $S_Q \otimes_{\mathcal{O},\iota} \mathbb{C} = \oplus_\pi \otimes_{v \in Q} X_v$, where X_v is the 1-dimensional space where U_{ϖ_v} acts via a lift of $\bar{\alpha}_v$. Since this space is spanned by ϕ_w, we see that Δ_v acts on S_Q via $\chi_2 = \chi_\alpha \circ \mathrm{Art}$. This completes the proof of the first part.

Finally, the second part is immediate from Proposition 5.4(2). □

Fix a place $v \in Q$. Since $\bar{\alpha}_v \neq \bar{\beta}_v$, by Hensel's lemma we may write

$$\operatorname{char} \rho_\varnothing^{\mathrm{mod}}(\mathrm{Frob}_v) = (X - A_v)(X - B_v)$$

for some $A_v, B_v \in \mathbb{T}_\varnothing$ with $A_v \equiv \bar{\alpha}_v$, $B_v \equiv \bar{\beta}_v \pmod{\mathfrak{m}_\varnothing}$.

Proposition 5.9. *We have an isomorphism $\prod_{v \in Q}(U_{\varpi_v} - B_v) : S_\varnothing \xrightarrow{\sim} S(U_{Q,0}, \mathcal{O})_{\mathfrak{m}_Q}$ (with the morphism being defined by viewing the source and target as submodules of $S(U_{Q,0}, \mathcal{O})_{\mathfrak{m}_\varnothing}$).*

Proof. We claim that it is enough to prove that the map is an isomorphism after tensoring with L, and an injection after tensoring with $\mathbb{F}$. To see this, write $X := S_\varnothing$, $Y := S(U_{Q,0}, \mathcal{O})_{\mathfrak{m}_Q}$, and write Q for the cokernel of the map $X \to Y$. Then X, Y are finite free $\mathcal{O}$-modules, and if the map $X \otimes L \to Y \otimes L$ is injective, then so is the map $X \to Y$, so that we have a short exact sequence $0 \to X \to Y \to Q \to 0$. Tensoring with L, we have $Q \otimes L = 0$. Tensoring with $\mathbb{F}$, we obtain an exact sequence $0 \to Q[\lambda] \to X \otimes \mathbb{F} \to Y \otimes \mathbb{F} \to Q \otimes \mathbb{F} \to 0$, so we have $Q[\lambda] = 0$. Thus $Q = 0$, as required.

In order to check that we have an isomorphism after tensoring with L, it is enough to check that the induced map $\prod_{v \in Q}(U_{\varpi_v} - B_v) : S_\varnothing \otimes_{\mathcal{O},\iota} \mathbb{C} \to S(U_{Q,0}, \mathcal{O})_{\mathfrak{m}_Q} \otimes_{\mathcal{O},\iota} \mathbb{C}$

is an isomorphism. This is easily checked: $S_\varnothing \otimes \mathbb{C} \cong \oplus_\pi \otimes_{v\in Q} (\chi_{1,v} \times \chi_{2,v})^{\mathrm{GL}_2(\mathcal{O}_{F_v})}$, and $(\chi_{1,v} \times \chi_{2,v})^{\mathrm{GL}_2(\mathcal{O}_{F_v})} = \mathbb{C}\phi_0$, where ϕ_0 is as in Exercise 4.7(3). Similarly, $S(U_{Q,0}, \mathcal{O})_{\mathfrak{m}_Q} \otimes_{\mathcal{O},\iota} \mathbb{C} = \oplus_\pi \otimes_{v\in Q} M_v$, where M_v is the subspace of $(\chi_{1,v} \times \chi_{2,v})^{U_0(v)}$ on which U_{ϖ_v} acts via a lift of $\bar{\alpha}_v$, which is spanned by ϕ_w. Since the natural map $(\chi_{1,v} \times \chi_{2,v})^{\mathrm{GL}_2(\mathcal{O}_{F_v})} \to (\chi_{1,v} \times \chi_{2,v})^{U_0(v)}$ sends $\phi_0 \mapsto \phi_1 + \phi_w$ (as $\phi_0(1) = \phi_0(w) = 1$), the result follows.

It remains to check injectivity after tensoring with $\mathbb{F}$. The kernel of the map, if nonzero, would be a nonzero finite module for the Artinian local ring $\mathbb{T}_\varnothing/\lambda$, and would thus have nonzero $\mathfrak{m}_\varnothing$-torsion, so it suffices to prove that the induced map

$$\prod_{v\in Q} (U_{\varpi_v} - B_v) : (S_\varnothing \otimes \mathbb{F})[\mathfrak{m}_\varnothing] \to S(U_{Q,0}, \mathcal{O})_{\mathfrak{m}_Q} \otimes \mathbb{F}$$

is an injection. By induction on $\#Q$, it suffices to prove this in the case that $Q = \{v\}$. Suppose for the sake of contradiction that there is a nonzero $x \in (S_\varnothing \otimes \mathbb{F})[\mathfrak{m}_\varnothing]$ with $(U_{\varpi_v} - \bar{\beta}_v)x = 0$. Since $x \in S_\varnothing \otimes \mathbb{F}$, we also have $T_v x = (\bar{\alpha}_v + \bar{\beta}_v)x$, and we will show that these two equations together lead to a contradiction.

Now, x is just a function $D^\times \backslash \mathrm{GL}_2(\mathbb{A}_F^\infty) \to \Lambda \otimes \mathbb{F}$, on which $\mathrm{GL}_2(\mathbb{A}_F^\infty)$ acts by right translation. If we make the action of the Hecke operators explicit, we find that there are g_i such that

$$U_v = \coprod_i g_i U_{Q,v}$$

and

$$T_v = \left(\coprod_i g_i \,\mathrm{GL}_2(\mathcal{O}_{F_v}) \right) \coprod \begin{pmatrix} 1 & 0 \\ 0 & \varpi_v \end{pmatrix} \mathrm{GL}_2(\mathcal{O}_{F_v}),$$

so that we have $\left(\begin{smallmatrix} 1 & 0 \\ 0 & \varpi_v \end{smallmatrix}\right)x = T_v x - U_{\varpi_v} x = \bar{\alpha}_v x$. Then $\left(\begin{smallmatrix} \varpi_v & 0 \\ 0 & 1 \end{smallmatrix}\right)x = w\left(\begin{smallmatrix} 1 & 0 \\ 0 & \varpi_v \end{smallmatrix}\right)wx = \bar{\alpha}_v x$, and $U_{\varpi_v} x = \sum_{a\in k(v)} \left(\begin{smallmatrix} \varpi_v & a \\ 0 & 1 \end{smallmatrix}\right)x = \sum_{a\in k(v)} \left(\begin{smallmatrix} 1 & a \\ 0 & 1 \end{smallmatrix}\right)\left(\begin{smallmatrix} \varpi_v & 0 \\ 0 & 1 \end{smallmatrix}\right)x = \#k(v)\bar{\alpha}_v x = \bar{\alpha}_v x$. But $U_{\varpi_v} x = \bar{\beta}_v x$, so $\bar{\alpha}_v = \bar{\beta}_v$, a contradiction. □

Set $S_Q^\square := S_Q \otimes_{R_Q^{\mathrm{univ}}} R_Q^\square$. Then we have $S_Q^\square/\mathfrak{a}_Q = S(U_{Q,0}, \mathcal{O})_{\mathfrak{m}_Q} \xrightarrow{\sim} S_\varnothing$, compatibly with the isomorphism $R_Q^\square/\mathfrak{a}_Q \xrightarrow{\sim} R_\varnothing^{\mathrm{univ}}$. Also, $S_Q^\square$ is finite free over $\mathcal{J}[\Delta_Q]$.

We now return to the Galois side. By Proposition 3.24, we can and do choose a presentation

$$R^{\mathrm{loc}}[\![x_1, \ldots, x_{h_Q}]\!] \twoheadrightarrow R_Q^\square,$$

where $h_Q = \#T + \#Q - 1 - [F : \mathbb{Q}] + \dim_{\mathbb{F}} H_Q^1(G_{F,T}, (\mathrm{ad}^0 \bar{\rho})(1))$, and

$$H_Q^1(G_{F,T}, (\mathrm{ad}^0 \bar{\rho})(1)) = \ker\bigl(H^1(G_{F,T}, (\mathrm{ad}^0 \bar{\rho})(1)) \to \oplus_{v\in Q} H^1(G_{k(v)}, (\mathrm{ad}^0 \bar{\rho})(1))\bigr).$$

The following result will provide us with the sets Q that we will use.

Proposition 5.10. *Let* $r = \max(\dim H^1(G_{F,T}, (\mathrm{ad}^0 \bar{\rho})(1)), 1 + [F : \mathbb{Q}] - \#T)$. *For each* $N \geq 1$, *there exists a set* Q_N *of places of* F *such that*:

- $Q_N \cap T = \varnothing$.
- *If* $v \in Q_N$, *then* $\bar{\rho}(\mathrm{Frob}_v)$ *has distinct eigenvalues* $\bar{\alpha}_v \neq \bar{\beta}_v$.
- *If* $v \in Q_N$, *then* $\#k(v) \equiv 1 \pmod{p^N}$.
- $\#Q_N = r$.
- $R^{\square}_{Q_N}$ (*respectively* $R^{\square,\prime}_{Q_N}$) *is topologically generated over* R^{loc} (*respectively* $R^{\mathrm{loc},\prime}$) *by* $\#T - 1 - [F:\mathbb{Q}] + r$ *elements.*

Proof. The last condition may be replaced by

- $H^1_{Q_N}(G_{F,T}, (\mathrm{ad}^0\,\bar{\rho})(1)) = 0$.

Therefore, it is enough to show that for each $0 \neq [\phi] \in H^1(G_{F,T}, (\mathrm{ad}^0\,\bar{\rho})(1))$, there are infinitely many $v \notin T$ such that:

- $\#k(v) \equiv 1 \pmod{p^N}$.
- $\bar{\rho}(\mathrm{Frob}_v)$ has distinct eigenvalues $\bar{\alpha}_v, \bar{\beta}_v$.
- $\mathrm{Res}[\phi] \in H^1(G_{k(v)}, (\mathrm{ad}^0\,\bar{\rho})(1))$ is nonzero.

This then gives us some set of places Q with the given properties, except that $\#Q$ may be too large; but then we can pass to a subset of cardinality r, while maintaining the injectivity of the map $H^1(G_{F,T}, (\mathrm{ad}^0\,\bar{\rho})(1)) \to \oplus_{v\in Q} H^1(G_{k(v)}, (\mathrm{ad}^0\,\bar{\rho})(1))$.

We will use the Chebotarev density theorem to do this; note that the condition that $\#k(v) \equiv 1 \pmod{p^N}$ is equivalent to v splitting completely in $F(\zeta_{p^N})$, and the condition that $\bar{\rho}(\mathrm{Frob}_v)$ has distinct eigenvalues is equivalent to asking that $\mathrm{ad}\,\bar{\rho}(\mathrm{Frob}_v)$ has an eigenvalue not equal to 1.

Set $E = \overline{F}^{\ker \mathrm{ad}\,\bar{\rho}}(\zeta_{p^N})$. We claim that we have $H^1(\mathrm{Gal}(E/F), (\mathrm{ad}^0\,\bar{\rho})(1)) = 0$. In order to see this, we claim firstly that $\zeta_p \notin \overline{F}^{\ker \mathrm{ad}\,\bar{\rho}}$. This follows from the classification of finite subgroups of $\mathrm{PGL}_2(\overline{\mathbb{F}}_p)$: we have assumed that $\mathrm{Im}\,\bar{\rho} \supseteq \mathrm{SL}_2(\mathbb{F}_p)$, and this implies that $\mathrm{Im}\,\mathrm{ad}\,\bar{\rho} = \mathrm{PGL}_2(\mathbb{F}_{p^s})$ or $\mathrm{PSL}_2(\mathbb{F}_{p^s})$ for some s, and in particular $(\mathrm{Im}\,\mathrm{ad}\,\bar{\rho})^{\mathrm{ab}}$ is trivial or cyclic of order 2. Since $p \geq 5$ and p is unramified in F, we have $[F(\zeta_p):F] \geq 4$, so $\zeta_p \notin \overline{F}^{\ker \mathrm{ad}\,\bar{\rho}}$, as claimed.

The extension $E/\overline{F}^{\ker \mathrm{ad}\,\bar{\rho}}$ is abelian, and we let E_0 be the intermediate field such that $\mathrm{Gal}(E/E_0)$ has order prime to p, while $\mathrm{Gal}(E_0/\overline{F}^{\ker \mathrm{ad}\,\bar{\rho}})$ has p-power order. Write $\Gamma_1 = \mathrm{Gal}(E_0/F)$, $\Gamma_2 = \mathrm{Gal}(E/E_0)$. Then the inflation-restriction exact sequence is in part

$$0 \to H^1(\Gamma_1, (\mathrm{ad}^0\,\bar{\rho})(1)^{\Gamma_2}) \to H^1(\mathrm{Gal}(E/F), (\mathrm{ad}^0\,\bar{\rho})(1)) \to H^1(\Gamma_2, (\mathrm{ad}^0\,\bar{\rho})(1))^{\Gamma_1},$$

so in order to show that $H^1(\mathrm{Gal}(E/F), (\mathrm{ad}^0\,\bar{\rho})(1)) = 0$, it suffices to prove that $H^1(\Gamma_1, (\mathrm{ad}^0\,\bar{\rho})(1)^{\Gamma_2}) = H^1(\Gamma_2, (\mathrm{ad}^0\,\bar{\rho})(1))^{\Gamma_1} = 0$.

In fact, we claim that $(\mathrm{ad}^0\,\bar{\rho})(1)^{\Gamma_2}$ and $H^1(\Gamma_2, (\mathrm{ad}^0\,\bar{\rho})(1))$ both vanish. For the first of these, note that Γ_2 acts trivially on $\mathrm{ad}^0\,\bar{\rho}$ (since E_0 contains $\overline{F}^{\ker \mathrm{ad}\,\bar{\rho}}$), but that $\zeta_p \notin E_0$ (as $[E_0 : \overline{F}^{\ker \mathrm{ad}\,\bar\rho}]$ is a power of p). For the second term, note that Γ_2 has prime-to-p order.

Suppose that $\#k(v) \equiv 1 \pmod p$, and that $\bar\rho(\mathrm{Frob}_v) = \left(\begin{smallmatrix} \bar\alpha_v & 0 \\ 0 & \bar\beta_v \end{smallmatrix}\right)$. Then $\mathrm{ad}^0\,\bar\rho$ has the basis $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$, $\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$, $\left(\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right)$ of eigenvectors for Frob_v, with eigenvalues 1, $\bar\alpha_v/\bar\beta_v$, $\bar\beta_v/\bar\alpha_v$ respectively. Consequently, we see that there is an isomorphism $H^1(G_{k(v)}, (\mathrm{ad}^0\,\bar\rho)(1)) \cong \mathbb{F}$ (since in general for a (pro)cyclic group, the first cohomology is given by passage to coinvariants), which we can write explicitly as $[\phi] \mapsto \pi_v \circ \phi(\mathrm{Frob}_v) \circ i_v$, where i_v is the injection of $\mathbb{F}$ into the $\bar\alpha_v$-eigenspace of Frob_v, and π_v is the Frob_v-equivariant projection onto that subspace.

Let σ_0 be an element of $\mathrm{Gal}(E/F)$ such that:

- $\sigma_0(\zeta_{p^N}) = \zeta_{p^N}$.
- $\bar\rho(\sigma_0)$ has distinct eigenvalues $\bar\alpha$, $\bar\beta$.

(To see that such a σ_0 exists, note that $\mathrm{Gal}(\overline{F}^{\ker\bar\rho}/F(\zeta_{p^N})\cap\overline{F}^{\ker\bar\rho})$ contains $\mathrm{PSL}_2(\mathbb{F}_p)$, and so we can choose σ_0 so that its image in this group is an element whose adjoint has an eigenvalue other than 1.) Let $\tilde E/E$ be the extension cut out by all the $[\phi] \in H^1(G_{F,T}, (\mathrm{ad}^0\,\bar\rho)(1))$. In order to complete the proof, it suffices to show that we can choose some $\sigma \in \mathrm{Gal}(\tilde E/F)$ with $\sigma|_E = \sigma_0$, and such that in the notation above, we have $\pi_{\sigma_0} \circ \phi(\sigma) \circ i_{\sigma_0} \neq 0$, because we can then choose v to have $\mathrm{Frob}_v = \sigma$ by the Chebotarev density theorem.

To this end, choose any $\tilde\sigma_0 \in \mathrm{Gal}(\tilde E/F)$ with $\tilde\sigma_0|_E = \sigma_0$. If $\tilde\sigma_0$ does not work, then we have $\pi_{\sigma_0} \circ \phi(\tilde\sigma_0) \circ i_{\sigma_0} = 0$. In this case, take $\sigma = \sigma_1\tilde\sigma_0$ for some $\sigma_1 \in \mathrm{Gal}(\tilde E/E)$. Then $\phi(\sigma) = \phi(\sigma_1\tilde\sigma_0) = \phi(\sigma_1) + \sigma_1\phi(\tilde\sigma_0) = \phi(\sigma_1) + \phi(\tilde\sigma_0)$, so $\pi_{\sigma_0} \circ \phi(\sigma) \circ i_{\sigma_0} = \pi_{\sigma_0} \circ \phi(\sigma_1) \circ i_{\sigma_0}$.

Note that $\phi(\mathrm{Gal}(\tilde E/E))$ is a $\mathrm{Gal}(E/F)$-invariant subset of $\mathrm{ad}^0\,\bar\rho$, which is an irreducible $\mathrm{Gal}(E/F)$-module, since the image of $\bar\rho$ contains $\mathrm{SL}_2(\mathbb{F}_p)$. Thus the $\mathbb{F}$-span of $\phi(\mathrm{Gal}(\tilde E/E))$ is all of $\mathrm{ad}^0\,\bar\rho(1)$, from which it is immediate that we can choose σ_1 so that $\pi_{\sigma_0} \circ \phi(\sigma_1) \circ i_{\sigma_0} \neq 0$. □

We are now surprisingly close to proving the main theorem! Write $h := \#T - 1 - [F:\mathbb{Q}] + r$, and $R_\infty := R^{\mathrm{loc}}[[x_1, \ldots, x_h]]$. For each set Q_N as above, choose a surjection $R_\infty \twoheadrightarrow R^\square_{Q_N}$. Let $\mathcal{J}_\infty := \mathcal{J}[[y_1, \ldots, y_r]]$. Choose a surjection $\mathcal{J}_\infty \twoheadrightarrow \mathcal{J}[\Delta_{Q_N}]$, given by writing $Q_N = \{v_1, \ldots, v_r\}$ and mapping y_i to $(\gamma_i - 1)$, where γ_i is a generator of Δ_{v_i}. Choose a homomorphism $\mathcal{J}_\infty \to R_\infty$ so that the composites $\mathcal{J}_\infty \to R_\infty \twoheadrightarrow R^\square_{Q_N}$ and $\mathcal{J}_\infty \to \mathcal{J}[\Delta_{Q_N}] \to R^\square_{Q_N}$ agree, and write $\mathfrak{a}_\infty := (\mathfrak{a}, y_1, \ldots, y_r)$. Then $S^\square_{Q_N}/\mathfrak{a}_\infty = S_\varnothing$, $R^\square_{Q_N}/\mathfrak{a}_\infty = R^{\mathrm{univ}}_\varnothing$.

Write $\mathfrak{b}_N := \ker(\mathcal{J}_\infty \to \mathcal{J}[\Delta_{Q_N}])$, so that $S^\square_{Q_N}$ is finite free over $\mathcal{J}_\infty/\mathfrak{b}_N$. Since all the elements of Q_N are congruent to 1 modulo p^N, we see that

$$\mathfrak{b}_N \subseteq ((1+y_1)^{p^N} - 1, \ldots, (1+y_r)^{p^N} - 1).$$

We can and do choose the same data for $R^{\mathrm{loc},'}$, in such a way that the two sets of data are compatible modulo λ.

Now choose open ideals $\mathfrak{c}_N \lhd \mathcal{J}_\infty$ such that:

- $\mathfrak{c}_N \cap \mathcal{O} = (\lambda^N)$.
- $\mathfrak{c}_N \supseteq \mathfrak{b}_N$.
- $\mathfrak{c}_N \supseteq \mathfrak{c}_{N+1}$.
- $\cap_N \mathfrak{c}_N = 0$.

(For example, we could take $\mathfrak{c}_N = ((1+X_{v,i,j})^{p^N}-1, (1+y_i)^{p^N}-1, \lambda^N)$.) Note that since $\mathfrak{c}_N \supseteq \mathfrak{b}_N$, $S^\square_{Q_N}/\mathfrak{c}_N$ is finite free over $\mathcal{J}_\infty/\mathfrak{c}_N$. Also choose open ideals $\mathfrak{d}_N \lhd R^{\mathrm{univ}}_\varnothing$ such that:

- $\mathfrak{d}_N \subseteq \ker(R^{\mathrm{univ}}_\varnothing \to \mathrm{End}(S_\varnothing/\lambda^N))$.
- $\mathfrak{d}_N \supseteq \mathfrak{d}_{N+1}$.
- $\cap_N \mathfrak{d}_N = 0$.

If $M \geq N$, write $S_{M,N} = S^\square_{Q_M}/\mathfrak{c}_N$, so that $S_{M,N}$ is finite free over $\mathcal{J}_\infty/\mathfrak{c}_N$ of rank equal to the $\mathcal{O}$-rank of $S_\varnothing$; indeed $S_{M,N}/\mathfrak{a}_\infty \xrightarrow{\sim} S_\varnothing/\lambda^N$. Then we have a commutative diagram

$$\begin{array}{ccccc} \mathcal{J}_\infty & \longrightarrow & R_\infty & \twoheadrightarrow & R^{\mathrm{univ}}_\varnothing/\mathfrak{d}_N \\ & & \circlearrowright & & \circlearrowright \\ & & S_{M,N} & \twoheadrightarrow & S_\varnothing/\mathfrak{d}_N \end{array}$$

where $S_{M,N}$, $S_\varnothing/\mathfrak{d}_N$ and $R^{\mathrm{univ}}_\varnothing/\mathfrak{d}_N$ all have finite cardinality. Because of this finiteness, we see that there is an infinite subsequence of pairs (M_i, N_i) such that $M_{i+1} > M_i$, $N_{i+1} > N_i$, and the induced diagram

$$\begin{array}{ccccc} \mathcal{J}_\infty & \longrightarrow & R_\infty & \twoheadrightarrow & R^{\mathrm{univ}}_\varnothing/\mathfrak{d}_{N_i} \\ & & \circlearrowright & & \circlearrowright \\ & & S_{M_{i+1},N_{i+1}}/\mathfrak{c}_{N_i} & \twoheadrightarrow & S_\varnothing/\mathfrak{d}_{N_i} \end{array}$$

is isomorphic to the diagram for (M_i, N_i).

Then we can take the projective limit over this subsequence, to obtain a commutative diagram

$$\begin{array}{ccccc} \mathcal{J}_\infty & \longrightarrow & R_\infty & \twoheadrightarrow & R^{\mathrm{univ}}_\varnothing \\ & & \circlearrowright & & \circlearrowright \\ & & S_\infty & \twoheadrightarrow & S_\varnothing \end{array}$$

where S_∞ is finite free over $\mathcal{J}_\infty$. Furthermore, we can simultaneously carry out the same construction in the $'$ world, compatibly with this picture modulo λ.

This is the key picture, and the theorem will now follow from it by purely commutative algebra arguments. We have (ultimately by the calculations of the dimensions of the local deformation rings in Theorems 3.28 and 3.31)

$$\dim R_\infty = \dim R'_\infty = \dim \mathcal{J}_\infty = 4\#T + r,$$

and since S_∞, S'_∞ are finite free over the power series ring $\mathcal{J}_\infty$ (from Proposition 5.8), we have

$$\operatorname{depth}_{\mathcal{J}_\infty}(S_\infty) = \operatorname{depth}_{\mathcal{J}_\infty}(S'_\infty) = 4\#T + r.$$

(This is the "numerical coincidence" on which the Taylor–Wiles method depends; see [Calegari and Geraghty 2018] for a further discussion of this point, and of a more general "numerical coincidence".) Since the action of $\mathcal{J}_\infty$ on S_∞ factors through R_∞, we see that

$$\operatorname{depth}_{R_\infty}(S_\infty) \geq 4\#T + r,$$

and similarly

$$\operatorname{depth}_{R'_\infty}(S'_\infty) \geq 4\#T + r.$$

Now, if $\mathcal{P} \lhd R'_\infty$ is a minimal prime in the support of S'_∞, then we see that

$$4\#T + r = \dim R'_\infty \geq \dim R'_\infty/\mathcal{P} \geq \operatorname{depth}_{R'_\infty} S'_\infty \geq 4\#T + r,$$

so equality holds throughout, and $\mathcal{P}$ is a minimal prime of R'_∞. But R'_∞ has a unique minimal prime, so in fact

$$\operatorname{Supp}_{R'_\infty}(S'_\infty) = \operatorname{Spec} R'_\infty.$$

By the same argument, we see that $\operatorname{Supp}_{R_\infty}(S_\infty)$ is a union of irreducible components of $\operatorname{Spec} R_\infty$. We will show that it is all of $\operatorname{Spec} R_\infty$ by reducing modulo λ and comparing with the situation for S'_∞.

To this end, note that since $\operatorname{Supp}_{R'_\infty}(S'_\infty) = \operatorname{Spec} R'_\infty$, we certainly have

$$\operatorname{Supp}_{R'_\infty/\lambda}(S'_\infty/\lambda) = \operatorname{Spec} R'_\infty/\lambda.$$

This implies that $\operatorname{Supp}_{R_\infty/\lambda}(S_\infty/\lambda) = \operatorname{Spec} R_\infty/\lambda$, by the compatibility between the two pictures. Thus $\operatorname{Supp}_{R_\infty}(S_\infty)$ is a union of irreducible components of $\operatorname{Spec} R_\infty$, which contains the entirety of $\operatorname{Spec} R_\infty/\lambda$. Since (by Theorem 3.38) the irreducible components of $\operatorname{Spec} R_\infty/\lambda$ are in bijection with the irreducible components of $\operatorname{Spec} R_\infty$, this implies that $\operatorname{Supp}_{R_\infty}(S_\infty) = \operatorname{Spec} R_\infty$. Then

$$\operatorname{Supp}_{R_\infty/\mathfrak{a}_\infty}(S_\infty/\mathfrak{a}_\infty) = R_\infty/\mathfrak{a}_\infty,$$

i.e., $\operatorname{Supp}_{R^{\text{univ}}_\varnothing} S_\varnothing = R^{\text{univ}}_\varnothing$, which is what we wanted to prove.

6. Relaxing the hypotheses

The hypotheses in our main theorem are not optimal. We will now briefly indicate the "easy" relaxations of the assumptions that could be made, and discuss the generalizations that are possible with (a lot) more work.

Firstly, it is possible to relax the assumption that $p \geq 5$, and that $\operatorname{Im} \bar{\rho} \supseteq \mathrm{SL}_2(\mathbb{F}_p)$. These assumptions cannot be completely removed, but they can be considerably relaxed. The case $p = 2$ is harder in several ways, but important theorems have been proved in this case, for example the results of Kisin [2009b] which completed the proof of Serre's conjecture.

On the other hand, the case $p = 3$ presents no real difficulties. The main place that we assumed that $p > 3$ was in the proof that the finite groups G_i in Section 5.3 have order prime to p; this argument could also break down for cases when $p > 3$ if we allowed p to ramify in F, which in general we would like to do. Fortunately, there is a simple solution to this problem, which is to introduce an auxiliary prime v to the level. This prime is chosen in such a way that all deformations of $\bar{\rho}|_{G_{F_v}}$ are automatically unramified, so none of the global Galois deformation rings that we work with are changed when we relax the conditions at v. The existence of an appropriate v follows from the Chebotarev density theorem and some elementary group theory; see Lemma 4.11 of [Darmon et al. 1997] and the discussion immediately preceding it.

We now consider the possibility of relaxing the assumption that $\operatorname{Im} \bar{\rho} \supseteq \mathrm{SL}_2(\mathbb{F}_p)$. We should certainly assume that $\bar{\rho}$ is absolutely irreducible, because otherwise many of our constructions don't even make sense; we always had to assume this in constructing universal deformation rings, in constructing the universal modular deformation, and so on. (Similar theorems have been proved in the case that $\bar{\rho}$ is reducible, in particular by Skinner and Wiles [1999], but the arguments are considerably more involved, and at present involve a number of serious additional hypotheses, in particular ordinarity — although see [Pan 2022] for a theorem without an ordinarity hypothesis.) Examining the arguments made above, we see that the main use of the assumption that $\operatorname{Im} \bar{\rho} \supseteq \mathrm{SL}_2(\mathbb{F}_p)$ is in the proof of Proposition 5.10. Looking more closely at the proof, the key assumption is really that $\bar{\rho}|_{G_{F(\zeta_p)}}$ is absolutely irreducible; this is known as the "Taylor–Wiles assumption". (Note that by elementary group theory, this is equivalent to the absolute irreducibility of $\bar{\rho}|_{G_K}$, where K/F is the unique quadratic subextension of $F(\zeta_p)/F$; in particular, over $\mathbb{Q}$ the condition is equivalent to the absolute irreducibility of $\bar{\rho}|_{G_{\mathbb{Q}(\sqrt{(-1)^{(p-1)/2}p})}}$, which is how the condition is stated in the original papers.)

Unfortunately this condition isn't quite enough in complete generality, but it comes very close; the only exception is certain cases when $p = 5$, F contains $\mathbb{Q}(\sqrt{5})$, and the projective image of $\bar{\rho}$ is $\mathrm{PGL}_2(\mathbb{F}_5)$. See [Kisin 2009c, (3.2.3)] for the definitive statement (and see the work of Khare and Thorne [2017] for some improvements in this exceptional case). If $\bar{\rho}$ is absolutely irreducible, but $\bar{\rho}|_{G_{F(\zeta_p)}}$ is (absolutely) reducible, it is sometimes possible to prove modularity lifting theorems, but considerably more work is needed (and there is no general approach in higher dimension); see [Skinner and Wiles 2001] in the ordinary case, which uses similar arguments to those of [Skinner and Wiles 1999], and also [Thorne 2016; Pan 2022].

The other conditions that we could hope to relax are the assumptions on $\rho|_{G_{F_v}}$ and $\rho_0|_{G_{F_v}}$ at places $v \mid p$. We've hardly discussed where some of these assumptions come from, as we swept most issues with p-adic Hodge theory under the carpet. There are essentially two problems here. First, we have assumed that p is unramified in F, that the Galois representations are crystalline, and that the gaps between the Hodge–Tate weights are "small"; this is the Fontaine–Laffaille condition. There is also the assumption that ρ, ρ_0 have the same Hodge–Tate weights. Both conditions can be considerably (although by no means completely) relaxed (of course subject to the necessary condition that ρ is geometric). As already alluded to above, very general results are available in the ordinary case (even in arbitrary dimension), in particular those of Geraghty [2019]. In the case that $F_v = \mathbb{Q}_p$ there are again very general results, using the p-adic local Langlands correspondence for $\mathrm{GL}_2(\mathbb{Q}_p)$; see in particular [Emerton 2011; Kisin 2009a; Pan 2022]. However, beyond this case, the situation is considerably murkier, and at present there are no generally applicable results.

6.1. ***Further generalizations.*** Other than the results discussed in the previous subsection, there are a number of obvious generalizations that one could hope to prove. One obvious step, already alluded to above, is to replace 2-dimensional representations with n-dimensional representations; we could also hope to allow F to be a more general number field. At present it seems to be necessary to assume that F is a CM field, as otherwise we do not know how to attach Galois representations to automorphic representations; but if F is CM, then automorphy lifting theorems analogous to our main theorem are now known (for arbitrary n), and we refer to [Calegari 2021] for both the history of such results and the state of the art.

Another natural condition to relax would be the condition that the Hodge–Tate weights are distinct; for example, one could ask that they all be equal, and hope to prove Artin's conjecture, or that some are equal, to prove modularity results for abelian varieties. The general situation where some Hodge–Tate weight occurs with multiplicity greater than 2 seems to be completely out of reach (because there is no known way to relate the automorphic representations expected to correspond to such Galois representations to the automorphic representations which contribute to the cohomology of Shimura varieties, which is the only technique we have for constructing the maps $R \to \mathbb{T}$), but there has been considerable progress for small dimensional cases, for which we again refer the reader to [Calegari 2021].

Finally, we would of course like to be able to dispose of the hypothesis that $\bar{\rho}$ is modular (that is, to dispose of ρ_0). This is the problem of Serre's conjecture and its generalizations, and has only been settled in the case that $F = \mathbb{Q}$ and $n = 2$. The proof in that case (by Khare and Wintenberger [2009a; 2009b] and Kisin [2009b]) makes essential use of modularity lifting theorems. The proof inductively reduces to the case that $p \leq 5$ and $\bar{\rho}$ has very little ramification, when direct arguments using discriminant bounds can be made. The more general modularity lifting theorems mentioned above make it plausible that the inductive steps could be generalized, but the base case of the induction seems specific to the case of $\mathrm{GL}_2 /\mathbb{Q}$, and proving the modularity of $\bar{\rho}$ in greater generality is one of the biggest open problems in the field.

Acknowledgements

These notes are based in part on a lecture course given by Richard Taylor at Harvard University in 2009, and I am extremely grateful to him both for the many things I learned from his course, and for giving me permission to present them here.

These notes were originally written for the 2013 Arizona Winter School, and I would like to thank the organizers (Bryden Cais, Mirela Çiperiani, Matthew Papanikolas, Rachel Pries, David Savitt, and Romyar Sharifi) for the invitation to speak, and the participants for their feedback. I would like to thank Ana Caraiani for encouraging me to belatedly publish these notes. In revising for publication I removed some material relating to the Arizona Winter School project, and for the most part removed the outdated section on more recent developments (the reader disappointed by this change should read Calegari's excellent survey [2021]). I have otherwise maintained the feel of the original notes.

I would also like to thank Kevin Buzzard, Frank Calegari, Ana Caraiani, Kęstutis Česnavičius, Matthew Emerton, Jessica Fintzen, Jeff Hatley, Florian Herzig, Christian Johansson, Keenan Kidwell, Alexander Kutzim, Jack Lamplugh, Heejong Lee, Tom Lovering, Judith Ludwig, James Newton, Sug Woo Shin and Jack Shotton for their helpful comments on earlier versions of these notes.

Finally, I am extremely grateful to the anonymous referee for their many helpful corrections and suggestions.

References

[Barnet-Lamb et al. 2011] T. Barnet-Lamb, D. Geraghty, M. Harris, and R. Taylor, "A family of Calabi–Yau varieties and potential automorphy II", *Publ. Res. Inst. Math. Sci.* **47**:1 (2011), 29–98. MR Zbl

[Barnet-Lamb et al. 2014] T. Barnet-Lamb, T. Gee, D. Geraghty, and R. Taylor, "Potential automorphy and change of weight", *Ann. of Math.* (2) **179**:2 (2014), 501–609. MR Zbl

[Berger 2004] L. Berger, "An introduction to the theory of p-adic representations", pp. 255–292 in *Geometric aspects of Dwork theory*, vol. 1, edited by A. Adolphson et al., de Gruyter, Berlin, 2004. MR Zbl

[Bhatt 2021] B. Bhatt, "Algebraic geometry in mixed characteristic", preprint, 2021. To appear in the proceedings of the 2022 ICM. arXiv 2112.12010

[Böckle 2013] G. Böckle, "Deformations of Galois representations", pp. 21–115 in *Elliptic curves, Hilbert modular forms and Galois deformations*, edited by H. Darmon et al., Birkhäuser, Basel, 2013. MR Zbl

[Brinon and Conrad 2009] O. Brinon and B. Conrad, "CMI summer school notes on p-adic Hodge theory", 2009, available at http://math.stanford.edu/∼conrad/.

[Bushnell and Henniart 2006] C. J. Bushnell and G. Henniart, *The local Langlands conjecture for* GL(2), Grundl. Math. Wissen. **335**, Springer, 2006. MR Zbl

[Buzzard 2012] K. Buzzard, "Potential modularity—a survey", pp. 188–211 in *Non-abelian fundamental groups and Iwasawa theory*, edited by J. Coates et al., London Math. Soc. Lecture Note Ser. **393**, Cambridge Univ. Press, 2012. MR Zbl

[Calegari 2021] F. Calegari, "Reciprocity in the Langlands program since Fermat's last theorem", preprint, 2021. arXiv 2109.14145

[Calegari and Geraghty 2018] F. Calegari and D. Geraghty, "Modularity lifting beyond the Taylor–Wiles method", *Invent. Math.* **211**:1 (2018), 297–433. MR Zbl

[Carayol 1994] H. Carayol, "Formes modulaires et représentations galoisiennes à valeurs dans un anneau local complet", pp. 213–237 in *p-adic monodromy and the Birch and Swinnerton-Dyer conjecture* (Boston, MA, 1991), edited by B. Mazur and G. Stevens, Contemp. Math. **165**, Amer. Math. Soc., Providence, RI, 1994. MR Zbl

[Choi 2009] S. H. Choi, *Local deformation lifting spaces of mod l Galois representations*, Ph.D. thesis, Harvard University, 2009, available at https://www.proquest.com/docview/304892280. MR

[Clozel 1990] L. Clozel, "Motifs et formes automorphes: Applications du principe de fonctorialité", pp. 77–159 in *Automorphic forms*, *Shimura varieties*, *and L-functions* (Ann Arbor, MI, 1988), vol. 1, edited by L. Clozel and J. S. Milne, Perspectives in Mathematics **10**, Academic Press, Boston, MA, 1990. MR Zbl

[Clozel et al. 2008] L. Clozel, M. Harris, and R. Taylor, "Automorphy for some l-adic lifts of automorphic mod l Galois representations", *Publ. Math. Inst. Hautes Études Sci.* 108 (2008), 1–181. MR Zbl

[Curtis and Reiner 1962] C. W. Curtis and I. Reiner, *Representation theory of finite groups and associative algebras*, Pure and Applied Mathematics **XI**, Interscience Publishers, New York, 1962. MR Zbl

[Darmon et al. 1997] H. Darmon, F. Diamond, and R. Taylor, "Fermat's last theorem", pp. 2–140 in *Elliptic curves*, *modular forms & Fermat's last theorem* (Hong Kong, 1993), edited by J. Coates and S. T. Yau, Int. Press, Cambridge, MA, 1997. MR Zbl

[Dickinson 2001a] M. Dickinson, "A criterion for existence of a universal deformation ring", pp. 339–343 in *Arithmetic algebraic geometry* (Park City, UT, 1999), edited by B. Conrad and K. Rubin, IAS/Park City Math. Ser. **9**, Amer. Math. Soc., Providence, RI, 2001. Appendix 1 of F. Gouvêa, "Deformations of Galois representations", pp. 233–406. MR

[Dickinson 2001b] M. Dickinson, "On the modularity of certain 2-adic Galois representations", *Duke Math. J.* **109**:2 (2001), 319–382. MR Zbl

[Emerton 2011] M. Emerton, "Local-global compatibility in the p-adic Langlands programme for $\mathrm{GL}_{2/\mathbb{Q}}$", 2011, available at http://www.math.uchicago.edu/∼emerton/pdffiles/lg.pdf.

[Fargues 2011] L. Fargues, "Motives and automorphic forms: the (potentially) abelian case", 2011, available at https://webusers.imj-prg.fr/~laurent.fargues/Motifs_abeliens.pdf.

[Flath 1979] D. Flath, "Decomposition of representations into tensor products", pp. 179–183 in *Automorphic forms, representations and L-functions* (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), edited by A. Borel and W. Casselman, Proc. Sympos. Pure Math., XXXIII **1**, Amer. Math. Soc., Providence, R.I., 1979. MR Zbl

[Fontaine 1994] J.-M. Fontaine, "Représentations l-adiques potentiellement semi-stables", pp. 321–347 in *Périodes p-adiques* (Bures-sur-Yvette, 1988), Astérisque **223**, Société Mathématique de France, Paris, 1994. MR Zbl

[Fontaine and Laffaille 1982] J.-M. Fontaine and G. Laffaille, "Construction de représentations p-adiques", *Ann. Sci. École Norm. Sup.* (4) **15**:4 (1982), 547–608. MR Zbl

[Fontaine and Mazur 1995] J.-M. Fontaine and B. Mazur, "Geometric Galois representations", pp. 41–78 in *Elliptic curves, modular forms, & Fermat's last theorem* (Hong Kong, 1993), edited by J. H. Coates and S.-T. Yau, Series in Number Theory **1**, International Press, Cambridge, MA, 1995. MR Zbl

[Gelbart 1975] S. S. Gelbart, *Automorphic forms on adèle groups*, Annals of Mathematics Studies **83**, Princeton University Press, Princeton, N.J., 1975. MR Zbl

[Geraghty 2019] D. Geraghty, "Modularity lifting theorems for ordinary Galois representations", *Math. Ann.* **373**:3-4 (2019), 1341–1427. MR Zbl

[Gruenberg 1967] K. Gruenberg, "Profinite groups", pp. 116–127 in *Algebraic Number Theory* (Proc. Instructional Conf., Brighton, 1965), edited by J. W. S. Cassels and A. Fröhlich, Thompson, Washington, D.C., 1967. MR Zbl

[Katz and Messing 1974] N. M. Katz and W. Messing, "Some consequences of the Riemann hypothesis for varieties over finite fields", *Invent. Math.* **23** (1974), 73–77. MR Zbl

[Khare and Thorne 2017] C. B. Khare and J. A. Thorne, "Automorphy of some residually S_5 Galois representations", *Math. Z.* **286**:1-2 (2017), 399–429. MR Zbl

[Khare and Wintenberger 2009a] C. Khare and J.-P. Wintenberger, "Serre's modularity conjecture, I", *Invent. Math.* **178**:3 (2009), 485–504. MR Zbl

[Khare and Wintenberger 2009b] C. Khare and J.-P. Wintenberger, "Serre's modularity conjecture, II", *Invent. Math.* **178**:3 (2009), 505–586. MR Zbl

[Kisin 2008] M. Kisin, "Potentially semi-stable deformation rings", *J. Amer. Math. Soc.* **21**:2 (2008), 513–546. MR Zbl

[Kisin 2009a] M. Kisin, "The Fontaine–Mazur conjecture for GL_2", *J. Amer. Math. Soc.* **22**:3 (2009), 641–690. MR Zbl

[Kisin 2009b] M. Kisin, "Modularity of 2-adic Barsotti–Tate representations", *Invent. Math.* **178**:3 (2009), 587–634. MR Zbl

[Kisin 2009c] M. Kisin, "Moduli of finite flat group schemes, and modularity", *Ann. of Math.* (2) **170**:3 (2009), 1085–1180. MR Zbl

[Mazur 1989] B. Mazur, "Deforming Galois representations", pp. 385–437 in *Galois groups over* $\mathbf{Q}$ (Berkeley, CA, 1987), edited by Y. Ihara et al., Math. Sci. Res. Inst. Publ. **16**, Springer, 1989. MR Zbl

[Mazur 1997] B. Mazur, "An introduction to the deformation theory of Galois representations", pp. 243–311 in *Modular forms and Fermat's last theorem* (Boston, MA, 1995), edited by G. Cornell et al., Springer, 1997. MR Zbl

[Milne 2006] J. S. Milne, *Arithmetic duality theorems*, 2nd ed., BookSurge, LLC, Charleston, SC, 2006. MR Zbl

[Pan 2022] L. Pan, "The Fontaine–Mazur conjecture in the residually reducible case", *J. Amer. Math. Soc.* **35**:4 (2022), 1031–1169. MR Zbl

[Pilloni 2008] V. Pilloni, "The study of 2-dimensional p-adic Galois deformations in the $\ell \neq p$ case", 2008, available at http://perso.ens-lyon.fr/vincent.pilloni/Defo.pdf.

[Ramakrishna 1993] R. Ramakrishna, "On a variation of Mazur's deformation functor", *Compositio Math.* **87**:3 (1993), 269–286. MR Zbl

[Serre 1979] J.-P. Serre, *Local fields*, Graduate Texts in Mathematics **67**, Springer, 1979. MR Zbl

[Shotton 2016] J. Shotton, "Local deformation rings for GL_2 and a Breuil–Mézard conjecture when $\ell \neq p$", *Algebra Number Theory* **10**:7 (2016), 1437–1475. MR Zbl

[Skinner and Wiles 1999] C. M. Skinner and A. J. Wiles, "Residually reducible representations and modular forms", *Inst. Hautes Études Sci. Publ. Math.* 89 (1999), 5–126. MR Zbl

[Skinner and Wiles 2001] C. M. Skinner and A. J. Wiles, "Nearly ordinary deformations of irreducible residual representations", *Ann. Fac. Sci. Toulouse Math.* (6) **10**:1 (2001), 185–215. MR Zbl

[Tate 1979] J. Tate, "Number theoretic background", pp. 3–26 in *Automorphic forms, representations and L-functions, II* (Corvallis, OR, 1977), edited by A. Borel and W. Casselman, Proc. Sympos. Pure Math. **33**, American Mathematical Society, Providence, RI, 1979. MR Zbl

[Taylor 2003] R. Taylor, "On icosahedral Artin representations, II", *Amer. J. Math.* **125**:3 (2003), 549–566. MR Zbl

[Taylor 2006] R. Taylor, "On the meromorphic continuation of degree two L-functions", *Doc. Math.* Extra Vol. (2006), 729–779. MR

[Taylor 2008] R. Taylor, "Automorphy for some l-adic lifts of automorphic mod l Galois representations, II", *Publ. Math. Inst. Hautes Études Sci.* 108 (2008), 183–239. MR Zbl

[Thorne 2016] J. A. Thorne, "Automorphy of some residually dihedral Galois representations", *Math. Ann.* **364**:1-2 (2016), 589–648. MR Zbl

Received 9 Feb 2022. Revised 4 Sep 2022.

TOBY GEE:

toby.gee@imperial.ac.uk
Mathematics Department, Imperial College London, London, United Kingdom

ESSENTIAL NUMBER THEORY

msp.org/ent

See inside back cover or msp.org/ent for submission instructions.

Essential Number Theory (ISSN 2834-4634 electronic, 2834-4626 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840 is published continuously online.

ENT peer review and production are managed by EditFlow® from MSP.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

https://msp.org/

ESSENTIAL NUMBER THEORY

2022 vol. 1 no. 1

The cubic case of Vinogradov's mean value theorem 1
D. R. HEATH-BROWN

Exceptional zeros, sieve parity, Goldbach 13
JOHN B. FRIEDLANDER and HENRYK IWANIEC

A note on Tate's conjectures for abelian varieties 41
CHAO LI and WEI ZHANG

A Diophantine problem about Kummer surfaces 51
WILLIAM DUKE

Quartic index form equations and monogenizations of quartic orders 57
SHABNAM AKHTARI

Modularity lifting theorems 73
TOBY GEE