p-adic analytic continuation of symmetric power functoriality

Author: Kenz Kallal (kallal@princeton.edu)
Advisor: Prof. Gaëtan Chenevier (gaetan.chenevier@math.cnrs.fr)

27 June 2022

Dedicated to the memory of Justin Se Geun Lim

Abstract

This M2 mémoire is broadly about the theme of p-adic variation of automorphic forms and Galois representations, and the connection between this theme and the Langlands functoriality conjecture. More specifically, it is about how techniques in the domain of p-adic interpolation of automorphic forms and Galois representations were applied in a recent landmark work of Newton and Thorne to prove that symmetric power functoriality holds for all holomorphic modular forms of level 1 and weight $k \geq 2$ if and only if it holds for a single one. Along the way, we attempt to give useful explanations of the basic theory of p-adic automorphic forms and eigenvarieties that underpins the arguments of Newton–Thorne.

There is no original content in this mémoire: it is purely an expository synthesis culled from various sources, especially the books and papers of Bellaïche, Buzzard, Loeffler, Chenevier, Bellaïche–Chenevier, Ye, Breuil, Breuil–Hellmann–Schraen, and Newton–Thorne.

Acknowledgements

> "Always two, there are; no more, no less: a master and an apprentice."
>
> ---
>
> Master Yoda to Master Windu,
> *Star Wars Episode I: The Phantom Menace*

I am deeply indebted to my M2 advisor Gaëtan Chenevier, who has extremely generously spent many hours patiently explaining many aspects of the Langlands program to me. I also thank my parents Hédi Kallal and Malika Zeghal, who have selflessly supported me in all aspects of my life for the past 22 years, as well as my family members Asmaa Abada, Arlette Artois, Mehdi Zeghal, and Mohamed Zeghal, for their warm welcome and hospitality during my stay in Paris.

Thanks are also due to Mark Kisin and Barry Mazur for many useful mathematical conversations about p-adic automorphic forms and eigenvarieties during my undergraduate education, and for their advice and help — as well as that of Michael Harris and Ahmed Abbes — in being able to come to France and study under the supervision of Professor Chenevier. I further thank Professor Abbes for sponsoring my visit to the Institut des Hautes Études Scientifiques via the Fulbright U.S. Student Program, and for his own contribution to my mathematical education via his advice regarding algebraic geometry (particularly his suggestion to read Raynaud's book [Ray1970] on henselian local rings which was quite useful in my study of the work of Bellaïche–Chenevier on pseudodeformations) and the 2022 Franco–Asian summer school on arithmetic geometry he organized at the Centre International des Rencontres Mathématiques in Luminy. The I.H.É.S. has provided an extremely good environment for me to carry out this mémoire, especially thanks to useful conversations with Ahmed Abbes, Eknath Ghate, Tongmu He, Inkang Kim, Yeuk Hay Joshua Lam, Graham Smith, and Dmitry Vaintrob. Outside of I.H.É.S., I equally thank Kevin Buzzard, Matthew Emerton, Hao Billy Lee, Siyan Daniel Li-Huerta, David Loeffler, Samuel Marks, Mohamed Moakher, Mikayel Mkrtchyan, Gal Porat, Loren Spice, Vijay Srinivasan, Zhixiang Wu, and Fan Zhou for many useful explanations.

I also thank my teachers at Université Paris–Sud XI: Benjamin Schraen for his course on algebraic number theory, Emanuele Macri for his course on algebraic geometry, and Vincent Pilloni and Olivier Schiffmann for their course on the moduli stack of vector bundles on a curve.

I was lucky to have very generous classmates who helped make these courses more enjoyable and instructive: Adib Abdollahi, Mahmoud Elimam, Virginio Fratianni, Yuanyang Jiang, Damien Simon, Reinier Sorgdrager, Yuchen Wu, and Jiangfan Yuan. In fact, I further thank Professor Pilloni, as well as Valentin Hernandez, for agreeing to serve on the jury for my M2 mémoire defense.

Finally, my studies in Paris this year were generously supported by the Sophie Germain scholarship of the Fondation Mathématique Jacques Hadamard and the Franco-American Fulbright Commission. I thank both organizations for funding this very useful educational experience, as well as the Princeton University Department of Mathematics and the United States National Science Foundation Graduate Fellowship Program for allowing me to defer the beginning of my graduate studies for a year in order to pursue this educational opportunity.

Contents

Abstract **ii**

Acknowledgements **iii**

1 Introduction **1**

1.1 Langlands functoriality . . . 1

1.2 Galois representations associated to automorphic forms . . . 4

1.2.1 Algebraic Hecke characters and Galois representations . . . 5

1.2.2 Hodge–Tate decomposition of Weil's p-adic representations associated to Algebraic Hecke characters, and the Fontaine–Mazur conjecture for GL_1 . . . 20

1.3 The Sato–Tate conjecture . . . 32

2 Eigenvarieties **48**

2.1 The concept of a p-adic overconvergent automorphic form . . . 51

2.1.1 Overconvergent modular forms, à la Katz and Coleman . . . 51

2.1.2 The level at p à la Katz and Coleman . . . 57

2.1.3 Weights and weight space . . . 59

2.1.4 Classical p-adic automorphic representations . . . 60

2.1.5 Overconvergent p-adic automorphic forms . . . 65

2.1.6 Interlude: the level at p à la Buzzard–Chenevier–Loeffler–Ye . . . 71

2.1.7 The Atkin–Lehner algebra . . . 72

2.1.8 Atkin–Lehner theory and refined automorphic representations . . . 73

2.2 Construction of Eigenvarieties, à la Emerton–Newton–Thorne . . . 75

2.2.1 Two alternative constructions . . . 75

2.2.2 Accumulation of classical points . . . 77

2.2.3 Classicality theorems, d'après Chenevier and Newton–Thorne . . . 78

3 Analytic continuation of symmetric power functoriality **80**

3.1 The trianguline variety . . . 80

3.2 Passing to definite unitary groups . . . 93

3.3 Bonus details on base-change 105

4 Ping–Pong **109**

References **122**

Chapter 1

Introduction

"凡事总须研究，才会明白"

鲁迅，《狂人日记》

1.1 | Langlands functoriality

In this introductory first section, we follow the exposition of Emerton [Eme2021] as well as private communication with Gaëtan Chenevier in explaining the basic yoga of the (Arthur–)Langlands conjectures. The purpose is to show how to predict the symmetric power functoriality conjecture which this mémoire is about, and not to do the full detail of the various constructions, which themselves could be the subject of an entire M2 mémoire.

Let G be a connected reductive algebraic group over a number field F. To this, one may attach an "L-group" ${}^LG$ (see [Bor1979, Cas2001]). It will not matter much for us what it is, since for $G = \mathrm{GL}_n/\mathbf{Q}$, depeding on the convention, the L-group can be taken to be $\mathrm{GL}_n(\mathbf{C})$.

The Langlands conjecture predicts that for G, H connected reductive groups over F, where H is quasi-split, a homomorphism ${}^LG \to {}^LH$ should induce a way of going from automorphic representations of G to automorphic representations of H. The requirement is that the relationship is mediated by L-functions of the automorphic representations for the two groups, which we now explain.

For a finite place $\mathfrak{p}$ of F, we say that G is *unramified* at $\mathfrak{p}$ if it is quasi-split over $F_\mathfrak{p}$ and splits over $F_\mathfrak{p}^{\mathrm{ur}}$.

Generalizing the concept of Satake parameters [Sat1963], Langlands [Lan1970] found a way to attach, to an arbitrary automorphic representation π for G, and finite place $\mathfrak{p}$ of F such that both G and π are unramified at $\mathfrak{p}$ a *Langlands parameter*, which is a $\widehat{G}$-conjugacy class $c_{\pi,\mathfrak{p}}$ in

$$\widehat{G} \rtimes \mathrm{Frob}_\mathfrak{p}.$$

Equivalently, it is a $\widehat{G}$-conjugacy class of homomorphisms

$$\varphi_{\pi,\mathfrak{p}} : W_{F_\mathfrak{p}} \times SU(2) \to {}^L G = \widehat{G} \rtimes \langle \mathrm{Frob}_\mathfrak{p} \rangle$$

that respect[1] the canonical map $W_{F_\mathfrak{p}} \to \langle \mathrm{Frob}_\mathfrak{p} \rangle$ and are themselves *unramified* in the sense that they are trivial on $I_{F_\mathfrak{p}} \times SU(2)$.

The point of the Langlands parameters is that the data of the Langlands parameter φ completely determines $\pi_\mathfrak{p}$. Being able to do this at all places $\mathfrak{p}$, including those in the set S of finite places where G and π are unramified, is the local Langlands conjecture (known for $G = \mathrm{GL}_n$ thanks to Harris– Taylor [HT2001] and Scholze [Sch2013]).

In any event, with the basic structure of how Langlands parameters work out of the way, we can give the full detail of the statement of the Langlands functoriality conjecture:

Conjecture 1.1.1 (Langlands functoriality)**.** Let G, H be connected reductive groups over a number field F, with H quasi-split. Let π be an automorphic representation for G, and let S be the set of finite places $\mathfrak{p}$ of F with the property that either G or H or π is not unramified at $\mathfrak{p}$. Suppose we are given an L-homomorphism $r :^L G \to {}^L H$. Then there exists an automorphic representation $r(\pi)$ for H which is unramified outside of S and such that $c_{r(\pi),\mathfrak{p}} = r(c_{\pi,\mathfrak{p}})$ for all $\mathfrak{p}$ outside of S.

One reason why Conjecture 1.1.1 and the Langlands parameters are useful is that they are related to L-functions of automorphic forms. In particular, the L-function of π as above outside of S with respect to some representation $\rho : {}^L G \to \mathrm{GL}_n$ is defined to be

$$L^S(s, \pi, \rho) := \prod_{\mathfrak{p} \notin S} \frac{1}{\det(1 - \mathrm{N}\mathfrak{p}^{-s} \rho(c_{\pi,\mathfrak{p}}))}.$$

In many concrete cases (e.g. those in the examples below), it is a straightforward exercise to check that this L-function corresponds to a more classical construction. Conjecture 1.1.1 may be restated as

Conjecture 1.1.2. Let H, G, π, S, r be as in Conjecture 1.1.1, and let ρ be an n-dimensional representation of $\widehat{H}$. Then there exists an automorphic representation $r(\pi)$ for H such that

$$L^S(s, \pi, \rho \circ r) = L^S(s, r(\pi), r).$$

This is useful because it relates, on the left hand side, an L-function whose analytic properties we might care about (for which the finite set S really doesn't matter), since it is not exactly the same as an actual automorphic L-function (which would be $L^S(s, \pi, \rho)$), to, on the right hand

[1]This condition is called being an *L-homomorphism*.

side, something which is an automorphic L-function, and therefore has good analytic properties by the general theory (e.g. [GJ1972, JS1976]).

Example 1.1.3. Let H be the algebraic group corresponding to the units of a definite quaternion algebra $D/\mathbf{Q}$, and let $G = \mathrm{GL}_2/\mathbf{Q}$. Since H is an inner form of G, there L-groups are the same. On the other hand, only G is quasi-split. So Conjecture 1.1.2 predicts that to each quaternionic modular form on D, one should be able to attach a modular form in a way that respects L-functions, but not always in the other direction. This is the Jacquet–Langlands correspondence, which was proved (like many other things) using the Selberg trace formula [JL1970, GJ1979]. Indeed, since a definite quaternion algebra is by definition ramified at ∞, by class field theory, it is ramified at at least one finite place, which means the modular forms in the image of the Jacquet-Langlands transfer cannot have level 1.

Example 1.1.4. Let G be the trivial group, and $H = \mathrm{GL}_n/\mathbf{Q}$. Then we can choose the L-group ${}^LG = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. Conjecture 1.1.2 predicts that the Artin L-function of any choice of Galois representation $\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_n(\mathbf{C})$ is actually automorphic, and therefore has analytic continuation and functional equation by [GJ1972]. In particular, the Artin conjecture is a consequence of Conjecture 1.1.2.

Example 1.1.5. Let $G = \mathrm{GL}_2/\mathbf{Q}$ and $H = \mathrm{GL}_n/\mathbf{Q}$, so that ${}^LG = \mathrm{GL}_2(\mathbf{C})$ and ${}^LH = \mathrm{GL}_n(\mathbf{C})$, and we can consider the map

$$\mathrm{Sym}^{n-1} : {}^LG \to {}^LH.$$

Conjecture 1.1.2 then predicts that for any modular eigenform f, the symmetric power L-function

$$L(s, \mathrm{Sym}^{n-1} f)$$

is actually automorphic.

The conjecture suggested by this last example is the *symmetric power functoriality conjecture for holomorphic modular forms.* It was proved in a recent paper by Newton and Thorne [NT2021]. The following theorem, which is what the first half of [NT2021] is about, is the main goal of this mémoire.

Theorem 1.1.6. *Let f be an eigenform of level 1 and weight $k \geq 2$, and let $n \geq 2$. If $L(s, \mathrm{Sym}^{n-1} f)$ is automorphic, then $L(s, \mathrm{Sym}^{n-1} g)$ is automorphic as well for all eigenforms g of level 1 and weight ≥ 2.*

The method is by technique of p-adic analytic continuation, many technical details of which will be explained in the subsequent chapters.

Later on in Conjecture 1.3.7, we will explain why the Sato–Tate conjecture is a further concrete reason as to why symmetric power functoriality is interesting.

1.2 | Galois representations associated to automorphic forms

Intertwined with functoriality is the *Arthur–Langlands conjecture*, which states that cuspidal automorphic forms on G should correspond to homomorphisms $\mathbf{L}_{\mathbf{Q}} \to {}^L G$, where $\mathbf{L}_{\mathbf{Q}}$ is the conjectural Langlands group. There are various technical difficulties to it, for example those related to Arthur's multiplicity formula, but suffice it to say for now that in some cases (and all the cases we care about here), part of the conjecture amounts to the existence of a Galois representation associated to an automorphic representation. For example, thanks to Deligne, Shimura, and Deligne–Serre [Del1971, DS1974], or alternatively by the Langlands–Kottwitz method [Sch2011], to an eigenform $f \in S_k(\Gamma_1(N), \mathbf{C})$, for all p not dividing N, there is a corresponding Galois representation

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

At p, ρ_f is always crystalline and has Hodge–Tate weights $k-1$ and 0 [Sai1997, Pan2020].

A massive project of generalization has recently been completed by the authors of the "Paris book project" [CH2013] and Caraiani [Car2012]. It goes as follows.

Theorem 1.2.1 (Chenevier–Harris–Caraiani). *Let $F/\mathbf{Q}$ be a CM field, and suppose π is an automorphic representation of $\mathrm{GL}_n(\mathbf{A}_F)$ which is cuspidal, regular algebraic, and conjugate self-dual. Then for any isomorphism $\iota : \overline{\mathbf{Q}}_p \to \mathbf{C}$, there exists a continuous semisimple representation*

$$r_{\pi,\iota} : G_F \to \mathrm{GL}_n(\overline{\mathbf{Q}}_p)$$

such that $\mathrm{WD}(r_{\pi,\iota})^{F\text{-}ss} \cong \mathrm{rec}^T_{F_v}(\iota^{-1}\pi_v)$ for all finite places v of F, where $\mathrm{rec}^T_{F_v}$ denotes the Tate normalization of the Local langlands correspondence [NT2021, p. 7]. *In particular, for any $\tau : F_v \to \overline{\mathbf{Q}}_p$, the τ-Hodge–Tate weights*[2] *of $r_{\phi,\iota}|_{G_{F_v}}$ are exactly*

$$\{\lambda_{\iota\tau,1} + (n-1), \cdots, \lambda_{\iota\tau,n}\},$$

where the $(\lambda_{\iota\tau,i})_{i=1,\dots,n}$ are the highest weights with respect to the upper-triangular Borel of the irreducible algebraic representation W of $(\mathrm{Res}_{F/\mathbf{Q}}\mathrm{GL}_n)_{\mathbf{C}}$ with the same infinitesimal character as π_∞.

A similar result holds for definite unitary groups, using the machinery of base change (see [NT2021, Corollary 1.3]).

In order to explain some basic concepts that will be used later, and to make up for the fact that we give no indication of a proof for Theorem 1.2.1, we now go ahead and explain the case

[2] to be explained later in this section

$n = 1$, where this amounts to Weil's recipe for constructing p-adic Galois characters from algebraic Hecke characters, and the determination of the Hodge–Tate weights of that character (for which we follow the proof in an appendix of [Ser1989]).

1.2.1 | Algebraic Hecke characters and Galois representations

The aforementioned construction due to Weil originally appeared in [Wei1956].

Weil's construction is a first entry in the theme of "transfering information at infinity to the finite places over p while transferring the coefficient field from $\mathbf{C}$ to $\overline{\mathbf{Q}}_p$," which is a main point of [Buz2004] and will feature prominently in the proceeding chapters. In doing it, we follow Chenevier's exercise [Che2010, Lecture 2, Problem 9], though we go a bit farther.

Fix a number field F, a rational prime p, an embedding $\iota_\infty : \overline{\mathbf{Q}} \to \mathbf{C}$ and an embedding $\iota_p : \overline{\mathbf{Q}} \to \overline{\mathbf{Q}}_p$. Let $\chi : \mathbf{A}_F^\times \to \mathbf{C}^\times$ be a Hecke character.

Lemma 1.2.2. *χ is unramified at all but finitely many places of F.*

Proof. This is the standard "no small subgroups" argument. Being a Lie group, $\mathbf{C}^\times = \mathrm{GL}_1(\mathbf{C})$ has no small subgroups. So let us choose a small open neighborhood $1 \in U \subset \mathbf{C}^\times$, which is small in the sense that it contains no nontrivial subgroup of $\mathbf{C}^\times$ (one way to prove the existence of U, i.e. the "no small subgroups property" for Lie groups in general is by using the fact that the exponential map $\mathfrak{g} \to G$ is an isomorphism near the identity). By continuity of χ, we know $\chi^{-1}(U)$ is an open neighborhood of 1 in $\mathbf{A}_F^\times$. By definition of the topology on $\mathbf{A}_F^\times$, this means that $\chi^{-1}(U)$ contains a subgroup of the form

$$H = \left(\prod_{v \in \Sigma} \mathcal{O}_{F_v}^\times\right) \times \prod_{v \in \Sigma'} H_v,$$

where Σ is a set of finite places of F containing all but finitely many of the places, Σ' is the complement of Σ in the set of all places of F, and H_v is some open neighborhood of 1 in $F_v^\times$ for each $v \in \Sigma'$. Finally, $\chi(H) \subset U$, and the fact that U is small enough to contain no nontrivial subgroups, implies that $\chi(H) = 1$. Since all but finitely many v are in Σ, it follows that $\chi_v|_{\mathcal{O}_{F_v}^\times} = 1$, as desired. Note that we did not use anything about $\mathbf{C}^\times$ other than that it is a Lie group. □

Thanks to Lemma 1.2.2, it is legitimate to write $\chi = \prod_v \chi_v$.

Lemma 1.2.3. *For all finite places v of F (not just the all but finitely many unramified ones), $\chi_v|_{\mathcal{O}_{F_v}^\times}$ has finite image in $\mathbf{C}^\times$.*

Proof. We do the same "no small subgroups" argument, but now we leverage also the fact that $\mathcal{O}_{F_v}^\times$ is compact. As before, let $U \subset \mathbf{C}^\times$ be an open neighborhood of 1 containing no nontrivial subgroup of $\mathbf{C}^\times$, and consider the open subset $\chi_v|_{\mathcal{O}_{F_v}^\times}^{-1}(U) \subset \mathcal{O}_{F_v}^\times$. The identity in $\mathcal{O}_{F_v}^\times$ has a fundamental system of neighborhoods $U_{F_v}^{(n)} = 1 + \mathfrak{p}_v^n$, $n \geq 1$, so we conclude that $\chi_v(U^{(n)}) = 1$ for sufficiently large n. And $U_{F_v}^{(n)}$ is an open subgroup of a compact group, so it is finite-index. Hence, χ kills a finite-index subgroup, which implies it has finite image. $\square$

Lemma 1.2.4. *Suppose that χ is algebraic with weights $\{a_\sigma\}_{\sigma: F \to \mathbf{C}}$. Then for all finite places v of F, the subfield of $\mathbf{C}$ generated by the image of χ_v is a finite extension of $\mathbf{Q} \subset \mathbf{C}$. In fact, there is a finite extension $E/\mathbf{Q}$, depending only on χ and the field F, such that the image of χ_v is contained in $E^\times$ for all finite finite places v of F.*

Proof. By Lemma 1.2.3, for any finite place w of F, $\chi_w(\mathcal{O}_{F_w}^\times)$ is a finite subgroup of $\mathbf{C}^\times$. Therefore, it lives in $\mathbf{Q}(\zeta_{N_w})^\times$, where $N_w = [\mathcal{O}_{F_w}^\times : \ker \chi_w|_{\mathcal{O}_{F_w}^\times}]$. By Lemma 1.2.2, all but finitely many of the N_w are 1, so in fact there is a fixed $N_\chi < \infty$ (the l.c.m. of all the N_w) such that

$$\chi_w(\mathcal{O}_{F_w}^\times) \subset \mathbf{Q}(\zeta_{N_\chi})^\times.$$

Since we want to deal with an arbitrary number field F, there is an obstruction of nontrivial class group (so there isn't necessarily a π_v such that $v(\pi_v) = 1$ and $w(\pi_v) = 0$ for all finite $w \neq v$, since that would mean $\mathfrak{p}_v = (\pi_v)$). For now we ignore that problem, but for the purposes of dealing with it, we rephrase the lemma as saying that the restriction of χ to the finite idèles has image contained in $E^\times$ for some finite extension $E/\mathbf{Q}$, where E depends only on F and χ. We claim that

$$\chi|_{\mathbf{A}_F^{\times,\mathrm{fin}}}\left(F^\times \cdot \prod_{v<\infty} \mathcal{O}_{F_v}^\times\right) \subset L(\zeta_{N_\chi})^\times,$$

where L is the Galois closure (viewed as a subfield of $\overline{\mathbf{Q}} \subset \mathbf{C}$) of $F/\mathbf{Q}$ and $F^\times$ acts on the finite idéles just by the diagonal embedding into the finite places. The reason we have to include L is because multiplying by $x \in F^\times$ inside the finite idéles is not obtained by "restriction to subset" from multiplying by x inside the full idéles for $x \in F^\times$, since that multiplication has an effect on the infinite places and therefore does not necessarily take finite idèles to finite idèles. Let $x \in F^\times$ and $\alpha \in \prod_{v<\infty} \mathcal{O}_{F_v}^\times$. Then (using the fact that χ is a Hecke character)

$$\chi((x\alpha_v)_{v<\infty}) = \chi((x\alpha)_v)\chi((x^{-1})_{v|\infty}) = \chi(\alpha) \prod_{\sigma: F \to \mathbf{C}} \epsilon_{v_\sigma}(\sigma(x^{-1}))\sigma(x^{-1})^{a_\sigma},$$

which is indeed in $L(\zeta_{N_\chi})$, since $\chi(\alpha) \in \mathbf{Q}(\zeta_{N_\chi})$ (that is what we proved in the first paragraph) and the second term is in L (the Galois closure of $F/\mathbf{Q}$ viewed in $\overline{\mathbf{Q}} \subset \mathbf{C}$), since it is a product of embeddings of $F \to \mathbf{C}$ taken at $x^{-1} \in F$. The ϵ_{v_σ}'s, which are 1 when σ is complex and either 1 or sign when σ is real, obviously don't enlarge the field since they will only add a multiplicative factor of ± 1.

Unfortunately, the class group is not necessarily trivial. Luckily, it is finite. Let

$$h = |\mathrm{Cl}(F)| = \left| \mathbf{A}_F^{\times,\mathrm{fin}} / \left(F^\times \cdot \prod_{v<\infty} \mathcal{O}_{F_v}^\times \right) \right|,$$

which is a finite number by finiteness of the class group. Every finite F-idéle α therefore has the property that $\alpha^h \in F^\times \cdot \prod_{v<\infty} \mathcal{O}_{F_v}^\times$, hence

$$\chi(\alpha)^h = \chi(\alpha^h) \in L(\zeta_{N_\chi})^\times.$$

Therefore, $\chi(\alpha)$ is an h-th root of an element of $L(\zeta_{N_\chi})$, which means it is algebraic over $\mathbf{Q}$. Letting $\alpha_1, \ldots, \alpha_h \in \mathbf{A}_F^{\times,\mathrm{fin}}$ be representatives for $\mathrm{Cl}(F)$, we know that every element $\gamma \in \mathbf{A}_F^{\times,\mathrm{fin}}$ is of the form $\alpha_i \beta$ for some $i = 1, \ldots, h$ and some $\beta \in F \cdot \prod_{v<\infty} \mathcal{O}_{F_v}^\times$, so we conclude that

$$\chi(\gamma) \in L(\zeta_{N_\chi h}, \chi(\alpha_i)^{1/h})^\times,$$

and therefore that the image of $\chi|_{\mathbf{A}_F^{\times,\mathrm{fin}}}$ lives inside the finite extension $L(\zeta_{N_\chi h}, \chi(\alpha_1)^{1/h}, \ldots, \chi(\alpha_h)^{1/h})$ of $\mathbf{Q}$, as desired. $\square$

From now on, χ denotes an algebraic Hecke character for F with weights $\{a_\sigma\}_{\sigma:F\to\mathbf{C}}$. From χ, Weil used "technique of transfering ∞-type to p-type" in order to produce a $\overline{\mathbf{Q}}_p^\times$-valued Hecke character which is trivial on the connected component of the ∞-component, namely

$$\eta_{\chi,p} : \mathbf{A}_F^\times \to \overline{\mathbf{Q}}_p^\times$$

given by

$$\eta_{\chi,p}((\alpha_v)_v) = \prod_{\substack{v|\infty \\ \text{real}}} \epsilon_v(\alpha_v) \prod_{v<\infty} \left[\iota_p \circ \iota_\infty^{-1} \circ \chi_v(\alpha_v)\right] \prod_{\sigma:F\to\overline{\mathbf{Q}}_p} \sigma(\alpha_{v_\sigma})^{a_{\iota_\infty\circ\iota_p^{-1}\circ\sigma}},$$

where σ is abuse of notation for the embedding $F_{v_\sigma} \to \overline{\mathbf{Q}}_p$ induced by σ.

Similarly to in the definition of the weight of an overconvergent automorphic form for GL_1 from [Buz2004], the point of the last term indexed by the embeddings $\sigma : F \to \overline{\mathbf{Q}}_p$ is that these embeddings are in bijection with the embeddings $F \to \mathbf{C}$ by composition on the left by $\iota_\infty \circ \iota_p^{-1}$. For such an embedding σ, the place v_σ of F is the one it induces (via the subspace topology from $\overline{\mathbf{Q}}_p$ obtained by embedding F in there with σ); so another way to write the product $\prod_{\sigma:F\to\overline{\mathbf{Q}}_p} \sigma(\alpha_{v_\sigma})^{a_{\iota_\infty\circ\iota_p^{-1}\circ\sigma}}$ would be $\prod_{v|p} \prod_{\sigma:F_v\to\overline{\mathbf{Q}}_p} \sigma(\alpha_v)^{a_{\iota_\infty\circ\iota_p^{-1}\circ\sigma\circ(F\to F_v)}}$. The character $\eta_{\chi,p}$ is well-defined because there are finitely many σ, and all but finitely many of the χ_v are trivial on $\mathcal{O}_{F_v}^\times$, by Lemma 1.2.2.

In fact[3], by Lemma 1.2.4, the image of $\eta_{\chi,p}$ lives in the compositum of the Galois closures in $\overline{\mathbf{Q}}_p$ of the various F_v for $v|p$ together with (taking another compositum) $\iota_p \circ \iota_\infty^{-1} \circ E_\chi$ (E_χ being the finite extension of $\mathbf{Q} \subset \mathbf{C}$ described in Lemma 1.2.4, where it was just called "E"). In particular, we can view $\eta_{\chi,p}$ as a character $F^\times \backslash \mathbf{A}_F^\times \to E^\times$ for some finite $E/\mathbf{Q}_p$ (E is bigger than the E_χ from before since it is an extension of $\mathbf{Q}_p$; the precise form of E_χ won't matter, so we will stick with this confusing notation).

Lemma 1.2.5. *The p-adic character $\eta_{\chi,p}$ factors through $F^\times \backslash \mathbf{A}_F^\times$.*

Proof. Let $\alpha \in F^\times$. It suffices to prove that $\eta_{\chi,p}(\alpha) = 1$. As we just mentioned, there is a finite list $v_1, \ldots, v_m$ of finite places of F such that $\chi_v(\alpha) \neq 1$, so that the product defining $\eta_{\chi,p}(\alpha)$ is well-defined. It is equal to

$$\eta_{\chi,p}(\alpha) = \prod_{\substack{v|\infty \\ \text{real}}} \epsilon_v(\sigma_v(\alpha)) \prod_{j=1}^{m} \iota_p \circ \iota_\infty^{-1} \circ \chi_v(\alpha) \prod_{\sigma: F \to \overline{\mathbf{Q}}_p} \sigma(\alpha)^{a_{\iota_\infty \circ \iota_p^{-1} \circ \sigma}}.$$

Since $\alpha \in F$ is algebraic over $\mathbf{Q}$, so is $\sigma(\alpha)$ for all embeddings $\sigma : F \to \overline{\mathbf{Q}}_p$, and therefore so is $\eta_{\chi,p}(\alpha)$ (it is a finite product of algebraic numbers; we were implictly doing this argument at the beginning of the proof of Lemma 1.2.4 to deduce that the image of $F^\times \cdot \prod_{v<\infty} \mathcal{O}_{F_v}^\times$ under χ was in $L(\zeta_{N_\chi})^\times$). Therefore, we may apply $\iota_\infty \circ \iota_p^{-1}$ to both sides to get the equality

$$\begin{aligned}
\iota_\infty \circ \iota_p^{-1}(\eta_{\chi,p}(\alpha)) &= \prod_{\substack{v|\infty \\ \text{real}}} \epsilon_v(\alpha_v) \prod_{j=1}^{m} \chi_{v_j}(\alpha) \prod_{\sigma: F \to \overline{\mathbf{Q}}_p} \iota_\infty \circ \iota_p^{-1} \circ \sigma(\alpha)^{a_{\iota_\infty \circ \iota_p^{-1} \circ \sigma}} \\
&= \prod_{\substack{v|\infty \\ \text{real}}} \epsilon_v(\alpha_v) \prod_{j=1}^{m} \chi_{v_j}(\alpha) \prod_{\tau: F \to \mathbf{C}} \tau(\alpha)^{a_\tau} \\
&= \chi(\alpha) \\
&= 1,
\end{aligned}$$

where we have used the fact that postcomposing with $\iota_\infty \circ \iota_p^{-1}$ provides a bijection beteen embeddings $F \to \mathbf{C}$ and embeddings $F \to \overline{\mathbf{Q}}_p$, and that the $\epsilon_v = \pm 1$. This implies that $\eta_{\chi,p}(\alpha) = 1$ because $\iota_\infty \circ \iota_p^{-1}$ is bijective where it is defined. We have therefore concluded that $\eta_{\chi,p}$ kills $F^\times$, as desired. ◻

As a consequence of the fact that we removed all the stuff from the infinite places (at least making it trivial on the connected component), we have

[3]It is a general fact from the literature, that I will prove later, that a Galois representation with codomain $\mathrm{GL}_n(\overline{\mathbf{Q}}_p)$ actually has image in $\mathrm{GL}_n(K)$ for some finite $E/\mathbf{Q}_p$, but we don't need to know it in general to know it for $\eta_{\chi,p}$.

Lemma 1.2.6. *The character $\eta_{\chi,p} : F^\times\backslash\mathbf{A}_F^\times \to E_\chi^\times \subset \overline{\mathbf{Q}}_p^\times$ also vanishes on the kernel of the global Artin reciprocity map*

$$\mathrm{rec}_F : F^\times\backslash\mathbf{A}_F^\times \to G_F^{\mathrm{ab}}.$$

Proof. By class field theory[4], $\ker \mathrm{rec}_F$ is the closure of the image of $\prod_{v \text{ real}} \mathbf{R}_{>0}^\times \times \prod_{v \text{ complex}} \mathbf{C}^\times$ in the idèle class group of F.

Since $\eta_{\chi,p}$ is trivial on the connected components of the archimedean completions of F (whose product is exactly the product we wrote down in the previous sentence), and on $F^\times$, and it is continuous, we deduce that it kills the kernel of rec_F, as desired. $\square$

In particular, $\eta_{\chi,p}$ factors through rec_F (since rec_F is surjective) to produce the unique abelian p-adic Galois character $\rho_{\chi,p}$ fitting in the diagram

$$\begin{tikzcd}
F^\times\backslash\mathbf{A}_F^\times \arrow[dr, "\mathrm{rec}_F"'] \arrow[drr, "\eta_{\chi,p}"] & & \\
G_F \arrow[r] \arrow[rr, bend right, "\rho_{\chi,p}"'] & G_F^{\mathrm{ab}} \arrow[r] & \overline{\mathbf{Q}}_p^\times
\end{tikzcd}$$

Remark 1.2.7. Recall: the image of $\rho_{\chi,p}$ is the same as the image of $\eta_{\chi,p}$, which itself is contained inside $E^\times$, where $E/\mathbf{Q}_p$ is a finite extension that we described explicitly.

[4] I think you should be able to do this directly (just by writing the kernel as the intersection of all open subgroups of the idèle class group using the Artin reciprocity law and the existence theorem and trying to show that this is the same as the connected component), but for some reason most likely linked to my level of competence, I could never get it to go through (Lang refuses to spoonfeed me the answer and instead cites Artin–Tate, but I could not find the argument in Artin–Tate, so I had to try and deduce it from whatever was in Lang right before he cites Artin–Tate). At the very least, here is the argument that I remember, which uses some more input: by [Lan1994, Ch. XI, §6, Theorem 6], $\ker \mathrm{rec}_F$ is infinitely divisible. Since G_F^{ab} is profinite, it has no nontrivial infinitely divisible elements (any finite quotient is killed by some integer), and so all infinitely divisible elements of $F^\times\backslash\mathbf{A}_F^\times$ are in $\ker \mathrm{rec}_F$, i.e. $\ker \mathrm{rec}_F$ is exactly the subgroup of infinitely divisible elements (I don't think I need to use this last sentence but it is neat).

Consider the image of $\ker \mathrm{rec}_F$ in the quotient of the idèle class group by $(F^\times\backslash\mathbf{A}_F^\times)^\circ$. That quotient is totally disconnected. In fact, the quotient $(F^\times\backslash\mathbf{A}_F^\times)/(F^\times\backslash\mathbf{A}_F^\times)^\circ$ is also compact. This is because it is equal to

$$(F^\times\backslash\mathbf{A}_F^\times)/\overline{\prod_{v \text{ real}} \mathbf{R}_{>0}^\times \prod_{v \text{ complex}} \mathbf{C}^\times}$$

(since we know explicitly what the connected component is, as we will do in the rest of the proof of Lemma 1.2.6) which is compact by compactness of $\mathbf{A}_F^{\times,1}$. So we are looking at a compact totally disconnected Hausdorff (since we quotiented by a closed subgroup) topological group. All such topological groups are profinite. Since $\ker \mathrm{rec}_F$ is infinitely divisible, as we have already argued, it must be sent to the identity in this quotient.

Again using that G_F^{ab} is profinite and hence totally disconnected, we know that $(F^\times\backslash\mathbf{A}_F^\times)/\ker \mathrm{rec}_F$ is totally disconnected. The connected component $(F^\times\backslash\mathbf{A}_F^\times)^\circ$ is therefore killed by rec_F (else there would be a nontrivial connected component of G_F^{ab}, namely the image of $(F^\times\backslash\mathbf{A}_F^\times)^\circ$).

Hence we have shown both desired inclusions between $\ker \mathrm{rec}_F$ and $(F^\times\backslash\mathbf{A}_F^\times)^\circ$.

Lemma 1.2.8. *The Galois character $\rho_{\chi,p}$ is unramified at all the finite places of F at which χ is unramified (i.e., $\chi_v|_{\mathcal{O}_{F_v}^\times} = 1$), except possibly the places above p.*

Proof. By local-global compatibility of class field theory (e.g. [Lan1994, Ch. XI, §4]), for any finite place v of F, the decomposition group $D_v \subset G_F^{\mathrm{ab}}$ (no choice needs to be made because of the "ab"; also we always have $D_v \cong G_{F_v}^{\mathrm{ab}}$) has the property that

$$\mathrm{rec}_F|_{F_v^\times} : F_v^\times \to D_v \cong G_{F_v}^{\mathrm{ab}}$$

is the local reciprocity map. In particular, it takes $\mathcal{O}_{F_v}^\times$ bijectively onto the inertia group $I_{F_v}^{\mathrm{ab}}$ (the local reciprocity map is identified with the inclusion of $F_v^\times = \mathcal{O}_{F_v}^\times \times \mathbf{Z}$ into its profinite completion $G_{F_v}^{\mathrm{ab}} \cong \mathcal{O}_{F_v}^\times \times \widehat{\mathbf{Z}}$ so even though the local reciprocity map is not surjective, it is when restricted to a map $\mathcal{O}_{F_v} \to I_{F_v}^{\mathrm{ab}}$ — the point is that $\mathcal{O}_{F_v}^\times$ is profinitely complete, and $F_v^\times$ is not). So if a finite place v has the property that $\chi|_{\mathcal{O}_{F_v}^\times} = 1$, then if v does not lie over p, we would have, for $\alpha \in \mathcal{O}_{F_v}^\times$,

$$\eta_{\chi,p}(\alpha) = \iota_p \circ \iota_\infty^{-1} \circ \chi_v(\alpha) = 1.$$

Therefore, Weil's associated Galois character $\rho_{\chi,p}$ is 1 on I_v^{ab} thanks to all the stuff we just said. Of course here I have abused notation by viewing $\rho_{\chi,p}$ as a character of G_F^{ab}, but this is okay because I_v surjects onto I_v^{ab} by the basic theory. □

Example 1.2.9 (Norm Hecke character and p-adic cyclotomic character)**.** Consider the Hecke character χ given by the norm $\|\|$ on the idèles of the general number field F. What is the corresponding Galois character ? First, we need to check that χ is algebraic. Indeed, it is given on $F_v^\times$ by $\alpha \mapsto |\alpha|_v$ for finite v (therefore unramified since the units have absolute value 1, not that it matters for the purposes of being algebraic), by

$$\alpha \mapsto |\alpha|_v = \mathrm{sgn}(\alpha)\alpha$$

for v real, and by

$$\alpha \mapsto |\alpha|_v = \alpha\overline{\alpha}$$

for v complex, where σ and $\overline{\sigma}$ are the complex conjugate pair of complex embeddings inducing the place v (recall that $|\alpha|_v$ means the square of the complex absolute value, since there is only one place for the two embeddings). So χ is algebraic with all the weights equal to 1. The corresponding p-adic character $\eta_{\chi,p}$ is given by

$$(\alpha_v)_v \mapsto \prod_{v \text{ real}} \mathrm{sgn}(\alpha_v) \prod_{v<\infty} |\alpha_v|_v \prod_{v|p} \prod_{\sigma: F_v \to \overline{\mathbf{Q}}_p} \sigma(\alpha_v) = \prod_{v \text{ real}} \mathrm{sgn}(\alpha_v) \prod_{v<\infty} |\alpha_v|_v \prod_{v|p} \mathrm{N}_{\mathbf{Q}_p}^{F_v} \alpha_v.$$

Note that in this case the image is all the way downstairs in $\mathbf{Q}_p$.

By definition of the global reciprocity map[5], Weil's associated p-adic Galois character $\rho_{\chi,p}$ has the property that for all finite places v except those over p (by Lemma 1.2.8, $\rho_{\chi,p}$ is unramified at all of these v so we can use the definition on π_v of the global reciprocity map for Galois groups of extensions of F unramified at v)

$$\rho_{\chi,p}(\mathrm{Frob}_v^{-1}) = \eta_{\chi,p}(\pi_v) = |\pi_v|_v = (\mathrm{N}v)^{-1}.$$

By the Čebotarev density theorem, this determines completely what $\rho_{\chi,p}$ is. In particular, if we cook up some other character $\rho' : G_F^{\mathrm{ab}} \to \mathbf{Q}_p$ with the same values on Frobenii for finite v not over p, then we know $\rho_{\chi,p} = \rho'$. Let ρ' be the p-adic cyclotomic character $G_F^{\mathrm{ab}} \to \mathbf{Z}_p^\times$. By definition of ρ' and the definition of the Frobenius elements as well as the explicit description of Galois groups of cyclotomic fields, for v not lying over p, we have

$$\mathrm{Frob}_v(\zeta_{p^n}) = \zeta_{p^n}^{\mathrm{N}v \mod p^n}.$$

and so $\rho'(\mathrm{Frob}_v) = \mathrm{N}v$. We conclude that $\rho_{\chi,p}$ is the p-adic cyclotomic character $G_F \to \mathbf{Z}_p^\times \subset \mathbf{Q}_p^\times$.

The calculation in Example 1.2.9 is not such a special case, as is shown by the last part of your exercise:

Lemma 1.2.10. *If F is totally real, then all algebraic Hecke characters for F are of the form $\|\cdot\|^a\eta$, where $a \in \mathbf{Z}$ and η is a finite-order Hecke character (i.e. a "Dirichlet–Hecke character").*

Proof. Let χ be such a character, with weights $\{a_\sigma\}_{\sigma:F\to\mathbf{C}}$. Since F is totally real, the unit group $\mathcal{O}_F^\times$ maps to $\mathbf{R}^{[F:\mathbf{Q}]}$ via the logarithm map

$$L : \alpha \mapsto (\log|\sigma(\alpha)|)_{\sigma:F\to\mathbf{R}},$$

and the Dirichlet unit theorem says that $L(\mathcal{O}_F^\times)$ is a full-rank lattice in the trace-0 subspace of $\mathbf{R}^{[F:\mathbf{Q}]}$. The kernel of L is the set of roots of unity in F, which is just ± 1, again because it is totally real. So Dirichlet's unit theorem is really giving us an identification

$$\mathcal{O}_F^\times \cong (\mathbf{Z}/2\mathbf{Z}) \times \mathbf{Z}^{[F:\mathbf{Q}]-1}.$$

For all but finitely many finite places v, we know that $\chi_v|_{\mathcal{O}_{F_v}^\times} = 1$ by Lemma 1.2.2. For the others, we know that χ_v is trivial on a finite-index subgroup of $\mathcal{O}_{F_v}^\times$ by Lemma 1.2.3. Consider

[5]In order to get the "right answer", we have to take the "geometric Frobenius" version of the global reciprocity map, which is defined (up to inertia at v) by $\pi_v \mapsto \mathrm{Frob}_v^{-1}$. Of course, this is the inverse of the usual convention (at least that of [Lan1994]) where you are supposed to take a uniformizer at v to $\mathrm{Frob}_{\mathfrak{p}}$ in the Galois group of any extension of F unramified at v

the diagonal map

$$\mathcal{O}_F^\times \to \prod_{v<\infty} \mathcal{O}_{F_v}^\times.$$

The preimage H in $\mathcal{O}_F^\times$ of $\prod_{v<\infty} \ker \chi_v|_{\mathcal{O}_{F_v}^\times}$ is a finite index subgroup of $\mathcal{O}_F^\times$, since $\prod_{v<\infty} \ker \chi_v|_{\mathcal{O}_{F_v}^\times}$ is a finite-index subgroup of $\prod_{v<\infty} \ker \mathcal{O}_{F_v}^\times$ by what we just wrote involving Lemma 1.2.2 and Lemma 1.2.3. By the (proof of the) structure theorem for finitely generated modules over a PID, there is a basis for the free part of $\mathcal{O}_F^\times$ such that we can write

$$L(H) \cong \prod_{i=1}^{[F:\mathbf{Q}]-1} m_i \mathbf{Z} \subset \prod_{i=1}^{[F:\mathbf{Q}]-1} \mathbf{Z} \cong L(\mathcal{O}_F^\times).$$

In particular, $L(H)$ is still a full-rank sublattice of the trace-zero subspace of $\mathbf{R}^{[F:\mathbf{Q}]}$ (we didn't need the structure theorem for this — could have just argued that the lattice is still cocompact).

The point is that we defined $H \subset \mathcal{O}_F^\times$ so that χ would kill it if we think of it as living in $\mathbf{A}_F^{\times,\text{fin}}$. If instead we think of $\alpha \in H$ diagonally embedded in $\mathbf{A}_F^\times$, we also know that χ kills it, since it is a Hecke character. This implies that χ kills the infinite part of α, i.e.

$$1 = \chi((\sigma(\alpha))_{\sigma:F\to\mathbf{R}} \times (1)_{v<\infty}) = \pm \prod_{\sigma:F\to\mathbf{R}} \sigma(\alpha)^{a_\sigma},$$

for all $\alpha \in H$. Taking the log of the absolute value, we get

$$0 = \sum_{\sigma:F\to\mathbf{R}} a_\sigma \log|\sigma(\alpha)| = \langle (a_\sigma)_{\sigma:F\to\mathbf{R}}, L(\alpha)\rangle.$$

Since this is true for all $\alpha \in H$, whose images under L span (over $\mathbf{R}$) the trace-0 subspace of $\prod_{\sigma:F\to\mathbf{R}} \mathbf{R}$, we conclude that the vector $(a_\sigma)_{\sigma:F\to\mathbf{R}}$ is orthogonal to the trace-zero subspace. Of course, the orthogonal complement of the trace-zero subspace is 1-dimensional and spanned by $(1,\dots,1)$. So we conclude that all the $a_\sigma \in \mathbf{Z}$ are equal to the same integer $a \in \mathbf{Z}$. Therefore, $\chi \cdot \|\cdot\|^{-a}$ is algebraic of weights all 0, i.e. finite-order, as desired. □

You mentioned in our meeting that it is a general theorem that algebraic Hecke characters always factor through $\mathrm{N}_{F'}^F : \mathbf{A}_F^\times \to \mathbf{A}_{F'}^\times$, where F' is the maximal CM subfield of F (that is, if F contains a CM subfield in the first place). Before I prove that, it makes sense to think first about whether Lemma 1.2.10 can be generalized to CM fields. Indeed, in the ways that affect the proof of Lemma 1.2.10, CM fields should only be a bit more complicated than totally real fields (at least according to the definition).

Let F be a CM field, and let F_0 be the maximal totally real subfield of F, so that F is totally imaginary of degree $[F:\mathbf{Q}] = 2r$ and F_0 is totally real of degree $[F_0:\mathbf{Q}] = r$.

Let $\sigma_1,\dots,\sigma_r$ be a set of complex embeddings of F such that there is exactly one σ_i per conjugate pair of embeddings. We consider the log maps $L : \mathcal{O}_F^\times \to \mathbf{R}^r$ (the i-th coordinate is $\log \circ |\sigma_i|^2 = \log \circ (\sigma_i \cdot \overline{\sigma}_i))$ and $L_0 : \mathcal{O}_{F_0}^\times \to \mathbf{R}^r$. The exact same argument as in the proof of Lemma 1.2.10 using the Dirichlet unit theorem shows (applied to L) that $a_{\sigma_i} + a_{\overline{\sigma_i}} \in \mathbf{Z}$ are all the same. The r real embeddings of F_0 are exactly the σ_i (they are invariant under conjugacy when restricted to F_0), and the Hecke character $\chi|_{\mathbf{A}_{F_0}^\times}$ is algebraic with weights $a_{\sigma_i} + a_{\overline{\sigma}_i}$, so the same argument as in the proof of Lemma 1.2.10 using the Dirichlet unit theorem shows that all of the $a_{\sigma_i} + a_{\overline{\sigma}_i}$ are the same for $i = 1, \ldots, r$. In other words, the totally real part of F gives us no new information. This gives us a nice constraint on the ∞-type of algebraic Hecke characters for F, but it is not a full classification of the algebraic Hecke characters like we got for F totally real.

Question 1.2.11. Let F be a CM field with $[F : \mathbf{Q}] = 2r$, let $\sigma_1, \overline{\sigma}_1, \ldots, \sigma_r, \overline{\sigma}_r$ be the various complex embeddings, and let $\{a_\sigma\}_{\sigma : F \to \mathbf{C}}$ be a collection of integers satisfying $a_\sigma + a_{\overline{\sigma}} = m$ for some fixed $m \in \mathbf{Z}$. Does there exist an algebraic Hecke character χ with weights equal to the $\{a_\sigma\}$?

For elliptic curves with complex multiplication (where we can use the statements of CM theory from [Sil1994, Ch. 2]), we can answer Question 1.2.11. Let F be an imaginary quadratic field, and let E be an elliptic curve[6] over a number field $L \supset F$ with complex multiplication by $\mathcal{O}_F$. Then there is an associated Hecke character $\xi_E : \mathbf{A}_L^\times \to \mathbf{C}^\times$. It is *defined* by

$$\xi_E(x) = \sigma_\infty(\alpha_E(x)) \mathrm{N}_F^L(x^{-1})_{v_{\sigma_\infty}},$$

where $\mathrm{N}_F^L(x^{-1}) \in \mathbf{A}_F^\times$ is the usual norm of the L-idèle x^{-1}, and σ_∞ denotes a choice of one of the two complex embeddings (fixed ahead of time); and $\alpha_E(x) \in F^\times$ denotes the unique element $\alpha_x \in F^\times$ such that

$$v_\mathfrak{p}(\alpha_x) = v_\mathfrak{p}(\mathrm{N}_F^L x)_\mathfrak{p}$$

for all finite prime $\mathfrak{p}$ of F, and for all fractional ideals $\mathfrak{a}$ of F, the action of $\mathrm{rec}_L(x)^{-1} \in \mathrm{Gal}(L^{\mathrm{ab}}/L)$ on $E(L^{\mathrm{ab}})$ restricts to $F/\mathfrak{a}$ to the morphism given by multiplication by $\alpha_x(\mathrm{N}_F^L x)^{-1}$. The multiplication by $(\mathrm{N}_F^L x)^{-1} \in \mathbf{A}_F^\times$ is defined via the decomposition

$$F/\mathfrak{a} = \bigoplus_\mathfrak{p} (F/\mathfrak{a})[\mathfrak{p}^\infty] \cong \bigoplus_\mathfrak{p} F_\mathfrak{p}/\mathfrak{a}_\mathfrak{p}.$$

The complex number $\xi_E(x)$ only depends on $\mathrm{N}_F^L x$ — the only potential problem in justifying this is the $\mathrm{rec}_L(x)$ that tells us how α_x restricts to $F/\mathfrak{a}$, but the restriction to $F/\mathfrak{a}$ only depends on the image of $\mathrm{rec}_L(x)$ in $\mathrm{Gal}(F^{\mathrm{ab}}/F)$, i.e. $\mathrm{rec}_K(\mathrm{N}_F^L(x))$. So (as a special case of the general

[6] Such an E always exists: take the quotient of $\mathbf{C}$ by the lattice in $\mathbf{C}$ given by $\mathcal{O}_F$ to get an elliptic curve over $\mathbf{C}$. This guy is defined over a number field (the j-invariant is algebraic; we don't need integrality so this is not deep [Sil1994, Proposition 2.1(b)]), and we can take this number field L to be larger in order to contain F (take the compositum with F), which is enough to make the full complex multiplication by $\mathcal{O}_K$ be defined over L by [Sil1994, Theorem 2.2(b)].

theorem that I promised to prove), $\xi_E(x)$ factors through $\mathrm{N}_F^L : \mathbf{A}_L^\times \to \mathbf{A}_F^\times$. Call the resulting character $\mathbf{A}_F^\times \to \mathbf{C}^\times$ ξ_F, so that $\xi_E = \xi_F \circ \mathrm{N}_F^L$. This is the one whose weights we are interested in (it only has two weights — one for each of the two conjugate complex embeddings of F). To figure out what the weights are, let $x \in L^\times$ be totally postive, so that any potential $\epsilon_{v_\sigma}(\sigma(x))$'s will be trivial (in other words, x is in the connected component of $(L \otimes \mathbf{R})^\times$, so we can figure out the weights of ξ_F just by looking at x). If we view x as being embedded in $\mathbf{A}_L^{\times,\mathrm{fin}}$, then we have $\mathrm{rec}_L(x) = 1$ (since the reciprocity map is trivial on $((L \otimes \mathbf{R})^\times)^\circ$), which implies that $\alpha_x = \mathrm{N}_F^L x \in F^\times$. On the other hand, since we are viewing x as being a finite idèle, the v_{σ_∞}-coordinate of $\mathrm{N}_F^L(x^{-1}) \in \mathbf{A}_F^{\times,\mathrm{fin}}$ is 1, and we deduce that

$$\xi_E(x) = \sigma_\infty(\alpha_x) = \sigma_\infty \circ \mathrm{N}_F^L(x).$$

It follows that that the algebraic Hecke character ξ_F satisfies $\xi_F(x) = \sigma_\infty(x)$ for x in an open subgroup of $F^\times \subset \mathbf{A}_F^{\times,\circ}$, and since ξ_F is trivial on $F^\times$ embedded in $\mathbf{A}_F^\times$, it is actually true that the ∞-type of ξ_F is $x \mapsto x^{-1}$. So we have produced an algebraic Hecke character ξ_F for F whose weights are -1 and 0. By changing our choice of σ_∞ to its conjugate (which simply has the effect of taking the conjugate of ξ_E), we can also produce an algebraic Hecke character for F of weights $(0, -1)$. Taking $\mathbf{Z}$-linear (multiplicative) combinations of of these two, or of one of them and $\|\cdot\|$, we see that all of the weight-tuples deemed possible by our previous discussion (which is just all pairs of integers since F has just two complex embeddings) can in fact be attained by algebraic Hecke characters associated to CM elliptic curves. So every algebraic Hecke character for an imaginary quadratic field F is given by a product of a finite-order character with a bunch of characters coming from elliptic curves with complex multiplication by F (the construction of which is explicit).

Remark 1.2.12. Though Question 1.2.11 is still useful (it is good to know whether these algebraic Hecke characters come from CM abelian varieties up to finite-order, e.g. for the purposes of Fontaine–Mazur conjecture for GL_1), the question of whether algebraic Hecke characters for a CM field F with weights satisfying $a_\sigma + a_{\overline{\sigma}} = m$ for a fixed $m \in \mathbf{Z}$ exist does not really require us to go in that direction. It's true that in the totally real case we had the particularly convenient character $\|\cdot\|$ that generated everything up to finite-order characters. But the point is that by finiteness of the class group, it suffices to define χ on $F^\times$ and on a finite-index subgroup $\prod_{v|\infty} F_v^\times \prod_{v<\infty} \mathcal{O}_{F_v}^\times$ (in a way that is compatible on the intersection); by finiteness of the class group of F, the subgroup of $\mathbf{A}_F^\times$ generated by this stuff is of finite-index, and hence χ can be extended to a global Hecke character in at least one way. Of course, we must define $\chi(F^\times) = 1$ (so that we end up with a Hecke character), so it remains to define it on a finite-index subgroup of $\prod_{v|\infty} F_v^\times \prod_{v<\infty} \mathcal{O}_{F_v}^\times$ so that it is trivial on the intersection with $F^\times$ and has the desired weights. What is the intersection with $F^\times$? The condition that an element of F is in $\mathcal{O}_{F_v}^\times$ for all finite v implies that this intersection is in $\mathcal{O}_F^\times$ (possibly smaller depending on the finite-index subgroup we choose). So we are looking to define a character χ on a finite-index subgroup of $\prod_{v|\infty} F_v^\times \times \prod_{v<\infty} \mathcal{O}_{F_v}^\times$ so that it is trivial on the intersection with $\mathcal{O}_F^\times$ and has ∞-type given by the $\{a_\sigma\}$ (all the σ are complex so there are no signs ϵ_v to worry about). As usual, such a χ (on this particular subset) must be given by $\prod_v \chi_v$, where for infinite v, $\chi_v(x_v) = x_v^{a_\sigma} \overline{x_v}^{a_{\overline{\sigma}}}$ and for finite v, χ_v is supposed to be trivial on the (finite-index) open subgroup of $\mathcal{O}_{F_v}^\times$ where we choose to define it. This completely determines how we will define χ, once we choose the open subgroups of $\mathcal{O}_{F_v}^\times$ (only finitely many of them proper) on which to define the χ_v to be defined and trivial. Our only constraint is that for all $\alpha \in \mathcal{O}_F^\times$ that happen to also be in all of those open subgroups, we need to have $\prod_\sigma \sigma(\alpha)^{a_\sigma} = 1$. The key point is that for $\alpha \in \mathcal{O}_F^\times$, the quantity $\prod_\sigma \sigma(\alpha)^{a_\sigma}$ is a priori a root of unity. We now justify this claim. Recall that if $z \in \mathbf{C}$ is an algebraic integer all of whose Galois conjugates have absolute value 1, then z is a root of unity[7]. So our first step should be to prove that $\prod_\sigma \sigma(\alpha)^{a_\sigma} \in S^1$ if $\alpha \in \mathcal{O}_F^\times$. For this, the condition that the $a_\sigma + a_{\overline{\sigma}} = m$ all coincide, together with the (easy part of the) Dirichlet unit theorem, is telling us that for all $\alpha \in \mathcal{O}_F^\times$,

$$0 = \sum_{i=1}^{r} \log |\sigma_i(\alpha)|^2 = \sum_{i=1}^{r} \log |\sigma_i(\alpha)^{a_{\sigma_i}} \overline{\sigma}_i(\alpha)^{a_{\overline{\sigma}_i}}| = \log \left| \prod_{\sigma: F \to \mathbf{C}} \sigma(\alpha)^{a_\sigma} \right|,$$

i.e. that

$$\left| \prod_{\sigma: F \to \mathbf{C}} \sigma(\alpha)^{a_\sigma} \right| = 1$$

Since F is a CM field, postcomposing by an element $g \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ commutes with complex conjugation, so it just permutes the unordered conjugate pairs of embeddings $\sigma : F \to \mathbf{C}$. Therefore,

$$\log \left| g \cdot \prod_{\sigma: F \to \mathbf{C}} \sigma(\alpha)^{a_\sigma} \right| = \sum_{i=1}^{r} \log |\sigma_i(\alpha)^{a_{g^{-1} \circ \sigma_i}} \overline{\sigma}_i(\alpha)^{a_{g^{-1} \circ \overline{\sigma}_i}}| = \sum_{i=1}^{r} \log |\sigma_i(\alpha)|^{2m} = 0$$

which implies that in fact all the Galois conjugates of $\prod_{\sigma: F \to \mathbf{C}} \sigma(\alpha)^{a_\sigma}$ have absolute value 1. Since $\alpha \in \mathcal{O}_F^\times$ are by definition algebraic over $\mathbf{Z}$, so is $\prod_{\sigma: F \to \mathbf{C}} \sigma(\alpha)^{a_\sigma}$. Therefore, this product is a root of unity. By Dirichlet's unit theorem again, $\mathcal{O}_F^\times$ is a finitely generated abelian group, so

[7] I learned this very nice fact from an exercise I did in Marcus' book on number fields a few years ago. Let $z \in \mathbf{C}$ be an algebraic integer all of whose Galois conjugates have absolute value 1. Then the minimal polynomial $f = \prod_{i=1}^{N} (X - z_i) \in \mathbf{Z}[X]$ ($z_1 = z$) of z, which is monic of degree N with coefficients in $\mathbf{Z}$, has i-th coefficient in $\mathbf{Z}$ of absolute value at most $\binom{N}{i}$. There are finitely many such polynomials (polynomials with $\mathbf{Z}$-coefficients of degree N where each coefficient has absolute value at most $\binom{N}{i}$). For every $n \geq 1$, we can consider the polynomial $f_n \in \mathbf{Z}[X]$ which is the minimal polynomial of z^n (which is an algebraic integer all of whose Galois conjugates have absolute value 1). The polynomial f_n has degree at most N and its i-th coefficient is of absolute value at most $\binom{\deg f_n}{i} \leq \binom{N}{i}$. Since the set $\{f = X^{\deg f} + a_{\deg f - 1} X^{\deg f - 1} + \cdots + a_0 \in \mathbf{Z}[X] : \deg f \leq N, a_i \leq \binom{N}{i}\}$ is finite, we conclude that the set of all Galois conjugates of all powers of z is finite, and hence that the set of powers of z is finite. This implies by the pigeonhole principle that z is a root of unity.

in fact there is a positive integer M such that

$$\prod_{\sigma:F\to\mathbf{C}} \sigma(\alpha)^{a_\sigma} \in \mu_M$$

for all $\alpha \in \mathcal{O}_F^\times$ (M can be taken as a maximum over a finite list of generators $\alpha : \mathcal{O}_F^\times$ of m such that $\prod_{\sigma:F\to\mathbf{C}} \sigma(\alpha)^{a_\sigma}$ is an m-th root of unity). Since $\mu_M \subset \mathbf{C}^\times$ is a finite group, it follows that there is a finite-index subgroup $H \subset \mathcal{O}_F^\times$ such that $\prod_\sigma \sigma(\alpha)^{a_\sigma} = 1$ for all $\alpha \in H$. By Chevalley's theorem [Che1951] on the congruence subgroup problem for $\mathcal{O}_F^\times$, there is a finite list of finite primes $\mathfrak{p}_1, \ldots, \mathfrak{p}_s$ of F and positive integers $f_1, \ldots, f_s$ such that

$$H \supseteq \mathcal{O}_F^\times \cap \prod_{\mathfrak{p}\notin\{\mathfrak{p}_i\}} \mathcal{O}_{F_v}^\times \times \prod_{i=1}^s (1+\mathfrak{p}^{f_i}).$$

Therefore, we conclude that there is an algebraic Hecke character χ for F which is given on $\prod_{v|\infty} F_v^\times \times \prod_{\mathfrak{p}\notin\{\mathfrak{p}_i\}} \mathcal{O}_{F_v}^\times \times \prod_{i=1}^s (1+\mathfrak{p}^{f_i})$ by being trivial on the finite parts and by $x_{v_\sigma} \mapsto x_{v_\sigma}^{a_\sigma} \overline{x_{v_\sigma}}^{a_{\overline{\sigma}}}$ for each complex embedding σ.

This concludes my discussion on constructing algebraic Hecke characters with prescribed weights over totally real and CM fields. The only remaining question seems to be whether the answer to Question 1.2.11 in the case of elliptic curves with complex multiplication by rings of integers of imaginary quadratic fields generalizes easily to abelian varieties with complex multiplication by rings of integers of CM fields.

Our computations with algebraic Hecke character associated to a CM elliptic curve also motivates the following fact.

Proposition 1.2.13. *Let L be a number field, and χ an algebraic Hecke character over L. If L contains a CM field, then let F be the maximal CM subfield of F. In this case, up to a finite-order character, χ factors through $\mathrm{N}_F^L : \mathbf{A}_L^\times \to \mathbf{A}_F^\times$. If L does not contain a CM field, then up to a finite-order character, χ is a power of $\|\cdot\|$.*

Proof. Let F_0 be the maximal totally real subfield of L. Then L contains a CM field if and only if it contains a square root of some totally negative element of F_0. In fact, if $a \in F_0$ is any totally negative element of F_0 such that $F_0(\sqrt{a}) \subset L$ in that colloquial sense, then $F_0(\sqrt{a})$ is the maximal[8] CM subfield of L.

In any event, we have an action of "complex conjugation" on F_0 and possibly on all of $F_0(\sqrt{a})$, if it exists, which is trivial on F_0 and acts by $\sqrt{a} \mapsto -\sqrt{a}$ on $F_0(\sqrt{a})$. The point is that

[8] Suppose b is another totally negative element of F_0 with a square root in L. Then the totally positive element ab would have a square root in L, which would itself have to be totally real. Since $F_0(\sqrt{ab})$ is totally real, by maximality of F_0, we conclude that ab has a square root in F_0, and hence that $F_0(\sqrt{a}) = F_0(\sqrt{b})$. The maximal CM subfield has F_0 as its totally real subfield, so any choice of a described in the text indeed works.

for any embeding $\sigma : F \to \mathbf{C}$ (real or complex), σ applied to the "complex conjugate" of an element α of F_0 or $F_0(\sqrt{a})$ is the same as $\overline{\sigma}(\alpha)$. In other words, for $x \in F_0(\sqrt{a})$, its "complex conjugation" is given by $\sigma^{-1}(\overline{\sigma}(x))$ for any choice of complex embedding σ (it doesn't matter which one, which is the point of being a CM field).

The key point in all of this is that the behavior of complex conjugation commuting in this way with the various complex embeddings $\sigma : L \to \mathbf{C}^\times$ is the defining property of $F_0(\sqrt{a})$ (or just F_0 if the former does not exist). More specifically: if it exists, then $F_0(\sqrt{a})$ has the property that $\sigma(F_0(\sqrt{a}))$ is the subfield of $\sigma(L)$ that is fixed by all elements of the commutator $[\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}), \{\overline{(\cdot)}\}]$ (a general element of this commutator in $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ doesn't necessarily act on $\sigma(L)$ since it is not necessarily Galois over $\mathbf{Q}$, but it is still Kosher to look at the fixed field since these are automorphisms of $\overline{\mathbf{Q}} \subset \mathbf{C}$ that still take $\sigma(L)$ to some other elements of $\overline{\mathbf{Q}}$). To prove this, just consider

$$\sigma(F_0(\sqrt{a})) \subset \sigma(L) \subset \overline{\mathbf{Q}} \subset \mathbf{C}.$$

If $\sigma(b) \in \sigma(L)$ with the property that $(g^{-1} \circ \overline{(\cdot)} \circ g)(\sigma(b)) = \overline{\sigma(b)}$ for all $g \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ (i.e. is fixed by the commutator $[\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}), \{\overline{(\cdot)}\}]$), then since $\sigma(F_0(\sqrt{a}))$ is also fixed by this commutator, we can translate this as saying that on $F_0(\sqrt{a}, b)$, the two complex embeddings $\overline{g \circ \sigma}$ and $g \circ \overline{\sigma}$ are the same for all $g \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. Since the complex embeddings of L are exhausted by the $g \circ \sigma$ for fixed σ and $g \in \mathrm{Gal}(\overline{\mathbf{Q}})$, this is saying that the automorphism $\tau^{-1} \circ \overline{\tau}$ of $F_0(\sqrt{a}, b)$ ("complex conjugation interpreted via the embedding τ") does not depend on $\tau : L \to \mathbf{C}$. This automorphism is nontrivial, for example because it takes $\sqrt{a}$ to $-\sqrt{a}$. We already said that this property was true for CM fields (obvious), and in fact it is equivalent to $F_0(\sqrt{a}, b)$ being a CM field. To see the other direction, just let $c = c_{F_0(\sqrt{a}, b)}$ be the automorphism given by $\tau^{-1}\overline{\tau}$ for all τ. By definition, $c^2 = \mathrm{id}$ (since complex conjugation satisfies this property). The fixed field of c is totally real, since it embeds via each τ into the subfield of $\mathbf{C}$ fixed by complex conjugation, i.e. $\mathbf{R}$. So we have produced a totally real subfield $K \subset F_0(\sqrt{a}, b)$ which is the fixed field of an involution and hence $[F_0(\sqrt{a}, b) : K] = 2$ (split it into $+1$ and -1-eigenspaces; the -1-eigenspace is there because c is nontrivial). On the other hand, $F_0(\sqrt{a}, b)$ is totally imaginary because $F_0(\sqrt{a})$ is. So $F_0(\sqrt{a}, b)$ is a CM field as claimed. Of course, this implies that $b \in F_0(\sqrt{a})$ by the fact that $F_0(\sqrt{a})$ is the maximal CM subfield of L, and hence $\sigma(F_0(\sqrt{a}))$ is, as claimed, the fixed field of $[\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}), \overline{(\cdot)}]$ in $\sigma(L)$.

If L contains no CM field, then the exact same argument shows that the maximal totally real subfield F_0 has the property that $\sigma(F_0)$ is the fixed field in $\sigma(L)$ of the exact same commutator subgroup (if $\sigma(b) \in \sigma(L)$ is fixed by that commutator, then we get again that $\tau^{-1} \circ \overline{\tau}$ does not depend on the choice of τ; if it is nontrivial then we get that $F_0(b) \subset L$ is CM, which is a contradiction, so we conclude that $F_0(b)$ is totally real and hence that $b \in F_0$ by the maximal property of F_0).

Let $\{a_\sigma\}$ be the weights of χ. By the same Dirichlet unit theorem arguments we have been giving, this implies that there is an $m \in \mathbf{Z}$ such that for all σ (including the real embeddings),

$$a_\sigma + a_{\overline{\sigma}} = m.$$

Applying the same Dirichlet unit theorem argument to $g \circ \chi$, which is algebraic of weights $b_\sigma = a_{g^{-1}\circ\sigma}$ we also know that

$$a_{g^{-1}\circ\sigma} + a_{g^{-1}\circ\overline{\sigma}} = m$$

for all $g \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ (the Galois group permutes the embeddings σ without necessarily preserving complex conjugate pairs).

Assume for now that L contains a CM field, so that $F_0(\sqrt{a})$ is the maximal CM subfield. We claim that a_σ depends only on $\sigma|_{F_0(\sqrt{a})}$. Let $\sigma, \sigma' : L \to \mathbf{C}$ such that they agree on $F_0(\sqrt{a})$. This means that there is a $g \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ which fixes $\sigma(F_0(\sqrt{a}))$ and satisfies $\sigma' = g \circ \sigma$. By the characterization of $F_0(\sqrt{a})$ as a fixed field explained in the previous paragraphs, we may assume for the purposes of what we want to deduce (by induction) that g restricts to $\sigma(L)$ to an embedding $\sigma(L) \to \mathbf{C}$ of the form

$$[h, \overline{(\cdot)}] = h^{-1} \circ \overline{(\cdot)} \circ h \circ \overline{(\cdot)},$$

and hence

$$\overline{h \circ \sigma'} = h \circ \overline{\sigma}.$$

We conclude that

$$a_\sigma = m - a_{\overline{\sigma}} = m - a_{h^{-1}\circ\overline{h\circ\sigma'}} = a_{h^{-1}\circ h\circ\sigma'} = a_{\sigma'},$$

as claimed. The exact same argument shows that if L contains no CM field, then a_σ depends only on the restriction of σ to the maximal totally real subfield $F_0 \subset L$.

Suppose that L contains no CM field, and let F_0 be the maximal totally real subfield. For each embedding $\tau : F_0 \to \mathbf{R}$, let $S_\tau = \{\sigma|\tau\}$ be the set of embeddings of L that agree with τ on F_0. This means that all the a_σ with $\sigma \in S_\tau$ are the same (this is the content of what we proved above), and that S_τ is stable under taking complex conjugates. Let a_τ be the common value of the a_σ for σ extending τ. In fact, the a_τ's are all the same and equal to $m/2$, thanks to the faact that S_τ is stable under taking complex conjugates and $a_\sigma + a_{\overline{\sigma}} = m$.

Then the Hecke character for L given by

$$\|\cdot\|^m_{\mathbf{A}^\times_{F_0}} \circ \mathrm{N}^L_{F_0}$$

is algebraic with weights equal to the $\{a_\sigma\}$. Indeed, if $x = (x_v)_v \in \mathbf{A}^\times_L$ with $x_v > 0$ for all real infinite places v and $x_v = 1$ at all the infinite places, then

$$\begin{aligned}\|N^L_{F_0}(x)\|^{m/2} &= \left(\prod_{\tau : F_0 \to \mathbf{R}} \left|\prod_{\sigma|\tau} x_\sigma\right|\right)^{a_\tau} \\ &= \prod_{\tau : F_0 \to \mathbf{R}} \prod_{\sigma|\tau} x_\sigma^{a_\tau} \\ &= \prod_{\sigma : L \to \mathbf{C}} x_\sigma^{a_\sigma}.\end{aligned}$$

Of course, for any place v of F_0, we have

$$\left|\prod_{w|v} \mathrm{N}^{L_w}_{(F_0)_v} x_w\right|_v = \prod_{w|v} \left|\mathrm{N}^{L_w}_{(F_0)_v} x_w\right|_v = \prod_{w|v} |x_w|_w$$

thanks to the normalization of the absolute values that we use to define $\|\cdot\|$; this means that our Hecke character $\|\cdot\|^{m/2}_{\mathbf{A}^\times_{F_0}} \circ \mathrm{N}^L_{F_0}$ can be written more succinctly as $\|\cdot\|^{m/2}_{\mathbf{A}^\times_L}$, and we have proved that this has the same weights as χ. It follows (from this and the fact that algebraic Hecke characters with all weights 0 are finite-order, as usual) that $\chi = \psi \cdot \|\cdot\|^{m/2}_{\mathbf{A}^\times_L}$ for some finite-order Hecke character ψ. This completes the case where L contains no CM subfield.

Now we consider the case where L contains a CM subfield, and $F = F_0(\sqrt{a})$ is the maximal such CM subfield. Previously in this letter (in Lemma 1.2.10 and the discussion after Question 1.2.11), I showed[9] that if F is a CM field, and $\{b_\tau\}_{\tau : F \to \mathbf{C}}$ a collection of integers satisfying $b_\tau + b_{\overline{\tau}} = m$ for all τ, then there is a Hecke character ξ for F that is algebraic with weights b_τ. In that case, since all the places in both F and L are complex, we don't have to think too much: if $x \in \mathbf{A}^\times_L$ has $x_v = 1$ for all finite v, then

$$\xi(\mathrm{N}^L_F x) = \xi\left(\left(\prod_{\sigma|\tau} x_\sigma\right)_{\tau : F \to \mathbf{C}}\right) = \prod_{\tau : F \to \mathbf{C}} \prod_{\sigma|\tau} x_\sigma^{b_\tau} = \prod_{\sigma : L \to \mathbf{C}} x_\sigma^{b_{\sigma|_F}}.$$

Setting the b_τ to a_σ (for any σ lying over τ) gives us ξ such that $\xi \circ \mathrm{N}^L_F$ is an algebraic Hecke character for L of weights equal to those of χ, and therefore $\chi = \psi \cdot (\xi \circ \mathrm{N}^L_F)$ for some finite-order ψ. This concludes. $\square$

[9] I have made no claims about whether this Hecke character is associated to a CM abelian variety — just that it exists, which I did prove after stating Question 1.2.11.

1.2.2 | Hodge–Tate decomposition of Weil's p-adic representations associated to Algebraic Hecke characters, and the Fontaine–Mazur conjecture for GL_1

Our investigations regarding the p-adic Galois character associated to an algebraic Hecke character by Weil showed that the cyclotomic character is prominent on the Galois side of this: in Lemma 1.2.10, we saw that up to finite-order characters, the powers of the cyclotomic character account for all the Galois characters coming from algebraic Hecke characters for totally real fields. We also saw that this remains true for any number field not containing any CM subfield in Proposition 1.2.13. For number fields containing a CM field, it is a priori unclear if there is an explicit way of understanding the corresponding Galois character (especially as I am not yet able to get them from CM abelian varieties), since the algebraic Hecke characters we constructed with prescribed weights for CM fields were not that explicit. However, in [Ser1989, Appendix to Ch. III] , it is shown that that in fact, being Hodge–Tate at the places above p is a necessary and sufficient condition for a p-adic Galois character to come from an algebraic Hecke character.

In our case, rather than simply having a 1-dimensional $\mathbf{Q}_p$-representation $G_{L_v} \to \mathbf{Q}_p^\times$, we have a character[10]

$$\rho : G_{L_v} \to E^\times,$$

where E is a finite extension of $\mathbf{Q}_p$. It is a representation on a 1-dimensional E-vector space, which is a $d = [E : \mathbf{Q}_p]$-dimensional $\mathbf{Q}_p$-vector space. It can be confusing at first how this should be considered to be Hodge–Tate, but the answer can be found in [Ser1989, Appendix to Ch. III, A.4]. This works for vector spaces of higher dimension over E (as is used constantly in [NT2021]), but we will just do it for the relevant case of dimension 1.

Let V be an abstract 1-dimensional E-vector space, so that ρ is reinterpreted as $G_{L_v} \to \mathrm{GL}(V)$. Also replace L_v with a general p-adic field K. Viewing V as a d-dimensional $\mathbf{Q}_p$-vector space, ρ gives it the structure of a d-dimensional p-adic Galois representation, together with the action of E that it already has. The Hodge–Tateness of V depends on the structure of $W = \mathbf{C}_K \otimes_{\mathbf{Q}} V$ as a d-dimensional $\mathbf{C}_K$-semilinear representation of G_K. But the "d dimensions" of this are deceptive, since W retains a $\mathbf{C}_{L_v}$-linear action of E that leaves the $\mathbf{C}_K$-coordinate alone. The point is that this breaks up into simultaneous eigenspaces whose systems of eigenvalues are given by the d embeddings $\sigma : E \to \overline{\mathbf{Q}}_p$. Writing $E = \mathbf{Q}_p(\alpha)$ for a primitive element $\alpha \in E$ with minimal polynomial $f_\alpha(X) \in \mathbf{Q}_p[X]$, we have

$$f_\alpha(X) = \prod_{\sigma : E \to \overline{\mathbf{Q}}_p} (X - \sigma(\alpha)),$$

[10]This is just the restriction to G_{L_v} of the global p-adic Galois character associated to some algebraic Hecke character.

and therefore

$$\begin{aligned}W &= \mathbf{C}_K \otimes_{\mathbf{Q}_p} V \\ &= \mathbf{C}_K \otimes_{\mathbf{Q}_p} \mathbf{Q}_p[X]/f_\alpha \\ &= \bigoplus_{\sigma: E \to \overline{\mathbf{Q}}_p} \mathbf{C}_K,\end{aligned}$$

where the E-action is seen explicitly on the σ-coordinate to be by $x \in E$ acting via multiplication by $\sigma(x)$. This direct sum decomposition of W also respects the Galois action in general: the σ-coordinate of $\sum_i c_i \otimes x_i$ is $\sum_i c_i \sigma(x_i)$, and the σ-coordinate of $g \cdot \sum_i c_i \otimes x_i = \sum_i g(c_i)\rho(g)x_i$ is $\sum_i g(c_i)\sigma \circ \rho(g)\sigma(x_i)$.

In other words, it is the usual Galois action on $\mathbf{C}_K$ except twisted by a multiplicative factor of $g \mapsto \sigma \circ \rho(g)$. So the study of the d-dimensional $\mathbf{C}_K$-semilinear representation W comes down to the study of the various $\mathbf{C}_K$-semilinear 1-dimensional representations (call them W_σ) given by $\sigma \circ \rho$ viewed as an element of $H^1(G_K, \mathbf{C}_K^\times)$. More precisely:

Lemma 1.2.14. *V (as defined above) is Hodge–Tate if and only if, for all $\sigma : E \to \overline{\mathbf{Q}}_p$, $\sigma \circ \rho \in H^1(G_K, \mathbf{C}_K^\times)$ is equivalent to χ^{n_σ} for some $n_\sigma \in \mathbf{Z}$.*

Proof. NB we have changed χ from meaning a Hecke character to meaning the p-adic cyclotomic character. We just saw that $W = \mathbf{C}_K \otimes_{\mathbf{Q}_p} V \cong \bigoplus W_\sigma$ where the W_σ has semilinear G_K-action given by $\sigma \circ \rho$, so certainly if $W_\sigma = \chi^{n_\sigma}$ in $H^1(G_K, \mathbf{C}_K^\times)$, then by the fact that $H^1(G, \mathrm{GL}_n(M))$ (where M has a G-action which induces in the natural way an action on $\mathrm{GL}_n(M)$) parametrizes 1-dimensional M-semilinear representations of G,

$$W \cong \bigoplus_\sigma W_\sigma \cong \bigoplus_\sigma \mathbf{C}_K(n_\sigma)$$

and hence V is Hodge–Tate with Hodge–Tate weights equal to the n_σ (or maybe the $-n_\sigma$ depending on the convention).

Conversely, if V is Hodge–Tate, we have two G_K-equivariant splittings into 1-dimensional (hence irreducible and indecomposable) $\mathbf{C}_K$-semilinear G_K-representations

$$W \cong \bigoplus_\sigma W_\sigma$$

and

$$W \cong \bigoplus_{i=1}^d \mathbf{C}_K(n_i).$$

By Schur's lemma (at least the part of it that still definitely holds in the semilinear setting, namely that a G_K-equivariant map between irreducibles is either 0 or an isomorphism) applied to

$$\mathbf{C}_K(n_i) \to \bigoplus_{i=1}^{f} \mathbf{C}_K(n_i) \cong \bigoplus_{\sigma} W_\sigma \to W_\tau$$

for each i, τ, we conclude that the W_σ are G_K-equivariantly isomorphic to some permutation of the $\mathbf{C}(n_i)$, as desired. □

We will ultimately want to check that the p-adic Galois character ρ coming from an algebraic Hecke character is Hodge–Tate. But by Weil's construction and local–global compatibility of class field theory, being algebraic will only tell us about what ρ looks like on on a small neighborhood of 1 in the inertia group of G_K. Luckily, for the purposes of Hodge–Tateness, this does not matter. This fact is proved in [Ser1989, Appendix to Ch. III, A.1] and also in [BC2009b, Theorem 2.4.6] by essentially the same technique.

Lemma 1.2.15. *Let $\rho : G_K \to \mathrm{GL}(V)$ be a p-adic Galois representation. Let K' be a finite extension of*[11] *K^{ur}, so that $G_{K'}$ is an finite-index closed (hence open) subgroup of I_K. The Hodge–Tateness and Hodge–Tate weights of ρ are independent of whether we look at ρ or $\rho|_{G_{K'}}$.*

Proof. From Ax–Sen–Tate theory, we know that $\mathbf{C}_K^{G_{K'}} = \widehat{K'}$. We will prove that the natural morphism of graded $\widehat{K'}$-vector spaces

$$\widehat{K'} \otimes_K \mathbf{D}_{\mathrm{HT},G_K}(V) \to \mathbf{D}_{\mathrm{HT},G_{K'}}(V)$$

(given just by multiplying by $\widehat{K'}$ in the $\mathbf{B}_{\mathrm{HT}}$-tensor-coordinate of $\mathbf{D}_{\mathrm{HT},G_K}(V)$ to get something still fixed by $G_{K'}$) is an isomorphism. This is enough, because it tells us that $\mathbf{D}_{\mathrm{HT},G_{K'}}(V)$ is a graded vector space of the same dimension and whose graded parts have the same dimension (albeit over different field) as those of $\mathbf{D}_{\mathrm{HT},G_K}(V)$.

We begin by proving the isomorphism in the case where K' is a finite Galois extension of K. This is essentially just Galois descent. In particular,

$$\mathbf{D}_{\mathrm{HT},G_{K'}}(V) = (\mathbf{B}_{\mathrm{HT}} \otimes_{\mathbf{Q}_p} V)^{G_{K'}}$$

has an action of G_K that factors through $G_K/G_{K'} = \mathrm{Gal}(K'/K)$, and so

$$\mathbf{D}_{\mathrm{HT},G_K}(V) = (\mathbf{B}_{\mathrm{HT}} \otimes_{\mathbf{Q}_p} V)^{G_K} = \mathbf{D}_{\mathrm{HT},G_{K'}}(V)^{\mathrm{Gal}(K'/K)}.$$

By Hilbert 90, for all $n \geq 1$ (and in particular n equal to the dimension of $\mathbf{D}_{\mathrm{HT},G_{K'}}(V)$ or of one of its graded pieces),

$$H^1(\mathrm{Gal}(K'/K), \mathrm{GL}_n(K')) = 1,$$

[11] Any finite extension of any unramified extension would also work, though of course the Galois group would cease to be an open subgroup of the inertia.

which means that the K'-semilinear $\mathrm{Gal}(K'/K)$-module $\mathbf{D}_{\mathrm{HT},G_{K'}}(V)$ (and, in fact, all of its graded pieces) is trivializable, hence

$$\mathbf{D}_{\mathrm{HT},G_K}(V) = \mathbf{D}_{\mathrm{HT},G_{K'}}(V)^{\mathrm{Gal}(K'/K)} \cong (K'^{\oplus n})^{\mathrm{Gal}(K'/K)} = K^{\oplus n}$$

which means that (either by applying the above to the graded pieces or just by remarking that the multiplication map respects the grading) indeed the multiplication map

$$K' \otimes \mathbf{D}_{\mathrm{HT},G_K}(V) \to \mathbf{D}_{\mathrm{HT},G_{K'}}(V)$$

is an isomorphism of graded K'-vector spaces. If K'/K is finite but not Galois, then consider its Galois closure $M \supset K' \supset K$. We know that the natural map

$$M \otimes \mathbf{D}_{\mathrm{HT},G_K}(V) \to \mathbf{D}_{\mathrm{HT},G_M}(V)$$

is an isomorphism of graded M-vector spaces, and we get the desired result just by taking $\mathrm{Gal}(M/K')$-fixed vectors on both sides.

Everything above works fine if K is replaced with K^{ur}, so it remains to check that it works with $K' = K^{\mathrm{ur}}$. The Galois group $G_{K^{\mathrm{ur}}}$ is the same as I_K, and $\mathrm{Gal}(K^{\mathrm{ur}}/K) = \mathrm{Gal}(\overline{k}/k) = G_k$, where k is the residue field of K. We must now do some "integral theory." In particular,

$$\mathbf{D}_{\mathrm{HT},G_{K^{\mathrm{ur}}}}(V)$$

is a graded $\widehat{K^{\mathrm{ur}}}$-vector space that comes with a $\widehat{K^{\mathrm{ur}}}$-semilinear $G_K/G_{K^{\mathrm{ur}}} = G_k$-action. We claim that it admits a G_k-invariant $\mathcal{O}_{K^{\mathrm{ur}}}$-lattice. To do this, just pick a random (not necessarily Galois-invariant) $\mathcal{O}_{K^{\mathrm{ur}}}$-lattice $\Lambda_0 \subset \mathbf{D}_{\mathrm{HT},G_{K^{\mathrm{ur}}}}(V)$. Writing the $\widehat{K^{\mathrm{ur}}}$-semilinear representation G_k-module $\mathbf{D}_{\mathrm{HT},G_{K^{\mathrm{ur}}}}(V)$ as a (continuous) 1-cocycle

$$\Xi : G_k \to \mathrm{GL}_d(\widehat{K^{\mathrm{ur}}}),$$

where the $\widehat{K^{\mathrm{ur}}}$-basis used to write down Ξ is a basis for Λ_0. The subgroup of $g \in G_k$ that stabilize Λ_0 is just $\Xi^{-1}(\mathrm{GL}_d(\mathcal{O}_{K^{\mathrm{ur}}}))$. Since $\mathrm{GL}_d(\mathcal{O}_{K^{\mathrm{ur}}})$ is open in $\mathrm{GL}_d(\widehat{K^{\mathrm{ur}}})$, it follows that Λ_0 is stabilized by a finite-index subgroup of G_k. Taking Λ to be the sum of the requisite finite number of G_k-translates of Λ_0, we can therefore produce a G_k-invariant lattice $\Lambda \subset \mathbf{D}_{\mathrm{HT},G_{K^{\mathrm{ur}}}}$. Our goal is to prove that

$$\widehat{K^{\mathrm{ur}}} \otimes_K \mathbf{D}_{\mathrm{HT},G_K} = \widehat{K^{\mathrm{ur}}} \otimes_K \mathbf{D}_{\mathrm{HT},G_{K^{\mathrm{ur}}}}^{G_k} \to \mathbf{D}_{\mathrm{HT},G_{K^{\mathrm{ur}}}}$$

is an isomorphism, which would follow after inverting π_K from showing that

$$\mathcal{O}_{K^{\mathrm{ur}}} \otimes_{\mathcal{O}_K} \Lambda^{G_k} \to \Lambda$$

is an isomorphism.

The proofs from [Ser1989] and [BC2009b] are slightly different from here, though they both essentially use technique of "successive p-adic approximation of cocycles" + Hilbert 90. Serre finishes by proving that $H^1(G_k, \mathrm{GL}_d(\mathcal{O}_{K^{\mathrm{ur}}}))$ is trivial, which he does directly by filtering $\mathrm{GL}_d(\mathcal{O}_{K^{\mathrm{ur}}})$ by subgroups of matrices congruent to the identity modulo higher and higher powers of π_K, using the fact that (thanks to Hilbert 90 and the usual compatibility of group cohomology with taking inverse limits) $H^1(G_k, \mathrm{GL}_d(\overline{k})) = H^1(G_k, M_{d\times d}(\overline{k})) = 0$.

I will explain what Brinon–Conrad do, which uses all the same stuff but I thought was more interesting. Continuing with the notation $d = \dim_{\widehat{K^{\mathrm{ur}}}} \mathbf{D}_{\mathrm{HT},G_{K^{\mathrm{ur}}}} = \mathrm{rk}_{\mathcal{O}_{\widehat{K^{\mathrm{ur}}}}} \Lambda$, we know that $\Lambda/\pi_K\Lambda$ is a $\mathcal{O}_{\widehat{K^{\mathrm{ur}}}}/(\pi_K) = \overline{k}$-vector space of dimension d. Since $\pi_K \in K$ is fixed by G_k, the d-dimensional $\overline{k}$-vector space $\Lambda/\pi_K\Lambda$ inherits a $\overline{k}$-semilinear G_k-action. For $[v] \in \Lambda/\pi_K\Lambda$, we have

$$\begin{aligned}\mathrm{Stab}_{G_k}([v]) &= \{g \in G_k : gv - v \in \pi_K\Lambda\} \\ &\supseteq \{g \in G_k : \Xi(g) \in I + M_{d\times d}(\pi_K \mathcal{O}_{\widehat{K^{\mathrm{ur}}}})\} \\ &= \Xi^{-1}(I + M_{d\times d}(\pi_K \mathcal{O}_{\widehat{K^{\mathrm{ur}}}})),\end{aligned}$$

where now Ξ stands for the 1-cocycle $G_k \to \mathrm{GL}_d(\mathcal{O}_{\widehat{K^{\mathrm{ur}}}})$ coming from a choice of basis of Λ that represents the $\mathcal{O}_{\widehat{K^{\mathrm{ur}}}}$-semilinear G_k-action on Λ. Since Ξ is continuous and $I + M_{d\times d}(\pi_K \mathcal{O}_{\widehat{K^{\mathrm{ur}}}})$ is open in $\mathrm{GL}_d(\mathcal{O}_{\widehat{K^{\mathrm{ur}}}})$, we conclude that the action of G_k on $\Lambda/\pi_K\Lambda$ has open stabilizers, i.e. that it is continuous for the discrete topology on $\Lambda/\pi_K\Lambda$. By Hilbert 90,

$$H^1(G_k, \mathrm{GL}_d(\overline{k})) = 1,$$

where this is continuous cohomology with the discrete topology on $\overline{k}$, so (thanks to the continuity we just proved) we have a G_k-equivariant isomorphism $\Lambda/\pi_K\Lambda \cong \overline{k}^{\oplus d}$. This already tells us the "mod π_K" version of the result:

$$\overline{k} \otimes_k (\Lambda/\pi_K\Lambda)^{G_k} \to \Lambda/\pi_K\Lambda$$

is an isomorphism.

To finish, we use additive Hilbert 90 to argue that $H^1(G_k, \Lambda/\pi_K\Lambda) = 0$, so by the long exact sequence,

$$\Lambda^{G_k}/\pi_K\Lambda^{G_k} = (\Lambda/\pi_K\Lambda)^{G_k},$$

which we just showed is a dimension-d k-vector space. Lifting a basis of this k-vector space to

Λ^{G_k}, we see that there are $e_1, \ldots, e_d \in \Lambda^{G_k}$ such that every element of Λ^{G_k} is within a multiple of π_K (that is, $\pi_K \cdot x$ for some $x \in \mathcal{O}_{\widehat{K^{\mathrm{ur}}}}$) of a $\mathcal{O}_{\widehat{K^{\mathrm{ur}}}}$-linear combination of the $e_1, \ldots, e_d$. By approximating x again by a linear combination of the e_i and repeating ad infinitum, we end up (by convergence of series whose elements go to zero) seeing that $e_1, \ldots, e_d$ actually span Λ^{G_k} over $\mathcal{O}_{\widehat{K^{\mathrm{ur}}}}$. Being now a finitely-generated torsion-free $\mathcal{O}_{\widehat{K^{\mathrm{ur}}}}$-module, Λ^{G_k} is free, and its rank is exactly d because that is the dimension of its reduction mod π_K. The map

$$\mathcal{O}_{\widehat{K^{\mathrm{ur}}}} \otimes_{\mathcal{O}_K} \Lambda^{G_k} \to \Lambda$$

is a map of free $\mathcal{O}_{\widehat{K^{\mathrm{ur}}}}$-modules of the same rank which is an isomorphism modulo π_K, so it is an isomorphism (e.g. by the same successive approximation arguments as before, since the map here is $\mathcal{O}_{\widehat{K^{\mathrm{ur}}}}$-linear), as desired. □

Let us consider an algebraic Hecke character $\xi : \mathbf{A}_F^\times \to \mathbf{C}^\times$ with weights $\{a_\sigma\}$. Denote Weil's corresponding p-adic Galois character by $\rho_{\xi,p} : G_F^{\mathrm{ab}} \to E^\times$, where E is a finite extension of $\mathbf{Q}_p$. We will allow E to be any such finite extension that contains the image of $\rho_{\xi,p}$ in $\overline{\mathbf{Q}}_p^\times$ (there doesn't seem to be a canonical choice of such a field except possibly the smallest one). Let v be a finite place of F over p. By local-global compatibility of class field theory and the definitions in Weil's construction, $\rho_{\xi,p}$ is given on an open subgroup H_v of $I_{F_v}^{\mathrm{ab}} \cong \mathcal{O}_{F_v}^\times$ (namely, the one on which ξ_v is trivial) by

$$\alpha \mapsto \prod_{\tau : F_v \to \overline{\mathbf{Q}}_p} \tau(\alpha)^{a_{\iota_\infty \circ \iota_p^{-1} \circ \tau \circ (F \to F_v)}} \in E^\times.$$

Let K be a finite extension of $\mathbf{Q}_p$ large enough to contain F_v and all the Galois conjugates of E, and also large enough so that the open subgroup $G_K \subset G_{F_v}$ has the property that $G_K \cap I_{F_v} = I_K$ lands inside $H \subset I_{F_v}^{\mathrm{ab}}$ under the projection to the abelianization $I_K \to I_K^{\mathrm{ab}}$ followed by the restriction map $I_K^{\mathrm{ab}} \to I_{F_v}^{\mathrm{ab}}$. We also might as well replace K with its Galois closure over $\mathbf{Q}_p$, in order to assume that $K/\mathbf{Q}_p$ is Galois (this is most likely not necessary but makes things a little clearer as usual). By Lemma 1.2.15, the Hodge–Tateness and Hodge–Tate weights of $\rho_{\xi,p}$ are the same as those of ρ_K, defined by

$$\rho_K : G_K \to G_{F_v} \to G_{F_v}^{\mathrm{ab}} \overset{\rho_{\xi,p}}{\to} E^\times.$$

Of course, ρ_K factors through G_K^{ab}, so a more reasonable way to write it is

$$\rho_K : G_K \to G_K^{\mathrm{ab}} \to G_{F_v}^{\mathrm{ab}} \overset{\rho_{\xi,p}}{\to} E^\times,$$

and to make things easier we can also abuse notation and consider ρ_K as a character $G_K^{\mathrm{ab}} \to G_{F_v}^{\mathrm{ab}} \overset{\rho_{\xi,p}}{\to} E^\times$ which restricts on $I_K^{\mathrm{ab}} \cong \mathcal{O}_K^\times$ to (using class field theory to identify the restriction maps between inertia groups with the norm between unit groups)

$$\mathcal{O}_K^\times \overset{\mathrm{N}_{F_v}^K}{\to} \mathcal{O}_{F_v}^\times \to E^\times.$$

By construction of K, $\mathrm{N}_{F_v}^K(\mathcal{O}_K^\times) \subset H \subset I_{F_v}^{\mathrm{ab}} \cong \mathcal{O}_{F_v}^\times$, so we can write via the identification $I_K^{\mathrm{ab}} \cong \mathcal{O}_K^\times$ that

$$\begin{aligned}
\rho_K|_{I_K^{\mathrm{ab}}}(x) &= \prod_{\tau: F_v \to \overline{\mathbf{Q}}_p} \tau\left(\mathrm{N}_{F_v}^K x\right)^{a_{\iota_\infty \circ \iota_p^{-1} \circ \tau \circ (F \to F_v)}} \\
&= \prod_{\tau: F_v \to \overline{\mathbf{Q}}_p} \tau\left(\prod_{\sigma \in \mathrm{Gal}(K/F_v)} \sigma(x)^{a_{\iota_\infty \circ \iota_p^{-1} \circ \tau \circ (F \to F_v)}}\right) \\
&= \prod_{\tau: F_v \to \overline{\mathbf{Q}}_p} \tau\left(\tau^{-1} \prod_{\substack{\sigma \in \mathrm{Gal}(K/\mathbf{Q}_p) \\ \sigma|_{F_v} = \tau}} \sigma(x)^{a_{\iota_\infty \circ \iota_p^{-1} \circ \tau \circ (F \to F_v)}}\right) \\
&= \prod_{\tau: F_v \to \overline{\mathbf{Q}}_p} \prod_{\substack{\sigma \in \mathrm{Gal}(K/\mathbf{Q}_p) \\ \sigma|_{F_v} = \tau}} \sigma(x)^{a_{\iota_\infty \circ \iota_p^{-1} \circ \tau \circ (F \to F_v)}} \\
&= \prod_{\sigma \in \mathrm{Gal}(K/\mathbf{Q}_p)} \sigma(x)^{a_{\iota_\infty \circ \iota_p^{-1} \circ \sigma|_{F_v} \circ (F \to F_v)}}.
\end{aligned}$$

So regardless of how big we choose K, the representation that we are interested in the Hodge–Tateness of is going to be a product of powers of embeddings of K into $\overline{\mathbf{Q}}_p$, where the powers only depend on the restriction of that embedding to F_v.

The key point in all of this is to observe that all such p-adic Galois characters that land in $E^\times$ can be rewritten as products of powers of the various $\tau^{-1}\mathrm{N}_{\tau E}^K$ for embeddings $\tau : E \to \overline{\mathbf{Q}}_p$. Serre has an argument using algebraic groups which I am perfectly comfortable with, but I will not reproduce his argument because (for the purposes of what I need to prove) that perspective does not seem to add anything useful and makes things more confusing for the purposes of explicitly determining the Hodge–Tate weights.

First, we rewrite, for $x \in K$ and an embedding $\tau : E \to \overline{\mathbf{Q}}_p$, the candidate "basis element" for our characters is

$$\tau^{-1}\mathrm{N}_{\tau E}^K(x) = \tau^{-1} \prod_{\sigma \in \mathrm{Gal}(K/\tau E)} \sigma(x) = \prod_{\substack{\sigma \in \mathrm{Gal}(K/\mathbf{Q}_p) \\ \sigma^{-1}|_E = \tau}} \sigma(x).$$

Since ρ_K lands in E, we know that for any $g \in \mathrm{Gal}(K/E)$, we have $g \circ \rho_K = \rho_K$, i.e. that

$$\prod_{\sigma \in \mathrm{Gal}(K/\mathbf{Q}_p)} \sigma(x)^{a_{\iota_\infty \circ \iota_p^{-1} \circ \sigma|_{F_v} \circ (F \to F_v)}} = \prod_{\sigma \in \mathrm{Gal}(K/\mathbf{Q}_p)} \sigma(x)^{a_{\iota_\infty \circ \iota_p^{-1} \circ g^{-1} \circ \sigma|_{F_v} \circ (F \to F_v)}}.$$

The notation is getting unwieldy, so let us call $b_\sigma := a_{\iota_\infty \circ \iota_p^{-1} \circ \sigma|_{F_v} \circ (F \to F_v)}$. The equation above being true for all $x \in K$ tells us that $b_\sigma = b_{\sigma'}$ if $\sigma, \sigma' \in \mathrm{Gal}(K/\mathbf{Q}_p)$ have the property that $g \circ \sigma = \sigma'$ for some $g \in \mathrm{Gal}(K/E)$, i.e. if $\sigma^{-1}|_E = (\sigma')^{-1}|_E$. For a given embedding $\tau : E \to \overline{\mathbf{Q}}_p$, we may define b_τ to be the common value of the b_σ such that $\sigma^{-1}|_E = \tau$. Then we get

$$\begin{aligned} \rho_K|_{I_K^{\mathrm{ab}}}(x) &= \prod_{\sigma \in \mathrm{Gal}(K/\mathbf{Q}_p)} \sigma(x)^{b_\sigma} \\ &= \prod_{\tau : E \to \overline{\mathbf{Q}}_p} \prod_{\substack{\sigma \in \mathrm{Gal}(K/\mathbf{Q}_p) \\ \sigma^{-1}|_E = \tau}} \sigma(x)^{b_\tau} \\ &= \prod_{\tau : E \to \overline{\mathbf{Q}}_p} \left(\prod_{\substack{\sigma \in \mathrm{Gal}(K/\mathbf{Q}_p) \\ \sigma^{-1}|_E = \tau}} \sigma(x) \right)^{b_\tau} \\ &= \prod_{\tau : E \to \overline{\mathbf{Q}}_p} \left(\tau^{-1} \mathrm{N}^K_{\tau E}(x) \right)^{b_\tau}. \end{aligned}$$

The numbers b_τ are common values of some subsets of the a_σ depending on how big E is (but, as we saw by the end, not on how big K is, which is good because we don't expect the Hodge–Tate weights to depend on the choice of K). If $E = \mathbf{Q}_p$, for example, then all the $\sigma \in \mathrm{Gal}(K/\mathbf{Q}_p)$ restrict to the same thing on E, and there is just one τ, so we may rewrite ρ_K as a power of the norm from K to $\mathbf{Q}_p$, and end up with the situation of Example 1.2.9 (which ends up showing that the Galois representation into $\mathbf{Q}_p^\times$ is Hodge–Tate with Hodge–Tate weight equal to that power). If $E = K$, on the other hand, then the b_τ are exactly the same as the a_σ's. Anyway, this all is consistent with the (eventual) fact that the b_τ are the Hodge–Tate weights of $\rho_{\xi,p}$, i.e. that $W = \mathbf{C}_K \otimes_{\mathbf{Q}_p} V$ (V being the 1-dimensional E-vector space that $\rho_{\xi,p}$ defines an action on) is a direct sum of $\mathbf{C}_K(b_\tau)$, where the $\mathbf{C}_K(b_\tau)$ is the W_τ that was defined in Lemma 1.2.14. In fact, this is what we prove now.

Proposition 1.2.16. *Let $\tau : E \to \overline{\mathbf{Q}}_p$ be an embedding. The Galois character $\chi_{\tau,E} : I_K^{ab} = G_{\widehat{K^{ur}}} \to E^\times$ given by $\alpha \mapsto \tau^{-1}\mathrm{N}^K_{\tau E}(\alpha)$ via the local class field theory identification $I_K^{ab} \cong \mathcal{O}_K^\times$ is Hodge–Tate with Hodge–Tate weight 1 in the τ-component and 0 in all the others (in the sense of Lemma 1.2.14).*

Proof. By Lemma 1.2.14, it suffices to check that $\tau \circ \chi_{\tau,E} = \chi$ in $H^1(G_{\widehat{K^{\mathrm{ur}}}}, \mathbf{C}_K^\times)$ and that $\sigma \circ \chi_{\tau,E} = 1$ in $H^1(G_{\widehat{K^{\mathrm{ur}}}}, \mathbf{C}_K^\times)$ for all $\sigma \neq \tau$. Thanks to the definition of $\chi_{\tau,E}$, this is the same as saying that (via the class field theory isomorphism $I_K^{\mathrm{ab}} \cong \mathcal{O}_K^\times$)

$$\big[x \mapsto \mathrm{N}^K_{\tau E}(x)\big] = \chi \in H^1(G_{\widehat{K^{\mathrm{ur}}}}, \mathbf{C}_K^\times), \qquad \big[x \mapsto \sigma \circ \mathrm{N}^K_{\tau E}(x)\big] = 1 \in H^1(G_{\widehat{K^{\mathrm{ur}}}}, \mathbf{C}_K^\times)$$

whenever $\sigma : \tau E \to \overline{\mathbf{Q}}_p$ is an embedding not equal to the inclusion $\tau E \to K$. By Lemma 1.2.15, if we can extend $x \mapsto \mathrm{N}^K_{\tau E}$ to all of G_K^{ab} then it will suffice to do this with $\widehat{K^{\mathrm{ur}}}$ replaced with K. In Galois–theoretic terms, this map is simply the one $I_K^{\mathrm{ab}} \to I_{\tau E}^{\mathrm{ab}}$ (i.e. $\mathcal{O}_K^\times \to \mathcal{O}_{\tau E}^\times$) given by the norm. We can't really use the same formula to extend to the profinite completion of $K^\times$, since that would land in the profinite completion of $(\tau E)^\times$ rather than $(\tau E)^\times$. Instead, we choose a uniformizer π of K, which induces a decomposition $K^\times \cong \mathcal{O}_K^\times \times \mathbf{Z}$ and therefore $G_K^{\mathrm{ab}} \cong K^\times \times \widehat{\mathbf{Z}}$ by class field theory. Define the Galois character

$$\chi_{\tau E,\pi} : G_K \to \mathcal{O}_{\tau E}^\times \subset (\tau E)^\times$$

by

$$G_K \to G_K^{\mathrm{ab}} \cong \mathcal{O}_K^\times \times \widehat{\mathbf{Z}} \to \mathcal{O}_K^\times \xrightarrow{\mathrm{N}^K_{\tau E}} \mathcal{O}_{\tau E}^\times.$$

We clearly have $\chi_{\tau E,\pi}|_{I_K^{\mathrm{ab}}} = \big[x \mapsto \mathrm{N}^K_{\tau E} x\big]$, so (as mentioned already) by Lemma 1.2.15, it suffices to check that

$$\chi_{\tau E,\pi} = \chi \in H^1(G_K, \mathbf{C}_K^\times), \qquad \sigma \circ \chi_{\tau E,\pi} = 1 \in H^1(G_K, \mathbf{C}_K^\times)$$

for every embedding $\sigma : \tau E \to \overline{\mathbf{Q}}_p$ not equal to the inclusion $\tau E \to K$. The key point now is that $\chi_{\tau E,\pi}$ is not just any Galois character: it is exactly the Tate module of the Lubin–Tate formal group $\mathscr{F}_\pi$ over $\mathcal{O}_{\tau E}$. This is essentially the statement of [Ser1967, Ch. 3, Theorem 3(c, e)], once one remembers that our Artin reciprocity map is the inverse of Serre's (which we did to accommodate Example 1.2.9) and therefore the extra inverse is not necessary. Indeed, $\mathfrak{F}_\pi(\mathfrak{m}_{\overline{\tau E}})[\pi_{\tau E}^n] \cong \mathcal{O}_{\tau E}/(\pi_{\tau E}^n)$ as $\mathcal{O}_E$-modules, so the Tate module is identified with

$$\varprojlim_n \mathfrak{F}_\pi(\mathfrak{m}_{\overline{\tau E}})[p^n] = \varprojlim_n \mathfrak{F}_\pi(\mathfrak{m}_{\overline{\tau E}})[\pi_{\tau E}^{e(\tau E|\mathbf{Q}_p)n}] = \varprojlim_n \mathcal{O}_{\tau E}/(\pi_{\tau E}^{e(\tau E|\mathbf{Q}_p)n}) = \mathcal{O}_{\tau E}.$$

[Ser1967, Ch. 3, Theorem 3(c)] (except with u instead of u^{-1} since our reciprocity map already has an inverse added) says that the $I_K^{\mathrm{ab}} \cong \mathcal{O}_K^\times$-action on this Tate module is given on the each coordinate of the inverse limit by multiplication by the restriction (i.e. norm) downstairs in $\mathcal{O}_E^\times$. [Ser1967, Ch. 3, Theorem 3(e)] says that the $\widehat{\mathbf{Z}} \subset G_K^{\mathrm{ab}}$-action is trivial. Putting this together, we indeed see that $\chi_{\tau E,\pi} : G_K \to \mathcal{O}_E^\times$ is the action of G_K on the Tate module of $\mathfrak{F}_\pi$.

In the sense of [Tat1967], we can look at the p-divisible group $\mathscr{G} = \mathfrak{F}_\pi[p^\infty]$, i.e. the connected p-divisible group corresponding to the divisible formal Lie group $\mathfrak{F}_\pi$ over $\mathcal{O}_{\tau E}$. The p-divisible group $\mathscr{G}$ is connected of height $d = [E : \mathbf{Q}_p]$, thanks to the fact that

$$\#\mathfrak{F}_\pi(\mathfrak{m}_{\overline{\tau E}})[p] = \#\mathfrak{F}_\pi(\mathfrak{m}_{\overline{\tau E}})[\pi_{\tau E}^{e(\tau E|\mathbf{Q}_p)}] = \#\left(\mathcal{O}_{\tau E}/(\pi_{\tau E}^{e(\tau E|\mathbf{Q}_p)})\right) = p^{e(\tau E|\mathbf{Q}_P)f(\tau E|\mathbf{Q}_p)} = p^d.$$

The Hodge–Tate decomposition for the Tate module of $\mathscr{G}$ (one of the main results of [Tat1967]) provides a G_K-equivariant isomorphism

$$\mathbf{C}_K \otimes_{\mathbf{Q}_p} T(\mathscr{G}) \cong \mathbf{C}_K \otimes_K \mathrm{Hom}_{\tau E}(t_{\mathscr{G}^\vee}(\tau E), K) \oplus \mathbf{C}_K(1) \otimes_K t_{\mathscr{G}}(K).$$

By [Tat1967, Proposition 3], $\mathscr{G}^\vee$ is of dimension $d-1$ (the height of $\mathscr{G}$ is d and the dimension is 1 by definition since it comes from a 1-dimensional formal Lie group over $\mathcal{O}_{\tau E}$). So this is a decomposition of a d-dimensional $\mathbf{C}_K$-semilinear representation of G_K into subspaces of dimensions $d-1$ and 1. Moreover, by functoriality, the Hodge–Tate decomposition is also τE-invariant, where τE is defined to act on both sides by the morphisms it induces via the $\mathcal{O}_{\tau E}^\times$-action on $\mathscr{G}$ (and $1/p$ acts just by multiplication by $1/p$, as is forced). The tangent space $t_{\mathscr{G}}(K)$ is defined to be the set of $\mathcal{O}_{\tau E}$-linear functions $\delta : \mathcal{O}_{\tau E}[[X]] \to K$ satisfying $\delta(fg) = f(0)\delta(g) + \delta(f)g(0)$. The $\mathcal{O}_{\tau E}^\times$-action on the formal group $\mathrm{Spf}\mathcal{O}_{\tau E}[[X]]$ is just by multiplying X by the units, so the induced action of $u \in \mathcal{O}_{\tau E}^\times$ on the tangent space is by taking δ to $u\delta$ (the key point being that δ only cares about linear terms). In other words, τE acts on the $\mathbf{C}_K(1) \otimes t_{\mathscr{G}}$-coordinate just by taking the canonical inclusion $\tau E \to K$ and using the K-vector space structure on $t_{\mathscr{G}}(K)$. That means that the simultaneous τE-eigenspace of $\mathbf{C}_K \otimes_{\mathbf{Q}_p} T(\mathscr{G})$ corresponding to the inclusion $\tau E \to K \to \overline{\mathbf{Q}}_p$ (i.e. the subspace $W_{\tau E \to K}$ from Lemma 1.2.14) is exactly $\mathbf{C}_K(1) \otimes_K t_{\mathscr{G}}(K)$. All the other simultaneous τE-eigenspaces (coming from all the embeddings $\tau E \to K$ other than the inclusion) are inside the other factor $\mathbf{C}_K \otimes_K \mathrm{Hom}_{\tau E}(t_{\mathscr{G}^\vee}(\tau E), K)$, since E acts on each direct summand individually (thanks to how the E-action is defined by being an induced action on tangent spaces).

First of all, the fact that the simultaneous eigenspace for the inclusion $\tau E \to K$ is equal to $\mathbf{C}_K(1) \otimes_K t_{\mathscr{G}}(K) = \mathbf{C}(1)$ (as a G_K-representation) tells us (by the proof of Lemma 1.2.14) $\chi_{\tau E,\pi} = \chi$ in $H^1(G_K, \mathbf{C}_K^\times)$. Similarly, the fact that if $\sigma : \tau E \to K$ is not the inclusion then the corresponding simultaneous eigenspace is in $\mathbf{C}_K \otimes$ (other thing with trivial Galois action) implies that this eigenspace is just a $\mathbf{C}_K$-line with the usual Galois action, and hence (it being equal to $\sigma \circ \chi_{\tau E,\pi}$ in $H^1(G_K, \mathbf{C}_K^\times)$ from the proof of Lemma 1.2.14) we conclude that $\sigma \circ \chi_{\tau E,\pi} = 1$ in $H^1(G_K, \mathbf{C}_K^\times)$, as desired. □

We have therefore concluded that the Weil's p-adic Galois characters corresponding to algebraic Hecke characters are Hodge–Tate, regardless of which coefficient field $E/\mathbf{Q}_p$ is chosen. To recap:

Corollary 1.2.17. *Let ξ be an algebraic Hecke character for F with weights $\{a_\sigma\}$, $\rho_{\xi,p} : G_F \to G_F^{ab} \to \overline{\mathbf{Q}}_p^\times$ Weil's corresponding p-adic Galois character, E some choice of finite extension of*

$\mathbf{Q}_p$ containing the image of $\rho_{\xi,p}$. Then $\rho_{\xi,p}$ is Hodge–Tate at each $v|p$ as a $[E:\mathbf{Q}_p]$-dimensional representation. Its Hodge–Tate weights at v are, as a set (the multiplicities will depend on the choice of E), equal to the set of numbers a_σ with $\iota_p \circ \iota_\infty^{-1} \circ \sigma$ inducing the place v on F.

Proof. By Lemma 1.2.15, we can consider instead the representation $\rho_K : G_K^{\mathrm{ab}} \to E^\times$ defined above. We proved that on $I_K^{\mathrm{ab}} \cong \mathcal{O}_K^\times$, it is defined by

$$x \mapsto \prod_{\tau : E \to \overline{\mathbf{Q}}_p} \left(\tau^{-1} \mathrm{N}^K_{\tau E}\right)^{b_\tau},$$

where b_τ is the common value of the a_σ over all $\sigma \in \mathrm{Gal}(K/\mathbf{Q}_p)$ such that $\sigma^{-1}|_E = \tau$. Letting V be the abstract 1-dimensional E-vector space that ρ_K defines a Galois action on, and $W = V \otimes_{\mathbf{Q}_p} \mathbf{C}_K$, byProposition 1.2.16, for any $\tau : E \to \overline{\mathbf{Q}}_p$, the simultaneous E-eigenspace W_τ is 1-dimensional (as always) and as a $\mathbf{C}_K$-semilinear G_K-module equals $\mathbf{C}_K(b_\tau)$ (as it is the b_τ-fold composite of something equal to $\mathbf{C}_K(1)$) in $H^1(G_K, \mathbf{C}_K^\times)$. □

In fact, the representations $G_F \to \overline{\mathbf{Q}}_p^\times$ which are Hodge–Tate at the places above p are exactly those that come from Weil's construction. First of all, being a representation into $\overline{\mathbf{Q}}_p^\times$ is the same as having coefficients in $E^\times$ for some finite extension $E/\mathbf{Q}_p$, as I learned from a paper of Breuil–Mézard [BM2002, Lemme 2.2.1.1] (a proof they say is due to Bost) after searching on the internet:

Lemma 1.2.18. *Let $\rho : G_K \to \mathrm{GL}_N(\overline{\mathbf{Q}}_p)$ be a continuous representation of the absolute Galois group G_K of a p-adic field K. Then there exists a finite extension $E/\mathbf{Q}_p$ such that $\rho(G_K) \subset \mathrm{GL}_N(E)$.*

Proof. The topological space G_K is compact and Hausdorff, so the Baire category theorem applies. In particular, if G_K has a cover consisting of a countable collection of closed sets C_i, then one of the C_i has an interior point. If C_i is a subgroup of G_K, then this implies that C_i is open.

By Krasner's lemma and compactness of $\mathbf{Z}_p$, there are countably many finite extensions $E/\mathbf{Q}_p$ inside a fixed algebraic closure (the usual argument involving Eisenstein polynomials implies there are only finitely many of given degree). Since $E \subset \overline{\mathbf{Q}}_p$ is closed and ρ is continuous, we get a countable covering of G_K by closed subgroups $\rho^{-1}(\mathrm{GL}_N(E_i))$, where $\{E_i\}_{i \in \mathbf{N}}$ is the set of finite extensions of $\mathbf{Q}_p$. One of these subgroups must be open by the previous paragraph. So there are finite extensions K'/K and $E/\mathbf{Q}_p$ such that $\rho(\mathrm{Gal}(\overline{K}/K')) \subset \mathrm{GL}_N(E)$. Letting $g_1, \ldots, g_n$ be a system of representatives in G_K for $G_K/\mathrm{Gal}(\overline{K}/K')$ (the key point being that there are finitely many), there is a finite extension $E'/\mathbf{Q}_p$ such that $\rho(g_j) \in \mathrm{GL}_N(E')$ for $j = 1, \ldots, n$, and hence the image of ρ lives inside $\mathrm{GL}_N(E \cdot E')$. □

Starting with a continuous representation $\rho : G_F \to E^\times$ which is Hodge–Tate at all $v|p$, it suffices to prove that it is algebraic in the p-adic sense at all $v|p$, i.e. that it is given (via the local class field theory isomorphism) on an open subgroup of $I_F^{\text{ab}} \cong \mathcal{O}_F^\times$ by $x \mapsto \prod_{\sigma: F_v \to \overline{\mathbf{Q}}_p} \sigma(x)^{a_\sigma}$ for some $\{a_\sigma\} \in \mathbf{Z}$. Once we have this it is obvious how to get back (bijectively) to the Hecke character side. To prove this, it mostly suffices to do the argument we just did in reverse. If $\rho : G_F \to E^\times$ is Hodge–Tate with Hodge–Tate weights $\{n_\sigma\}_{\sigma: E \to \overline{\mathbf{Q}}_p}$ at $v|p$, then by Lemma 1.2.15 and the fact that algebraicity only cares about an open subgroup of $I_{F_v}^{\text{ab}}$, we can consider ρ as instead coming from G_K where $K/\mathbf{Q}_p$ meets all the same conditions as before: contains all the Galois conjugates of E, contains F_v, is Galois over $\mathbf{Q}_p$. We can therefore consider the $E^\times$-valued character of G_K given by

$$\rho \cdot \left(\prod_{\tau : E \to \overline{\mathbf{Q}}_p} \tau^{-1} \circ \chi_{\tau, E}^{n_\tau} \right)^{-1},$$

which is Hodge–Tate but with all of its Hodge–Tate weights equal to zero. In other words, by Lemma 1.2.14,

$$\tau \circ \left(\rho \cdot \left(\prod_{\tau : E \to \overline{\mathbf{Q}}_p} \tau^{-1} \circ \chi_{\tau, E}^{n_\tau} \right)^{-1} \right) = 1$$

in $H^1(G_{F_v}, \mathbf{C}_{F_v}^\times)$ for all $\tau : E \to \overline{\mathbf{Q}}_p$. It is proven in [Ser1989, Appendix to Ch. III, A.3][12] that this implies that $\rho \cdot \left(\prod_{\tau : E \to \overline{\mathbf{Q}}_p} \tau^{-1} \circ \chi_{\tau, E}^{n_\tau} \right)^{-1}$ must be 1 on an open subgroup of $I_{F_v}^\times$, i.e. that ρ is algebraic (as we know that the character $\tau^{-1} \circ \chi_{\tau,E}^{n_\tau} = \tau^{-1} \mathrm{N}_{\tau E}^K$ is algebraic from our previous formulas expressing it on $I_K^{\text{ab}} \cong \mathcal{O}_K^\times$ as a product of powers of certain embeddings of K).

Remark 1.2.19. We have now proved a fairly remarkable fact: the algebraic Hecke characters for a number field F are in bijection (via Weil's construction) with the Hodge–Tate representations $G_F \to \overline{\mathbf{Q}}_p^\times$. Combined with our stuff from the previous section regarding Hecke characters factoring through the norm to the maximal CM subfield, we almost have the Fontaine–Mazur conjecture for GL_1. Indeed, if a Galois representation $\rho : G_F \to \overline{\mathbf{Q}}_p^\times$ is potentially semistable at all $v|p$, then it is Hodge–Tate. We don't lose any information by now forgetting the potentially semistable assumption, because by the p-adic monodromy theorem (probably not really necessary in the 1-dimensional case) and the fact that de Rham equals Hodge–Tate in 1 dimension [BC2009b, Example 6.3.9], actually it is equivalent to Hodge–Tate. By what we just proved, ρ comes from an algebraic Hecke character for F, which factors through the maximal CM subfield of F. If F has no maximal CM subfield, then the Hecke character is just $\|\|$ up to finite order, and the corresponding Galois character is (up to finite order) the cyclotomic character, which of course comes from geometry. In the case where F has a maximal CM subfield $F_0(\sqrt{a})$, we would need to know that the Galois characters associated to algebraic Hecke characters for CM fields come from geometry. An affirmative answer to the full Question 1.2.11 would give us this, as it would

[12] I don't want to do the details for this one, but it is essentially taking p-adic logarithms and applying [Tat1967, Theorem 2].

give us those characters as a subquotient of the Tate module of an abelian variety.

1.3 | The Sato–Tate conjecture

This section is about the Sato–Tate conjecture and the relationship with the symmetric power functoriality conjecture for holomorphic modular forms Example 1.1.5. We begin with the standard generalities on equidistribution (see for example [Elk2019]).

Proposition 1.3.1. *Let G be a compact Lie group and X the set of conjugacy classes of G, endowed with the quotient measure from G. Denote by μ the pushforward to X of the Haar measure on G via the quotient map $G \to X$, normalized so that $\mu(G) = \mu(X) = 1$.*

A sequence $\{x_n\} \in X$ is equidistributed with respect to μ if and only if for every nontrivial irreducible character χ of G,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \chi(x_i) = 0.$$

Proof. The sequence x_n being equidistributed on X with respect to μ is equivalent to having

$$\int_X f\, d\mu = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} f(x_i) \tag{1.1}$$

for all $f \in \mathscr{C}(X)$ (if we don't take this to be the definition of equidistribution, it is easy to show this is equivalent to the more intuitive one by looking at characteristic functions of open sets and using that their $\mathbf{C}$-span is dense in $\mathscr{C}(X)$ under the sup norm, for example [compactness of X makes this straightforward]).

The forwards direction is straightforward: just plug in each χ for $f \in \mathscr{C}(X)$, and use the fact that by definition of the pushforward measure, the left hand side of (1.1) equals

$$\int_G (f \circ (G \to X))(g)\, dg,$$

where dg is the Haar measure on G. When χ is plugged in, we know (after using the letter χ to denote both the function on G and the function on X) that this is equal to

$$\int_G \chi(g)\, dg = \begin{cases} 0, & \text{if } \chi \neq 1 \\ 1, & \text{if } \chi = 1 \end{cases}$$

(by the exact same argument as for finite groups, using the invariance of the Haar measure and integration rather than summation). So (1.1) (true from the assumption of equidistribution) implies the conclusion of the forward direction of the claimed result.

For the reverse direction, since the left hand side of (1.1) is tautologically equal to $\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \chi(x_i) = 1$ when $\chi = 1$ (using the assumption that $\int_G 1\, dg = 1$), it suffices to show that (1.1) being true for all irreducible characters (implied by the hypothesis of the reverse direction by what we just did) implies that (1.1) is true for all $f \in \mathscr{C}(X)$.

It is a consequence (which I will explain next) of the Peter–Weyl theorem that the $\mathbf{C}$-span of the irreducible characters of G is dense in $\mathscr{C}(X)$ under the sup norm. Once this is established, we are done, because (1.1) being true for all χ implies it is true for all $\mathbf{C}$-linear combinations of the χ (both sides being linear in the choice of function in $\mathscr{C}(X)$). Those linear combintions being dense in $\mathscr{C}(X)$ under the sup norm, for every $f \in \mathscr{C}(X)$ and $\varepsilon > 0$, there is a $g \in \mathscr{C}(X)$ for which (1.1) holds and for which $\|f - g\|_{L^\infty} < \epsilon$ (take g to be a linear combination of the χ close to f under the sup norm). Together, these two facts about g imply that the two sides of (1.1) are within 2ε of each other (thanks $\int_G 1\, dg = 1$); since $\epsilon > 0$ is arbitrary, we conclude that (1.1) holds for all $g \in \mathscr{C}(X)$ and therefore $\{x_i\}$ is equidistributed with respect to μ, as long as (1.1) is true for all $\chi \neq 1$.

The only thing left to justify is the claim that the $\mathbf{C}$-span of the irreducible characters of G is dense in $\mathscr{C}(X)$. Part of the Peter–Weyl theorem says that the matrix coefficients of G are dense in $\mathscr{C}(G)$. This part of the theorem essentially follows from Stone–Weierstrass. We need to deduce from this that every continuous class function on G can be uniformly approximated by $\mathbf{C}$-linear combinations of characters of irreducible representations. Let $f \in \mathscr{C}(X)$ and $\varepsilon > 0$ be arbitrary. By the part of the Peter–Weyl theorem mentioned above, there is a matrix coefficient $\psi : x \mapsto u(\pi(x)v)$ for G (here $\pi : G \to \mathrm{GL}(V)$ is a finite-dimensional representation, $v \in \pi$ and $u \in \pi^\vee$) such that $\|f - \psi\|_{L^\infty} < \varepsilon$. To produce a class function from ψ (which will hopefully approximate f well because ψ does, and will hopefully be a $\mathbf{C}$-linear combination of characters of irreducible representations of G), one can use the usual technique of averaging: consider the element $\varphi \in \mathscr{C}(G)$ given by

$$\varphi(x) = \int_G u(\pi(gxg^{-1})v)\, dg.$$

By invariance of the Haar measure, φ is a continuous class function on G. We have (thanks to having normalized the Haar measure such that $\int_G 1\, dg = 1$ and f being a class function)

$$\begin{aligned}
\|f - \varphi\|_{L^\infty} &= \left\| x \mapsto \int_G \big(f(x) - u(\pi(gxg^{-1})v)\big)\, dg \right\|_{L^\infty} \\
&= \left\| x \mapsto \int_G \big(f(gxg^{-1}) - u(\pi(gxg^{-1})v)\big)\, dg \right\|_{L^\infty} \\
&\leq \int_G \|x \mapsto f(gxg^{-1}) - u(\pi(gxg^{-1})v)\|_{L^\infty}\, dg \\
&\leq \int_G \|x \mapsto f(x) - u(\pi(x)v)\|_{L^\infty}\, dg
\end{aligned}$$

$$= \|f - \psi\|_{L^\infty} < \varepsilon$$

so it remains to confirm that φ is not only a class function, but a $\mathbf{C}$-linear combination of characters of irreducible representations. To do this, write $\pi = \bigoplus \pi_i$ as a direct sum of irreducible representations (G is compact and V is finite-dimensional so the usual Maschke argument works). Then the matrix coefficient ψ may be rewritten in terms of this decomposition:

$$\psi(x) = \sum u_i(\pi_i(x)v_i),$$

where $u = \sum u_i$ ($u_i \in \pi_i^\vee$) and $v = \sum v_i$ ($v_i \in \pi_i$). So $\varphi(x) = \sum_i \int_G u_i(\pi_i(gxg^{-1})v_i)\, dg$, and we just need to check that the i-th term here is a $\mathbf{C}$-multiple of χ_{π_i}. For any fixed x, the linear operator on π_i given by $v \mapsto \int_G \pi_i(gxg^{-1})v\, dg$ is G-intertwining (by invariance of Haar measure), so by Schur's lemma (this was the point of decomposing into irreducible π_i), it is of the form $v \mapsto \alpha_{i,x}v$, where $\alpha_{i,x} \in \mathbf{C}$. Therefore, again by $\int_G 1\, dg = 1$, and passing the linear functionals u_i outside the integral sign (allowed because of how integrating vector-valued functions is defined),

$$\varphi(x) = \sum_i u_i(\alpha_x v_i)\, dg = \sum_i u_i(v_i)\alpha_{i,x}.$$

What is the relationship with the character χ_i of π_i ? It is that

$$\chi_i(x) = \mathrm{Tr}\pi_i(x) = \int_G \mathrm{Tr}\pi_i(gxg^{-1})\, dg = \mathrm{Tr}\int_G \pi_i(gxg^{-1})\, dg = (\dim \pi_i)\alpha_{i,x},$$

(again using $\int_G 1\, dg = 1$, the fact that χ_i is a class function, and the theory of integration of vector-valued functions, treating $\pi_i(gxg^{-1})$ as a $(\dim \pi_i)^2$-dimensional vector). Hence,

$$\varphi(x) = \sum_i \frac{u_i(v_i)}{\dim \pi_i}\chi_i(x)$$

for all $x \in G$, which means that f can indeed be approximated uniformly by $\mathbf{C}$-linear combinations of characters of irreducible representations, completing the proof. □

Now we explain the relationship with L-functions. The technique of how to deduce equidistribution (in the form given by Proposition 1.3.1) from properties of L-functions follows the exact same lines as how we usually deduce results about asymptotic behavior of prime-counting functions from the analytic properties of the appropriate L-functions [Elk2019, `psi.pdf`, `chebi.pdf`, `pnt.pdf`, `pnt_q.pdf`]. In particular, we get something out of the Wiener–Ikehara tauberian theorem for the logarithmic derivative of appropriate L-function (or if we have a zero-free region that goes sufficiently far into the critical strip we can use the more explicit Perron integration technique on the logarithmic derivative and get a more explicit error term depending on what we know about the zeros of the L-function; this works for arbitrary number fields, as is classical for the prime number theorem and was worked out by Lagarias–Odlyzko and Serre for the Čebotarev density theorem [LO1977, Ser1981]). The asymptotic result that we actually want will follow by partial summation (in the same way that asymptotics for the Chebyshev ψ-function are equivalent to asymptotics for the prime-counting π-function by partial summation, and analogously in the proof of Dirichlet's theorem or more generally the Čebotarev density theorem). Indeed, the Čebotarev density theorem for finite extensions is an equidistribution theorem about a sequence of conjugacy classes (the Frobenii at unramified downstairs places) of a finite (Lie) group. It makes sense that we should be able to use the same techniques as in that proof to also deduce more general results about equidistribution of sequences indexed by places of a number field with values in a general compact Lie group.

We are forced to explain this in general, so that we can develop in full detail not only the example with the equidistribution of angles but also the Sato–Tate conjecture and its relationship with symmetric power L-functions.

Let F be a number field, G a compact Lie group, and X the space of conjugacy classes of G, as above. Let Σ be a subset of the set of finite places of F. We are interested in whether a given collection $\{x_v\}$ of elements of X is equidistributed with respect to μ (the pushforward of the Haar measure on G). In order to ask whether the $\{x_v\}_{v\in\Sigma}$ are equidistributed, we need to define some ordering on Σ. The obvious thing to do is to order them by their norm. Since the number of places of fixed norm is bounded by a constant that only depends on F (namely $[F:\mathbf{Q}_p]$), the question of equidistribution is unaffected by which one of the various orderings of the $v \in \Sigma$ such that $v \leq w$ is the same as $\mathrm{N}v \leq \mathrm{N}w$. So we pick an arbitrary such ordering, the point being that the only thing that matters in our truncated averages will be a sum over v such that $\mathrm{N}v \leq T$, where we will send $T \to \infty$. Since the sum will be divided by T, and has a bounded number of fewer terms compared to any truncated average computed over any subset of the v with $\mathrm{N}v \leq T+1$ that contains all of the v with $\mathrm{N}v \leq T$, the truncated averages over $\{v \in \Sigma : \mathrm{N}v \leq T\}$ going to zero as $T \to \infty$ implies the same thing for all of the truncated averages.

By Proposition 1.3.1, $\{x_v\}$ is equidistributed with respect to μ if and only if

$$\lim_{T\to\infty} \frac{1}{\#\{v \in \Sigma : \mathrm{N}v \leq T\}} \sum_{\substack{v\in\Sigma \\ \mathrm{N}v\leq T}} \chi(x_v) = 0$$

for all nontrivial irreducible characters χ of G. We want to build an L-function which is supposed to tell us about this, which we can basically do by direct analogy to Artin and/or Hecke L-functions (the former being for where the x_v are Frobenii in some finite Galois group and the latter being for where the x_v are images of uniformizers π_v in a ray class group or maybe some other quotient of the idéle class group that happens to be isomorphic to S^1 [these are quotients by the kernel of surjective unitary Hecke characters]; of course these [x_v being Frobenii and being images in ray class groups] amount to the same thing when you restrict to the case where the Galois group of the upstairs field over F is abelian by class field theory).

With particular attention to doing the exact same thing as Artin L-functions in the non-abelian case: for an irreducible representation ρ of G, define

$$L_\Sigma(s,\rho) := \prod_{v\in\Sigma} \frac{1}{\det(\mathrm{id}_\rho - (\mathrm{N}v)^{-s}\rho(x_v))}.$$

If this function has good properties (not always obvious or known) for all ρ, then we can crank the usual handle:

Proposition 1.3.2. *Suppose that for all nontrivial irreducible representations ρ of G, $L(s,\rho)$ converges to a nowhere-vanishing holomorphic function on $\{s \in \mathbf{C} : \Re(s) > 1\}$ which extends meromorphically to $\{s \in \mathbf{C} : \Re(s) \geq 1\}$ again without zeroes or poles. Suppose also that $L(s,1)$ converges for $\Re(s) > 1$ with meromorphic continuation to $\{s \in \mathbf{C} : \Re(s) \geq 1\}$ such that there are no zeroes and the only pole is a simple pole at $s = 1$. Then the $\{x_v\}$ are equidistributed with respect to μ.*

Proof. Let ρ be an irreducible representation of G. For each $v \in \Sigma$, let $\{\lambda_v^{(i)}\}_{i=1}^{\dim\rho}$ be the roots (listed with multiplicity) of the characteristic polynomial of $\rho(x_v)$. Using the Euler product (in our case the definition) of $L(s,\rho)$, we have

$$\begin{aligned}
\frac{L'}{L}(s,\rho) &= \frac{d}{ds}\log\prod_{v\in\Sigma}\prod_{i=1}^{\dim\rho}\frac{1}{1-\lambda_v^{(i)}(\mathrm{N}v)^{-s}} \\
&= -\sum_{v\in\Sigma}\sum_{i=1}^{\dim\rho}\frac{d}{ds}\log\left(1-\lambda_v^{(i)}(\mathrm{N}v)^{-s}\right) \\
&= -\sum_{v\in\Sigma}\sum_{i=1}^{\dim\rho}\frac{\lambda_v^{(i)}\log(\mathrm{N}v)(\mathrm{N}v)^{-s}}{1-\lambda_v^{(i)}(\mathrm{N}v)^{-s}} \\
&= -\sum_{v\in\Sigma}\sum_{m\geq 1}\log(\mathrm{N}v)(\mathrm{N}v)^{-ms}\sum_{i=1}^{\dim\rho}(\lambda_v^{(i)})^m \\
&= -\sum_{v\in\Sigma}\sum_{m\geq 1}\log(\mathrm{N}v)(\mathrm{N}v)^{-ms}\chi(x_v^m).
\end{aligned}$$

where χ is the character of ρ. Here we have used the unitary trick to see that $|\lambda_v^{(i)}| = 1$ and therefore the geometric series manipulation is allowed. This is valid in the region where the Dirichlet series for $L(s,\rho)$ converges, which is at least for $\Re(s) > 1$.

For the exact same reason that the contribution to $\psi(x)$ of higher prime powers is negligible in [Elk2019, `chebi.pdf`, p. 3], we really only need to be interested in the $m = 1$ terms here. Following the model of [Elk2019, `chebi.pdf`] rather than what is done by Serre in [Ser1989, appendix to Ch. I, A.3], we continue without doing anything to L'/L.

The Dirichlet series we just wrote down for $\frac{L'}{L}(s,\rho)$ cannot be directly plugged into the Wiener–Ikehara theorem, because of the $\chi(x_v)$, which are not nonnegative real numbers. So we need to bound our Dirichlet series by something with good convergence and positive real coefficients. This is where the hypothesis about the "zêta function" $L(s,1)$ comes in useful:

$$-(\dim\rho)\frac{L'}{L}(s,1) = \sum_{v\in\Sigma}\sum_{m\geq 1}\log(\mathrm{N}v)(\mathrm{N}v)^{-ms}(\dim\rho).$$

As noted above, we are guaranteed that this Dirichlet series converges for $\Re(s)>1$, and since $\chi(x_v^m)$ is the sum of $\dim\rho$ eigenvalues, all of absolute value 1, its coefficients are upper bounds for the absolute values of the coefficients of $\frac{L'}{L}(s,\rho)$.

The hypothesis that $L(s,\rho)$ converges to a holomorphic function on $\Re(s)>1$ with holomorphic continuation to $\Re(s)\geq 1$ with no zeros in that region implies that $\frac{L'}{L}(s,\rho)$ has no poles in the region $\Re(s)\geq 1$. Similarly, the hypothesis that $L(s,1)$ converges on $\Re(s)>1$ with meromorphic continuation to $\Re(s)\geq 1$ with no zeros anywhere and a simple pole at $s=1$ implies that $-(\dim\rho)\frac{L'}{L}(s,1)$ has a meromorphic continuation to $\Re(s)\geq 1$ such that the only pole is a simple pole at $s=1$.

In this situation, the Wiener–Ikehara Tauberian theorem (in the statement in [Lan1994, Ch. XV, §3, Theorem 1], we are setting $f(s)=\frac{L'}{L}(s,1)$ and $g(s)=\frac{L'}{L}(s,\rho)$) tells us all about the asymptotic behavior of the partial sums of coefficients of $\frac{L'}{L}(s,\rho)$. In particular, the fact that $\frac{L'}{L}(s,\rho)$ has no poles for $\Re(s)\geq 1$ implies that for $T>0$,

$$\sum_{v\in\Sigma}\sum_{\substack{m\geq 1\\(\mathrm{N}v)^m\leq T}}(\log(\mathrm{N}v))\chi(x_v^m)=o(T) \tag{1.2}$$

as $T\to\infty$. It is at this point that we choose to get rid of the terms with $m\geq 2$. For any fixed m, the absolute value of the contribution of the terms with that value of m is

$$\begin{aligned}\left|\sum_{\substack{v\in\Sigma\\ \mathrm{N}v\leq T^{1/m}}}(\log N(v))\chi(x_v^m)\right| &\leq \sum_{\substack{v\in\Sigma\\ \mathrm{N}v\leq T^{1/m}}}(\log N(v))\,|\chi(x_v^m)|\\ &\leq (\dim\rho)\sum_{\substack{v\in\Sigma\\ \mathrm{N}v\leq T^{1/m}}}\log N(v)\\ &\leq (\dim\rho)\#\{v\in\Sigma : \mathrm{N}v\leq T^{1/m}\}\log T^{1/m}\\ &\leq (\dim\rho)[F:\mathbf{Q}]\frac{1}{m}T^{1/m}\log T.\end{aligned}$$

Therefore, the sum of these contributions over all m is of absolute value $\ll_{F,\rho} (\log T)\sum_{m=2}^{\log_2 T} \frac{1}{m} T^{1/m}$, because the terms of (1.2) all have $T \geq 2^m$, i.e. $m \leq \log_2 T$. The $m = 2$ term is $\ll T^{1/2}\log T$, and the $m \geq 3$ terms all add up to $\ll T^{1/3}(\log T)^2 \ll T^{1/2}\log T = o(T)$. Combined with (1.2), we conclude that the asymptotic holds without any of the $m \geq 2$ terms, i.e. that

$$\sum_{\substack{v\in\Sigma \\ \mathrm{N}v\leq T}} (\log(\mathrm{N}v))\chi(x_v) = o(T) \tag{1.3}$$

as $T \to \infty$.

This implies the desired result by partial summation: the thing we want to estimate as $T \to \infty$ is (by partial summation, i.e. integration by parts for the Riemann–Stieltjes integral)

$$\begin{aligned}
\sum_{\substack{v\in\Sigma \\ \mathrm{N}v\leq T}} \chi(x_v) &= \int_{3/2}^{T} \frac{1}{\log x} d\left(\sum_{\substack{v\in\Sigma \\ \mathrm{N}v\leq x}} \log(\mathrm{N}v)\chi(x_v)\right) \\
&= \frac{1}{\log T}\sum_{\substack{v\in\Sigma \\ \mathrm{N}v\leq T}} \log(\mathrm{N}v)\chi(x_v) + \int_{3/2}^{T} \frac{1}{x\log(x)^2} \sum_{\substack{v\in\Sigma \\ \mathrm{N}v\leq x}} \log(\mathrm{N}v)\chi(x_v)\,dx \\
&= \frac{1}{\log T} o(T) + \int_{3/2}^{T} \frac{1}{x\log(x)^2} o(x)\,dx.
\end{aligned}$$

For sufficiently large x, the $o(x)$ in the second term is at most x. So as $T \to \infty$, we conclude that

$$\sum_{\substack{v\in\Sigma \\ \mathrm{N}v\leq T}} \chi(x_v) = o\left(\frac{T}{\log T}\right) + O\left(\frac{T}{\log(T)^2}\right) = o\left(\frac{T}{\log T}\right).$$

By Wiener–Ikehara applied just to $-\frac{L'}{L}(s,1)$, and the same argument applied again to ignore the $m \geq 2$ terms, since this function has its only pole at $s = 1$, which is simple of residue 1, we know that

$$\#\{v \in \Sigma : \mathrm{N}v \leq T\} = \frac{T}{\log T} + o\left(\frac{T}{\log T}\right)$$

(the "prime number theorem for number fields"). It follows that

$$\frac{1}{\#\{v \in \Sigma : \mathrm{N}v \leq T\}} \sum_{\substack{v\in\Sigma \\ \mathrm{N}v\leq T}} \chi(x_v) \to 0$$

as $T \to \infty$ whenever χ is the character of a nontrivial irreducible representation of G, i.e. that the x_v are equidistributed on X, by Proposition 1.3.1. $\square$

Example 1.3.3. If we are looking for an unramified unitary Hecke character χ for $\mathbf{Q}(i)$, since

$$1 = \mathrm{Cl}(\mathbf{Q}(i)) = \mathbf{Q}(i)^\times \backslash \mathbf{A}_{\mathbf{Q}(i)}^\times / \left(\mathbf{C}^\times \times \prod_{\mathfrak{p}<\infty} \widehat{\mathbf{Z}[i]}_{\mathfrak{p}}^\times \right),$$

it suffices to specify χ on $\mathbf{C}^\times$ (its unramifiedness means it is trivial on $\widehat{\mathbf{Z}[i]}_{\mathfrak{p}}^\times$; that is the definition). The unitary characters of $\mathbf{C}^\times$ are exactly those of the form $\alpha \mapsto (\alpha/|\alpha|)^k$ for $s \in \mathbf{Z}$. But not all of these work, because there is the additional restriction that χ is 1 on $\mathbf{Q}(i)^\times$. Since $\mathbf{Q}(i)^\times \cap \mathbf{C}^\times \times \prod_{\mathfrak{p}<\infty} \widehat{\mathbf{Z}[i]}_{\mathfrak{p}} = \{\pm 1, \pm i\}$, we are all set as long as $i^k = 1$, i.e. $4|k$. This is where our χ might come from. This Hecke character also has the nice property that for each finite prime $\mathfrak{p}$ of $\mathbf{Q}(i)$ above a rational prime p, if $\pi_{\mathfrak{p}}$ has $v_{\mathfrak{p}}(\pi_{\mathfrak{p}}) = 1$ and $v_{\mathfrak{q}}(\pi_{\mathfrak{p}}) = 0$ for all finite $\mathfrak{q} \neq \mathfrak{p}$ (that is, $\mathfrak{p} = (\pi_{\mathfrak{p}})$, which is okay for us because the class group is trivial), then

$$\chi_{\mathfrak{p}}(\pi_{\mathfrak{p}}) = \chi((\alpha_v)_v),$$

where $\alpha_v = \pi_{\mathfrak{p}}^{-1}$ whenever $v \neq \mathfrak{p}$ and $\alpha_v = 1$ when $v = \mathfrak{p}$ (this is just the definition of $\chi_{\mathfrak{p}}$ combined with the fact that $\pi_{\mathfrak{p}} \in \mathbf{Q}(i)^\times$). Since $\pi_{\mathfrak{p}}^{-1} \in \widehat{\mathbf{Z}[i]}_{\mathfrak{q}}^\times$ for all finite $\mathfrak{q} \neq \mathfrak{p}$, and χ is unramified, we conclude that $\chi_{\mathfrak{p}}(\pi_{\mathfrak{p}}) \in S^1$ is precisely the inverse of the fourth power of the argument of $\pi_{\mathfrak{p}}$, viewed as an element of $S^1 \subset \mathbf{C}^\times$. The point is that this is really an angle associated to $\mathfrak{p}$, as the fourth power assures us that it will not depend on the choice of generator $\pi_{\mathfrak{p}}$, which we are allowed to push around by $\pm 1, \pm i$. The information contained in this fourth-power thing is the same as the data of the argument of the generator of $\mathfrak{p}$ that lives in the first quadrant. For each finite prime $\mathfrak{p}$ of $\mathbf{Q}(i)$, we let $x_{\mathfrak{p}} \in S^1$ be $\chi(\pi_{\mathfrak{p}})$. The nontrivial irreducible (unitary) characters of S^1 are just the self-maps $S^1 \to S^1 \subset \mathbf{C}^\times$ given by $z \mapsto z^k$ for integers $k \neq 0$, so for any such character η, the truncated averages we are interested in thanks to Proposition 1.3.1 are exactly equal to

$$\frac{1}{\#\{\mathfrak{p} : \mathrm{N}\mathfrak{p} \leq T\}} \sum_{\substack{\mathfrak{p} \\ \mathrm{N}\mathfrak{p} \leq T}} \eta(x_{\mathfrak{p}}) = \frac{1}{\#\{\mathfrak{p} : \mathrm{N}\mathfrak{p} \leq T\}} \sum_{\substack{\mathfrak{p} \\ \mathrm{N}\mathfrak{p} \leq T}} \chi^k(\pi_{\mathfrak{p}}).$$

Of course, χ^k is just another nontrivial unramified unitary Hecke character for $\mathbf{Q}(i)$, namely the one that is trivial on $\widehat{\mathbf{Z}[i]}_{\mathfrak{p}}$ for all finite $\mathfrak{p}$, and equal to $\alpha \mapsto \left(\frac{\alpha}{|\alpha|}\right)^{4k}$ on $\mathbf{C}^\times$. That this does not depend on the choice of $\pi_{\mathfrak{p}}$ is immediate from this explicit form and also from the fact that χ is unramified. The unitary Hecke character χ^k is surjective onto S^1 when restricted to $\mathbf{A}_{\mathbf{Q}(i)}^{\times,1}$, since this subset contains $S^1 \subset \mathbf{C}^\times$, and the $4k$-th power map from S^1 to itself is surjective when $k \neq 0$. In the notation that we defined above in the general setup relating L-functions to equidistribution, we have

$$L(s,(\cdot)^k) = \prod_{\mathfrak{p}<\infty} \frac{1}{1-\chi^k(\pi_{\mathfrak{p}})(\mathrm{N}\mathfrak{p})^{-s}},$$

which is the exact same thing as $L^{\text{Hecke}}(s,\chi^k)$. The unitary Hecke character χ^k being surjective onto S^1 when restricted to $\mathbf{A}^{\times,1}_{\mathbf{Q}(i)}$, we know by the general theory of Hecke L-functions (see [Lan1994, Ch. XV, §4, Theorem 3], essentially the 3-4-2 trick) that the series defining $L(s,(\cdot)^k)$ converges to a holomorphic function on $\Re(s)>1$, which extends holomorphically to $\Re(s)\geq 1$ without any zeroes there either (there are no zeros on $\Re(s)>1$ by the general fact about infinite products of the form that $L(s,\rho)$ is given by [SS2003, Ch. 5, Proposition 3.1]). Since $L(s,1)$ would just be the usual Dedekind zeta function for $\mathbf{Q}(i)$, the hypotheses for Proposition 1.3.2 are met, and we conclude that the values $\chi(\pi_{\mathfrak{p}})$ are equidistributed in S^1 for the Haar measure ($G=S^1$ is abelian so X and G are the same, and μ is the same as the Haar measure).

The finite primes $\mathfrak{p}|p$ are either split, ramified, or inert. The inert ones are those for which $\mathfrak{p}$ is the only prime lying over p, in which case $\mathrm{N}\mathfrak{p}=p^2$. The split ones are those for which $\mathfrak{p},\overline{\mathfrak{p}}$ are distinct and are the only two primes lying over p. There is only one ramified prime, namely $(1+i)|2$, which we can safely ignore because its contribution will always go to zero (because 1 is a finite number).

In fact, we can also ignore the contribution of the inert primes. Their contribution to the sum from Proposition 1.3.1 is

$$\left|\lim_{T\to\infty}\frac{1}{\{\mathfrak{p}|:\mathrm{N}\mathfrak{p}\leq T\}}\sum_{\substack{\mathfrak{p}|p\text{ inert}\\ p^2\leq T}}\chi(\pi_{\mathfrak{p}})^k\right| \leq \lim_{T\to\infty}\frac{\#\{\mathfrak{p}|p\text{ inert }:p^2\leq T\}}{\#\{p:p\leq T\}} \leq \lim_{T\to\infty}\frac{\#\{p:p^2\leq T\}}{\#\{p:p\leq T\}} = 0$$

by the prime number theorem. So the equidistribution of the $\chi(\pi_{\mathfrak{p}})$ over all $\mathfrak{p}$ implies the equidistribution of the $\chi(\pi_{\mathfrak{p}})$ over just the $\mathfrak{p}|p$ which are split. In other words: if we order the rational primes in the usual way, and consider only those of the form $p=4k+1$, so that $p=a_p^2+b_p^2$ where $\{a_p,b_p\}\in\mathbf{N}^2$ is uniquely determined by p, then if we let $\theta_p^{(1)}=\arctan(b_p/a_p)$ and $\theta_p^{(2)}=\arctan(a_p/b_p)$, the sequence $\theta_3^{(1)},\theta_3^{(2)},\theta_5^{(1)},\ldots$ is equidistibuted on $[0,\pi/2]$. If we want to rewrite this in a way more "intrinsic" to p, one way would be to just say that the smaller θ_p among the two is equidistributed in $[0,\pi/4]$ (this sequence has just one element instead of two per p and contains the same information since $\theta_p^{(1)}$ and $\theta_p^{(2)}$ are the first quadrant representatives of $\pi_{\mathfrak{p}}$ and $\pi_{\overline{\mathfrak{p}}}$ so one of them determines the other in an explicit enough way [$\theta_p^{(2)}=\frac{\pi}{2}-\theta_p^{(1)}$] as $\pi_{\overline{\mathfrak{p}}}$ has a representative in the fourth quadrant which is the same as the complex conjugate of a first-quadrant representative of $\pi_{\mathfrak{p}}$). In a more symmetric fashion: an unordered pair $\{a,\overline{a}\}$ for $a\in S^1\subset\mathbf{C}$ is determined by $a+\overline{a}=2\Re(a)\in[-2,2]$. The pushforward measure on $[-2,2]$ of the Haar measure on S^1 under $\theta\mapsto 2\cos\theta$ (as I have computed by taking the derivative of arccos) can be normalized to be $\frac{dx}{\pi\sqrt{4-x^2}}$. The fact that $\chi(\pi_{\mathfrak{p}})$ (which lives on S^1 and only depends on the choice of $\mathfrak{p}$) is equidistributed on S^1 is equivalent to $2\Re(\chi(\pi_{\mathfrak{p}}))$ being equidistributed on $[-2, 2]$ with the measure mentioned above, since if something is in the image of $\{x_{\mathfrak{p}}\}$, then both possible preimages are in the $\{x_{\mathfrak{p}}\}$. Moreover, $2\Re(\chi(\pi_{\mathfrak{p}}))$ does not depend on the choice of $\pi_{\mathfrak{p}}$ (that was the point of χ) nor the choice of $\mathfrak{p}$ lying over p (that was the point of mapping to $[-2, 2]$), so we can really deduce that a naturally constructed sequence whose elements are naturally indexed by the rational primes $p = 4k+1$ (rather than one that is twice as big), namely

$$p \mapsto \frac{2}{p^2}\Re((a_p + b_p i)^4) = \frac{2a_p^4 - 6a_p^2 b_p^2 + 2b_p^4}{p^2},$$

is equidistributed on $[-2, 2]$ with respect to $\frac{dx}{\pi\sqrt{4-x^2}}$, and this is equivalent to the original equidistribution statement involving all the primes of $\mathbf{Q}(i)$. Indeed, this is a quantity that does not depend on the order or sign of a and b, so a_p, b_p can be chosen arbitrarily among the $(a_p, b_p) \in \mathbf{Z}^2$ such that $a_p^2 + b_p^2 = p$. Later in this letter (also copying what you wrote on the board at some point) I will write this in a more reasonable way as the equidistribution of traces of Frobenii in an induced Galois representation

Now let us consider a much more involved case of equidistribution, namely the Sato–Tate conjecture. Fix a rational prime ℓ. Let E be an elliptic curve over a number field F, and let Σ be the set of finite places v of F such that E has good reduction and v does not lie over ℓ (this eliminates only finitely many places of v so it makes no difference from the perspective of equidistribution). For all finite places v (though we will only care about those in Σ), let $\kappa_v = \mathcal{O}_F/\mathfrak{p}_v$.

Thanks to our two conditions, the easy direction of the Néron–Ogg–Shafarevich criterion [Sil2009, VII.4.1] tells us that the global ℓ-adic Galois representation $V_\ell := \mathbf{Q}_\ell \otimes T_\ell E$ (that is, $\rho : G_F \to \mathrm{GL}_2(\mathbf{Q}_\ell) \cong \mathrm{GL}(V_\ell)$) is unramified at all $v \in \Sigma$. Moreover, either by the classical explicit calculations that work for elliptic curves, or by the Weil conjectures, the Frobenii $\rho(\mathrm{Frob}_v)$, which are conjugacy classes in $\mathrm{GL}_2(\mathbf{Q}_\ell)$, satisfy the following property:

Theorem 1.3.4 (Hasse–Weil). *The characteristic polynomial $p_v(X)$ of $\rho(\mathrm{Frob}_v)$ (a priori a quadratic polynomial with coefficients in $\mathbf{Q}_p$) actually has coefficients in $\mathbf{Q}$. More precisely,*

$$p_v(X) = X^2 - (\mathrm{N}v + 1 - \#E_v(k_v))X + \mathrm{N}v,$$

where E_v denotes the mod-$\mathfrak{p}_v$-reduction of E.

Sketch of proof. The constant term is what it is just because $\det \circ \rho$ is the global ℓ-adic cyclotomic character (thanks to the fact that the Weil pairing respects the Galois action [Sil2009, III.8.1]), or because of the Riemann hypothesis part of the Weil conjectures. The X-coefficient (i.e.

computation of the trace) is what it is by writing

$$\mathrm{Tr}\rho(\mathrm{Frob}_v) = 1 + \det \rho(\mathrm{Frob}_v) - \det(\mathrm{id} - \rho(\mathrm{Frob}_v))$$

(this is a general fact about 2×2 matrices), this time using the fact [Sil2009, III.8.2] that the dual isogeny to an endomorphism f acts on the Tate module as the adjoint (with respect to the Weil pairing) of f to identify those determinants with degrees of morphisms F_v and $\mathrm{id} - F_v$ on E_v (F_v being the Frobenius morphism that acts on $\overline{k_v}$-points by $\mathrm{N}v$-th powers of the coordinates, which induces the action of $\rho(\mathrm{Frob}_v)$ on $T_\ell E = T_\ell E_v$) and then using the fact that $\mathrm{id} - F_v$ is a separable morphism [Sil2009, III.5.5], so its degree tells us about the number of $\overline{k}_v$-points in its kernel, i.e. the number of k_v-points on E_v. Alternatively, the trace computation is the Lefschetz fixed-point formula part of the Weil conjectures, in which case you have to unwrap what the Frobenius action is on the 1-dimensional $\overline{\mathbf{Q}}_\ell$-vector spaces $H^0_{\text{ét}}(E_v, \overline{\mathbf{Q}}_\ell)$ and $H^2_{\text{ét}}(E_v, \overline{\mathbf{Q}}_\ell)$. □

Thanks to Theorem 1.3.4, the data of the conjugacy classes $\rho(\mathrm{Frob}_v)$ in $GL_2(\mathbf{Q}_\ell)$ is of great interest, because if we understand them, then we understand the number of points in the mod-$\mathfrak{p}_v$-reductions of E for $v \in \Sigma$. For any choice of Frob_v, we know from Theorem 1.3.4 that the two generalized eigenvalues of $\rho(\mathrm{Frob}_v)$ are actually in $\overline{\mathbf{Q}} \subset \overline{\mathbf{Q}}_\ell$ (choose an embedding ι_ℓ), and that they are complex conjugates (choose an embedding of $\overline{\mathbf{Q}} \to \mathbf{C}$ (choose an embedding ι_∞), after which the generalized eigenvalues will still satisfy $p_v(X)$) of absolute value $\sqrt{\mathrm{N}v}$. To each $v \in \Sigma$, we can therefore attach the conjugacy class x_v of $SU(2)$ whose eigenvalues are[13]

$$\frac{\alpha_v}{\sqrt{\mathrm{N}v}}, \frac{\overline{\alpha_v}}{\sqrt{\mathrm{N}v}}$$

where α_v and its complex conjugate $\overline{\alpha_v}$ are the generalized eigenvalues of $\rho(\mathrm{Frob}_v)$ (here we abuse notation by using α_v to denote what is really $\iota_\infty \circ \iota_\ell^{-1}(\alpha_v)$). By the spectral theorem, the conjugacy classes of $SU(2)$ are determined by the choice of diagonal matrix whose entries are two complex conjugate elements of S^1, so it certainly makes sense to send our conjugacy classes $\rho(\mathrm{Frob}_v)$ to these particular diagonal ones.

Since the set of conjugacy classes of $SU(2)$ is in bijection with $[-2, 2]$ under the trace map $\mathrm{Tr} : [\mathrm{diag}(\alpha, \overline{\alpha})] \mapsto \alpha + \overline{\alpha} = 2\Re(\alpha)$, asking about equidistribution in $SU(2)$ of the $\rho(\mathrm{Frob}_v)$ is the same as asking about equidistribution under the pushforward of the chosen measure (which is the Haar measure as usual) under Tr of the traces of the Frobenii. One reason this is interesting is that by Theorem 1.3.4, understanding the distribution of $\mathrm{Tr}\rho(\mathrm{Frob}_v)$ as v varies (ordering by $\mathrm{N}v$ as usual) is the same as understanding the distribution of the error term in the Hasse–Weil estimate $E_v(k_v) \approx \mathrm{N}v + 1$.

[13] Note the similarity to Example 1.3.3, where we had two complex numbers in S^1 which were complex conjugates, namely $\frac{\pi_\mathfrak{p}}{\sqrt{p}}$ and its complex conjugate, and we were also interested in the distribution of this unordered pair of complex conjugates, which we did by taking their sum.

I did the exercise of computing the pushforward of the Haar measure on $SU(2)$ under Tr. If I had used polar coordinates on $SU(2)$, I would have ended up with a very simple computation. Unfortunately, I only realized this halfway through the computation I started.

Lemma 1.3.5. *The pushforward of the Haar measure on $SU(2)$ under $\mathrm{Tr} : SU(2) \to [-2, 2]$ is, up to a constant factor that we multiply by to make it a probability measure, $\frac{1}{2\pi}\sqrt{4 - x^2}\, dx$.*

Proof. $SU(2)$, via the usual identification with the unit Hamilton quaternions (identified with a pair of complex numbers by $a + bi + cj + dk \mapsto (a + bi, c + di)$), can be identified with $S^3 \subset \mathbf{C}^2$. The matrix (left-)action of $\alpha \in SU(2)$ on $\mathbf{C}^2$ coincides on S^3 with the one you get by identifying elements of S^3 with the corresponding element of $SU(2)$ and multiplying on the left by α, which I checked by explicitly writing down the isomorphism $SU(2) \cong S^3$. By definition, the action of $\alpha : SU(2)$ induces isomorphisms $T_x\mathbf{C}^2 \to T_{\alpha(x)}\mathbf{C}^2$ for each $x \in \mathbf{C}^2$ which preserve the natural Riemannian metric on $\mathbf{C}^2$ (the Euclidean one coming from $\mathbf{C}^2 \cong \mathbf{R}^4$, i.e. just the standard inner product at each point). If we equip $S^3 \subset \mathbf{C}^2$ with the metric induced from that inclusion, then we see that the action of $SU(2)$ on $SU(2) = S^3$ also preserves this metric (since you get the induced metric just by restricting from $T_x\mathbf{C}^2$ to T_xS^3, which is the same way that you get the pushforward of $\alpha \in SU(2)$ acting on S^3 rather than all of $\mathbf{C}^2$). But S^3 has a volume form coming from that metric, which induces a measure on S^3. The fact that $SU(2)$ preserves the metric implies that it preserves this measure when left-multiplying. By the fact that compact groups are unimodular, this makes the measure we just defined a Haar measure on $SU(2)$. So, to write down a Haar measure on $SU(2)$, all we need to do is explicitly write down the volume form on S^3 coming from the induced metric.

Once we have some good local coordinates for S^3 (for example, polar coordinates, which I mistakenly did not use), we could just use the fact that in local coordinates, the volume form induced by the metric g_{ij} is

$$\sqrt{\det g}\, dx^1 \wedge dx^2 \wedge dx^3,$$

and if the local coordinates are $\varphi : U \to S^3$ (U an open subset of $\mathbf{R}^3$), then the *definition* of the induced metric g_{ij} from the inclusion $S^3 \to \mathbf{R}^4$ with the Euclidean δ_{ij} metric is

$$\begin{aligned} g_{ij} &= g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) \\ &= \delta\left(\varphi^*\left(\frac{\partial}{\partial x^i}\right), \varphi^*\left(\frac{\partial}{\partial x^j}\right)\right) \\ &= \frac{\partial \varphi^\alpha}{\partial x^i}\frac{\partial \varphi^\beta}{\partial x^j}\delta_{\alpha\beta}. \end{aligned}$$

This is all well and good, but I did it a different way before I had thought rigorously about what the right volume form is. Suppose we have some local coordinates $\varphi : U \to S^3$ which is a local diffeomorphism almost everywhere on S^3. Then we get a local diffeomorphism almost everywhere (for example definitely not hitting the measure-zero set $\{0\}$) $\psi : U \times \mathbf{R}_{>0} \to \mathbf{R}^4$ given by $(x, r) \mapsto r\varphi(x)$. In the local coordinates provided by U, the volume form we want will just be $i_X(\psi^*\omega)|_{U\times\{1\}}$, where i_X is the interior derivative along the vector field X on $U \times \mathbf{R}_{>0}$ which is given at (u, r) by the unit normal $(0, 1) \in T_{(u,r)}(U \times \mathbf{R}_{>0})$, and $\omega = dx^1 \wedge dx^2 \wedge dx^3 \wedge dx^4$ is the Euclidean volume form on $\mathbf{R}^4$. It is easy to check, up to sign, that this is the same volume form as the pullback via φ of the volume form on S^3 coming from the Euclidean-induced metric: both of them give us the same definition of "positively-oriented covolume-1". Indeed, for all $p \in S^3$, the set of bases X_1, X_2, X_3 of T_pS^3 such that $(i_{\psi^*X}(\omega)|_{S^3})_p(X_1, X_2, X_3) = 1$ is, by definition of the interior derivative, the same as the set of bases such that $\omega(\psi^*X, X_1, X_2, X_3) = 1$, which (since ψ^*X is the outward-pointing *unit* normal to S^3 in $\mathbf{R}^4$) is the same as the set of bases that are actually positively[14]-oriented bases of covolume 1 for T_pS^3 under the Euclidean inner product obtained by restricting from $T_p\mathbf{R}^4$. This is exactly what the intrinsic definition of the volume form induced by a metric is (up to sign): that the volume form ω_g induced by the metric g_{ij} on an n-dimensional oriented Riemannian manifold is equal to $e_1^\vee \wedge \cdots e_n^\vee$ for any positively-oriented covolume-1 frame $e_1, \ldots, e_n$ (which can be assumed to be orthonormal) is just because in the coordinates $x^1, \ldots, x^n$ we can write $\frac{\partial}{\partial x^i} = Me_i$ for some $n \times n$ matrix M (M varies along the manifold just like everything else), in which case

$$e_1^\vee \wedge \cdots \wedge e_n^\vee = (\det M)\, dx^1 \wedge \cdots \wedge dx^n = \sqrt{\det g}\, dx^1 \wedge \cdots \wedge dx^n =: \omega_g$$

where the calculation of $\det M = \sqrt{\det g}$ is because the matrix g_{ij} is defined, in terms of the abstract inner product that the data of the metric is, by

$$g_{ij} = \left\langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \right\rangle_g = \langle Me_i, Me_j\rangle_g = \langle e_i, M^\top Me_j\rangle_g = (M^\top M)_{ij}.$$

In fact, this is why the volume form is defined the way it is with the $\sqrt{\det g}$.

Having justified the equivalence of the two volume forms, it remains only to write down an appropriate $\varphi : U \to S^3$ and compute the volume form $i_X(\psi^*\omega)|_{U\times\{1\}}$. By doing it this way, we avoid computing all the inner products of the previous method, but have to compute a 4×4 determinant rather than a 3×3[15]. Since $S^3 \subset \mathbf{C}^2$ is the set of $(a + bi, c + di)$ such that $a^2 + b^2 + c^2 + d^2 = 1$, and so $a^2 + b^2, c^2 + d^2 \leq 1$, a very natural choice of coordinates on $SU(2) = S^3$ is by

$$\varphi : U = [0, \pi/2] \times [0, 2\pi] \times [0, 2\pi] \to S^3$$

,

$$(\theta, \psi_1, \psi_2) \mapsto (e^{i\psi_1} \cos\theta, e^{i\psi_2} \sin\theta).$$

[14] Up to sign: so this could be supposed to be "negatively-oriented."

[15] In reality I don't think I saved any time; it's just that this was the first thing I came up with and did in my notes; then in this letter I had to justify the equivalence which definitely made it not worth it in terms of time.

The Jacobian of the resulting $\psi : U \times \mathbf{R}_{>0} \to \mathbf{R}^3$ is then

$$\begin{vmatrix} \cos\psi_1\cos\theta & -r\cos\theta\sin\psi_1 & 0 & -r\cos\psi_1\sin\theta \\ \sin\psi_1\cos\theta & r\cos\theta\cos\psi_1 & 0 & -r\sin\psi_1\sin\theta \\ \cos\psi_2\sin\theta & 0 & -r\sin\theta\sin\psi_2 & r\cos\psi_2\cos\theta \\ \sin\psi_2\sin\theta & 0 & r\sin\theta\cos\psi_2 & r\sin\psi_2\cos\theta \end{vmatrix} = r^3\sin\theta\cos\theta,$$

so the form $\psi^*\omega = \psi^*(dx^1 \wedge \cdots \wedge dx^4) = r^3 \sin\theta\cos\theta\, dr \wedge d\theta \wedge d\psi_1 \wedge d\psi_2$. Its interior product with $X = (0,1)$, restricted to the locus $r = 1$, is $\sin\theta\cos\theta d\theta \wedge d\psi_1 \wedge d\psi_2$. This is the Haar measure for $SU(2)$ in the coordinates we have provided. If I had chosen polar coordinates, it would now be obvious what the pushforward under Tr is, but instead we have to do more computation. Let μ be the measure we just wrote down on $SU(2)$. The measure $\mathrm{Tr}^*\mu$ on $[-2,2]$ is given by, for $[a,b] \subset [-2,2]$ with WLOG $[a,b] \subset [0,2]$,

$$\begin{aligned}
\mathrm{Tr}^*\mu((a,b)) &= \iiint_{\substack{\theta,\psi_1,\psi_2 \\ 2\cos\psi_1\cos\theta\in(a,b)}} \sin\theta\cos\theta\, d\theta d\psi_1 d\psi_2 \\
&= 2\pi \iint_{\substack{\theta,\psi_1 \\ 2\cos\psi_1\cos\theta\in(a,b)}} \sin\theta\cos\theta\, d\theta d\psi_1 \\
&= 4\pi \int_{-2}^{2} \int_{\cos\theta\in(a/y,b/y)} \sin\theta\cos\theta\, d\theta\, \frac{-1}{\sqrt{4-y^2}}\, dy \\
&= 4\pi \int_a^b \int_{a/y}^{1} x\, dx\, \frac{1}{\sqrt{4-y^2}}\, dy + 4\pi \int_b^2 \int_{a/y}^{b/y} x\, dx\, \frac{1}{\sqrt{4-y^2}}\, dy \\
&= 2\pi \int_a^b \left(1 - \frac{a^2}{y^2}\right) \frac{1}{\sqrt{4-y^2}}\, dy + 2\pi \int_b^2 \left(\frac{b^2}{y^2} - \frac{a^2}{y^2}\right) dx\, \frac{1}{\sqrt{4-y^2}}\, dy \\
&= 2\pi \int_a^b \frac{dy}{\sqrt{4-y^2}} - 2\pi a^2 \int_a^2 \frac{dy}{y^2\sqrt{4-y^2}} + 2\pi b^2 \int_b^2 \frac{dy}{y^2\sqrt{4-y^2}} \\
&= 2\pi \left[\arcsin\left(\frac{b}{2}\right) - \arcsin\left(\frac{a}{2}\right)\right] - \frac{\pi}{2} a\sqrt{4-a^2} + \frac{\pi}{2} b\sqrt{4-b^2} \\
&= \pi \int_a^b \sqrt{4-x^2}\, dx.
\end{aligned}$$

Renormalizing to make it a probability measure, we get the desired $c\mathrm{Tr}^*\mu = \frac{1}{2\pi}\sqrt{4-x^2}\, dx$ for some constant c. $\square$

Remark 1.3.6. Note that even though in either case we are taking the pushforward under the trace map of a Haar measure in order to state in the simplest way an equidistribution result about eigenvalues of Frobenii that come in complex conjugate pairs, the measure we get on $[-2,2]$ in Lemma 1.3.5 is completely different than the one we got in Example 1.3.3. This is because even when we rewrite Example 1.3.3 in terms of a Galois representation formed by $\mathrm{Ind}_{G_{\mathbf{Q}(i)}}^{G_{\mathbf{Q}}}$, the image of that Galois representation will be inside of $S^1 \subset SU(2)$ (the diagonal matrices that represent all the conjugacy classes), whereas the image of the Galois representation we are interested in here (at least in the non-CM case[16]) is too big, so we have to consider the Haar measure on all of $SU(2)$ rather than just S^1. Indeed, the pushforward of the Haar measure on $SU(2)$ to S^1 (viewed as the space of conjugacy classes) is not the Haar measure $\frac{1}{2\pi}d\theta$ on S^1. Instead, it is $\frac{2}{\pi}\sin^2\theta\, d\theta$, which I computed just by pulling back the volume form $\sqrt{4-x^2}\,dx$ on $[-2,2]$ under the trace map $S^1 \to [-2,2]$, $\alpha \mapsto \alpha + \overline{\alpha} = 2\cos\theta$.

Anyway, now that we have measures on the space of conjugacy classes of $SU(2)$, and, what is equivalent, the "Sato–Tate measure" on $[-2,2]$ obtained by pushforward under the trace of that measure and computed to be the "semicircle" measure in Lemma 1.3.5, I can state the Sato–Tate conjecture:

Conjecture 1.3.7 (Sato–Tate)**.** Fix F, E, ℓ, Σ, ρ as above, and suppose that E does not have CM. For $v \in \Sigma$, let $x_v = \mathrm{Tr}\rho(\mathrm{Frob}_v)$, where $\rho(\mathrm{Frob}_v)$ is considered as a conjugacy class in $SU(2)$ as explained above. Then the sequence $\{x_v\} \in [-2,2]$, ordered by $\mathrm{N}v$, is equidistributed with respect to the Sato–Tate measure, that is, $\frac{1}{2\pi}\sqrt{4-x^2}\,dx$.

By the L-function machinery, it makes sense that properties of L-functions would be related to Conjecture 1.3.7. The key point is that the symmetric power L-functions appear because the irreducible representations of $SU(2)$ are exactly the matrix actions on the symmetric powers of $\mathbf{C}^2$.

To be more precise: by Proposition 1.3.2, the L-functions we should be interested in for the purposes of Conjecture 1.3.7 are the "$L(s,\rho)$" obtained from the sequence $\{x_v\}$ for all irreducible representations $\rho = \mathrm{Sym}^k\mathbf{C}^2$. By definition of the x_v,

$$L(s, \mathrm{Sym}^k\mathbf{C}^2) = \prod_{v\in\Sigma} \frac{1}{\det(\mathrm{id} - (\mathrm{N}v)^{-s}\cdot \mathrm{Sym}^k \frac{\rho(\mathrm{Frob}_v)}{\sqrt{\mathrm{N}v}})},$$

where here $\frac{\rho(\mathrm{Frob}_v)}{\sqrt{\mathrm{N}v}}$ is just standing for the conjugacy class in $SU(2)$ that we were discussing earlier (we can just pick an arbitrary element, e.g. a diagonal one), and ρ is $V_\ell(E)$ rather than an irreducible representation of $SU(2)$. If $\alpha_v, \overline{\alpha_v} \in S^1$ are the two eigenvalues of $\rho(\mathrm{Frob}_v)/\sqrt{\mathrm{N}v}$, then its Sym^k is the $(k+1)$-dimensional $\mathbf{C}$-vector space of degree-k homogeneous elements of $\mathbf{C}[X,Y]$, and since $\rho(\mathrm{Frob}_v)/\sqrt{\mathrm{N}v}$ can be assumed to be $\mathrm{diag}(\alpha_v, \overline{\alpha_v})$, hence acting on X by

[16] This explains why the Sato–Tate conjecture is only the way it is for non-CM elliptic curves; in the CM case the measure, being obtained by pushforward from S^1 rather than all of $SU(2)$, is different and one is essentially done by a consequence of the main theorem of complex multiplication [Sil1994, II.10.5] which says that the Galois representation we are interested in is, as in Example 1.3.3, obtained by induction from $G_{\mathbf{Q}(i)}$ (though the Galois representation corresponding to the Hecke character from Example 1.3.3 via Weil's construction is *never* the one associated to a CM elliptic curve, as we will see).

α_v and Y by $\overline{\alpha_v}$; so its Sym^k is also diagonal, with eigenvalues equal to

$$\alpha_v^m \overline{\alpha_v}^{k-m} = \alpha_v^{2m-k}, m = 0, \ldots, k.$$

So we have in fact

$$L(s, \mathrm{Sym}^k \mathbf{C}^2) = \prod_{v \in \Sigma} \prod_{m=0}^{k} \frac{1}{1 - (\mathrm{N}v)^{-s} \alpha_v^{2m-k}}.$$

Granting that E has no CM, this L-function does not seem to be easily written in terms of Hecke characters; so we need some other strategy to show that it has the analytic properties required by Proposition 1.3.2 (we would need to do this even if E did have CM, it's just that we would then be able to apply [Sil1994, II.10.5] and relate the situation to the situation of L-functions of Hecke characters which are well-understood). The point is that $L(s, \mathrm{Sym}^k \mathbf{C}^2)$ is essentially the same as the L-function of the k-th symmetric power of the G_F-representation that is the rational ℓ-adic Tate module of E: the latter L-function is defined to be the Artin L-function

$$\begin{aligned} L(s, \mathrm{Sym}^k \rho) &:= \prod_{v \in \Sigma} \frac{1}{\det(\mathrm{id} - (\mathrm{N}v)^{-s} \mathrm{Sym}^k \rho(\mathrm{Frob}_v))} \\ &= \prod_{v \in \Sigma} \prod_{m=0}^{k} \frac{1}{1 - (\mathrm{N}v)^{-s} (\sqrt{\mathrm{N}v})^k \alpha_v^{2m-k}} \\ &= L\left(s - \frac{k}{2}, \mathrm{Sym}^k \mathbf{C}^2\right). \end{aligned}$$

If we can prove (e.g. in some form of symmetric power functoriality) that $L(s, \mathrm{Sym}^k \rho)$ equals the L-function of a cuspidal automorphic representation of $\mathrm{GL}_{k+1}(\mathbf{A}_F)$, then it is a general fact due to Jacquet–Shalika [JS1976] that it does not vanish on the line $\Re(s) = 1$. We would then conclude that $L(s, \mathrm{Sym}^k \mathbf{C}^2)$ does not vanish for $\Re(s) \geq 1 - k/2$ for all $k \geq 1$. For $k = 0$ we clearly have a zeta function, so we would have a valid proof of Conjecture 1.3.7 by Proposition 1.3.2.

Remark 1.3.8. In [Ser1989, I-26], Serre claims that I have a sign error – it should be $s + k/2$ in the input to the L function at the bottom of the display equation above, and $\Re(s) \geq 1 + k/2$ for the place where $L(s, \mathrm{Sym}^k \mathbf{C}^2)$ does not vanish. I mention this here since differing from Serre means there is certainly a mistake in my work.

Chapter 2

Eigenvarieties

"Many of the truths we cling to depend greatly on our own point of view."

Obi–Wan Kenobi to Luke Skywalker,
Star Wars Episode VI: Return of the Jedi

Eigenvarieties are rigid analytic spaces that parametrize p-adic families of automorphic forms. The concept was introduced by Coleman–Mazur [CM1998], who constructed (for $p > 2$) the *eigencurve* (equidimensional of dimension 1 as the name suggests), a reduced rigid analytic space $\mathscr{E}_{p,1}$ over $\mathbf{Q}_p$ containing a Zariski-dense subset of $\mathbf{C}_p$-points (the "classical modular locus" of [CM1998, Definition 1]) that is in natural bijection with the set of normalized eigenforms of finite slope and level $\Gamma_1(p^{m\geq 1})$ [CM1998, Theorem F]. In fact, $\mathscr{E}_{p,1}(\mathbf{C}_p)$ is itself in natural bijection with the set of "normalized finite-slope overconvergent modular forms of tame level 1" [CM1998, Theorem E]. The construction of Coleman–Mazur relied in a crucial way on earlier work of Coleman [Col1997b] on p-adic families of modular forms. The main technical point of that work was to generalize the theory of compact operators on p-adic Banach spaces [Ser1962] to Banach spaces over the affinoid algebra of weight space.

More generally, Buzzard [Buz2007, part II] constructed, for any rational prime p and any tame level N, a generalization of the eigencurve parametrizing normalized finite-slope p-adic overconvergent eigenforms of tame level N. The construction of eigenvarieties parametrizing p-adic families of automorphic forms other than holomorphic modular forms has grown into an important and useful industry, having notably been done for Hilbert modular forms [KL2005, AM2004, Pil2013], forms of GL_2/F compact modulo center at infinity [Buz2007, Part III] for $F/\mathbf{Q}$ totally real (e.g. quaternionic Hilbert modular forms), forms of $\mathrm{GL}_n/\mathbf{Q}$ compact modulo center at infinity [Che2004], and for reductive groups G such that $G^{\mathrm{der}}(\mathbf{R})$ satisfies the Harish-Chandra condition [Urb2011].

Of particular importance to this mémoire is Emerton's landmark work [Eme2006b], which works in the special cases that we will require (namely the case of the Coleman–Mazur eigencurve and the case of eigenvariety for definite unitary groups).

In broad strokes, there are typically two ways of thinking about an eigenvariety. We use the "C" vs. "D" notation in order to be consistent with [CM1998] and [Buz2007]:

(1) The "C" eigencurve of [CM1998, §6] is defined as a Zariski-closed subspace of $\mathfrak{X}^{\mathrm{rig}} \times \mathcal{T}$, where $\mathfrak{X} = \mathrm{Spf} R^{\mathrm{univ}}$ is a formal scheme coming from a universal deformation ring (so that its Berthelot rigid generic fiber[1] is a rigid space parametrizing some collection of pseudorepresentations), and $\mathcal{T}$ is a space intended to parametrize nonzero Atkin–Lehner Hecke eigenvalues (typically it is a product of copies of $\mathbf{G}_m$ or the character variety of a torus depending on the exact viewpoint). In the case of the Coleman–Mazur eigencurve, this is essentially classical: by Atkin–Lehner theory, this is the information that determines a modular form — the oldforms for $\Gamma_0(pN)$ associated to a given newform for $\Gamma_0(N)$ have the same associated Galois representation, but they are distinguished by their Atkin–Lehner U_p-eigenvalue. Indeed, once the "C" eigenvariety is projected down to $\mathfrak{X}^{\mathrm{rig}}$, one obtains the "infinite fern" of [GM1998, Che2011], the many transverse intersections of which come from Coleman families interpolating the various eigenforms projecting to the same point in $\mathfrak{X}$.

(2) The "D" eigencurve of [CM1998, §7] is constructed without any Galois-theoretic input, and it is axiomitized by Buzzard in [Buz2007]. Essentially, once one has a system of Banach modules on weight space equipped with a Hecke action where some distinguished operator U (playing the role of Atkin–Lehner U_p operator) for which everything is finite-slope is compact (see also [Che2004] for a nice description of the setup), one can construct the "spectral variety," which is a Fredholm variety given by the vanishing locus of the characteristic series of U. The eigenvariety is then constructed by taking a affinoid cover of the spectral variety over which the characteristic series of αU (for all α in the Hecke algebra) is divisible by a polynomial $Q(T)$, proving this cover is admissible (always [Buz2007, Theorem 4.6] or some variant), and defining the eigenvariety over an affinoid $\mathrm{Sp} A$ in the cover to be the Sp of the image of the Hecke algebra in the endomorphism algebra of the finite-type module of overconvergent forms killed by $T^{\deg Q} Q(T^{-1})$. It turns out that this construction can be glued to form the eigenvariety D, which is locally finite flat in the source over weight space (this is [CM1998, Theorem C] and follows directly from the construction we just described).

Correspondingly, there are two approaches: one can construct the "D" eigencurve using just the spectral eigenvariety machine of [Buz2007], and then embed it in $\mathfrak{X}^{\mathrm{rig}} \times \mathcal{T}$ by constructing a universal pseudocharacter on D (e.g. by using [Che2014, Example 2.32] to interpolate the

[1] see [dJ1998, §7]

construction of Galois representations associated to a Zariski-dense subset of classical automorphic forms, if such a construction is available). Alternatively, one can construct the "C" eigencurve directly as being the Zariski-closure of the classical modular locus in $\mathfrak{X}^{\mathrm{rig}} \times \mathcal{T}$, and prove separately that the classical points are dense in D in order to show that they coincide (e.g. via the technique of [BC2009a, Proposition 7.2.8]). Note that either way, these approaches both require proving that the classical points are Zariski-dense in the natural rigid space that contains them. This is NOT always true, as remarked by Ash–Pollack–Stevens [AIS2014] in the case of $\mathrm{GL}_n/\mathbf{Q}$ and by Calegari–Mazur [CM2009] in the case of GL_2/K where $K/\mathbf{Q}$ is imaginary quadratic (as is remarked in the introduction to [Urb2011]). However, such a density result is true in the cases required by [NT2021], namely $\mathrm{GL}_2/\mathbf{Q}$ and definite unitary groups, and the proof of this fact using the representation-theoretic techniques of Emerton [Eme2006b] will be a focus of this chapter. We also remark here (before doing the full detail) that Emerton's construction is very nice because it provides a global version of the construction of the "D" eigencurve (at least in the sense that all the local constructions and gluing are hidden in the theory of relative Sp, which is taken care of by [Con2006]).

Before starting with the full detail, we finally remark that the "D" eigenvariety construction is very useful for proving geometric properties of eigenvarieties: for example the fact that the eigenvariety is equidimensional of certain relative dimension over weight space is typically a direct consequence of the "D" construction. In addition, there is a well-known conjecture of Coleman–Mazur–Buzzard–Kilford to the effect that eigenvarieties are supposed to admit a particularly nice structure over boundary annuli of weight space [LWX2017, Conjecture 1.2]. Progress on this conjecture in special cases [BK2005, LWX2017, Ye2019, JN2019] has typically used the "D" eigenvariety together with explicit computations of slopes using the Newton polygon of the characteristic series of U_p (indeed, it was motivated by such computations). We mention this conjecture in particular because the special case of it that was proved by Buzzard–Kilford [BK2005] for $\mathscr{E}_{2,1}$ is used in a crucial way in the "ping-pong" argument of [NT2021].

While the "C" eigenvariety construction on its own does not *a priori* give any useful information in the same way as the "D" eigenvariety construction does, keeping that perspective is ultimately equally useful to the application to symmetric power functoriality in [NT2021]. Indeed, the entire point is that the question of symmetric power functoriality can be reinterpreted via the "C" construction as being a question about comparing the image of the GL_2-eigenvariety under the obvious Sym^{n-1} morphism (obvious because it is obvious how to define it on the ambient space $\mathfrak{X}^{\mathrm{rig}} \times \mathcal{T}$) and the GL_n-eigenvariety (except of course there is not yet any useful notion of "GL_n eigenvariety" and so a definite unitary group of rank n must be used). Moreover, this perspective allows for the systematic use of p-adic Hodge theory to study the eigenvariety, particularly via comparison with the trianguline variety, which is also central to the arguments of [NT2021].

Having a global object that parametrizes p-adic automorphic forms (that is, the entire eigenvariety, as opposed to just an understanding of how to interpolate locally in families, especially the "C" interpretation of it), and understanding the global structure of that object (especially via the "D" interpretation), has been for at least two decades a promising approach to attacking the Langlands conjectures via technique of p-adic analytic continuation, as was pointed out in [Kis2003, Bel2019]. An early example of this philosophy is the work of Chenevier [Che2005], which provided a general framework for p-adic interpolation of cases of Langlands functoriality. The recent breakthrough of Newton–Thorne [NT2021] on symmetric power functoriality that is the main subject of this mémoire appears to be one of the first fully successful instances of the technique of p-adic interpolation being applied to make progress on the Langlands conjectures themselves.

2.1 | The concept of a p-adic overconvergent automorphic form

2.1.1 | Overconvergent modular forms, à la Katz and Coleman

Assume[2] for this entire section that $p \geq 5$ is a rational prime and that $N \geq 3$ is a positive integer not divisible by p. The original approach of Coleman–Mazur [CM1998] used as the main input the geometric theory of overconvergent modular forms developed by Katz [Kat1973] and Coleman [Col1997b, Col1996, Col1997a] (see also Gouvea [Gou1988] and the AWS notes of Calegari [Cal2013]). This section follows these standard reference in sketching the main point of the theory of overconvergent modular forms. In the style of [Buz2004] (except with more details), this is meant to motivate the main features and predict some important phenomena of the general theory of overconvergent automorphic forms that will be in reality the viewpoint of the remainder of this thesis.

In my view, the starting point of the geometric theory is really an observation from Serre's paper [Ser1973] that takes the point of view of p-adic families of q-expansions, namely that for $p \geq 5$[3], the normalized Eisenstein series E_{p-1} has q-expansion satisfying $E_{p-1} \equiv 1 \mod p$ (Clausen von Staudt congruence[Ser1973, §1.1(d)]), and therefore E_{p-1} is invertible as a p-adic modular form, as

$$\frac{1}{E_{p-1}} = \lim_{m\to\infty} E_{p-1}^{p^m-1}$$

[2]In reality, the papers of Katz and Coleman deal with basically all choices of p, N, but this requires making the exposition somewhat more technical, due to needing to pass to higher N (in order to have a representable moduli problem) and due to the ad hoc choice of lift of Hasse invariant in the cases $p = 2, 3$ (where E_{p-1} is not a modular form). Neither of these technicalities will be relevant in the representation-theoretic viewpoint, so we ignore them completely in this section.

[3]For $p = 2, 3$ you can still get the theory to work, though it becomes a bit more technical because you need to consider large N and take fixed points by a group action. All of this is done in full detail in [Kat1973].

(this is exactly the argument implicit in [Ser1973, §1.4(c), exemple]). Since E_{p-1} is a lift to characteristic zero of the Hasse invariant on elliptic curves in characteristic p (see [Cal2013, §1] or [Kat1973, §2.1] — this is also a consequence of the Clausen–von Staudt congruence and basic theory of the Hasse invariant), this motivates part of the definition of Katz's modular definition of p-adic modular forms [Kat1973, §2.2]: a p-adic modular form should be considered as a rule defined only[4] on elliptic curves over p-adically complete $\mathbf{Z}_p$-algebras which have ordinary reduction at p. The same kind of reasoning is also present in a slightly different form in [Gou1988, p. 2]. For any complete field $K/\mathbf{Q}_p$, we should therefore define the space of p-adic modular forms of level $\Gamma_1(N)$ over K to be

$$M_k^\dagger(\Gamma_1(N), K, 0) := H^0(X_1(N)(0)_K, \omega^k),$$

where $X_1(N)(0)_K$ is the base-change via $\mathbf{Q}_p \to K$ of the affinoid (by [Col1989, Lemma 3.30]) region of the rigid analytification $X_1(N)^{\mathrm{rig}}_{\mathbf{Q}_p}$ defined by $v_p(E_{p-1}) \leq 0$, and ω^k is the rigid-analytification of the usual line bundle. As it turns out (see maybe [KM1985]), $X_1(N)(0)_K$ is connected. These p-adic modular forms coincide with those of Serre [Ser1973] via the q-expansion principle (that is, evaluation at the appropriate Tate curve): see e.g. [Col1996, Theorem 9.1] (which itself relies on the equivalence between Katz and Serre p-adic modular forms [Kat1973, Theorem 4.5.1]) for the full detail.

However, the theory of p-adic modular forms is usually restricted to the so-called *overconvergent* modular forms, which are those that can be defined on a much larger rigid subspace than just the ordinary locus, namely the locus $X_1(N)(v)$ defined by $v(E_{p-1}) \leq v$, where v is allowed to vary within the interval[5]

$$I_m := \left(0, \frac{p^{2-m}}{p+1}\right) \subset \mathbf{R}$$

where m will usually be related to the level at p. Though Katz uses the language of formal schemes (see [Abb2010, Ray1974, BL1993]), Coleman [Col1996, §2] remarks that Katz [Kat1973, §2.9] shows that if $r \in \mathcal{O}_K$ with $v(r) = v$, his space of modular forms $S(\mathcal{O}_K, r, N, k) \otimes K$ coincides with Coleman's space of v-overconvergent modular forms, namely

$$M_k^\dagger(\Gamma_1(N), K, v) := H^0(X_1(N)(v)_K, \omega^k).$$

With higher level at p, say $\Gamma_1(Np^m)$ with $m \geq 1$, the locus in the $\mathbf{Q}_p$-rigid analytification of $X_1(Np)$ defined by $v(E_{p-1}) \geq v$ is not connected. One must then define $X_1(Np)(v)$ to be

[4]Of course, classical modular forms like E_{p-1} can be defined everywhere, but the point is that we want to consider E_{p-1}^{-1}, being a limit of q-expansions of modular forms, as a p-adic modular form, and this one will only be defined on the ordinary locus.

[5]The reason for this particular interval is due to the theory of the canonical subgroup, which will be partially explained later in this section and then completely in Section 2.1.2.

the connected component of ∞ in this locus, and then define $M_k^\dagger(\Gamma_1(Np), K, v)$ in the exact same way as $M_k^\dagger(\Gamma_1(N), K, v)$ except with "N" replaced with "Np." For a given $v \in I_m$, there are much fewer v-overconvergent modular forms than there are p-adic modular forms. For example, the point of the previous paragraph was that E_{p-1}^{-1} is a p-adic modular form that is not r-overconvergent for any $r \in I_m$, as $p/(p+1) < 1$. The point is that restricting to overconvergent modular forms allows for a much more manageable spectral theory of the Atkin–Lehner U_p-operator on the relevant Banach spaces of forms. Following Calegari [Cal2013, §2.3.1], we first explain why the spectrum of U_p is much too complicated if we consider all finite-slope p-adic modular forms at once. Start with the "p-deprived Ramanujan Δ-function

$$g := \Delta - V_p U_p \Delta = \sum_{(n,p)=1} \tau(n) q^n \in M_{12}(\Gamma_1(p), \mathbf{Z}_p).$$

This satisfies

$$U_p g = 0,$$

i.e. g is of infinite slope. The point is not that this is bad in and of itself (we were definitely going to exclude the infinite-slope forms anyway). Rather, the point is that if we do not care at all about restrictions on the radius of overconvergence, then we can use this to generate overconvergent eigenforms of all sorts of U_p-eigenvalues: for any $\lambda \in \mathbf{C}_p$ satisfying $|\lambda|_p < 1$, we can consider

$$f_\lambda := \sum_{n \geq 0} \lambda^n V_p^n g.$$

Since $|\lambda|_p < 1$, and by definition of V_p on q-expansions (see [Cal2013, Exercise 2.3.2] or [Ser1973]), this indeed defines a p-adic modular form in $M_{12}^\dagger(\Gamma_1(p), \mathcal{O}_{\mathbf{C}_p}, 0)$. Of course, we have no control over how much it overconverges, since applying V_p could be making overconvergence worse at each step. The point is that

$$\begin{aligned}
U_p f_\lambda &= \sum_{n \geq 0} \lambda^n U_p V_p^n g \\
&= U_p g + \sum_{n \geq 1} \lambda^n (U_p V_p)(V_p^{n-1} g) \\
&= 0 + \sum_{n \geq 1} \lambda^n V_p^{n-1} g \\
&= \lambda \sum_{n \geq 0} \lambda^n V_p^n g \\
&= \lambda f_\lambda,
\end{aligned}$$

i.e. we have constructed a p-adic modular form $f_\lambda \in M_{12}^\dagger(\Gamma_1(p), \mathcal{O}_{\mathbf{C}_p}, 0)$ with U_p-eigenvalue λ for any λ in the open unit ball of $\mathbf{C}_p$. This is very bad: it means that U_p does not have discrete spectrum when it acts on the Banach spaces of p-adic modular forms, and hence it won't be compact (see [Dwo1962, §2]) and we cannot apply the theorems of p-adic functional analysis (e.g. those of [Ser1962, Col1997b, Buz2007] mentioned in the introduction that are crucially used to construct eigenvarieties).

On the other hand, if we restrict to the K-Banach space of r-overconvergent p-adic modular forms $M_k^\dagger(\Gamma_1(N), K, v)$ for $v \in I_m \cap \mathbf{R}_{>0}$, then U_p will be compact. The remainder of this section will be devoted to sketching the proof of this fact. It crucially depends on the modular interpretation of U_p (to be explained later) together with the following KEY FACT [Kat1973, Theorem 3.1]:

Theorem 2.1.1 (Lubin–Katz). *Let R_0 be a complete DVR of residue characteristic p, let $r \in R_0$ with $v(r) \in I_1$, and let $(E/R, \alpha_N, Y)$ be a tuple of data*[6] *consisting of*

1. *An elliptic curve $\pi : E \to \mathrm{Spec} R$ (in the sense of* [Kat1973, §1.0]*, where R is a p-adically complete R_0-algebra).*
2. *$\alpha_N : \mu_{N,R} \to E[N]$ a level structure at N.*
3. *$Y \in H^0(E, (\Omega^1_{E/R})^{\otimes(1-p)})$ such that $Y \cdot E_{p-1}(E) = r$.*[7]

There exists a unique way to associate to every such tuple of data an order-p finite flat R-subgroup scheme $\mathscr{K}_{(E,\alpha_N)} \subset E$ depending only on the isomorphism class of the data (E, α_N) such that

1. *The formation of $\mathscr{K}_{(E,\alpha_N)}$ commutes with arbitrary change of p-adically complete R_0-algebras $R \to R'$.*
2. *If*[8] *$p/r \in R_0$ vanishes in R, then $\mathscr{K}$ is the kernel of the relative Frobenius*[9] *$E \to E^{(p)} := E \times_{R, x \mapsto x^p} R$.*
3. *If E/R is the Tate curve $\mathrm{Tate}(q^N)$ over $R = (R_0/p^M R_0)((q))$, then $\mathscr{K}$ is the subgroup*[10] *μ_p.*

[6] This is the first time we are writing down the full detail of what this data is, but previously in this section we explained why it is exactly the modular data that Katz's definition of p-adic overconvergent modular forms is defined on (except for the additional possibility of extra level structure at p). The purpose here is just to explain the main idea of why the theory of overconvergent modular forms is the way it is, before moving on to a more general framework which will be more representation-theoretic and involve none of the modular viewpoint of Katz.

[7] Here the notation makes sense because the value of a weight-$(p-1)$ Katz modular form such as E_{p-1} on a test object E is a global section of $(\Omega^1_{E/R})^{\otimes(p-1)}$.

[8] N.B. $p/r \in R_0$ since R_0 is a DVR and thanks to the assumption that $v(r) \in I_1$

[9] See [SGA3.I, exposé $\mathrm{VII_A}$, §4]

[10] Recall that the subgroup μ_p of the Tate curve just refers to the canonical inclusion $\mu_p \subset \mathbf{G}_m \to \mathrm{Tate}(q^N)$ (of course you have to work through the explicit formulas of the Tate uniformization, e.g. in [Kat1973, Appendix A1.2] if you want to write it down explicitly)

Proof. See [Kat1973, Ch. III] or [Col2005]. In the latter paper, Coleman explicitly computes the isomorphism class of the canonical subgroup as a finite flat R-group scheme, and finds that it can be written using the "$G_{a,b}$" classification of [TO1970] in terms of $v(E_{p-1}(E))$. □

The finite flat R-subgroup scheme $\mathscr{K}_{(E,\alpha_N)} \subset E$ is called the *canonical subgroup* attached to the given test object. It provides a canonical way of lifting the relative Frobenius from characteristic p to characteristic 0, defined by quotienting out by $\mathscr{K}$ (the resulting morphism will be called "Frob" but we will not really need the notation that much). Taking such a quotient is a well-defined operation: see for example [Ray1967].

As we just mentioned above, one reason for studying the canonical subgroup is that the Atkin–Lehner U_p operator acting on forms of level $\Gamma_1(Np^m)$, $m \geq 1$, can be interepreted in a modular fashion so that it involves the quotient by $\mathscr{K}$, and this will allow us to prove that U_p acts compactly on the Banach spaces of overconvergent p-adic modular forms. Another reason is that it provides the reason for why the level at p is distinguished in the theory of p-adic modular forms, as will be explained in Section 2.1.2.

Before discussing the level at p, let us first state the second KEY FACT that underpins the compactness of U_p acting on spaces of overconvergent modular forms, namely the analysis of the fibers of the Frob morphism.

Theorem 2.1.2 (Lubin–Katz)**.** *Let Ω be a an algebraically closed complete field of characteristic zero and residue characteristic p, and let E be an elliptic curve over $\mathcal{O}_\Omega$ such that $v(E_{p-1}(E/\mathcal{O}_\Omega, \omega)) \in I_2 \cap \mathbf{R}_{>0}$ for any[11] choice of $\mathcal{O}_\Omega$-basis ω of $H^0(E, \Omega^1_{E/\mathcal{O}_\Omega})$. Then there are exactly p elliptic curves $E^{(i)}/\mathcal{O}_\Omega$, $i = 1, \ldots, p$ such that there is an $\mathcal{O}_\Omega$-isomorphism $E \cong E^{(i)}/\mathscr{K}_{E^{(i)}}$. In fact, the curve $E^{(i)}$ can be constructed explicitly as the quotient of E by the i-th order-p finite flat $\mathcal{O}_\Omega$-subgroup scheme of E which is NOT $\mathscr{K}_E$. Moreover,*

$$v(E_{p-1}(E^{(i)}, \omega^{(i)})) = \frac{1}{p} v(E_{p-1}(E, \omega)),$$

where $\omega^{(i)}$ is an $\mathcal{O}_\Omega$-basis element of $H^0(E^{(i)}, \Omega^1_{E^{(i)}/\mathcal{O}_\Omega})$.

Proof. See [Kat1973, Theorem 3.10.7(5)]. □

We now state the modular definition of the U_p operator (here we follow [Kat1973], but note that via the explicit description of the $E^{(i)}$ in Theorem 2.1.2, this definition agrees with that of [Cal2013]):

Definition 2.1.3. Let R_0 be a complete DVR of residue characteristic p and field of fractions K of characteristic 0. Let $f \in S(R_0, r, N, k)$ with $v(r) \in I_2$. Define $U_p f \in M_k^\dagger(\Gamma_1(N), K, r)$ to be

[11] Obviously the choice of ω does not make a difference if we are just considering the valuation of E_{p-1}.

given on test objects by

$$
\begin{aligned}
(U_p f)(E/R, \alpha_N, Y) &:= \frac{1}{p}\mathrm{Tr}_{\mathrm{Frob}}(f)(E/R, \alpha_N, Y) \\
&:= \frac{1}{p}\sum_{i=1}^{p} f((E \times_{R_0} \mathcal{O}_{\mathbf{C}_K})^{(i)}, \widetilde{\alpha_N}, Y),
\end{aligned}
$$

where $\widetilde{\alpha_N}$ denotes the obvious base-change-then-take-preimage-under-quotient-by-$\mathscr{K}_{E^{(i)}}$ of the level structure α_N.

By evaluation on Tate curves (it is essentially done in [Kat1973, p. 22-23], except in this case we remember that we ignore the canonical subgroup, which is μ_p by Theorem 2.1.1 — so only the terms coming from the "H_i" of [Kat1973, p. 22] appear) shows that Definition 2.1.3 is a good definition: it agrees with the classical definition of Atkin–Lehner, namely

$$\sum_{n\geq 0} a_n q^n \mapsto \sum_{n\geq 0} a_{pn} q^n.$$

Combining Definition 2.1.3 and Theorem 2.1.2, we obtain the following crucial result, which is the fundamental reason for compactness of U_p when it is restricted to overconvergent modular forms:

Corollary 2.1.4. *For $r \in I_2 \cap \mathbf{R}_{>0}$, $U_p : M_k^\dagger(\Gamma_1(N), K, v) \to M_k^\dagger(\Gamma_1(N), K, v)$ actually factors through*

$$M_k^\dagger(\Gamma_1(N), K, v) \to M_k^\dagger(\Gamma_1(N), K, pv) \to M_k^\dagger(\Gamma_1(N), K, v),$$

where the second map is restriction from $X_1(N)(pv)_K$ to $X_1(N)(v)_K$, and the first map is continuous

Proof. The only part that doesn't follow directly is the continuity of the first map. This is not so hard to see from the definitions either: see [Gou1988, Corollary II.3.7]. □

As this is often stated, "U_p improves overconvergence." This is the central phenomenon in the theory of p-adic automorphic forms, and it will always be the fact that leads to compactness of the Atkin–Lehner action on overconvergent automorphic forms, and thus to the entire theory of eigenvarieties as described at the beginning of this chapter. Moreover, the statement and proof of Corollary 2.1.4 essentially go through as stated if N is replaced with Np^m, $m \geq 1$ (see [Gou1988]), but one has to keep track of the level structure at p which could have some nontrivial interaction with Frob — so we have left the analysis of that situation out in order to simplify the exposition and especially to avoid writing down anything false[12].

[12] I have undoubtedly already failed in this endeavor.

For good measure, we formally state the compactness and briefly sketch the proof of how it follows from Corollary 2.1.4 (this is essentially an abridged version of something done in full detail in [Col1997b, §B.3] and [Gou1988]):

Corollary 2.1.5. *For $m \geq 1$ and $v \in I_{\max(2,m)} \cap \mathbf{R}_{>0}$, $U_p : M_k^\dagger(\Gamma_1(Np^m), K, v) \to M_k^\dagger(\Gamma_1(Np^m), K, v)$ is a compact morphism of K-Banach spaces.*

Proof. By using the usual technique of multiplication by weight-k p-deprived Eisenstein series (see [CM1998, §2.2], [Col1997b, §B.1]), one has a non-Hecke-equivariant isomorphism

$$M_k^\dagger(\Gamma_1(Np^m), K, v) \cong M_0^\dagger(\Gamma_1(Np^m), K, v).$$

The good thing is that this isomorphism does respect the restriction maps, so this comes down to the compactness of the restriction map from the affinoid algebra of $X_1(Np^m)(pv)$ to the affinoid algebra of $X_1(Np^m)(v)$. This is [Gou1988, Corollary I.2.9]. □

The proof of [Gou1988, Corollary I.2.9] essentially proceeds by choosing an explicit basis. In the representation-theoretic context this will also be the fastest way of proving compactness of the Atkin–Lehner action, and it will be much easier in practice because the affinoids $X_1(Np^m)(v)$ will be replaced with much more concrete spaces. We also remark here that [Col1997b, A5.2] provides a fairly general way of proving that a restriction map on affinoid algebras is compact, though checking the hypotheses in the case of the modular curve will ultimately require the same kind of considerations as in the proof of [Gou1988, Corollary I.2.9].

By studying how the isomorphism $M_\kappa^\dagger(\Gamma_1(Np^m), K, v) \cong M_0^\dagger(\Gamma_1(Np^m), K, v)$ constructed via the restricted Eisenstein family E_κ (for general possibly non-classical weights κ) intertwines U_p, Coleman [Col1997b, §B.3] proved that the characteristic series of U_p is itself (coefficient-by-coefficient) an analytic function of the weight, and constructed his p-adic families of modular forms.

2.1.2 | The level at p à la Katz and Coleman

One might find it suspicious that in Definition 2.1.3, the U_p operator may be defined on modular forms of level $\Gamma_1(N)$, whereas in the classical Atkin–Lehner theory the level will jump up to $\Gamma_1(Np)$ in general. The point here is that the level at p (the p being the same as in the "p-adic" of the rings that everything is defined over) is very special, essentially due to the theory of the canonical subgroup. For example, Serre [Ser1973, §3] observed that a classical form of level $\Gamma_0(p)$ was in fact p-adically a form for $\mathrm{SL}_2(\mathbf{Z})$. Using the modular perspective of Katz and the theory of the canonical subgroup, this is easy to generalize[13] as follows.

[13] Serre's proof is essentially an obfuscated way of rewriting the modular proof while writing everything in terms of q-expansions.

Lemma 2.1.6. *Let $m \geq 1$. For any $v \in I_m$, there is a canonical injection of K-Banach spaces*

$$M_k(\Gamma_1(N) \cap \Gamma_0(p^m), K) \to M_k^\dagger(\Gamma_1(N), K, v).$$

Proof. Interpreting the classical modular forms on the left hand side as Katz modular forms, simply define the image of the classical form f to be

$$(E, \alpha_N, Y) \mapsto f(E, \alpha_N, \mathscr{K}^m_{(E,\alpha_N)}, Y),$$

where $\mathscr{K}^m_{(E,\alpha_N)}$ denotes the order-p^m canonical subgroup of E, which exists by inductive application of Theorem 2.1.1 since $v \in I_m$. □

In fact, using the same kind of technique with the canonical subgroup, one can further generalize Lemma 2.1.6 to conveniently understand how p-adic modular forms of level $\Gamma_1(Np^m)$ are tied together for all m.

Theorem 2.1.7. *For $v \in I_m$, The overconvergent locus $X_1(Np)(v)$ is actually the quotient of $X_1(Np^m)(v)$ by the action of the diamond operators in $(1 + p\mathbf{Z}_p)/(1 + p^m\mathbf{Z}_p)$.*

Proof. Recall that $X_1(Np^m)(v)$ is defined to be the connected component of the locus of $X_1(Np^m)$ cut out by $v(E_{p-1}) \leq v$. It turns out [KM1985] that this connected component may be described explicitly in a way that depends only on the relationship between the level structure at p (see the definitions in [Col1997b, §B.2], [CM1998, §2.1]) and the canonical subgroups of various p-power orders. In short, the fiber over a point in $X_1(Np)(v)$ consists of level structure on the same elliptic curve that only differ at p^m, but whose images are all equal to the canonical subgroup of order p^m and whose images on μ_p are all equal to the canonical subgroup of order p (and which of course all agree on μ_p). The elements of the fiber are therefore acted on transitively by diamonds that permute μ_{p^m} but act trivially on μ_p, which justifies the claim. □

The point of Theorem 2.1.7 is that in the theory of p-adic modular forms, even if we plan on studying the forms of level $\Gamma_1(Np^m)$ for all $m \geq 1$, we can get away with only looking at the Banach spaces $M_k^\dagger(\Gamma_1(Np), K, v)$. More specifically, fix some $m \geq 1$, and a Nebentypus character $\chi : (\mathbf{Z}/p^m\mathbf{Z})^\times$ of conductor m. Multiplication by Eisenstein series $E_{(k,\chi)}$ provides an identification of K-Banach spaces

$$M_0^\dagger(\Gamma_1(Np), K, v) \cong M_k^\dagger(\Gamma_1(Np^m), K, v; \chi\tau^{-k}),$$

at least if χ is defined over K (so that the Eisenstein series we need is actually defined on the locus we wrote down) and $v \in I_m$ (here τ is the p-adic Teichmüller character). This is why in the theory of overconvergent modular forms, we simply consider a single fixed tame level N, and allow the level at p to be arbitrarily deep, keeping track of that level by simply keeping track of the Nebentypus character χ. The Nebentypus character becomes part of the data of the weight, which we discuss next.

2.1.3 | Weights and weight space

As early as Serre [Ser1973], it was observed that for the purpose of p-adic interpolation of modular forms, it was best to think of the weight and the Nebentypus at p as being part of the same data, which the families of modular forms live over. This makes some sense as soon as we consider the weight and Nebentypus of the restrcted Eisenstein family, which was done in [Ser1973] (see also [CM1998, §2.2]). In general, given a character

$$\kappa : \mathbf{Z}_p^\times \to \mathbf{C}_p^\times,$$

we can form the Eisenstein series E_κ which is an overconvergent modular form of weight-character κ, and multiplication by E_κ provides us with the definition of the space $M_\kappa^\dagger(N, \mathbf{C}_p, v)$ of v-overconvergent p-adic modular forms of tame level N (see [CM1998, §2.4] for the full detail of how to define it). We remark that although this was the original definition and was sufficient for the construction and proof of basic properties of the eigencurve, Pilloni [Pil2013] and Andreatta–Iovita–Stevens [AIS2014] have constructed sheaves that the elements of $M_\kappa^\dagger(N, \mathbf{C}_p, v)$ can be viewed as sections of. These constructions are very useful: for example Ye [Ye2020] used them to provide a new[14] proof of the conjecture of Coleman–Mazur to the effect that the eigencurve satisfies the valuative criterion for properness.

There is a natural rigid space $\mathcal{W}$ over $\mathbf{Q}_p$ that parametrizes p-adic families of continuous homomorphisms $\kappa : \mathbf{Z}_p^\times \to \mathbf{C}_p^\times$. It is the same as what [Buz2004] calls $\mathrm{Hom}(\mathbf{Z}_p^\times, \mathbf{G}_m)$. The structure of $\mathcal{W}$ is fairly easy to understand: it is just a disjoint union of $p-1$ open discs. The reason is that (still taking $p > 2$ for convenience)

$$\mathbf{Z}_p^\times \cong \mu_{p-1} \times (1 + p\mathbf{Z}_p),$$

and $1 + p\mathbf{Z}_p \cong \mathbf{Z}_p$ via the p-adic logarithm and exponential, so the $p-1$ discs are indexed by choice of image of a primitive $(p-1)$-th root of unity, and the discs themselves parametrize the choice of where to send a topological generator of $1 + p\mathbf{Z}_p$. This choice of topological generator is usually taken to be $\exp(p)$ or $1 + p$. It must satisfy $\kappa(1+p)^{p^m} \to 1$ as $m \to \infty$, which is why the possible choices of $\kappa(1+p)$ are parametrized by an open unit disc around 1.

Remark 2.1.8. When $p = 2$ it doesn't get that much harder: the radius on which log and exp produce an isomorphism gets smaller, so we have to replace 2 by 4 in a bunch of places: $\mathbf{Z}_2^\times \cong (\mathbf{Z}/4\mathbf{Z})^\times \times (1 + 4\mathbf{Z}_2)$, and the distance from the center of a given disc making up $\mathcal{W}$ is given by $|\kappa(5) - 1|$. We make this remark because the ping-pong arguments of Newton and

[14] The original proof having been given by Diao and Liu [DL2016] using (φ, Γ)-module techniques.

Thorne take place on the 2-adic eigencurve so it is necessary to understand completely what weight space is.

The classical weights (k, χ), $k \geq 2$, $\chi : (\mathbf{Z}/p^m\mathbf{Z})^\times \to \mathbf{C}^\times \cong \overline{\mathbf{Q}}_p^\times$ consisting of a weight and Nebentypus character can be thought of as elements of $\mathcal{W}(\mathbf{C}_p) = \mathrm{Hom}(\mathbf{Z}_p^\times, \mathbf{C}_p^\times)$, namely

$$z \mapsto z^{k-2}\chi(z).$$

This is how one forms the classical p-deprived Eisenstein series $E_{(k,\chi)}$ (though it is important to heed the warning that some sources replace "$k-2$" with "k"). The classical weights (k, m) (closed points of $\mathcal{W}$ corresponding to the Galois orbit of $\mathbf{C}_p$-points of the form (k, χ) where χ is of conductor m) are actually trés Zariski-dense in $\mathcal{W}$, in the sense of [Che2005]. This is not so hard to check: trés Zariski density is just asking that $\{(k, m)\}$ accumulates at itself. Given a particular (k, χ) of conductor m, which corresponds to

$$(1+p)^k \zeta_{p^m(p-1)} - 1 \in \mathcal{W}(\mathbf{C}_p),$$

in an arbitrary open affinoid containing (k, χ), there are plenty of other classical points, namely those of the form (k', χ) with k' congruent to k modulo a high power of p. In any event, the basic lesson here is that Nebentypus characters of deeper conductor at p bring a classical point closer to the center of weight space, and in the absence of a Nebentypus character, weights k p-adically close to zero bring a classical point closer to the center of weight space.

Note that the point of χ being of conductor m is that it is constant on $1 + p^m\mathbf{Z}_p$, and hence locally constant (by multiplicativity). More generally, as all continuous characters $\mathbf{Z}_p^\times \to \mathbf{C}_p^\times$ are locally analytic, we can consider the locus in $\mathcal{W}(\mathbf{C}_p)$ consisting of weight-characters which are analytic on $1 + p^m\mathbf{Z}_p$. It is straightforward (using the standard explicit reasoning for why the continuous characters are locally analytic) to verify that this locus is actually the set of $\mathbf{C}_p$-points of the affinoid open $\mathcal{W}_m \subset \mathcal{W}$ given by the condition that the distance from the center of weight space is less than $p^{\frac{-1}{p^m(p-1)}}$. See for example [Che2010, Lecture 7, Proposition 3.5] for the proof, though Chenevier's convention for what m is is slightly different from ours.

2.1.4 | Classical p-adic automorphic representations

So far in this chapter, we have sketched the basic idea of how p-adic modular forms live in families parametrized by the data of their weight and Nebentypus character at p. This suggests that in general, for the theory of automorphic forms and representations for a particular reductive group to be interpolated p-adically, one should consider the local data at the infinite places at the same time as the local data at the p-adic places. I have nothing intelligent to say about the conceptual reason for this that goes past the following paragraph from [Eme2006b, §3]:

"Common conditions that arise in the p-adic theory of automorphic representations, such as an automorphic representation π being ordi-nary or of non-critical slope, are not local at p; they depend on a comparison of invariants obtained from the local factors at the infinite places (typically, an infinity type) and from the local factor at p (such as a Satake parameter). "

To be more specific (and explain in detail Emerton's claim about things like ordinaryness and non-criticality of slope): if f is a new eigenform of level $\Gamma_1(N)$ and weight $k \geq 2$, then the data of the weight is just the data of the infinity-type $\pi_{f,\infty}$, and the data at p is the data of the Satake parameters, i.e., the data of two unramified characters χ_1, χ_2 of $\mathbf{Q}_p^\times$ such that $\pi_{f,p}$ is the parabolic induction of $\chi_1 \otimes \chi_2$. The data of $a_p(f)$, and hence the data of whether f is ordinary (or indeed any condition depending on the slope of f) depends on both pieces of information: it is

$$p^{\frac{k-1}{2}}(\chi_1(p) + \chi_2(p))$$

(see for example [LW2012, §2.2]).

For this reason, in the general theory of p-adic interpolation of automorphic representations, it is always good idea to "shift the data at ∞ over to the p-adic places." This was perhaps first written down by Buzzard [Buz2004], who did the example of GL_1, where p-adic families of Hecke characters trivial at connected components of places at ∞ could be defined, and a Galois-theoretic criterion for classicality result could be deduced from the relationship between algebraic Hecke characters and Hodge–Tate Galois characters [Ser1989].

We now describe Emerton's general formalism for how to modify an automorphic representation or automorphic by "transferring the data at ∞ to p" to make it amenable to p-adic interpolation. This is slightly different from the conventions of Loeffler [Loe2011], Buzzard [Buz2004], Chenevier [Che2004], and Bellaïche–Chenevier [BC2009a], but it is not hard to go between the two notions.

The basic concept underlying the definition of a classical p-adic automorphic form is to consider automorphic forms valued in some finite-dimensional (locally) algebraic representation, which allows you to isolate the data at infinity and make it "p-adic" via a choice of isomorphism $\iota : \overline{\mathbf{Q}}_p \cong \mathbf{C}$.

Let us now be more specific. Let G be a reductive algebraic group over a number field $F/\mathbf{Q}$ that is totally split at p, satisfying the conditions that $G(F_\infty)$ is compact (or at least compact modulo center) and that there exists an $n \geq 1$ such that $G \times_F F_v \cong \mathrm{GL}_{n/F_v} = \mathrm{GL}_{n/\mathbf{Q}_p}$ for all $v|p$. In fact, we also assume that $G(F_\infty)$ is connected, but this is not really necessary (it just allows us to not think about the group "π_0" in [Eme2006b]). The first hypothesis ensures that the shimura variety associated to G is (or, in the case of "compact modulo center", can be assumed to be) zero-dimensional — this is very convenient because it means that automorphic forms can be viewed as simply being tuples of elements in the module they are valued in (indeed, this is a main reason why the explicit computations of [Buz2004, LWX2017, WXZ2017, Ye2019] can be done). It is also convenient in analyzing the space of automorphic forms for a few reasons: it means that everything is cuspidal, so there is no continuous spectrum; it means that there is no need for the theory of $(\mathfrak{g}, K)$-modules, as the $G(F_\infty)$-action clearly preserves K-finiteness (so the local archimedean component can be thought of as a bona fide representation of $G(F_\infty)$), and all the automorphic forms are automatically square-integrable. The second hypothesis can be weakened considerably (see [Loe2011]), but it is all we will need here. For example, G could be the rank-n definite unitary group attached to a CM extension of $\mathbf{Q}$ that is totally split at p.

We now follow Bellaïche [Bel2021][15] in explaining how to isolate a particular ∞-type. Let $\mathcal{A}(G)$ be the space of automorphic forms on G. Thanks to the hypothesis of compactness at infinity, there are none of the usual subtleties: we have, as a representation of $G(\mathbf{A}_F)$,

$$\mathcal{A}(G) = \bigoplus_{\pi} m(\pi)\pi, \tag{2.1}$$

where π ranges over all the automorphic representations of G. It is a theorem due to Harish-Chandra [HC1968] that the numbers $m(\pi)$ are finite. The fact that $\mathcal{A}(G)$ decomposes is a consequence of the fact that $\mathcal{A}_{\text{cusp}}(G) = \mathcal{A}(G)$ (thanks to the compactness assumption) and the theorem of Gelfand–Graev–Piatetski-Shapiro [GfGPS1990] to the effect that $\mathcal{A}_{\text{cusp}}(G) \subset \mathcal{A}_{\text{disc}}(G)$. The automorphic representation π decomposes (by Flath's tensor product theorem) as

$$\pi \cong \bigotimes_{v}{}' \pi_v,$$

and our goal is to isolate any particular choice of $\pi_\infty = \otimes_{v|\infty}\pi_v$, which is always just an irreducible (finite-dimensional by Peter–Weyl theorem) representation of the compact Lie group $G(F_\infty)$.

It follows from Schur's lemma (as for now the coefficient field is $\mathbf{C}$) and the decomposition in (2.1) that

$$\bigoplus_{\substack{\pi \\ \pi_\infty \cong W}} m(\pi)\pi = (\mathcal{A}(G) \otimes W^\vee)^{G(F_\infty)} \otimes W. \tag{2.2}$$

For π with archimedean component W, the $G(\mathbf{A}_F)$-module $\pi \otimes W^\vee)^{G(F_\infty)}$ is the same as π except with trivial archimedean component.

It is immediate from the definition that $(\mathcal{A}(G) \otimes W^\vee)^{G(F_\infty)} \cong \mathcal{A}(G, W^\vee)$ as $G(\mathbf{A})$-modules, where $\mathcal{A}(G, W^\vee)$ is as follows:

Definition 2.1.9 (Modular forms valued in $G(F_\infty)$-module)**.** Let W be a representation of the

[15] WARNING: this material cannot be found in the published version of Bellaïche's eigenbook. Here we are referring to the draft on his website.

Lie group $G(F_\infty)$ as above. Then an *algebraic automorphic form valued in W* is a function

$$f : G(F)\backslash G(\mathbf{A}_F) \to W$$

which is K-finite for some (any) compact open $K \subset G(\mathbf{A}_F^\infty)$ and such that $f(gh) = h \cdot f(g)$ for all $h \in G(F_\infty)$, where $G(F_\infty) \subset G(\mathbf{A}_F)$ in the usual way.

Thanks to the condition that elements of $\mathcal{A}(G, W^\vee)$ are $G(F_\infty)$-equivariant on the right, restricting to $G(\mathbf{A}_F^\infty)$ does not lose any information, and provides the isomorphism of the following lemma.

Lemma 2.1.10. *$\mathcal{A}(G, W)$ is isomorphic as a $G(\mathbf{A}_F^\infty)$-module to the space of functions $f : G(\mathbf{A}_F^\infty) \to W$ such that for all $\gamma \in F$ and $g \in \mathbf{A}_F^\infty$,*

$$f(\gamma g) = \gamma_\infty^{-1} \cdot f(g),$$

and f is K-finite for some open compact subgroup $K \subset \mathbf{A}_F^\infty$.

In any event, (2.1) and (2.2) imply

Theorem 2.1.11. *As representations of $G(\mathbf{A}_F) = G(\mathbf{A}_F^\infty) \times G(F_\infty)$,*

$$\mathcal{A}(G) \cong \bigoplus_W \bigoplus_{\substack{\pi \\ \pi_\infty \cong W}} m(\pi)\pi \cong \bigoplus_W \mathcal{A}(G, W^\vee) \otimes W,$$

so understanding the spaces of automorphic forms described in Lemma 2.1.10 for each irreducible representation W of $G(F_\infty)$ is the same as understanding $\mathcal{A}(G)$.

Now it is time to make the definition p-adic, following [Eme2006b]. Assume further that π_∞ comes from an irreducible representation of $G \times_F \mathbf{C}$ via the natural map $G(F_\infty) \to G(\mathbf{C})$ (this is what the "allowable" hypothesis of [Eme2006b, Definition 3.1.3] boils down to in our case). The point of asking for this is that W is then uniquely determined by the infinitesimal character of π_∞ (for example by highest-weight theory of representations of reductive groups over algebraically closed field), but more importantly because then via an isomorphism $\iota : \overline{\mathbf{Q}}_p \cong \mathbf{C}$, the representation W may be viewed as an algebraic representation $\iota^{-1}W$ of $G \times_F \overline{\mathbf{Q}}_p \cong \mathrm{GL}_{n/\overline{\mathbf{Q}}_p}$ with coefficients in $\overline{\mathbf{Q}}_p$. By restriction to $F_\mathfrak{p} = \mathbf{Q}_p$-points for some $\mathfrak{p}|p$, we can (and from now on always will) regard $\iota^{-1}W$ as a finite-dimensional algebraic representation of the locally $\mathbf{Q}_p$-analytic group $\mathrm{GL}_n(\mathbf{Q}_p)$. Let $E/\mathbf{Q}_p$ be a finite extension so that $\iota^{-1}W$ and $\iota^{-1}\pi^\infty$ can be defined over E. There is no harm in considering both of these representations as being defined over E rather than over $\mathbf{C} \cong \overline{\mathbf{Q}}_p$: the descent to E is unique, by [Eme2006b, 3.1.4], which itself is a standard application of the basic theory of Hecke algebras (see e.g. [JL1970]). Having made this switch, we make the obvious change to the perspective on module-valued automorphic forms of Lemma 2.1.10:

Definition 2.1.12. Let V be a finite-dimensional algebraic representation of $\mathrm{GL}_n(\mathbf{Q}_p)$ with coefficients in a finite extension $E/\mathbf{Q}_p$ and fix a prime $\mathfrak{p}|p$. Then define the space $\mathcal{A}(G, V)$ of *algebraic automorphic forms valued in* V to be the set of

$$f : G(\mathbf{A}_F^\infty) \to V$$

such that for all $\gamma \in G(F)$ and $g \in G(\mathbf{A}_F^\infty)$,

$$f(\gamma g) = \gamma_\mathfrak{p} f(g)$$

and such that f is K-invariant for some compact open $K \subset G(\mathbf{A}_F^\infty)$.

Thanks to the assumption that $G(F_\infty)$ is connected, our Definition 2.1.12 is clearly equivalent to Emerton's [Eme2006b, Definition 3.2.1] (note the "locally constant" hypothesis in Emerton's definition). It is equivalent to the definitions of [Buz2004, Loe2011, Ye2019] by taking f to $g \mapsto g_\mathfrak{p}^{-1} \cdot f(g)$.

Putting all of this together, the whole point is the following theorem, to the effect that (despite the shifting of the action from ∞ to $\mathfrak{p}$) the space $\mathcal{A}(G, \iota^{-1}W)$ is just a p-adic model of $\mathcal{A}(G, W)$.

Theorem 2.1.13. *Let W be a finite-dimensional irreducible complex representation of $G(F_\infty)$ that factors through $G(F_\infty) \to G(\mathbf{C})$. Then there is a $G(\mathbf{A}_F^\infty)$-equivariant isomorphism*

$$\mathcal{A}(G, W) \cong \mathcal{A}(G, \iota^{-1}W) \otimes_{E,\iota} \mathbf{C},$$

where as above $E/\mathbf{Q}_p$ is a finite extension chosen so that $\iota^{-1}W$ is defined over E.

Proof. This is a straightforward generalization of the argument in [DT1994, p. 443]. The isomorphism is defined by taking $f \in \mathcal{A}(G, W)$ to the element of $\mathcal{A}(G, \iota^{-1}W) \otimes \mathbf{C}$ given by

$$g \mapsto \iota(g_\mathfrak{p}^{-1}) \cdot f(g) \in W \cong (\iota^{-1}W) \otimes_{E,\iota} \mathbf{C}$$

where the "g" in $f(g)$ is thought of as being in $\mathbf{A}_F \supset \mathbf{A}_F^\infty$, and ι is abuse of notation for the the isomorphism $G(\overline{\mathbf{Q}}_p) \cong \mathrm{GL}_n(\overline{\mathbf{Q}}_p) \cong \mathrm{GL}_n(\mathbf{C}) \cong G(\mathbf{C})$ induced by the actual isomorphism $\iota : \overline{\mathbf{Q}}_p \to \mathbf{C}$ that was chosen ahead of time. In particular, we use the fact that for $\gamma \in G(F)$, γ_∞ and $\iota(\gamma_\mathfrak{p})$ have the same image in $G(\mathbf{C})$, since $E \supset F_\mathfrak{p} \supset F$. □

From the representation-theoretic point of view on $\mathcal{A}(G)$, Definition 2.1.16 provides the following p-adic model of Theorem 2.1.11:

Corollary 2.1.14. *For an allowable algebraic representation W of G, let*

$$\mathcal{A}(G)_W := \bigoplus_{\substack{\pi \\ \pi_\infty \cong W}} m(\pi)\pi \cong \mathcal{A}(G, W^\vee) \otimes W.$$

Then
$$\mathcal{A}(G)_W \cong \big(\mathcal{A}(G, \iota^{-1}W^\vee) \otimes_E \iota^{-1}W\big) \otimes_{E,\iota} \mathbf{C}.$$

Proof. This is a direct consequence of the isomorphism of Definition 2.1.16 together with the usual facts from commutative algebra:

$$\begin{aligned}(\mathcal{A}(G, \iota^{-1}W^\vee) \otimes_{E,\iota} \mathbf{C}) \otimes_{\mathbf{C}} W &\cong (\mathcal{A}(G, \iota^{-1}W^\vee) \otimes_{E,\iota} \mathbf{C}) \otimes_{\mathbf{C}} (\mathbf{C} \otimes_{E,\iota} \iota^{-1}W) \\ &\cong (\mathcal{A}(G, \iota^{-1}W^\vee) \otimes_E \iota^{-1}W) \otimes_{E,\iota} \mathbf{C},\end{aligned}$$

which concludes. □

Finally, from the perspective of representation theory, again by [Eme2006b, Lemma 3.1.4] applied to the various π_f's thought of as being defined over E (as this is part of the hypothesis on E), Corollary 2.1.14, and (2.2), we have

Corollary 2.1.15. *For W as above, we have*

$$\mathcal{A}(G, \iota^{-1}W^\vee) \otimes \iota^{-1}W \cong \bigoplus_{\substack{\pi \\ \pi_\infty \cong W}} \left(\sideset{}{'}\bigotimes_{\substack{v<\infty \\ v \neq \mathfrak{p}}} \pi_v \right) \otimes \pi_{\mathfrak{p}} \otimes_E \iota^{-1}W.$$

as representations of $G(\mathbf{A}_F^\infty)$

This was the full detail of the motivation for Emerton's [Eme2006b, Definition 3.1.15] of classical $\mathfrak{p}$-adic automorphic representations, which we now repeat for convenience:

Definition 2.1.16. Let π be an automorphic representation of G such that π_∞ is allowable as above. Then the *classical $\mathfrak{p}$-adic automorphic representation associated to π* is the $G(\mathbf{A}_f)$-module

$$\widetilde{\pi} := \left(\sideset{}{'}\bigotimes_{\substack{v<\infty \\ v \neq \mathfrak{p}}} \pi_v \right) \otimes_E \pi_{\mathfrak{p}} \otimes_E \iota^{-1}W,$$

where $\iota^{-1}W$ is as usual considered as an algebraic E-representation of $\mathrm{GL}_n(\mathbf{Q}_p) \cong G(F_{\mathfrak{p}})$.

Remark 2.1.17. Of course, $\widetilde{\pi}$ is not in general an honest automorphic representation: the local component at $\mathfrak{p}$, namely $\pi_{\mathfrak{p}} \otimes_E \iota^{-1}W$, is not even necessarily a smooth representation of $G(F_{\mathfrak{p}})$ in general.

2.1.5 | Overconvergent p-adic automorphic forms

In the previous Section 2.1.4, we explained how to construct $\mathfrak{p}$-adic versions of the classical automorphic forms and representations for G, and showed that (as Emerton explains is desirable) they are the same as the usual automorphic representations except the data at infinity is transfered to $\mathfrak{p}$. This section is about the general technique of how to interpolate the resulting systems of Hecke eigenvalues in p-adic families. There are at least two basic approaches:

1. The perspective of Buzzard [Buz2004, Buz2007], Chenevier [Che2004], and Loeffler [Loe2011]: you copy the definition (Definition 2.1.12) of $\mathcal{A}(G, V)$ where V is an algebraic reprentation of $G(F_{\mathfrak{p}}) \cong \mathrm{GL}_n(\mathbf{Q}_p)$, except you allow V to be a locally analytic representation instead. The radius of convergence of this analytic action is what tells you how overconvergent an automorphic form is.

2. The perspective of Emerton [Eme2006b], in which an "Eichler–Shimura"-type result is used: one views the systems of eigenvalues of the classical forms as being inside of the cohomology of some Shimura variety, and then finds a way to p-adically complete this cohomology to pick up the systems of eigenvalues that come from the overconvergent automorphic forms. In Emerton's perspective, many questions can be neatly resolved by simply applying the general theory of locally analytic representations of locally p-adic analytic groups to this "completed cohomology", though Emerton had to expend some effort [Eme2017, Eme2006a] to sufficiently complete that general theory.

It was shown in Loeffler's thesis (see [Loe2011, §3]) that in fact the two viewpoints essentially pick up the same data (except for usually the first approach requires asking for some Iwahori-fixed vectors and therefore only allows one to construct some central zone of the eigenvariety[16].

Let us first begin by laying the groundwork of Emerton's approach under the same hypotheses on G as in Section 2.1.4. In this situation, the Shimura varieties associated to G are 0-dimensional. Indeed, by compactness of $G(\mathbf{R})$, for all choice of compact open level structure $K \subset G(\mathbf{A}_F)$,

$$|Y(G, K)| = |G(F)\backslash G(\mathbf{A}_F)/K| < \infty,$$

where $Y(G, K)$ is the level-K Shimura variety for G, namely the double coset space $G(F)\backslash G(\mathbf{A}_F)/K$.

By the same argument as in [Buz2004, p. 10ff], if the level $K \subset G(\mathbf{A}_F^\infty)$ is small enough, then

$$g_i^{-1} G(F) g_i \cap K = 1$$

[16]Chenevier has told me that this defect of the Buzzard–Chenevier–Loeffler approach is not consequential. I have not worked out the details, but it makes sense: one can just take the union over deeper and deeper Iwahori subgroups, and eventually pick up any information about the global structure of the eigenvariety that one is interested in by doing what amounts essentially to looking at everything over increasingly wide subsets of weight space. Presumably one can even construct the full eigenvariety by taking an increasing union of these. Moreover, in the next section we will follow [Ye2019] in proving that an analog of Lemma 2.1.6 (that is, the appropriate generalization of [Buz2004, Lemma 4(4)]) holds, and hence for many purposes, especially things that just rely on understanding slopes, it really is enough to just understand things for a single choice of Iwahori).

for all $g_i = g_1, \ldots, g_n$ representatives in $G(\mathbf{A}_F)$ of $Y(G, K)$. Therefore, for any V, $\mathcal{A}(G, V)^K$ (easy to view as the space of V-valued p-adic automorphic forms) is canonically isomorphic to $V^{\otimes n}$. This is convenient for computations and sometimes even in conceptual arguments.

To have a version of Eichler–Shimura, we need to have some local system on the Shimura variety $Y(G, K)$:

Definition 2.1.18 (Definition 2.2.3 of [Eme2006b])**.** Let $K = K_\mathfrak{p} K^\mathfrak{p} \subset G(\mathbf{A}_F^\infty)$ be a compact open subgroup, and M a $\mathbf{Q}_p$-representation of $K_\mathfrak{p} \subset G(F_\mathfrak{p}) \cong \mathrm{GL}_n(\mathbf{Q}_p)$. Define the $\mathbf{Q}_p$-local system[17] on $Y(G, K)$

$$\mathcal{V}_M^{(K)} := (M \times (G(F)\backslash G(\mathbf{A}_F^\infty)))/K.$$

Moreover, for any tame level $K^\mathfrak{p}$, define

$$H^0(K^\mathfrak{p}, \mathcal{V}_M) := \varinjlim_{K_\mathfrak{p}} H^0(Y(G, K_\mathfrak{p} K^\mathfrak{p}), \mathcal{V}_M^{(K_\mathfrak{p} K^\mathfrak{p})}).$$

The point of Definition 2.1.18 is the following Lemma 2.1.19, which replaces the Eichler–Shimura isomorphism (see for example [Con, Shi1994]) in our situation in which $Y(G, K)$ is zero-dimensional (of course it is much easier than the Eichler–Shimura isomorphism).

Lemma 2.1.19 (Proposition 3.2.2 of [Eme2006b])**.** *For all finite-dimensional $\mathbf{Q}_p$-representations W of $G(F_\mathfrak{p}) \cong \mathrm{GL}_n(\mathbf{Q}_p)$, there is a $G(\mathbf{A}_F^\infty)$-equivariant isomorphism*

$$\mathcal{A}(G, W)^{K^\mathfrak{p}} \cong H^0(K^\mathfrak{p}, \mathcal{V}_W).$$

Proof. We follow the proof of [Eme2006b, Proposition 3.2.2] (which is the same as this except slightly more general because Emerton does not assume that $G(F_\infty)$ is connected; also Emerton does not appear to be consistent in his definition of $\mathcal{V}_W$ so if we want to be consistent with Definition 2.1.18, we must also incorporate the transformation of the proof of [Eme2006b, 2.2.4]). The isomorphism is given by sending $f \in \mathcal{A}(G, W)^{K^\mathfrak{p} K_\mathfrak{p}}$ to

$$g \mapsto (g_\mathfrak{p}^{-1} f(g), g),$$

which defines a bona fide element of $H^0(Y(K^\mathfrak{p} K_\mathfrak{p}), \mathcal{V}_W)$ because for $\gamma \in G(F)$, $g \in G(\mathbf{A}_F^\infty)$,

$$(g_\mathfrak{p}^{-1} \gamma_\mathfrak{p}^{-1} \cdot f(\gamma g), \gamma g) \sim (\gamma_\mathfrak{p}^{-1} \cdot f(g), g).$$

It is then straightforward to verify the desired properties. □

As mentioned at the beginning of this section, Emerton's approach requires us to p-adically complete this thing. Doing so requires choosing a $\mathbf{Z}_p$-lattice in W and taking the induced

[17] In the full generality, Emerton has the coefficients in E where V is defined over E. But in our situation, V is a representation of $\mathrm{GL}_n(\mathbf{Q}_p)$ valued in $\overline{\mathbf{Q}}_p$, and is hence automatically defined over $E = \mathbf{Q}_p$ as $\mathrm{GL}_n(\mathbf{Q}_p)$ is split.

$\mathbf{Z}_p$-local system on the various $Y(G, K_\mathfrak{p}K^\mathfrak{p})$, as in [Eme2006b, Definition 2.2.9]. In reality, thanks essentially to Corollary 2.1.14, we expect to only really need to consider the case where $W = \mathbf{Q}_p$ is the trivial representation of G. In this case, there is an obvious choice of lattice, $\mathbf{Z}_p$, to complete with respet to. More generally, it is still convenient to consider an arbitrary finite extension $E/\mathbf{Q}_p$ and the lattice $\mathcal{O}_E$ (so that we will then have the space of p-adic automorphic forms over E and be able to define the eigenvariety over E).

Definition 2.1.20. For $n \geq 0$ (though only $n = 0$ is relevant to our situation[18]), define the n-th *completed cohomology* associated to the data of G and a tame level $K^\mathfrak{p} \subset G(\mathbf{A}_F^{\infty,\mathfrak{p}})$ and a $E[G(F_\mathfrak{p})]$-module W to be

$$\widetilde{H}^n(K^\mathfrak{p}, \mathcal{V}_W) := E \otimes_{\mathcal{O}_E} \varprojlim_{s} \varinjlim_{K_\mathfrak{p}} H^n(Y(K_\mathfrak{p}, K^\mathfrak{p}), \mathcal{V}_{W_0}/p^s),$$

where W_0 is an arbitrary choice of $\mathcal{O}_E$-lattice in W, and $K_\mathfrak{p}$ runs over the compact opens of $G(F_\mathfrak{p}) \cong \mathrm{GL}_n(\mathbf{Q}_p)$ that stabilize W_0, so that the $\mathcal{O}_E$-local system $\mathcal{V}_{W_0}$ can be defined in the exact same way as Definition 2.1.18. Emerton [Eme2006b, Lemma 2.2.8] shows that this is well-defined in the sense that it does not depend on the choice of (separated) lattice W_0. The completed cohomology $\widetilde{H}^n(K^\mathfrak{p}, \mathcal{V}_W)$ is an E-Banach space that comes with the structure of a continuous representation of $G(F_\mathfrak{p})$.

The only completed cohomology that we will actually need is

$$\widetilde{H}^0(K^\mathfrak{p}, E),$$

which Emerton [Eme2006b, Definition 3.2.3] calls the "space of $\mathfrak{p}$-adic automorphic forms of tame level $K^\mathfrak{p}$." This name makes sense, for example by [Loe2011, Proposition 3.10.1], but also[19] because $\widetilde{H}^0(K^\mathfrak{p}, E)$ is precisely the set of continuous functions

$$G(F)\backslash G(\mathbf{A})/K^\mathfrak{p} \to E,$$

and the $F_\mathfrak{p}$-analytic vectors of $\widetilde{H}^0(K^\mathfrak{p}, E)$ (considered as usual as a continuous representation of the locally $\mathbf{Q}_p = F_\mathfrak{p}$-analytic group $\mathrm{GL}_n(\mathbf{Q}_p) \cong G(F_\mathfrak{p})$) consist of all such functions which are locally analytic on $G(F_\mathfrak{p})$-cosets.

The key point is that the space $\widetilde{H}^0(K^\mathfrak{p}, E)$ p-adically interpolates the classical $\mathfrak{p}$-adic automorphic representations of G (those of Definition 2.1.16), for the exact same structural reason

[18] We remark, however, that the Coleman–Mazur eigencurve can also be recovered by setting $G = \mathrm{GL}_2$ (which does NOT satisfy the hypotheses we put on G), setting $n = 1$, and using Eichler–Shimura, as is done in [Eme2006b, §4].

[19] This key fact is the first thing Emerton mentions after the definition, and the first thing that Professor Chenevier told me in his office when I told him the definition, but it is NOT mentioned in the discussion of [NT2021] on completed cohomology.

as Corollary 2.1.15 and friends:

Proposition 2.1.21 (Proposition 3.2.4 of [Eme2006b]). *The locally algebraic irreducible closed $G(\mathbf{A}_F^\infty)$-subrepresentations of $\widetilde{H}^0(K^\mathfrak{p}, E)$ are precisely the classical $\mathfrak{p}$-adic classical automorphic representations of $G(\mathbf{A}_F^\infty)$ coming from automorphic representations π with π^∞ definable over E.*

Proof. By definition ([Eme2017, Definition 4.2.1, Proposition-Definition 4.2.6]), a locally algebraic closed subrepresentation of the E-Banach space $\widetilde{H}^0(K^\mathfrak{p}, E)$ must live inside $\widetilde{H}^0(K^\mathfrak{p}, E)_{W\text{-loc. alg.}}$ for some finite-dimensional irreducible algebraic representation W of $G \times_F F_\mathfrak{p}$. The key technical input (which we will not prove) is [Eme2006b, Corollary 2.2.25], which says that the natural map

$$H^0(Y(K^\mathfrak{p}), \mathcal{V}_{W^\vee}) \otimes_E W \to \widetilde{H}^0(K^\mathfrak{p}, E)_{W\text{-loc. alg.}}$$

is an isomorphism. Therefore, the W-locally algebraic vectors provide, by Corollary 2.1.15 and Lemma 2.1.19, exactly the direct sum of the classical $\mathfrak{p}$-adic automorphic representations coming from automorphic representations π with $\pi_\infty \cong W$. □

The Banach space $\widetilde{H}^0(K^\mathfrak{p}, E)$ contains all sorts of vectors which are not locally algebraic, and these will provide the "flesh" of the eigenvariety. It helps (at least psychologically) to know that the overconvergent automorphic forms defined along the lines of Buzzard, Chenevier, and Loeffler can essentially be viewed as living inside the locally analytic vectors of this Banach space, and that the systems of Hecke eigenvalues can be viewed as living inside the locally analytic Jacquet module thereof.

Let $G_0 \subset G(F_\mathfrak{p}) \cong \mathrm{GL}_n(\mathbf{Q}_p)$ be a compact open subgroup which is decomposable in the sense of [Loe2011, Definition 2.2.1]. For example, G_0 could be an Iwahori subgroup

$$G_0 := \begin{pmatrix} 1 + p^e\mathbf{Z}_p & \mathbf{Z}_p & \mathbf{Z}_p & \cdots & \mathbf{Z}_p \\ p^e\mathbf{Z}_p & 1 + p^e\mathbf{Z}_p & \mathbf{Z}_p & \cdots & \mathbf{Z}_p \\ & & \ddots & & \\ p^e\mathbf{Z}_p & p^e\mathbf{Z}_p & \cdots & p^e\mathbf{Z}_p & 1 + p^e\mathbf{Z}_p \end{pmatrix}$$

or any of the subgroups "$\Gamma(\underline{c})$" or "$\Gamma_0(\underline{c})$" of [Ye2019] for group-like $\underline{c}$. Then for $r \geq 0$, let G_r be the subgroup of G_0 given by the image of the exponential map on $\pi^{r-1}\mathfrak{g}_0$, where $\mathfrak{g}_0$ is the $\mathbf{Z}_p$-lattice in the Lie algebra $\mathfrak{g}$ of $\mathrm{GL}_n(\mathbf{Q}_p)$ that exponentiates to G_0 to begin with (more explicitly, you just add r to all the exponents). Similarly to as in [Loe2011] except with M replaced with T (as we have no reason to consider arbitrary parabolics), define $N_r, T_r, \overline{N}_r$ to be defined just like G_r except using the decomposition $\mathfrak{g} = \mathfrak{n} \oplus \mathfrak{t} \oplus \overline{\mathfrak{n}}$ into positive and negative and zero weights with respect to the upper-triangular borel. Obviously this is just taking about the torus, upper-triangular, and lower-triangular unipotent radicals in the Iwahori decomposition of G_r. As usual, we also let $B_r = T_r N_r$ be the upper-triangular Borel of G_r.

Definition 2.1.22 (Definition 3.7.1 of [Loe2011]). For an r-locally analytic weight-character $\chi : T_0 \to E^\times$ (meaning that χ is analytic when restricted to T_r; the typical situation is that $T_0 = (\mathbf{Z}_p^\times)^n$ and $T_r = (1 + p^r\mathbf{Z}_p)^n$ for $r \geq 1$), the *r-overconvergent automorphic forms for G* of tame level $K^\mathfrak{p}$ and level G_0 at $\mathfrak{p}$ and weight-character χ are simply

$$\mathcal{A}(G, (\mathrm{Ind}_B^{G_0}\chi)_{G_k\text{-an}})^{K^\mathfrak{p}G_0},$$

where this is defined slightly differently from Definition 2.1.12: the space $\mathcal{A}(\cdots)$ is here defined to be the set of

$$f : G(F)\backslash G(\mathbf{A}_F) \to (\mathrm{Ind}_B^{G_0}\chi)_{G_k\text{-an}}$$

satisfying $f(gu) = u_\mathfrak{p}^{-1}f(g)$ for all $u \in G_0K^\mathfrak{p}G(F_\infty)$. As mentioned previously, this is equivalent to Definition 2.1.12 via an easy transformation, but it is more convenient because u belonging to the level structure will always have $u_\mathfrak{p} \in G_0$, which is necessary for it to have a well-defined action on any $f(g)$'s which live in a locally analytic representation of G_0.

The point is that when χ is locally algebraic, it should be considered as a classical weight (it will look like $t\chi$ where t stands for $(x_1, \ldots, x_n) \mapsto \prod x_j^{t_j}$ and χ is a locally constant "Nebentypus" character). This is completely analogous to the situation of Section 2.1.3. We finally remark that Definition 2.1.22 is the direct generalization of the definition of [Buz2004] for overconvergent quaternionic modular forms, thanks to [Loe2011, Proposition 2.2.4], which identifies the locally analytic induction here with a space of locally analytic functions (with specified radius of convergence) on $\overline{N_0}$. For $n = 2$ this is just $\mathbf{Z}_p$, which is exactly what is going on in [Buz2004].

We finally state Loeffler's comparison between the Buzzard–Chenevier–Loeffler Definition 2.1.22 and Emerton's completed cohomology Definition 2.1.20.

Proposition 2.1.23 (Proposition 3.10.1 of [Loe2011]). *The space of overconvergent p-adic automorphic forms of level $G_0K^\mathfrak{p}$ (but arbitrary radius of overconvergence r) of weight-character χ is isomorphic to*

$$\left(\widetilde{H}^0(K^\mathfrak{p}, E)_{\mathbf{Q}_p\text{-loc. an.}} \otimes \chi\right)^{B_0}.$$

Moreover, this isomorphism is $G(\mathbf{A}_F^{\mathfrak{p},\infty})$-equivariant.

Proof. The isomorphism is given by taking an automorphic form f in the sense of Definition 2.1.22, that is, a function

$$f : G(\mathbf{A}_F) \to (\mathrm{Ind}_{B_0}^{G_0}\chi)_{\mathbf{Q}_p\text{-loc. an.}} \cong \mathscr{C}^{\text{loc. an.}}(\overline{N_0}, \chi),$$

and sending it to the element of $\widetilde{H}^0(K^\mathfrak{p}) \otimes \chi$ given by

$$g \mapsto f(g)(1).$$

For the verification that this works, see the full detail in [Loe2011, Proposition 3.10.1]. Note that Loeffler's conventions are not quite the same as ours — our Borels are conjugate to his because we want things to conform to the concrete situation of [Buz2004]. □

2.1.6 | Interlude: the level at p à la Buzzard–Chenevier–Loeffler–Ye

In this section, we explain the analog to Lemma 2.1.6 for automorphic forms on G, to the effect that the level at p can be made deeper without changing the space of automorphic forms, at the cost of changing the radius of overconvergence. This is important because it justifies in many situations the ability to compute the slopes of modular forms for a fixed Iwahori level at p in order to deduce a global fact about the spectral variety and therefore the entire eigenvariety. This kind of thing is a key technical observation for justifying the computations of [LWX2017] and [Ye2019] towards the Coleman–Mazur–Buzzard–Kilford conjecture about the structure of the eigenvariety near the boundary of weight space. Though technically orthogonal to the ultimate goal of this mémoire (as the special case of Coleman–Mazur–Buzzard–Kilford that is used is the paper [BK2005] that does it for the eigencurve and not for any group compact at infinity), we still discuss it briefly in order to round out the discussion in full generality of the main phenomena described in [Buz2004]. In particular, this is the direct generalization to higher-rank groups of [Buz2004, Lemma 4(4)].

We will follow [Ye2019, Proposition 4.1.2], where this is done for locally algebraic weights (but observe that it clearly works just as well for locally analytic weights). Ye's proof has the advantage of generality and hindsight over that of Buzzard, and is considerably easier to follow.

For $r \geq 1$, let $\Gamma_0(p^r)$ be the usual compact open subgroup of $\mathrm{GL}_n(\mathbf{Z}_p)$, namely the one congruent modulo p^r to the upper-triangular unipotent radical in $\mathrm{GL}_n(\mathbf{Z}/p^r\mathbf{Z})$.

Proposition 2.1.24 (Proposition 4.1.2 of [Ye2019])**.** *Let c, d, e be positive integers with $d \leq e$ and $c+d-e \geq 1$. Let χ be a $(c+d-e)$-locally analytic character of T_0, and $K^\mathfrak{p}$ a tame level. Then the c-overconvergent automorphic forms on G of weight-character χ and level $K^\mathfrak{p}\Gamma_0(p^d)$ is Hecke-equivariantly isomorphic to the $(c+d-e)$-overconvergent automorphic forms on G of weight-character χ and level $K^\mathfrak{p}\Gamma_0(p^e)$.*

Proof. The c-overconvergent forms of level $K^\mathfrak{p}\Gamma_0(d)$ are the elements of

$$\left(\mathrm{Hom}(G(F)\backslash G(\mathbf{A}_F^\infty), E) \otimes_E \mathrm{Ind}_{B_0}^{\Gamma_0(p)}(\chi)_{c\text{-loc.an.}}\right)^{K^\mathfrak{p}\Gamma_0(p^e)}$$

which are furthermore invariant under some set of representatives $\alpha_1, \ldots, \alpha_n \in \Gamma_0(p^d)$ for $\Gamma_0(p^e)\backslash\Gamma_0(p^d)$. Ye then argues that the restriction map from the $\{\alpha_i\}$-invariant tuples of elements of $\mathscr{C}^{c\text{-loc. an.}}(p\mathbf{Z}_p^{n(n-1)/2}, E)$ indexed by $G(F)\backslash G(\mathbf{A}_F)/K^\mathfrak{p}\Gamma_0(p^e)$ to tuples of elements of $\mathscr{C}^{(c+d-e)\text{-loc. an.}}(p^{e-d}\mathbf{Z}_p^{n(n-1)/2}, E)$ indexed by the same double coset space is actually an isomorphism, by constructing the explicit isomorphism. Of course,

$$\mathscr{C}^{(c+d-e)\text{-loc. an.}}(p^{e-d}\mathbf{Z}_p^{n(n-1)/2}, E) \cong \mathscr{C}^{(c+d-e)\text{-loc. an.}}(p\mathbf{Z}_p^{n(n-1)/2}, E)$$

and (after checking the Hecke-equivariance, which we omit) we see that the target consists of $(c+d-e)$-overconvergent forms of level $\Gamma_0(p^e)$. $\square$

2.1.7 | The Atkin–Lehner algebra

In the above sections, we completely omitted any discussion of the Hecke action at $\mathfrak{p}$. As discussed at the beginning of this chapter, it is the Atkin–Lehner operators that we expect to be compact, and this is the crucial feature that allows for the construction of the "D" eigenvariety using the eigenvariety machine of Buzzard [Buz2007]. As in Section 2.1.1, the reason why they are compact is that they improve overconvergence. In fact, the automorphic definition of what an Atkin–Lehner Hecke operator at $\mathfrak{p}$ is is essentially engineered to make this true. Let T be the diagonal torus of $G(F_\mathfrak{p}) \cong \mathrm{GL}_n(\mathbf{Q}_p)$, and take T_0 as before. In a fairly general context (i.e. without our hypotheses that force $G(F_\mathfrak{p})$), the Atkin–Lehner operators are defined to be some representatives z of T/T_0 which satisfy $|\alpha(z)| > 1$ for all simple roots $\alpha \in X^*(T)$. This is all well and good, but in our context this can be said completely explicitly: T/T_0 has a system of representatives given by the diagonal matrices whose diagonal entries are integer powers of p . The Atkin–Lehner operators are defined as follows

Definition 2.1.25 (Very special case of Definition 2.4.3 of [Loe2011])**.** Define the set of Atkin–Lehner elements of T to be

$$\Sigma^{++} := \left\{ \begin{pmatrix} p^{\alpha_1} & & & \\ & p^{\alpha_2} & & \\ & & \ddots & \\ & & & p^{\alpha_n} \end{pmatrix} : \alpha_i > \alpha_j \text{ for all } i < j \right\}.$$

For $z \in \Sigma^{++}$ and $r \geq 0$, the Atkin–Lehner action of z on the space of r-overconvergent automorphic forms of weight χ on G is defined (by translating via the isomorphism $(\mathrm{Ind}_{B_0}^{G_0}\chi)_{G_r\text{-an}} \cong \mathscr{C}^{r\text{-loc.an.}}(\overline{N_0}, \chi)$) by taking

$$\mathscr{C}^{r\text{-loc.an.}}(\overline{N_0}, \chi) \to \mathscr{C}^{r\text{-loc.an.}}(\overline{N_0}, \chi),$$

$$f \mapsto f \circ (n \mapsto z^{-1}nz).$$

It is straightforward to check that this is well-defined (in fact it would even be well-defined for $z \in \Sigma^{+}$, which is the same as Σ^{++} except we only require $\alpha_i \geq \alpha_j$ for $i \leq j$), as $z^{-1}(\cdot)z$ preserves affinoid polydiscs around 0 in the affinoid whose points are $\overline{N}_0 \cong \mathbf{Z}_p^{n(n-1)/2}$. Since (as discussed in Section 2.1.5) the r-automorphic forms are just finite tuples of elements of these function spaces, we just define the Hecke action componentwise[20]. The point of asking $z \in \Sigma^{++}$ is the following:

Proposition 2.1.26. *For $z \in \Sigma^{++}$, the corresponding Atkin–Lehner action on overconvergent modular forms for G improves overconvergence, and is therefore compact.*

Proof. This is true for the same reason that $z \in \Sigma^{+}$ define well-defined operators on the r-overconvergent forms (i.e. do not worsen overconvergence). In particular, conjugation by z multiplies the various entries of $n \in \overline{N}_0$ by $p^{\alpha_i - \alpha_j}$ for various $i < j$. Since $z \in \Sigma^{++}$, this will in fact always multiply by a positive integer power of p, and therefore will take r-locally analytic functions to $(r+1)$-locally analytic functions. In other words, z improves overconvergence. $\square$

Remark 2.1.27. In November, Chenevier told me that many of these things could be made simpler if one takes the representation-theoretic perspective of locally analytic induction rather than the perspective of locally analytic functions on conjugate unipotent radical. However, we remark here that all the proofs we have provided here regarding the basic phenomena for overconvergent automorphic forms have blatantly disregarded Chenevier's advice: both Proposition 2.1.26 and Proposition 2.1.24, while fairly simple conceptually, use explicit computations with those analytic functions valued only on $\overline{N}_0$. I'm not sure whether this means I have missed a key point; probably not, as the map from one to the other is just restriction to $\overline{N}_0$.

Finally, now that we have defined the Atkin–Lehner action, we add (but do not bother proving, as usual) that all the prime-to-$\mathfrak{p}$-Hecke-equivariant isomorphisms above are also equivariant under the Atkin–Lehner operators.

2.1.8 | Atkin–Lehner theory and refined automorphic representations

In this final section, we go back to the theory of modular forms and explain the relevance of *accessibly refined automorphic representations* to the theory of systems of Hecke eigenvalues.

In Atkin–Lehner theory (see for instance [Bel2021]), the point is that for a new eigenform $f(z) \in S_k(\Gamma_1(N), \mathbf{C})$, the resulting oldforms $f(z), f(pz) \in S_k(\Gamma_1(Np), \mathbf{C})$ have the same system of Hecke eigenvalues for T_ℓ, $(\ell, Np) = 1$, and U_ℓ, $\langle \ell \rangle$ $\ell | N$, and therefore have the same attached p-adic Galois representation (which is just the same as that of f), but they are NOT the same. Therefore, the choice of a level-Np system of eigenvalues for the Hecke algebra

$$\mathbf{C}[\{U_\ell, \langle \ell \rangle\}_{\ell | N} \cup \{T_\ell\}_{(\ell, Np)=1} \cup \{U_p\}]$$

[20] Of course the finiteness is not crucial here: we can just compose any function valued in $\mathscr{C}^{r\text{-loc.an.}}(\overline{N_0}, \chi)$ with the map coming from z

occuring in the modular forms is the same as a choice of Galois representation that occurs there, plus the data of U_p-eigenvalue. By explicitly computing the characteristic polynomial of the matrix of U_p acting on the span of $f(z)$ and $f(pz)$, it is easy to check that this eigenvalue must be one of the two roots of

$$X^2 - a_pX + \epsilon_M(p)p^{k-1},$$

where a_p is the T_p-eigenvalue of f and ϵ_M is the Nebentypus character of f. In general it is not known whether this polynomial always has two distinct roots, but the question can typically be completely avoided by just looking at forms with slope not equal to $(k-1)/2$. The forms of slope equal to $(k-2)/2$ are also typically avoided, because all the newforms of level Np, $m \geq 1$ have slope equal to that (see for example [Ogg1969, Lemma 4(c)]). When these two bad cases, we have the "twin forms" that Gouvêa–Mazur [Maz1997, GM1998] use to produce their "infinite fern" inside Galois deformation spaces: the point is that when two different modular forms induce the same p-adic Galois representation, studying the local geometry of the image of the eigenvariety inside the relevant Galois deformation space (and in particular proving the transversality of the branches coming from the twin forms) can allow one to prove that the modular points are Zariski-dense in the Galois deformation space. See [GM1998, Böc2001, Che2011, Che2013] for more details on this. While we will use the concept of twin forms in the ping-pong, the geometry of the infinite fern itself will not be relevant, so we do not provide any details.

Rather, the point of this section is to clarify the representation-theoretic interpretation of the choice of U_p-eigenvalue in terms of accessible refinements. Given the newform $f \in S_k(\Gamma_1(N), \mathbf{C})$, the attached automorphic representation π_f of $\mathrm{GL}_2(\mathbf{Q})$ has $\pi_{f,p}$ equal to an unramified (irreducible) principal series $\pi(\chi_1, \chi_2)$ for some smooth unramified character $\chi_1 \otimes \chi_2 : T \to \mathbf{C}^\times$ [LW2012]. Since it is irreducible, we have

$$\pi(\chi_1, \chi_2) \cong \pi(\chi_2, \chi_1),$$

for example by [BZ1977]. Thus the data of π also comes with two choices, namely the choice of how to order χ_1 and χ_2. Since $\chi_1(p)$ and $\chi_2(p)$ have the property that when you multiply them by $p^{(k-1)/2}$, they are also the roots of the Satake polynomial

$$X^2 - a_pX + \epsilon_M(p)p^{k-1},$$

this choice of ordering is naturally equivalent to the choice of ordering of the two U_p-eigenvalues in the oldform space generated by f. This is all a special case of the general definition

Definition 2.1.28. An *accessible refinement* at $\mathfrak{p}$ of an automorphic representation π is a choice of smooth character $\chi_1 \otimes \chi_2 : T(\mathbf{Q}_p) \to \mathbf{C}^\times$ such that there is an embedding $\pi_{\mathfrak{p}} \to \mathrm{Ind}_{B(\mathbf{Q}_p)}^{\mathrm{GL}_n(\mathbf{Q}_p)}(\chi_1 \otimes \chi_2)$. Equivalently (by the adjunction between parabolic induction and Jacquet module, e.g. in [BZ1976]), it is a choice of $\chi_1 \otimes \chi_2$ which is a subquotient of the Jacquet module $J_B(\pi_p)$.

The point of an accessible refinement is that it is a convenient automorphic representation-theoretic way of capturing the Atkin–Lehner eigenvalues (beyond what we just did, see for instance [Ye2019, Lemma 4.5.2], [BC2009a, Ch. 6]).

By the fact that the Atkin–Lehner action can be read off the Jacquet module, the classical eigenforms parametrized by the eigencurve, namely the p-stablized newforms of [Eme2006b, Definition 4.4.1], are in canonical bijection with the accessibly refined automorphic representations (π, χ), where χ_1 is unramified, and $\chi_1(p)$ is directly related (up to some constant power of p that will depend on the normalization conventions for induction and the Jacquet module anyway) to the choice of U_p-eigenvalue. The fact that χ_1 must be unramified encapsulates the fact that for $m \geq 2$, the only relevant forms of level $\Gamma_1(Np^m)$ are the newforms — correspondingly, for an automorphic representation of level $\Gamma_1(Np^{m-1})$, there is only one accessible refinement (χ_1, χ_2) with χ_1 unramified.

For groups compact at infinity, there is no need for the extra restriction on χ_1

2.2 | Construction of Eigenvarieties, à la Emerton–Newton–Thorne

2.2.1 | Two alternative constructions

Take G as in the previous section. Emerton's definition of the tame level-$K^{\mathfrak{p}}$ eigenvariety for G is as follows:

Definition 2.2.1 (Definition 0.6 of [Eme2006b])**.** Let $\mathcal{T}$ be the rigid space over $\mathbf{Q}_p$ parametrizing p-adic characters of $T(\mathbf{Q}_p)$, let $\mathbf{T}^{\mathfrak{p}} := \mathbf{T}^{\mathfrak{p},\mathrm{sph}}$ be the spherical[21] Hecke algebra for $K^{\mathfrak{p}}$ away from $\mathfrak{p}$, and fix an isomorphism $\iota : \overline{\mathbf{Q}}_p \to \mathbf{C}$. The eigenvariety $\mathscr{E}(G, K^{\mathfrak{p}})$ is the Zariski-closure in

$$(\mathrm{Spec}\mathbf{T}^{\mathfrak{p}})^{\mathrm{rig}} \times \mathcal{T}$$

of the set of systems of eigenvalues of refined $\mathfrak{p}$-adic classical automorphic representations, that is, the set of points in

$$\mathrm{Hom}(\mathbf{T}^{\mathfrak{p}}, \overline{\mathbf{Q}}_p) \times \mathrm{Hom}(T(\mathbf{Q}_p), \overline{\mathbf{Q}}_p^{\times})$$

such that there exists an accessibly refined automorphic representation (π, χ) of G such that $\pi_\infty \cong W$ is allowable such that the first coordinate is equal to the system of eigenvalues of $\mathbf{T}^{\mathfrak{p}}$ acting on $\pi^{K^{\mathfrak{p}}}$, and the second coordinate is $\chi \cdot \psi$, where ψ is the highest weight of $\iota^{-1}W$ with respect to the upper-triangular borel of $G(F_{\mathfrak{p}})$.

[21]This means you forget all the finitely-many places away from $\mathfrak{p}$ at which $K^{\mathfrak{p}}$ is not a hyperspecial maximal compact.

Emerton also defines another rigid subspace of $(\mathrm{Spec}\mathbf{T}^{\mathfrak{p}})^{\mathrm{rig}} \times \mathcal{T}$ by using his Jacquet module for locally analytic representations. In particular, recall that the whole space of locally analytic p-adic automorphic forms

$$\widetilde{H}^0(K^{\mathfrak{p}}, E)_{\mathbf{Q}_p\text{-loc. an.}}$$

is an E-Banach space that admits a locally analytic action of the locally $\mathbf{Q}_p$-analytic group $G(F_{\mathfrak{p}}) \cong \mathrm{GL}_n(\mathbf{Q}_p)$. In fact, it is admissible in the sense of [Eme2017, Definition 6.1.1], by [Eme2006b, Theorem 2.2.22]. Therefore, taking the Jacquet module (defined in generality by Emerton in [Eme2006a]) with respect to the upper-triangular Borel B of $\mathrm{GL}_n(\mathbf{Q}_p)$, we obtain an essentially admissible (see [Eme2017, §6.4]) locally analytic representation of the diagonal torus $T(\mathbf{Q}_p)$.

$$J_B\left(\widetilde{H}^0(K^{\mathfrak{p}}, E)_{\mathbf{Q}_p\text{-loc. an.}}\right).$$

By definition of essential admissibility (see [Eme2006b, Proposition 2.3.2]), this module is the same as a coherent sheaf on $\mathcal{T}$. In analogy with the eigenvariety machine of [Buz2007], one can take the image $\mathcal{A}$ of $\mathbf{T}^{\mathfrak{p}}$ in the endomorphism ring of this sheaf. Then consider the relative Sp (defined for example in [Con2006])

$$\mathrm{Sp}_{\mathcal{T}}(\mathcal{A}).$$

Given that the Jacquet module is supposed to pick up all the systems of eigenvalues of overconvergent forms (see [Loe2011, 3.10.3], which is in any event a consequence of Proposition 2.1.23), and given the construction of [Buz2007], we can maybe expect this to be the same as $\mathscr{E}(G, K^{\mathfrak{p}})$ via the natural closed embedding

$$\mathrm{Sp}_{\mathcal{T}}(\mathcal{A}) \to (\mathrm{Spec}\mathbf{T}^{\mathfrak{p}})^{\mathrm{rig}} \times \mathcal{T}.$$

First, we note that one of the inclusions is easy:

Proposition 2.2.2. $\mathrm{Sp}_{\mathcal{T}}(\mathcal{A}) \supset \mathscr{E}(G, K^{\mathfrak{p}})$ *as analytic subsets of* $(\mathrm{Spec}\mathbf{T}^{\mathfrak{p}})^{rig} \times \mathcal{T}$

Proof. This is a direct consequence of the combination of Proposition 2.1.21, [Eme2006b, Proposition 2.3.3(3)] (which is itself a direct consequence of the definition), the left-exactness of Emerton's Jacquet module functor ([Eme2006a, Theorem 4.2.32]), the fact that Emerton's Jacquet module coincides with the classical one in the case of smooth representations, and the fact that the Hecke action can be read off of the Jacquet module. □

Therefore, to prove the conjectured equality, it suffices to show that the classical points (i.e. those systems of eigenvalues coming from accessibly refined classical $\mathfrak{p}$-adic automorphic representations) are Zariski-dense in $\mathrm{Sp}_{\mathcal{T}}(\mathcal{A})$. This was technically unknown at the time of [Eme2006b], butwe will prove it using the technique of (at least locally) explicit comparison with Buzzard's "D" eigenvariety.

2.2.2 | Accumulation of classical points

The point of how to prove the accumulation property of the classical points in $\mathrm{Sp}_{\mathcal{T}}(\mathcal{A})$ is to observe that the points that are classical because of numerical non-criticality are already enough: not only do they accumulate at all classical points, but actually at all points of locally algebraic (classical) weight. The entire technique for how to do this is borrowed from [BHS2017], although one must technically redo the arguments since in that paper they are done for the "patched eigenvariety."

Technically, we cannote deduce any relationship between the eigenvariety $\mathrm{Sp}_{\mathcal{T}}(\mathcal{A})$ and the "D" eigenvariety constructed by Buzzard [Buz2007] and Loeffler [Loe2011] until we know the accumulation of classical points. However, we can still abstractly observe parallels between the constructionsthat will help us prove things. The main thing to do is to find the spectral variety in the Jacquet module construction. In particular, the restriction morphism $\nu : \mathcal{T} \to \mathcal{W}$, where $\mathcal{W} = \mathrm{Hom}(T(\mathbf{Z}_p), \mathbf{G}_m)$, together with evaluation at a choice of $z \in \Sigma^{++}$ (which shouldn't matter but we take the obvious choice of $z = \mathrm{diag}(p^{n-1}, p^{n-2}, \ldots, p, 1)$), provides the composition

$$\mathrm{Sp}_{\mathcal{T}}(\mathcal{A}) \to \mathcal{T} \to \mathcal{W} \times \mathbf{G}_m$$

This identifies $\mathrm{Sp}_{\mathcal{T}}(\mathcal{A})$ with a cover of the Fredholm variety $\mathcal{Y}_z$ for the (compact by Proposition 2.1.26) z. Take the admissible affinoid cover $\{U_i\}_{i\in I}$ of the Fredholm variety whose existence is guaranteed by [Buz2007, Theorem 4.6]. This cover has the property that the image of U_i in $\mathcal{W}$ is an open affinoid W_i, and U_i is a finite cover of W_i. By the finiteness of the map $\mathrm{Sp}_{\mathcal{T}}(\mathcal{A}) \to \mathcal{T} \to \mathcal{W} \times \mathbf{G}_m$, we may pull back the U_i's to an admissible affinoid cover $\{V_i\}_{i\in I}$ of $\mathrm{Sp}_{\mathcal{T}}(\mathcal{A})$, in which case V_i is finite over W_i. It follows from a general lemma in rigid geometry (see [Taï2016, Lemma 2.1.2]) that (the connected components of) the V_i's provide a basis of affinoids for the canonical topology on $\mathrm{Sp}_{\mathcal{T}}(\mathcal{A})$, and hence for every $z_0 \in \mathrm{Sp}_{\mathcal{T}}(\mathcal{A})$, they provide a neighborhood basis for z_0.

Proposition 2.2.3. *Fix $z_0 \in \mathrm{Sp}_{\mathcal{T}}(\mathcal{A})$ such that $\kappa(z_0)$, namely the projection to $\mathcal{W}$, is locally algebraic. Fix some $V = V_i$ in the affinoid neighborhood basis of z_0 described above. Then the classical points are Zariski-dense in V.*

Proof. We will construct a Zariski-dense subset of classical points in V by constructing a subset whose associated refinements are numerically non-critical in the sense of [NT2021, Definition 2.9, Remark 2.21].

The valuation of the value at p of one of the n coordinates of the $\mathcal{T}$-coordinate is by definition an analytic function on the affinoid V. Since affinoids behave as if they are compact, these valuations are all bounded, say by some number B_U

By definition of numerically non-critical, and by finiteness of the map from V to W_i, this means we just need to find Zariski-dense set of classical weights $(k_1, \ldots, k_n)$ in W_i such that the

k_i's are increasing and $k_{i+1} - k_i$ is larger than the bound B_U. There are many of these weights, and in fact they can be made to accumulate at $\kappa(z_0)$ by adding large powers of p to the integer weights defining the algebraic part of $\kappa(z_0)$ (the smooth part can just be taken to match that of $\kappa(z_0)$). Indeed, done appropriately, this can make the $(k_{i+1} - k_i)$'s all positive and larger than B_U^n.

In any event, now that we have these numerically non-critical points which are Zariski-dense in V, we use the numerical non-criticality criterion for classicality [Eme2006a, Theorem 4.4.5]. One must compare the notion of numerical non-criticality from [NT2021, Definition 2.9] to Emerton's notion of non-critical slope, and this requires some fiddling with modulus characters. But we choose not to make this explicit here due to lack of space and also the fact that even if the correct definition of numerical non-criticality was off by a few modulus characters here and there, the above boudnedness argument would still work. □

Since the classical points need to have locally algebraic weight (by definition), we conclude the desired result:

Theorem 2.2.4. *The classical points are Zariski-dense and self-accumulating in* $\mathrm{Sp}_{\mathcal{T}}(\mathcal{A})$*. In other words,* $\mathrm{Sp}_{\mathcal{T}}(\mathcal{A}) = \mathscr{E}(G, K^{\mathfrak{p}})$*.*

We are also allowed to assume that $\mathscr{E}$ is reduced. In fact, it is already true (see [NT2021, Proposition 2.22(1)]), but it does not harm us to just take the associated reduced E-rigid space.

2.2.3 | Classicality theorems, d'après Chenevier and Newton–Thorne

The classicality result [Eme2006a, Theorem 4.4.5] that was used to prove Theorem 2.2.4 uses Verma module techniques. The following classicality theorem, which we will use in the next chapter to deduce the analytic continuation of symmetric power functoriality, also does. For the reason of lack of space, we do not give any real details of the proof.

Proposition 2.2.5 (Lemma 2.30 of [NT2021])**.** *Let $z \in \mathscr{E}(G, K^{\mathfrak{p}})$ be a point whose associated p-adic Galois representation is absolutely irreducible at all the places above p, and whose $\mathcal{T}$-coordinate δ is regular and locally algebraic. If every triangulation of the associated p-adic Galois representations at places above p are all non-critical in the sense of* [BC2009a, Definition 2.4.5], *then z is classical.*

Proof. The proof is fully explained in [NT2021, Lemma 2.30]. There are several ingredients, but fundamentally it is an argument about twin/companion points. One uses the analytic continuation of triangulations from [KPX2014] together with [BHS2017, Lemma 2.11] (the proof of which also uses [KPX2014]), to show that the algebraic part of δ is strictly dominant. The other technical input is the theory of [OS2015], which provides a way (a functor called $\mathcal{F}_P^G$) of taking locally analytic induction of representations of Lie algebras in the BGG category $\mathcal{O}$

[Hum2008], together with an adjunction from [Bre2015] between $\mathcal{F}_P^G$ and Emerton's Jacquet module. This allows you to study the subquotients of $J_B(\widetilde{H}^0(K^{\mathfrak{p}}, E))_{\mathbf{Q}_p\text{-loc.an.}}$ using Verma module techniques. By explicit description of the Jordan–Holder factors of Verma modules, the fact that the Jordan–Holder factor indexed by $w = 1$ always appears with multiplicity 1, one argues that the only elements of $\widetilde{H}^0(K^{\mathfrak{p}}, E))_{\mathbf{Q}_p\text{-loc.an.}}$ with such a parameter actually come from the $w = 1$ Jordan–Holder factor, and are therefore locally algebraic. □

The argument of Proposition 2.2.5 is quite similar to that of [Che2011, Proposition 4.2], where Verma module techniques and analytic continuation of p-adic Hodge theory data is also used. Chenevier's version uses weaker information about the analytic continuation from the famous paper of Kisin [Kis2003], since [KPX2014] was not yet available. On the other hand, it uses stronger Lie-theoretic information, namely the BGG resolution (whereas no deep information about Verma modules at all is used in the proof of Proposition 2.2.5 — everything can be found in the introductory chapters of [Hum2008]). By incorporating the BGG resolution argument into the proof of Proposition 2.2.5, the result could most likely be made slightly stronger.

Chapter 3

Analytic continuation of symmetric power functoriality

> "These p-adic Hodge theorists seemed to me like an order of monks, who were able to reveal the hidden design of a tapestry by examining it one thread at a time."
>
> Mark Kisin, [Kis2019]

This chapter is about the proof of the following theorem [NT2021, Corollary 2.33]. As usual, let $\iota : \overline{\mathbf{Q}}_p \to \mathbf{C}$ be an isomorphism.

Theorem 3.0.1. *Let z_0, z_0' be classical points on the eigencurve that live on the same irreducible component of $\mathscr{E}_{\mathbf{C}_p}$. Suppose that these come from refined automorphic representations (π_0, χ_0), (π_0', χ_0') for GL_2, which further satisfy*

1. *The refinements χ_0, χ_0' are numerically non-critical and n-regular (see* [NT2021, Definition 2.23]*).*
2. *The Zariski-closures of the images of $r_{\pi_0,\iota}, r_{\pi_0',\iota}$ on $G_{\mathbf{Q}_p}$ contain SL_2.*

Then $\mathrm{Sym}^{n-1} r_{\pi_0,\iota}$ is automorphic if and only if $\mathrm{Sym}^{n-1} r_{\pi_0,\iota}$ is.

3.1 | The trianguline variety

The argument is based on looking at an irreducible component of an eigenvariety inside an irreducible component of the trianguline variety (to be defined), and using the usual things about trianguline deformations of Galois representations (together with the key input of vanishing of adjoint Selmer group, which is the main theorem of previous work of Newton–Thorne [NT2020]) to prove that in fact the two things are smooth of the same dimension and locally isomorphic at the points we are interested in. We do NOT discuss [NT2020] at all in this mémoire.

Remark 3.1.1. In [NT2021], everything is made slightly more complicated in the preliminary section because there is no assumption on the splitting type of p in the CM field F. In the actual situation that this must be applied in, p is totally split in F, so the fields "$F_{\tilde{v}}$, $F_{\tilde{v}}$" are really just $\mathbf{Q}_p$ (at least this is my understading of the situation). This is particularly nice because it means that we can technically get away with only using the theory of (φ, Γ)-modules for $G_{\mathbf{Q}_p}$, i.e. the stuff that is proved in your book with Bellaïche [BC2009a], rather than needing additional input from Nakamura and others. It is also convenient (though doesn't really simplify any of the arguments)

The theory that underlies all of this is the link between the local geometry of the pseudocharacter varieties and deformation theory, which is developed in your paper on pseudocharacters [Che2014] (implicitly the pseudocharacters we use here have always been the so-called "determinants", which we need in order to deal with pseudodeformations of a residual pseudocharacter over a characteristic p field — it would be incorrect to use the pseudocharacters that I actually know about).

The following theorem is essentially what is stated and proved in your work [Che2014] (the only difference is that the $W(k)$ is replaced with $\mathcal{O}_E$ where E is allowed to be anything that accommodates the given residual character, making this statement slightly more general, but the proof is identical):

Theorem 3.1.2 (Chenevier)**.** *Let E be a finite extension of $\mathbf{Q}_p$ with residue field k_E, and fix a conjugate self-dual pseudocharacter $\overline{\tau}$ of $G_{F,S}$ with coefficients in k_E.*

1. *The functor from the category of complete local Noetherian $\mathcal{O}_E$-algebras with residue field k_E to* Set *given on objects by*

   $A \mapsto$ *the set of continuous conjugate self-dual pseudocharacters $G_{F,S} \to A$ lifting $\overline{\tau}$*

   is representable by the "universal pseudodeformation ring" $R(\overline{\tau})$, a complete local Noetherian $\mathcal{O}_E$-algebra.

2. *The functor from the category of E-rigid spaces to* Set *given by*

   $\mathcal{Y} \mapsto$ *the set of residually constant pseudocharacters $G_{F,S} \to \mathcal{O}(\mathcal{Y})$ with residual pseudocharacter $\overline{\tau}$*

   is represented by the E-rigid space

   $$\mathfrak{X}_{\overline{\tau}} := (\mathrm{Spf} R(\overline{\tau}))^{rig},$$

   where $(-)^{rig}$ denotes the rigid generic fiber in the sense of Berthelot.

3. *The functor from the category of E-rigid spaces to* Set *given by*

$$\mathcal{Y} \mapsto \text{the set of pseudocharacters} G_{F,S} \to \mathcal{O}(\mathcal{Y}) \text{ of dimension } n$$

is represented by the E-rigid space

$$\mathfrak{X}_n := \bigsqcup_{\overline{\tau}} (\operatorname{Spf} R(\overline{\tau}))^{rig},$$

where the τ run over the set of all residual determinants of dimension n.

4. *For any closed point $x \in \mathfrak{X}_n$ of residue field $\kappa(x)$ (a finite extension of E), the completed local ring $\widehat{\mathcal{O}}_{\mathfrak{X}_n,x}$ represents the functor*[1] *from the category of complete local Noetherian $\kappa(x)$-algebras with residue field $\kappa(x)$ to* Set *given by*

$$A \mapsto \text{the set of conjugate self-dual pseudocharacters } G_{F,S} \to A \text{ lifting } D_x,$$

where $D_x : G_{F,S} \to \kappa(x)$ is the pseudocharacter corresponding to the point x.

The same theorem is also true if we replace $G_{F,S}$ with $G_{\mathbf{Q}_p}$ and get rid of the words "conjugate self dual" everywhere. For any place $v|p$ of F^+, denote by $\mathfrak{X}_{n,v}$ the rigid space representing the functor of analytic families of pseudocharacters of $G_{F_{\tilde{v}}} = G_{\mathbf{Q}_p}$. Obviously the space itself does not depend on the choice of v, but the restriction map[2] $\mathfrak{X}_n \to \mathfrak{X}_{n,v}$ does depend on v.

Definition 3.1.3. For any place $v|p$ of F^+, let $\mathfrak{X}_{n,v}^{\mathrm{irr}}$ be the *absolutely irreducible locus* in $\mathfrak{X}_{n,v}$, i.e. the locus on which the universal pseudocharacter $G_{\mathbf{Q}_p} \to \mathcal{O}(\mathfrak{X}_{n,v})$ is absolutely irreducible (you proved that it is an open subspace in [Che2014]). Define the open subspace $\mathfrak{X}_n^{p-\mathrm{irr}} \subset \mathfrak{X}_n$ via the pullback

$$\begin{array}{ccc} \mathfrak{X}_n^{p-\mathrm{irr}} & \longrightarrow & \prod_{v|p} \mathfrak{X}_{n,v}^{\mathrm{irr}} \\ \downarrow & & \downarrow \\ \mathfrak{X}_n & \longrightarrow & \prod_{v|p} \mathfrak{X}_{n,v} \end{array}$$

i.e. by imposing an absolute irreducibility condition at each $v|p$.

The absolute irreducibility condition is useful because in that case the deformations of a pseudocharacter are the same thing as a deformations of the unique irreducible representation

[1] I think it is kind of interesting how deformation theory comes up in two ways in this type of result: once in the sense of characteristic-0 lifts of a positive characteristic objects, in order to define the rigid space we are interested in, and then the second time in the sense of lifts from characteristic zero to characteristic zero in order to deal with the local geometry of this rigid space.

[2] Given via the functorial description by the $\mathfrak{X}_n$-family of pseudocharacters of $G_{F_{\tilde{v}}}$ given by restricting to $G_{F_{\tilde{v}}} \subset G_{F,S}$ the universal pseudocharacter.

that corresponds to it (we can even use the pseudocharacters of [BC2009a] rather than determinants since we really only care about the equal characteristic zero case, in which case this is a standard fact, a consequence e.g. of the fact that complete local rings are henselian).

As a consequence of this and the usual identification between the tangent space of Galois deformation functor and first Galois cohomology of adjoint representation, we have

Lemma 3.1.4. *For a closed point $z \in \prod_{v|p} \mathfrak{X}^{irr}_{n,v}$ with corresponding tuple of isomorphism classes of Galois representations $(\rho_{z,v})_{v|p}$,*

$$T_z \prod_{v|p} \mathfrak{X}^{\mathrm{irr}}_{n,v} \cong \bigoplus_{v|p} H^1(G_{\mathbf{Q}_p}, \mathrm{ad}\rho_{z,v}).$$

A closed point $z \in \mathfrak{X}^{p-\mathrm{irr}}_n$ is certainly absolutely irreducible, since it is absolutely irreducible when restricted to $G_{\mathbf{Q}_p} = G_{F_{\tilde{v}}} \subset G_{F,S}$. This is convenient, but note that the computation of $T_z\mathfrak{X}^{p-\mathrm{irr}}_n$ is not completely obvious in the same way as Lemma 3.1.4, because in Theorem 3.1.2(4) the pseudodeformations must also be conjugate self-dual. For this reason, it is convenient[3] to introduce the group scheme $\mathcal{G}_n$ of [CHT2008, §2].

Obviously the definition might as well be made over $\mathbf{Z}$, but we will only have use for the version defined over E (i.e. the base change to E of the $\mathbf{Z}$-version).

Definition 3.1.5. Define the E-group scheme

$$\mathcal{G}_n := (\mathrm{GL}_n \times \mathrm{GL}_1) \rtimes \{1, J\}$$

where J acts on $\mathrm{GL}_n \times \mathrm{GL}_1$ by $(g, \mu) \mapsto (\mu \cdot g^{\top,-1}, \mu)$. Let $\nu : \mathcal{G}_n \to \mathrm{GL}_1$ be the character[4] defined on the $(\mathrm{GL}_n \times \mathrm{GL}_1) \times \{1\}$-component by projection to GL_1, and on the other component by the negation of the projection to GL_1.

The point of all this is that n-dimensional conjugate self-dual Galois representations (and deformations thereof) are supposed to be related to homomorphisms into $\mathcal{G}_n$. In fact, we have:

Lemma 3.1.6 (Clozel–Harris–Taylor)**.** *Let E'/E be a finite extension, and let $\rho : G_{F,S} \to \mathrm{GL}_n(E')$ be an absolutely irreducible representation such that $\rho^\vee \epsilon^{1-n} \cong \rho^c$. Then up to $\mathrm{GL}_n(\overline{\mathbf{Q}}_p)$-conjugacy, there is exactly one homomorphism*

$$\widetilde{\rho} : G_{F^+,S} \to \mathcal{G}_n(E')$$

[3]though unlikely really necessary — at this point I think Newton and Thorne are just showing off

[4]To check that ν is a group scheme homomorphism: by the universal property of semidirect products, it suffices to prove that

$$\nu(J \cdot (g, \mu)) = \nu(J)\nu((g, \mu))\nu(J)^{-1}$$

which is true because the left hand side is $\nu(\mu \cdot g^{\top,-1}, \mu) = \mu$ and the right hand side is also μ since $\nu(J) = -1$.

satisfying $\nu \circ \widetilde{\rho} = \epsilon^{1-n}\delta^a_{F/F^+}$ *for some* $a \in \{0,1\}$ *and* $\widetilde{\rho}^{-1}((\mathrm{GL}_n(E') \times \mathrm{GL}_1(E')) \times \{1\}) = G_{F,S}$[5]*; in fact the value of* a *is uniquely determined by* ρ*.*

Proof. This is some combination of various lemmas in [CHT2008, §2], except I will use slightly more concrete language.

Let us fix a basis for $(E')^{\oplus n}$ for the n-dimensional vector space involved here, and use it to produce a matrix B for the $E'[G_{F,S}]$-linear isomorphism $\rho^\vee\epsilon^{1-n} \to \rho^c$. By Schur's lemma, and the absolute irreducibility assumption, the matrix B is unique up to scaling by element of $(E')^\times$. Furthermore, taking the transpose of B, we have an isomorphism

$$B^\top : \rho^{c,\vee} \to \rho\epsilon^{n-1},$$

which (by writing down explicitly what the equivariance condition means) is also an isomorphism

$$B^\top : \rho^\vee\epsilon^{1-n} \to \rho^c,$$

i.e. $B^\top$ satisfies exactly the same $E'[G_{F,S}]$-linearity condition as B, and hence there is some $\alpha \in (E')^\times$ such that

$$B = \alpha B^\top.$$

Taking the transpose of both sides, we obtain $\alpha^2 = 1$, i.e. $\alpha = \pm 1$. Both possibilities are possible, and the value of α (obviously uniquely determined by ρ) will affect the value of a.

In any event, the $E'[G_{F,S}]$-linearity of B gives us the condition

$$\rho(g)^{\top,-1}\epsilon(g)^{1-n} = B^{-1}\rho(cgc)B \tag{3.1}$$

for every $g \in G_{F,S}$. Furthermore, since $\epsilon(c) = -1$, we can rewrite the condition $B = \pm 1 B^\top$ as

$$B = (-1)^b \cdot (-\epsilon^{n-1}(c))B^\top \tag{3.2}$$

where $b \in \{0,1\}$ is uniquely determined by ρ (b depends on α as well as the parity of n, and thus ultimately only on ρ).

Anyhow, the homomorphism $\widetilde{\rho} : G_{F^+,S} \to \mathcal{G}_n(E')$ a priori has to satisfy

$$\widetilde{\rho}(g) = (\rho(g), \epsilon(g)^{1-n}, 1)$$

for all $g \in G_{F,S}$. This is a perfectly well-defined homomorphism so far, and our only job is to think about how it can be extended to $G_{F^+,S}$. The extension is determined by its value on

[5] All such homomorphisms will be assumed to satisfy this last condition, without any comment.

complex conjugation $c \in G_{F^+,S} \setminus G_{F,S}$. By hypothesis, the value on complex conjugation is

$$\widetilde{\rho}(c) = (A, (-1)^a(-\epsilon(c)^{1-n}), J) = (A, (-1)^{a-n}, J),$$

for some $A \in \mathrm{GL}_n(E')$. In order for this to define an actual homomorphism, it must satisfy

$$1 = \widetilde{\rho}(c)^2 = (AA^{\top,-1}(-1)^a(-\epsilon(c)^{1-n}), 1, 1),$$

i.e.

$$A = (-1)^a(-\epsilon(c)^{1-n})A^{\top},$$

and it must have the property that for $g \in G_{F,S}$, the two possible ways of writing down $\widetilde{\rho}(gc)$ agree:

$$\begin{aligned}\widetilde{\rho}(gc) &= \widetilde{\rho}(g)\widetilde{\rho}(c) \\ &= (\rho(g), \epsilon(g)^{1-n}, 1) \cdot (A, (-1)^a(-\epsilon(c)^{1-n}), J) \\ &= (\rho(g)A, (-1)^a(-\epsilon(gc)^{1-n}), J)\end{aligned}$$

and

$$\begin{aligned}\widetilde{\rho}(gc) &= \widetilde{\rho}(c)\widetilde{\rho}(cgc) \\ &= (A, (-1)^a(-\epsilon(c)^{1-n}), J) \cdot (\rho(cgc), \epsilon(g)^{1-n}, 1) \\ &= (A\rho(cgc)^{\top,-1}\epsilon(g)^{1-n}, (-1)^a(-\epsilon(gc)^{1-n}), J),\end{aligned}$$

so the only condition is

$$\rho(g)A = A\rho(cgc)^{\top,-1}\epsilon(g)^{1-n},$$

i.e. A is $E'[G_{F,S}]$-equivariant from $\rho^{c,\vee}\epsilon^{1-n}$ to ρ, which is the same as being from $\rho^{\vee}\epsilon^{1-n}$ to ρ^c. By Schur's lemma, the only possibility is for $A = B$ up to a multiplicative constant, and, by Equation (3.2), the only option that can happen is $b = a$.

The only thing it remains to verify is that all the possible $\widetilde{\rho}$, defined by $\widetilde{\rho} = (\beta B, (-1)^a(-\epsilon(c)^{1-n}), J)$ for $\beta \in (E')^{\times}$, are all conjugate under $\mathrm{GL}_n(\overline{\mathbf{Q}}_p)$. Indeed, if we conjugate by $M \in \mathrm{GL}_n(\overline{\mathbf{Q}}_p)$, i.e. $(M, 1, 1) \in \mathcal{G}_n(\overline{\mathbf{Q}}_p)$, we end up with something whose value at g is

$$(M\rho(g)M^{-1}, \epsilon(g)^{1-n}, 1).$$

This won't even be a valid $\tilde{\rho}$ unless $M \in \overline{\mathbf{Q}}_p^{\times}$, since ρ is absolutely irreducible (by Shur's lemma, the centralizer of the image of ρ is just the scalar matrices). And if $M = m \in \overline{\mathbf{Q}}_p^{\times}$, the effect on

$\widetilde{\rho}(c) = (B, (-1)^b(-\epsilon(c)^{1-n}), J)$ is to take it to

$$(m^2 B, (-1)^b(-\epsilon(c)^{1-n}), J).$$

Since $\overline{\mathbf{Q}}_p$ is algebraically closed, this accounts for all scalar multiples of B, as desired. $\square$

Since complete local rings are henselian [Ray1970], by the standard theorems (e.g. [BC2009a, Ch. 1]), the conjugate self-dual deformations of the pseudocharacter corresponding to a given closed point $z \in \mathfrak{X}_n^{p-\mathrm{irr}}$ all correspond to a unique conjugate self-dual deformation of the corresponding irreducible representation $\rho_z : G_{F,S} \to \mathrm{GL}_n(\overline{\mathbf{Q}}_p)$. By generalizing Lemma 3.1.6 to the case of deformations, Newton–Thorne show that these conjugate self-dual deformations correspond to deformations of $\widetilde{\rho}_z$ (defined in the obvious way), which I will explain later. First, I define all of these things fully and explain why the tangent space of the deformation functor for a $\widetilde{\rho}$ can be written in the usual way using Galois cohomology.

Definition 3.1.7. For a homomorphism $\tilde{\rho} : G_{F^+,S} \to \mathcal{G}_n(E')$ satisfying $\nu \circ \tilde{\rho} = \epsilon^{1-n}\delta^a_{F/F^+}$, define:

- The deformation functor $D_{\tilde{\rho}}$ from the category of complete local Noetherian E'-algebras with residue field E' to Set defined on objects by

$$D_{\tilde{\rho}}(A) := \{\sigma : G_{F^+,S} \to \mathcal{G}_n(A) \text{ lifting } \widetilde{\rho} : \nu\circ\sigma = \epsilon^{1-n}\delta^a_{F/F^+}\}/\ker(\mathrm{GL}_n(A) \to \mathrm{GL}_n(E'))$$

- The adjoint representation $\mathrm{ad}\tilde{\rho}$ is defined via the adjoint action of $\mathcal{G}_n(E')$ on the E'-points of the Lie algebra of GL_n. It is easy to see that $\mathrm{GL}_n(E') \subset \mathcal{G}_n(E')$ acts by the usual adjoint action (conjugation of matrices), $\mathrm{GL}_1(E')$ acts trivially, and J acts by $x \mapsto -x^\top$.

The following is supposed to be true:

Lemma 3.1.8. *Let $\tilde{\rho} : G_{F^+,S} \to \mathcal{G}_n(E')$ be a homomorphism satisfying $\nu \circ \tilde{\rho} = \epsilon^{1-n}\delta^a_{F/F^+}$. Then*

$$D_{\tilde{\rho}}(E'[\varepsilon]) \cong H^1(G_{F^+,S}, \mathrm{ad}\tilde{\rho}).$$

Proof. Exactly the same argument as in [Che2010, Lecture 3, Proposition 3.3]) can be made to work; you just need to think a little bit more precisely about how the non-identity component acts in the adjoint action (since in that case you cannot just think about everything as a matrix). $\square$

Anyway, I do not think that this identification in terms of cohomology is actually useful, and the H^1 here might as well just be another name for the tangent space. In any event, we now carry out the final step of identifying this deformation functor with the pseudodeformation functor that we know is related to $T_z\mathfrak{X}_n^{p-\mathrm{irr}}$ thanks to Theorem 3.1.2(4):

Lemma 3.1.9. *Let $z \in \mathfrak{X}_n^{p-irr}$ be a closed point of residue field $\kappa(z)$. Then we know from* [BC2009a, Ch. 1] *that the corresponding $\kappa(z)$-valued pseudocharacter comes from a unique absolutely irreducible conjugate self-dual representation $\rho_z : G_{F,S} \to \mathrm{GL}_n(\kappa(z))$, and from* Lemma 3.1.6 *that there is a (unique up to scalar constants modifying the image of complex conjugation) corresponding $\widetilde{\rho_z} : G_{F^+,S} \to \mathcal{G}_n(\kappa(z))$. Then the Zariski tangent space $T_z\mathfrak{X}_n^{p-irr}$ may be computed from the deformation theory of this homomorphism:*

$$T_z\mathfrak{X}_n^{p-irr} \cong H^1(G_{F^+,S}, \mathrm{ad}\widetilde{\rho_z}).$$

Proof. Since complete local rings are henselian, [BC2009a, Lemma 1.4.3] and Theorem 3.1.2(4) tell us that $T_z\mathfrak{X}_n^{p-\mathrm{irr}}$ is isomorphic to the tangent space of the conjugate self-dual deformation functor of ρ_z (since in particular the conjugate self-dual pseudodeformation functor of $\mathrm{Tr}\rho_z$ is isomorphic to the conjugate self-dual deformation functor of ρ_z). We claim that this deformation functor is in turn isomorphic to $D_{\widetilde{\rho_z}}$, which would imply the claim thanks to Lemma 3.1.8. This is what is done in [NT2021, Lemma 2.12], which is easy for us now because we have essentially spelled out all the missing details from [CHT2008] in our proof of Lemma 3.1.6.

There is an obvious map from $D_{\tilde{\rho_z}}$ to the conjugate self-dual deformation functor of ρ_z, given by restricting to $G_{F,S}$ and projecting to GL_n (this results in something conjugate self-dual thanks to the analysis in Lemma 3.1.6 — the proof of this part works just as well over an arbitrary ring).

- Injectivity: Let $\sigma_1, \sigma_2 \in D_{\tilde{\rho_z}}(A)$ for some complete Noetherian local $\kappa(z)$-algebra A with residue field $\kappa(z)$. Suppose that $\pi_{\mathrm{GL}_n} \circ \sigma_1|_{G_{F,S}}$ and $\pi_{\mathrm{GL}_n} \circ \sigma_2|_{G_{F,S}}$ are conjugate under some $M \in \ker(\mathrm{GL}_n(A) \to \mathrm{GL}_n(\kappa(z)))$. We need to deduce that σ_1, σ_2 are also conjugate by M, and the only thing left is to show that $\sigma_1(c), \sigma_2(c)$ are conjugate by M. After conjugating one of them by M, we can assume that $\pi_{\mathrm{GL}_n} \circ \sigma_1|_{G_{F,S}} = \pi_{\mathrm{GL}_n} \circ \sigma_2|_{G_{F,S}} =: \tau$, and consider the operator $(\pi_{\mathrm{GL}_n} \circ \sigma_1(c))(\pi_{\mathrm{GL}_n} \circ \sigma_2(c))^{-1} \in \mathrm{GL}_n(A)$. Again by the same reasoning as in the proof of Lemma 3.1.6 (but extended to arbitrary base ring), $\pi_{\mathrm{GL}_n} \circ \sigma_i(c)$ are $A[G_{F,S}]$-equivariant isomorphisms $\tau^\vee \epsilon^{1-n} \to \tau^c$, and therefore $(\pi_{\mathrm{GL}_n} \circ \sigma_1(c))(\pi_{\mathrm{GL}_n} \circ \sigma_2(c))^{-1}$ centralizes the image of a residually absolutely irreducible representation of $G_{F,S}$ valued in $\mathrm{GL}_n(A)$. In this situation, Schur's lemma in the form of [CHT2008, Lemma 2.1.8] (easy exercise in induction on length of Artinian algebras) shows that there is an $\alpha \in A^\times$ such that $\sigma_1(c) = \alpha\sigma_2(c)$. But since σ_1, σ_2 are both lifts to A of the same $\widetilde{\rho}_z$, we know that $\alpha \equiv 1 \mod \mathfrak{m}_A$. This is very nice, because since A is henselian, it implies α has a square root in $1 \mod \mathfrak{m}_A \subset \ker(\mathrm{GL}_n(A) \to \mathrm{GL}_n(\kappa(z)))$, which we can then further conjugate σ_2 by to get σ_1.

- Surjectivity is easier: as soon as we have a conjugate self-dual deformation of ρ_z with coefficients in A, we can construct the coresponding element of $D_{\tilde{\rho_z}}(A)$ by the same procedure as Lemma 3.1.6.

□

Anyhow, since the identifications work in the exact same way, the natural diagram

$$\begin{array}{ccc} T_z\mathfrak{X}_n^{p-\mathrm{irr}} & \longrightarrow & H^1(G_{F^+,S}, \mathrm{ad}\widetilde{\rho}_z) \\ \downarrow & & \downarrow \\ T_{(\mathrm{res} z)_{v|p}} \prod_{v|p} \mathfrak{X}_{n,v}^{\mathrm{irr}} & \longrightarrow & \bigoplus_{v|p} H^1(G_{\mathbf{Q}_p}, \mathrm{ad}\rho_z|_{G_{F_v}}) \end{array}$$

commutes.

The key point in all of this is to analyze the relationship between the eigenvariety and the "trianguline locus" both sitting inside of the character variety $\mathfrak{X}_n \times \mathcal{T}_n$, which we will be able to do inside the open locus $\mathfrak{X}_n^{p-\mathrm{irr}} \times \mathcal{T}_n^{\mathrm{reg}}$ where we can interpret the points as being isomorphism classes of conjugate self-dual n-dimensional absolutely irreducible representations of $G_{F,S}$ together with the data of some characters which are meant to play the role of a triangulation at each p-adic place, and therefore think about the local geometry of both of these objects in terms of deformation theory.

Of course, thinking about the geometry of the trianguline locus will require us to use the results of [KPX2014] regarding spreading-out of triangulations, which in turn requires us to start out with a globally-defined analytic family of p-adic Galois reprensentations. Moreover, $\mathfrak{X}_{n,v}^{\mathrm{irr}}$ is not necessarily equipped with a global universal family $G_{\mathbf{Q}_p} \to \mathrm{GL}_n(\mathcal{O}(\mathfrak{X}_{n,v}^{\mathrm{irr}}))$: the theory from [Che2014, §4.2] only guarantees the existence of a global Azumaya algebra $\mathcal{A}$ (i.e. the universal Cayley–Hamilton quotient) over $\mathfrak{X}_{n,v}^{\mathrm{irr}}$ such that there is a universal representation $G_{\mathbf{Q}_p} \to \mathcal{A}^\times$ inducing pointwise the pseudocharacters in $\mathfrak{X}_{n,v}^{\mathrm{irr}}$ via taking reduced trace — there is no reason for $\mathcal{A}$ to be globally split. On the bright side, for any closed point $z \in \mathfrak{X}_{n,v}^{\mathrm{irr}}$, the absolute irreducibility of z implies that the stalk $\mathcal{A}_z$ splits (as $\mathcal{O}_{\mathfrak{X}_{n,v}^{\mathrm{irr}}}$ is a henselian local ring), and hence there is an affinoid subdomain $\mathcal{U}$ of $\mathfrak{X}_{n,v}^{\mathrm{irr}}$ containing z and carrying a universal representation $G_{\mathbf{Q}_p} \to \mathrm{GL}_n(\mathcal{O}(\mathcal{U}))$. The exact same argument applies to the global character variety $\mathfrak{X}_{n,v}^{p-\mathrm{irr}}$. We will use the existence of these universal representations locally on the pseudocharacter varieties without comment.

Now it is finally time to make precise what we meant by "trianguline locus":

Definition 3.1.10. Define the *(local) trianguline locus* $\Delta_v \subset \mathfrak{X}_{n,v}^{\mathrm{irr}} \times \mathcal{T}_{n,v}$ to be the Zariski closure of the set of points (ρ, δ) such that ρ is trianguline of parameter δ. We will only be interested in the intersection of Δ_v with affinoid open subdomains of $\mathfrak{X}_{n,v}^{\mathrm{irr}} \times \mathcal{T}_{n,v}$ small enough to have an associated universal representation (in order to use [KPX2014]) and δ is regular (in order to relate the local geometry of the trianguline locus to deformation theory).

Of course, the purpose of all of this is that the refined automorphic representations that define classical points on our eigenvarieties will live in $\mathfrak{X}_n \times \mathcal{T}$ and be trianguline at all places of F^+ over p, and so we will ultimately be interested in the global version:

Definition 3.1.11. Define the *(global) trianguline locus* in $\mathfrak{X}_{n,v}^{\mathrm{irr}} \times \mathcal{T}_n$ to be

$$\Delta := \bigcap_{v|p} \mathrm{res_v}^{-1}(\Delta_v).$$

For the same reasons as before, we are only interested in this construction when intersected wit affinoid opens $\mathcal{U}$ where $\mathcal{U} \subset \mathfrak{X}_n^{p-\mathrm{irr}} \times \mathcal{T}_{n,v}^{\mathrm{reg}}$ and $\mathcal{U}$ is small enough to admit a universal representation. In fact, since we will be interested in understanding the local geometry of Δ based on the local geometry of the various Δ_v, we will further ask that $\mathcal{U}$ is contained in $\cap_{v|p}\mathrm{res_v}^{-1}(\mathcal{U}_v)$, where $\mathcal{U}_v$ are nice affinoid opens as described in the definition of Δ_v.

Anyway, we can now start with understanding the local geometry of Δ_v inside the regular locus.

Proposition 3.1.12. *Let $v|p$ be a place of F^+, and let $z \in \Delta_v \cap \mathfrak{X}_{n,v}^{irr} \times \mathcal{T}_{n,v}^{reg}$ be a closed point corresponding to a pair (ρ_z, δ_z) such that ρ_z is trianguline of parameter δ_z. Then*

$$T_z\Delta_v \subset H^1_{tri,\delta_v}(G_{\mathbf{Q}_p}, \mathrm{ad}\rho_z),$$

where this H^1_{tri,δ_v} denotes the set of $\kappa(z)[\varepsilon]$-valued deformations that are trianguline of parameter δ_v and is viewed as a subspace of

$$T_z(\mathfrak{X}_{n,v}^{irr} \times \mathcal{T}_{n,v}^{reg}) = H^1(G_{\mathbf{Q}_p}, \mathrm{ad}\rho_z) \oplus T_{\delta_z}\mathcal{T}_{n,v}^{reg}$$

(though NOT necessarily in the obvious way — there is additional information about the tangent direction at the parameter δ_z that is involved here)

Proof. First of all, the statement of the result is fairly intuitive: the tangent space of the trianguline locus should be interpreted as trianguline elements of the ambient tangent space. Since the map from the δ_z-trianguline deformation functor to the deformation functor of ρ_z is injective and relatively representable ([BC2009a, Proposition 2.3.6, 2.3.6], using in a crucial way the regularity hypothesis), we can realize $H^1_{\mathrm{tri},\delta_z}(G_{\mathbf{Q}_p}, \mathrm{ad}\rho_z)$ as the tangent space of an actual universal deformation ring R_{ρ_z,δ_z} which admits a surjection from $\widehat{\mathcal{O}}_{\mathfrak{X}_{n,v}\times\mathcal{T}_{n,v}}$ (as this is the universal deformation ring of ρ_z by Theorem 3.1.2). Therefore, the desired existence of the dotted arrow in the diagram

$$\begin{tikzcd} T_z(\mathfrak{X}_{n,v} \times \mathcal{T}_{n,v}) & T_z\Delta_v \arrow[l, hook'] \arrow[dl, dotted] \\ H^1_{\mathrm{tri},\delta_z}(G_{\mathbf{Q}_p}, \mathrm{ad}\rho_z) \arrow[u, hook] & \end{tikzcd}$$

would be implied by the existence of the dotted arrow in the diagram

$$\begin{tikzcd}
\widehat{\mathcal{O}}_{\mathfrak{X}_{n,v}\times\mathcal{T}_{n,v},z} \arrow[r, two heads] \arrow[d, two heads] & \widehat{\mathcal{O}}_{\Delta_v,z} \\
R_{\rho_z,\delta_z} \arrow[ur, dotted] &
\end{tikzcd}$$

To argue that this is true, we use [KPX2014], the point being that the desired dotted line is essentially given by taking a global triangulation of the universal representation on $\mathcal{U}\cap\Delta_v$ for a small open affinoid subdomain $\mathcal{U}$ of $\mathfrak{X}_{n,v}\times\mathcal{T}_{n,v}^{\mathrm{reg}}$ containing z. Indeed, if such a global triangulation actually existed, then we would immediately have the desired dotted line by the universal property of R_{ρ_z,δ_z} (as the local triangulation would give us a $\widehat{O}_{\Delta_v,z}$-valued trianguline deformation[6] of ρ_z of residual parameter δ_z), and the fact that the diagram commutes is then a consequence of the universal property of $\widehat{\mathcal{O}}_{\mathfrak{X}_{n,v}\times\mathcal{T}_{n,v},z}$ (itself being identified with a certain universal deformation ring thanks to Theorem 3.1.2).

However, the results of [KPX2014] are not so strong as to automatically give us such a global triangulation of the universal representation near z. Instead, this triangulation takes place over a rigid space Δ_v' equipped with a proper birational morphism f to $\Delta_v\cap\mathcal{U}$. Indeed, [KPX2014] constructs this $f:\Delta_v'\to\Delta_v\cap\mathcal{U}$ by first taking the normalization and then a series of proper birational morphisms defined locally by birational projective morphisms of schemes (e.g., blowups), and in particular [KPX2014, Corollary 6.3.10] provides a filtration of the global (φ,Γ)-module $\mathbf{D}^{\dagger}_{\mathrm{rig},\Delta_v'}(f^*(\rho^u_{\mathcal{U}\cap\Delta_v}))$ by coherent submodules that become a bona fide global triangulation[7] over the preimage in Δ_v' of the set of points of $\mathcal{U}\cap\Delta_v$ which are trianguline of the given parameter (part of the content of the corollary from [KPX2014] is that this is the set of points of a Zariski open subdomain), and furthermore restrict to the actual unique triangulation of the image under f. The KEY POINT here is that this global triangulation over an open set of Δ_v', thanks to the fact that this open set contains all preimages of z, provides us (again via the universal property of R_{ρ_z,δ_z}) with a morphism

$$R_{\rho_z,\delta_z}\to\prod_{z_i'\in T}\widehat{\mathcal{O}}_{\Delta_v',z_i'},$$

where T is any finite set of preimages of z in Δ_v' (technically in order to do this we need to make sure that the residue fields of all the points are the same, which we can accomplish just by increasing the size of the base field E so that all of the points we are interested in have residue field E). The point is that (again regardless of the choice of $T\neq\emptyset$) the natural diagram

[6] N.B.: here we are also using the fact that the set of points of $\Delta_v\cap\mathcal{U}$ that are actually trianguline is open in $\Delta_v\cap\mathcal{U}$ whenever $\mathcal{U}$ is an affinoid open of the regular locus. This is part of the content of the main theorems of [KPX2014], or, in the context of our hypothetical discussion about the existence of a local triangulation, implicitly part of the assumptions.

[7] Recall that the coherent submodules defininig a bona fide global triangulation must furthermore be direct summands; this is only guaranteed over a particular open set.

$$\begin{tikzcd}
\widehat{\mathcal{O}}_{\mathfrak{X}_{n,v}\times\mathcal{T}_{n,v},z} \arrow[r, two heads] \arrow[d, two heads] & \widehat{\mathcal{O}}_{\Delta_v,z} \arrow[r] & \prod_{z_i'\in T}\widehat{\mathcal{O}}_{\Delta_v',z_i'} \\
R_{\rho_z,\delta_z} \arrow[urr] & &
\end{tikzcd}$$

commutes (easy to check using universal property of $\widehat{\mathcal{O}}_{\mathfrak{X}_{n,v}\times\mathcal{T}_{n,v},z}$ as a deformation ring, from Theorem 3.1.2). To prove the existence of the dotted arrow, it therefore suffices to choose T such that the natural map

$$\widehat{\mathcal{O}}_{\Delta_v,z} \to \prod_{z_i'\in T}\widehat{\mathcal{O}}_{\Delta_v',z_i'}$$

is injective (injective maps of local rings are monomorphisms in the category of local rings).

To do this, we just construct T by starting with the fiber $\{\tilde{z}_1,\ldots,\tilde{z}_m\}$ of z in the normalization of $\mathcal{U}\cap\Delta_v$, and then just arbitrarily choosing a preimage of each $\tilde{z}_i$ (recall from above how Δ_v' was constructed). The point is that

$$\widehat{\mathcal{O}}_{\Delta_v,z} \to \prod_{i=1}^{m}\widehat{\mathcal{O}}_{\widetilde{\Delta_v\cap\mathcal{U}},\tilde{z}_i},$$

where $\widetilde{\Delta_v\cap\mathcal{U}}$ denotes the normalization, is injective (taking normalization is compatible with taking completed stalks, and every reduced local ring injects into its normalization); and the remaining coordinatewise maps

$$\widehat{\mathcal{O}}_{\widetilde{\Delta_v\cap\mathcal{U}},\tilde{z}_i} \to \widehat{\mathcal{O}}_{\Delta_v',z_i'}$$

are injective because a projective birational morphism of reduced schemes is injective on stalks (obvious as the stalks all live in the field of rational functions), and this injectivity is then preserved by taking completions (use [EGA, EGA I, Corollaire 3.9.8], which applies because the stalk of the base scheme has completion which is a domain thanks to the fact that affinoid algebras are excellent[8] and all the schemes we are looking at in between the normalization and Δ_v' are normal by assumption[9]). □

Remark 3.1.13. Note that the hypotheses of [KPX2014, Corollary 6.3.10] include the assumption that the pointwise triangulations you start out with are "strictly trianguline," i.e. unique with the given parameter. This might be slightly weaker than "regular," but in any event without [KPX2014] it seems doubtful that Theorem 3.0.1 can be proved with similar methods without the regularity hypotheses.

Armed with this description of the local trianguline locus, we are able to deduce some strong information about the relationship between the global trianguline locus and the eigenvariety

[8] according to [FvdP2004] this is somewhere in [BGR1984] but I have only found it in [BKKN1967]. Given this reference is in German, no wonder I did not know this useful fact beforehand...

[9] c.f. MSE186547

near nice enough classical points. This uses as a KEY INPUT the main theorem of [NT2020].

Proposition 3.1.14. *Let (π, χ) be an accessibly refined automorphic representation of $G_n(\mathbf{A}_{F^+})$ that has the priviledge of providing a classical point z on the eigenvariety $\mathscr{E}_n$ (this is basically a condition on the level of π). Suppose furthermore that z is non-critical and regular, and that $r_{\pi,\iota}|_{G_{F_{\tilde{v}}}}$ is irreducible for all $v|p$. Then the following hold:*

1. *The global trianguline locus Δ is regular at z.*
2. *$\mathscr{E}_n$ is locally isomorphic to Δ near z via the canonical inclusion of $\mathscr{E}_n$ into $\mathfrak{X}_n \times \mathcal{T}_n$.*

Proof. Define

$$H^1_{\mathrm{tri},\delta_z}(G_{F^+,S}, \mathrm{ad}\widetilde{r_{\pi,\iota}})$$

to be the subset of $H^1(G_{F^+,S}, \mathrm{ad}\widetilde{r_{\pi,\iota}}) = D_{\widetilde{r_{\pi,\iota}}}(\kappa(z)[\varepsilon])$ constisting of elements which are trianguline deformations of $r_{\pi,\iota}|_{G_{F_{\tilde{v}}}}$ when restricted to each $v|p$, i.e. (recall Lemma 3.1.9 and the commutative diagram that follows),

$$H^1_{\mathrm{tri},\delta_z}(G_{F^+,S}, \mathrm{ad}\widetilde{r_{\pi,\iota}}) := \bigcap_{v|p} \mathrm{res}_v^{-1} H^1_{\mathrm{tri},\delta_{z,v}}(G_{\mathbf{Q}_p}, \mathrm{ad} r_{\pi,\iota}|_{G_{F_{\tilde{v}}}}).$$

By Lemma 3.1.9 and Proposition 3.1.12 (using the regularity hypothesis in a crucial way here), the Zariski tangent space to the global trianguline locus Δ at z is contained in $H^1_{\mathrm{tri},\delta_z}(G_{F^+,S}, \mathrm{ad}\widetilde{r_{\pi,\iota}})$ considered as a subspace of $T_z(\mathfrak{X}_n^{\text{p-irr}} \times \mathcal{T}_n^{\mathrm{reg}}) \cong H^1(G_{F^+,S}, \mathrm{ad}\widetilde{r_{\pi,\iota}}) \oplus T_z\mathcal{T}_n^{\mathrm{reg}}$. Now we apply the non-criticality hypothesis to obtain via [BC2009a, Proposition 2.3.4] that the natural map

$$H^1_{\mathrm{tri},\delta_z}(G_{F^+,S}, \mathrm{ad}\widetilde{r_{\pi,\iota}}) \to T_{\delta_z|_{\mathbf{Z}_p^\times}}\mathcal{W}_n$$

is injective (elements of the kernel have triangulation of parameter living in $E \subset E[\varepsilon]$, are hence de Rham at each $\tilde{v}$, and therefore count as elements of[10] $H^1_f(G_{F^+}, \mathrm{ad}\widetilde{r_{\pi,\iota}})$, which vanishes by the main theorem of [NT2020]). It follows from our two observations that

$$\dim T_z\Delta \leq \dim \mathcal{W}_n.$$

But since we have all sorts (Zariski-dense-set worth) of classical points accumulating at z in Δ, there is an affinoid open neighborhood $\mathcal{V} \subset \mathscr{E}_n$ contained in Δ, and hence

$$\dim_z \Delta \geq \dim \mathscr{E}_n.$$

Combining this with the basic fact that $\dim \mathscr{E}_n = \dim \mathscr{W}_n$ and the general fact that $\dim T_z\Delta \geq \dim_z \Delta$, we finally see that Δ is regular at z of the same dimension as $\mathscr{E}_n$. Therefore, in some

[10] The reason we can write "f" instead of "g" here is that $r_{\pi,\iota}$ is automatically generic (any N is automatically zero) thanks to [Car2012]

open affinoid neighborhood $\mathcal{V}'$ of z, Δ has exactly one irreducible component passing through z. We conclude from this and the fact that $\mathcal{V} \to \Delta$ is a closed embedding that $\mathcal{V}$ also has only one irreducible component passing through z, and that $\mathscr{E}_n$ is locally isomorphic to Δ near z. □

Applying this theorem to an irreducible component of $\mathscr{E}_n$ and an irreducible component of $\mathscr{E}_2$ mapping into the same deformation space except with symmetric powers, we may deduce Theorem 3.0.1. This is done in the next section.

3.2 | Passing to definite unitary groups

The first field that we are supposed to construct is this one:

Lemma 3.2.1. *There exists an abelian CM extension $F'/\mathbf{Q}$ such that every prime dividing Np splits in F', $[(F')^+ : \mathbf{Q}]$ is even, and $F'/(F')^+$ is everywhere unramified.*

Proof. At first, I thought that this would be an easy consequence of class field theory, but indeed I found that this proved harder than it first seemed. Instead, the technique of explicit construction using compositum of quadratic fields, which you explained in your office a few weeks ago, works fine (though note that what you said the second time around wasn't quite enough, the point being that F' needs to be *everywhere* unramified over $(F')^+$, not just at the primes above $\ell | Np$; in fact what I ended up doing is closer to what I said than to the Krasner's lemma argument you gave, it's just that I accidentally reversed a divisibility and caused a true thing to seem false).

Our field F' will be of the form $\mathbf{Q}(\sqrt{a}, \sqrt{b})$ where $a > 0$ and $b < 0$ are squarefree. This is nice because it is automatically CM abelian over $\mathbf{Q}$, and the maximal totally real subfield is quadratic extension of $\mathbf{Q}$. We just need to guarantee that $\mathbf{Q}(\sqrt{a})$ and $\mathbf{Q}(\sqrt{b})$ are split over all rational primes $\ell | Np$, and that $\mathbf{Q}(\sqrt{a}, \sqrt{b})$ is everywhere unramified over $\mathbf{Q}(\sqrt{a})$. In particular, it cannot be the case that there are rational primes ℓ that ramify in $\mathbf{Q}(\sqrt{b})$ but not in $\mathbf{Q}(\sqrt{a})$, since then any prime $\mathfrak{p} | \ell$ of $\mathbf{Q}(\sqrt{a})$ would have to ramify in $\mathbf{Q}(\sqrt{a}, \sqrt{b})$. Therefore (modulo shenanigans at 2 that we will be able to ignore by assuming WLOG that N is even), a must be divisible by b (this is the divisibility that I accidentally reversed in your office two days ago). We therefore change notation to $b = -d$ and $a = -dn$, where $d, n > 0$ are positive integers that we will choose.

Let $d > 0$ be a squarefree positive integer such that the Legendre symbol

$$\left(\frac{-d}{\ell} \right) = 1$$

for all primes $\ell | Np$. In fact, to avoid problems at 2, I also ask that $-d \equiv 1 \mod 4$. This is a consistent system of congruences, since at $\ell = 2$ the first condition just says that d is odd. Such a d exists by Sunzi's remainder theorem (together with, say, Dirichlet's theorem on primes in arithmetic progressions). The point is that (by the Dedekind–Kummer criterion) every $\ell | Np$ then splits in $\mathbf{Q}(\sqrt{-d})$. We then choose a squarefree positive integer $n > 0$ such that $n \equiv -1 \mod 4Np$ and $(n, d) = 1$. This is possible again for example by Dirichlet's theorem on primes in arithmetic progression. Then nd satisfies exactly the same congruence conditions that we asked $-d$ to satisfy, and therefore all $\ell | 2Np$ also split in $\mathbf{Q}(\sqrt{nd})$. By the surjectivity of Galois restriction maps restricted to decomposition groups, we conclude that all rational primes $\ell | 2Np$ split in

$$F' := \mathbf{Q}(\sqrt{nd}, \sqrt{-d}).$$

Moreover, we can rewrite F' as a compositum of linearly disjoint quadratic fields with coprime discriminant[11]

$$F' = \mathbf{Q}(\sqrt{-n}, \sqrt{-d}) = \mathbf{Q}(\sqrt{-n}) \cdot \mathbf{Q}(\sqrt{-d}),$$

from which it follows that

$$\Delta_{F'/\mathbf{Q}} = \Delta^2_{\mathbf{Q}(\sqrt{-n})/\mathbf{Q}} \Delta^2_{\mathbf{Q}(\sqrt{-d})}.$$

By the formula for relative discriminants in towers, we conclude that

$$\mathrm{N}_{\mathbf{Q}}^{(F')^+} \Delta_{F'/(F')^+} = \Delta_{F'/\mathbf{Q}} \Delta(F')^+/\mathbf{Q}^{-2} = \frac{n^2 d^2}{n^2 d^2} = 1$$

and hence that F' is everywhere unramified over its maximal totally real subfield, as desired. $\square$

Apologies for all the waffle in the proof of Lemma 3.2.1; as we now both agree, it is a very easy exercise.

Remark 3.2.2. The construction in Lemma 3.2.1 is also interesting because it allows us to construct a large family of real quadratic fields with nontrivial class group (consistent with Cohen–Lenstra).

Now I describe the eigenvariety notation to be used. Let $\mathcal{C}$ be an irreducible component of the cuspidal tame level N Coleman–Mazur eigencurve $\mathscr{E}$, let $\mathcal{W}$ be the usual weight space

$$\mathcal{W} := \mathrm{Hom}((\mathbf{Z}/N\mathbf{Z})^\times \times \mathbf{Z}_p^\times, \mathbf{G}_m(-))$$

in the sense of [Buz2004], and let $W \subset \mathcal{W}$ be the connected component containing the image of $\mathcal{C}$ under the weight map $\kappa : \mathscr{E} \to \mathcal{W}$ (i.e. the one corresponding to the discrete part of the nebentypus-weight-character[12] $\varepsilon_N \varepsilon_p : (\mathbf{Z}/N\mathbf{Z})^\times \times (\mathbf{Z}/q\mathbf{Z})^\times \to \overline{\mathbf{Q}}_p^\times \cong \mathbf{C}^\times$ that all the forms parametrized by $\mathcal{C}$ share). As you remember, at some point I was confused about the following definition due to the strange wording in [NT2021], so I now make it really explicit. Let

[11] the requirement that the discriminants are coprime is where we are happy that we required everything to be split at 2 and $(n, d) = 1$.

[12] Here q is the usual notation, as in [LWX2017], for p when p is odd and 4 when $p = 2$

$\varepsilon_N : (\mathbf{Z}/N\mathbf{Z})^\times \to \overline{\mathbf{Q}}_p^\times \cong \mathbf{C}^\times$ be the constant N-nebentypus of $\mathcal{C}$. Then by the classical theory of Galois representations associated to modular forms, the universal pseudocharacter over $\mathcal{C}$ has determinant varying over $\mathcal{C}$ given by

$$(\det \rho)_z = (\varepsilon_N \circ (G_{\mathbf{Q}} \to \mathrm{Gal}(\mathbf{Q}(\zeta_N)/\mathbf{Q}) \cong (\mathbf{Z}/N\mathbf{Z})^\times)) \cdot (\kappa(z) \cdot (x \mapsto x)) \circ \chi_{\text{cyclo}},$$

where $\chi_{\text{cyclo}} : G_{\mathbf{Q}} \to \mathbf{Z}_p^\times$ is the p-adic cyclotomic character and $z \in \mathcal{C}$. Note that this family $G_{\mathbf{Q}} \to \mathcal{O}(\mathcal{C})^\times$ factors through the weight map κ, as it comes from the family $\chi' : G_{\mathbf{Q}} \to \mathcal{O}(W)^\times$ given by the exact same formula (except $\kappa(z)$ is replaced now by the variable parametrizing W).

Definition 3.2.3. The character $\chi : G_{\mathbf{Q}} \to \mathcal{O}(W)^\times$ is defined to be just like χ', except we get rid of the extra "$x \mapsto x$" to make it cleaner. In other words, χ is $\chi' \cdot (x \mapsto x^{-1}) \circ \chi_{\text{cyclo}} = \chi' \cdot \chi_{\text{cyclo}}^{-1}$

Question 3.2.4. Doesn't this mean there is a typo — in [NT2021, Theorem 2.33] it should say "the determinant of the universal pseudocharacter over $\mathcal{C}$ is $\epsilon\chi$" rather than $\epsilon^{-1}\chi$? Of course I suppose what Newton–Thorne say would be correct if the convention for a classical weight k form z is that $\kappa(z) = x \mapsto x^k$ rather than x^{k-2}.

In the technical constructions that follow, the key point about χ that we will need to use is not really its full definition, but rather that it is unramified away from Np, and globally on W of finite order (hence "potentially unramified") when restricted to the inertia I_ℓ for any $\ell | N$ (in fact the order on inertia is globally on W bounded by $\phi(N)$).

Lemma 3.2.5. *There is a finite étale morphism of rigid spaces $\eta : \widetilde{W} \to W$ and an analytic family of characters $\psi : G_{F'} \to \mathcal{O}(\widetilde{W})^\times$ with the following properties:*

1. *ψ is unramified at all but finitely many places of F'*
2. *For each place $v|p$ of $(F')^+$, ψ is unramified at at least one of the two places $\tilde{v}|v$ of F'*
3. *$\psi\psi^c = \eta^*(\chi|_{G_{F'}})$*

Proof. Define the set $\widetilde{S}_p$ as a subset of the set of places of F' lying over p, by making an arbitrary choice of one out of the two $\tilde{v}|v$ for each place $v|p$ of $(F')^+$. Therefore, $\{\tilde{v}|p\} = \widetilde{S}_p \sqcup \widetilde{S}_p^c$. We will satisfy (2) by constructing ψ to be unramified at all places in $\widetilde{S}_p$. Now define the family of continuous characters

$$L : \prod_{\tilde{v}|p} \mathcal{O}_{F'_{\tilde{v}}}^\times \to \mathcal{O}(W)^\times$$

by

$$(u_{\tilde{v}})_{\tilde{v}|p} \mapsto \prod_{\tilde{v} \in \widetilde{S}_p^c} \chi|_{G_{F'}}^{-1} \circ \mathrm{Art}_{F'_{\tilde{v}}}(u_{\tilde{v}}).$$

This is clearly the type of thing that we want, but we need to extend it to an analytic family of Hecke characters in order to extract a Galois character by class field theory. We show this is possible (up to possible finite étale cover of W) by using the theory of rigid character varieties established in [Buz2004]. If $u \in \mathcal{O}_{(F')^+}^\times \subset \prod_{\tilde{v}|p} \mathcal{O}_{F'_{\tilde{v}}}^\times$, for any $\tilde{v}|p$, the two local components $u_{\tilde{v}} \in \mathcal{O}_{F'_{\tilde{v}}}^\times, u_{\tilde{v}^c} \in \mathcal{O}_{F'_{\tilde{v}^c}}^\times$ both come (via the canonical embeddings/isomorphisms) from the same element $u_v \in \mathcal{O}_{(F')^+_v}^\times$, where v is the place of $(F')^+$ below $\tilde{v}$ and $\tilde{v}^c$ (in particular u_v is the image of u under the canonical embedding into $(F')^+_v$). By the compatibility of Galois restriction with field norms under local class field theory, and the fact that all places $v|p$ of $(F')^+$ are split in F', we have

$$\begin{aligned}
L(u) &= \prod_{\tilde{v} \in \widetilde{S}_p^c} \chi(\mathrm{Art}_{F'_{\tilde{v}}}(u_{\tilde{v}}))^{-1} \\
&= \prod_{\tilde{v} \in \widetilde{S}_p^c} \chi\left(\mathrm{Art}_{(F')^+_v}\left(\mathrm{N}_{(F')^+_v}^{F'_{\tilde{v}}} u_{\tilde{v}}\right)\right)^{-1} \\
&= \prod_{v|p} \chi(\mathrm{Art}_{(F')^+_v}(u_v))^{-1} \\
&= \prod_{v|p} \chi\left(\mathrm{Art}_{\mathbf{Q}_p}\left(\mathrm{N}_{\mathbf{Q}_p}^{(F')^+_v} u_v\right)\right)^{-1} \\
&= \chi\left(\mathrm{Art}_{\mathbf{Q}_p}\left(\prod_{v|p} \mathrm{N}_{\mathbf{Q}_p}^{(F')^+_v} u_v\right)\right)
\end{aligned}$$

(in all of this there is ambiguity up to conjugation by an element of $G_{\mathbf{Q}}$ for the decomposition group of $G_{\mathbf{Q}}$ in which the local Artin map lands, but it doesn't matter since χ is a character). Therefore, if $u \in \mathcal{O}_{(F')^+}^\times$ satisfies $\mathrm{N}_{\mathbf{Q}}^{(F')^+} u = 1$, then $L(u) = 1 \in \mathcal{O}(W)^\times$ (the product of local norms over p is just the p-component of the global norm). The kernel of $\mathrm{N}_{\mathbf{Q}}^{(F')^+}|_{\mathcal{O}_{(F')^+}^\times}$ being a finite-index subgroup of $\mathcal{O}_{(F')^+}^\times$ (since the image is contained in the finite group $\mathcal{O}_{\mathbf{Q}}^\times = \{\pm 1\}$), which itself is a finite-index subgroup of $\mathcal{O}_{F'}^\times$ by Dirichlet's unit theorem ($F'/\mathbf{Q}$ is CM), is therefore a finite-index subgroup of $\mathcal{O}_{F'}^\times$ contained in the kernel of L. By Chevalley's theorem on congruence subgroups [Che1951, Théorème 1] (the same one I used in my letter to you back in October), there is an ideal $\mathfrak{m} = \prod_{w<\infty} \mathfrak{p}_w^{m_w}$ of F' (from now on w always ranges over places of F') such that

$$\ker L \supset \mathscr{U}_{\mathfrak{m}} \cap \mathcal{O}_{F'}^\times,$$

where $\mathscr{U}_{\mathfrak{m}} \subset \mathbf{A}_{F'}^\times$ is the compact open subgroup

$$\mathscr{U}_{\mathfrak{m}} := \left(\prod_{\substack{w<\infty \\ m_w>0}} (1 + \pi_w^{m_w} \mathcal{O}_{F'_w})\right) \times \left(\prod_{\substack{w<\infty \\ m_w=0}} \mathcal{O}_{F'_w}^\times\right) \times \left(\prod_{w|\infty} F_w^\times\right).$$

Let $\mathscr{U}^p$ be the away-from-p part of $\mathscr{U}_{\mathfrak{m}}$ (defined the same way except for $w|p$ the coordinates must all be 1), and define the profinite groups

$$H' = \left(\overline{\mathscr{U}_{\mathfrak{m}} \cap \mathcal{O}_{F'}^{\times}}\right)\diagdown\left(\prod_{w|p} \mathcal{O}_{F'_w}^{\times}\right)$$
$$H = \left((F')^{\times}\diagdown \mathbf{A}_{F'}^{\times}\right)\diagup \mathscr{U}^p.$$

Both H and H' are abelian profinite groups that admit a finite direct sum of copies of $\mathbf{Z}_p$ as an open subgroup: indeed, this is already true of $\prod_{w|p} \mathcal{O}_{F'_w}^{\times}$, so it is true of H'; and H' is embedded in H as a closed subgroup via the canonical embedding $\prod_{w|p} \mathcal{O}_{F'_w} \to \mathbf{A}_{F'}^{\times}$ (which is well defined after taking quotients because elements of $\mathscr{U}_{\mathfrak{m}} \cap \mathcal{O}_{F'}^{\times} \subset \prod_{w|p} \mathcal{O}_{F'_w}^{\times} \subset \mathbf{A}_{F'}^{\times}$ are in $\mathcal{O}_{F'}^{\times} \cdot \mathscr{U}^p$), but in fact H' has finite index in H, by the finiteness of ray class groups (see e.g. [Lan1994]), so it is an open subgroup of H — in particular we have shown that H and H' are both abelian profinite groups with $\mathbf{Z}_p^{\oplus n}$ as an open subgroup, i.e. they are both of the form $\mathbf{Z}_p^{\oplus n} \times$ (finite abelian group). So by [Buz2004, Lemma 2(iv)], the restriction map

$$r : \mathrm{Hom}(H, \mathbf{G}_m(-)) \to \mathrm{Hom}(H', \mathbf{G}_m(-))$$

is a finite étale morphism of rigid spaces. The point of what we did above was that our original family of characters $L : \prod_{w|p} \mathcal{O}_{F'_w}^{\times} \to \mathcal{O}(W)^{\times}$ vanishes on $\overline{\mathscr{U}_{\mathfrak{m}} \cap \mathcal{O}_{F'}^{\times}}$, and therefore factors through to a W-family of characters of H', i.e. a rigid analytic morphism $W \to \mathrm{Hom}(H', \mathbf{G}_m(-))$. Taking the fibered product with r, we get the pullback square

$$\begin{tikzcd} W' \arrow[r, "\eta"] \arrow[d] & W \arrow[d] \\ \mathrm{Hom}(H, \mathbf{G}_m(-)) \arrow[r, "r"] & \mathrm{Hom}(H', \mathbf{G}_m(-)) \end{tikzcd}$$

the base-changed morphism η being finite étale because r is. The base-changed morphism $W' \to \mathrm{Hom}(H, \mathbf{G}_m)$ is then an extension of L to all of H, in the sense that it defines (by universal property of these Hom spaces) a continuous W'-family of characters

$$L' : \mathbf{A}_{F'}^{\times} \to \mathcal{O}(W')^{\times}$$

which is 1 on $(F')^{\times}$ and $\mathscr{U}^p$, and satisfies

$$L'|_{\prod_{w|p} \mathcal{O}_{F'_w}^{\times}} = \eta^*(L).$$

By global class field theory, L' gives rise to a Galois character

$$\lambda : G_{F'} \to \mathcal{O}(W')^{\times}.$$

The conjugate

$$\lambda^c : G_{F'} \to \mathcal{O}(W')^\times$$

can also be given via class field theory from $L' \circ c$. Without loss of generality, we shrink the modulus $\mathfrak{m}$ so that m_v is invariant under c, hence we can assume that $L' \circ c$ is still trivial on $\mathscr{U}^p$. By local-global compatibility of class field theory, and the definition of L, we conclude that

$$\eta^*(\chi|_{G_{F'}} \circ \mathrm{Art}_{F'}) \cdot L' \cdot (L' \circ c)|_{\prod_{w|p} \mathcal{O}_{F'_w}^\times} = 1.$$

Moreover, since χ is (globally on W) unramified away from Np, and all the ramification over primes dividing N comes from a ray class character modulo N (the N-part of the nebentypus), we can furthur shrink $\mathfrak{m}$ so that χ is also trivial on $\mathscr{U}^p$, and hence the character

$$\eta^*(\chi|_{G_{F'}} \circ \mathrm{Art}_{F'}) \cdot L' \cdot (L' \circ c) \to \mathcal{O}(W')^\times$$

is trivial on $(F')^\times$, $\mathscr{U}^p$, and $\prod_{w|p} \mathcal{O}_{F'_w}^\times$.By finiteness of ray class groups and class field theory, we have concluded that $\eta^*(\chi|_{G_{F'}})\lambda\lambda^c$ is globally on W' of finite order. Since the character variety of a finite group is discrete, by replacing W' with any connected component W' (doesn't change the fact that $\eta|_{W'}$ is finite étale since W is connected), we can assume that $\eta^*(\chi|_{G_{F'}})\lambda\lambda^c$ is actually the constant family given by a finite-order character $G_{F'} \to (E')^\times$ (for some finite $E'/\mathbf{Q}_p$). In other words, λ^{-1} does the trick, up to the trivial W'-family corresponding to some finite-order character. It also follows from its definition and class field theory that λ is unramified almost everywhere and unramified at each place in $\widetilde{S}_p$. So we are reduced to the constant case, where this becomes a more straightforward exercise in class field theory: Newton–Thorne cite something somewhat more general in [BLGGT2014], but I will just do it.

To ease the notation, let us relabel $\omega := \eta^*(\chi|_{G_{F'}})\lambda\lambda^c : G_{F'} \to (E')^\times$ (abusing notation to replace the constant family over W' by the single field-valued character it is pulled back from). Since G'_F has index 2 in $G_{(F')^+}$, we can extend ω to $\widetilde{\omega} : G_{(F')^+} \to (E')^\times$ by setting $\omega(c) = 1$. This corresponds via class field theory to the Hecke character $\widetilde{\omega} \circ \mathrm{Art}_{(F')^+}$, and the fact that $\widetilde{\omega}$ extends ω implies (by compatibility via class field theory of Galois restriction and field norms) that

$$\omega \circ \mathrm{Art}_{F'} = \left(\widetilde{\omega} \circ \mathrm{Art}_{(F')^+}\right) \circ \mathrm{N}^{F'}_{(F')^+}.$$

In particular, if we can extend the finite-order (hence ray-class) character $\widetilde{\omega} \circ \mathrm{Art}_{(F')^+}$ to a ray-class character $\widetilde{\widetilde{\omega}}$ on $\mathbf{A}_F^\times$ which is unramified at each place in $\widetilde{S}_p$, then this simply translates to

$$\omega \circ \mathrm{Art}_{F'} = \widetilde{\widetilde{\omega}} \cdot \left(\widetilde{\widetilde{\omega}} \circ c\right)$$

and so the Galois character corresponding to $\widetilde{\widetilde{\omega}}$ finishes the job (after further base-changing W' to the coefficient field of this extension, which of course might have to be bigger than E'). It remains to see that we can actually extend $\widetilde{\omega} \circ \mathrm{Art}_{(F')^+}$ in such a fashion. This is explained in [HSBT2010, Lemma 2.2] I think, but it is not hard anyway: the point is that

$$(F')^\times \cdot \mathbf{A}^\times_{(F')^+} \cdot \prod_{w|\infty} (F')^\times_w \cdot \prod_{w|p} \mathcal{O}^\times_{F'_w} \cdot \mathscr{K}^p,$$

where $\mathscr{K}^p$ is a compact open of $\mathbf{A}^{p,\infty,\times}_{F'}$ such that $\mathscr{K}^p \cap \mathbf{A}_{(F')^+} = \mathrm{N}^{F'}_{(F')^+}\mathscr{U}^p$, is a finite-index subgroup of $\mathbf{A}^\times_{F'}$, to which $\widetilde{\omega} \circ \mathrm{Art}_{(F')^+}$ can be extended via the requirement of being trivial on $(F')^\times$, $\prod_{w|\infty}(F')^\times_w$, $\prod_{w|p} \mathcal{O}^\times_{F'_w}$, and $\mathscr{K}^p$ because all of four of these sets, when intersected with $\mathbf{A}^\times_{(F')^+}$, give something on which $\widetilde{\omega} \circ \mathrm{Art}_{(F')^+}$ is trivial. We now have several choices of extension $\widetilde{\widetilde{\omega}}$ all the way up to $\mathbf{A}^\times_{F'}$ (if we allow coefficients in some finite E''/E), and thanks to what we just did, any one of those is guaranteed to produce a Galois character $\varphi : G_{F'} \to (E'')^\times$ which is unramified at the places over p and almost all of the other places, such that $\eta^*(\chi|_{G_{F'}})\lambda\lambda^c = \varphi\varphi^c$ (where now $\eta : \widetilde{W} := W' \times_{\mathbf{Q}_p} E'' \to W$ is the final choice of finite étale morphism), and hence $\varphi\lambda^{-1}$ is the desired character ψ. □

Remark 3.2.6. The constructions made in the papers cited by Newton–Thorne in the proof above are more complicated than what we did, because they need to account for the full detail of Weil's construction of p-adic Hodge–Tate Galois character corresponding to algebraic Hecke character. But in our situation, all of the Hodge–Tate weights are zero, so no procedure of "transfer of algebraic weights from ∞ to p" was necessary for us and the constructions were more transparent.

We retain the notation ψ for the character of Lemma 3.2.5.

Lemma 3.2.7. *There exists a solvable Galois CM extension $F/\mathbf{Q}$ containing F' such that*

1. *F/F^+ is everywhere unramified*
2. *Every rational prime dividing Np splits in F*
3. *$\psi|_{G_F}$ is unramified at every finite place of F not lying over Np.*

Proof. This one I prove using class field theory rather than technique of explicit construction. By its construction, ψ has the property that for all finite places w not lying over p, there is a (finite-index) open subgroup $H_w \subset I_w$ such that $\psi|_{H_w} = 1$. Let S be the finite set of places w of F' not lying over Np such that $H_w \neq I_w$. By class field theory, there exists an abelian extension F'' of $(F')^+$ which is split over all the infinite places of $(F')^+$ and over all the finite places lying over any rational prime $\ell | Np$, and such that $F := F'' \cdot (F')^+$ has inertia group at places above w contained in H_w for all $w \in S$ (use local-global compatibility of class field theory, continuity of idèle norm, and [AT2009, Ch. X, Theorem 4]; in particular note that the set of local conditions we have asked for is finite, since ψ is a priori only ramified at finitely many places). Then F

satisfies (1) [formal consequence of the same property for $F'/(F')^+$ since F'', F' are linearly disjoint] (2) [primes that split in two fields split in the compositum] and (3) [by construction]. It is also a CM extension [this was the point of asking $F''/(F')^+$ is split at infinity, i.e. that F'' is totally real].

However, there is no reason for the extension $F/\mathbf{Q}$ we have constructed to be Galois. We fix this by replacing F'' with its Galois closure over $\mathbf{Q}$. Since splitting types are unchanged upon taking the conjugate field by elements of $G_{\mathbf{Q}}$, $F''/(F')^+$, and since the new F'' contains the old one, the ramification properties above still hold, with $F''/\mathbf{Q}$ and hence $F/\mathbf{Q}$ Galois. In fact, $F/\mathbf{Q}$ is solvable. The reason is that $F''/(F')^+$ is a compositum of various $G_{\mathbf{Q}}$-conjugates of the old F'', which is abelian over $(F')^+$ since each of those conjugates are. This concludes. $\square$

The three lemmas above provide the base-change infrastructure we need to do the analytic continuation arguments in [NT2021, Theorem 2.33]. As usual, let S be the set of primes of F^+ dividing Np, and S_p the subset of those lying over p. For $v \in S$, denote $\tilde{v}, \tilde{v}^c$ the two places of F above v. We have the usual rigid space

$$\mathcal{T}_2 := \mathrm{Hom}((F_{\tilde{v}}^\times)^2, \mathbf{G}_m(-))$$

that is used to parametrize parameters of triangulations of 2-dimensional conjugate self-dual representations. Similarly, let

$$\mathcal{T}_0 := \mathrm{Hom}(\mathbf{Q}_p^\times/\mathbf{Z}_p^\times, \mathbf{G}_m(-))$$

be the relevant space of parameters for the Coleman–Mazur eigencurve $\mathscr{E}_0$ of tame level N. In reality, we will base-change by the finite étale morphism η and work with

$$\widetilde{\mathcal{T}_0} := \mathcal{T}_0 \times_W \widetilde{W}$$

in order to accommodate the family of characters constructed in Lemma 3.2.5. Finally, let $\mathfrak{X}_0$ be the usual $\mathbf{Q}_p$-rigid space parametrizing 2-dimensional pseudocharacters of $G_{\mathbf{Q}}$ unramified outside Np (meaning that the inertia outside Np is contained in the kernel defined in the usual way, or, what is equivalent, pseudocharacters of $G_{\mathbf{Q},Np}$), and let $\mathfrak{X}_2$ be the usual $\mathbf{Q}_p$-rigid space parametrizing conjugate-self-dual 2-dimensional pseudocharacters of $G_{F,S}$. The point is that we then have a well-defined morphism (Galois restriction, as should correspond to base change on the automorphic side)

$$b : \mathfrak{X}_0 \times_{\mathbf{Q}_p} \widetilde{\mathcal{T}_0} \to \mathfrak{X}_2 \times_{\mathbf{Q}_p} \mathcal{T}_2$$

given by

$$(\tau, \mu) \mapsto \left(\tau|_{G_{F,S}} \otimes \psi_{\tilde{\kappa}(\mu)}^{-1}, \left(\mu_v \circ \mathrm{N}_{\mathbf{Q}_p}^{F_{\tilde{v}}} \cdot \psi_{\tilde{\kappa}(\mu)}^{-1}|_{G_{F_{\tilde{v}}}} \circ \mathrm{Art}_{F_{\tilde{v}}} \right)_{v \in S_p} \right).$$

The norms on the side of μ being due to the fact that norms correspond under class field theory to Galois restriction.

Anyhow, to conclude the analysis, we need to have a full understanding of triangulations of 2-dimensional Galois representations.

Lemma 3.2.8. *Let $\rho : G_{\mathbf{Q}} \to \mathrm{GL}_2(\overline{\mathbf{Q}}_p)$ be a de Rham representation with distinct Hodge–Tate weights $k_1 < k_2$ with $\mathrm{WD}(\rho)^{ss} \cong \chi_1 \oplus \chi_2$, with $\chi_i : W_{\mathbf{Q}_p} \to \overline{\mathbf{Q}}_p^\times$ distinct continuous characters. In this situation, there are only finitely many triangulations of ρ. In fact:*

1. *If ρ is not potentially cristalline, then it has only one triangulation. There is a canonical way to choose the ordering of χ_1 and χ_2 so that this triangulation is numerically non-critical of parameter*
$$(x^{-k_1}\chi_1 \circ \mathrm{Art}_{\mathbf{Q}_p}, x^{-k_2}\chi_2 \circ \mathrm{Art}_{\mathbf{Q}_p}).$$

2. *If ρ is potentially crystalline and indecomposable, then ρ has two triangulations, of parameters*
$$(x^{-k_1}\chi_1 \circ \mathrm{Art}_{\mathbf{Q}_p}, x^{-k_2}\chi_2 \circ \mathrm{Art}_{\mathbf{Q}_p})$$
and
$$(x^{-k_1}\chi_2 \circ \mathrm{Art}_{\mathbf{Q}_p}, x^{-k_2}\chi_1 \circ \mathrm{Art}_{\mathbf{Q}_p}).$$
Both are numerically non-critical

3. *If ρ is decomposable as $\rho \cong \psi_1 \oplus \psi_2$, where ψ_1, ψ_2 are Hodge–Tate of weights k_1, k_2, then ρ has two triangulations: a critical one of parameter*
$$(x^{-k_2}\chi_2 \circ \mathrm{Art}_{\mathbf{Q}_p}, x^{-k_1}\chi_1 \circ \mathrm{Art}_{\mathbf{Q}_p})$$
and a non-critical one of parameter
$$(x^{-k_1}\chi_1 \circ \mathrm{Art}_{\mathbf{Q}_p}, x^{-k_2}\chi_2 \circ \mathrm{Art}_{\mathbf{Q}_p}).$$

Proof. In all of the cases, the point is to consider the filtered $(\varphi, N, G_{\mathbf{Q}_p})$-module
$$\mathbf{D}_{\mathrm{pst}}(\rho) \cong \mathbf{D}_{\mathrm{pst}}(\mathbf{D}^\dagger_{\mathrm{rig}}(\rho)),$$
and use the fact that a triangulation of ρ is the same as a filtration of this by subobjects in the category of filtered $(\varphi, N, G_{\mathbf{Q}_p})$-modules [Ber2008], i.e. a choice of 1-dimensional subobject. As is typical, the key point of the analysis will be to sort out the cases when this subobject is weakly admissible (i.e. comes from a subrepresentation of ρ), which is why the hypotheses are organized via reducibility phenomena of ρ.

In the situation that ρ is not potentially cristalline, the monodromy operator N is nontrivial. Therefore, in $\mathbf{D}_{\text{pst}}(\rho)$, there is exactly one subobject, namely $\ker N$. Moreover (now I carry out standard exercise in basic theory of Weil–Deligne representations), the fact that $(r, N) := \mathrm{WD}(\rho)^{\text{Frob-ss}} = \chi_1 \oplus \chi_2$ with nontrivial N means that (letting e_1, e_2 be spanning vectors for the lines acted on by χ_1, χ_2, well-defined up to scalars as χ_1, χ_2 are distinct, and using $N \neq 0$ to assume WLOG that $N(e_2) \neq 0$ — this is what provides the canonical ordering of χ_1, χ_2)

$$r(g)N(e_2) = p^{-\alpha(g)}\chi_2(g)N(e_2)$$

by definition of what a Weil–Deligne representation is. In particular, $W_{\mathbf{Q}_p}$ acts by scalars on $N(e_2)$, so since N is nilpotent, we can assume (by scaling appropriately) $N(e_2) = e_1$, and $\chi_1 = \chi_2 \cdot (|\cdot|_{\mathbf{Q}_p} \circ \mathrm{Art}_{\mathbf{Q}_p}^{-1})$. In particular, φ is already semisimple since it has distinct eigenvalues[13]

$$\varphi_i = \chi_i(\mathrm{Art}_{\mathbf{Q}_p}(p)), \qquad i = 1, 2.$$

In fact, $\varphi_2 = p\varphi_1$. Since $N \neq 0$, $\mathbf{D}_{\text{pst}}(\rho)$ has just a single nonzero subobject, namely the one spanned by e_1. The fact that $\mathbf{D}_{\text{pst}}(\rho)$ is an admissible filtered $(\varphi, N, G_{\mathbf{Q}_p})$-module is an additional condition on the relationship between φ and the Hodge–Tate weights. The slopes of the Newton polygon are just $v_p(\varphi_1)$ and $v_p(\varphi_2) + 1$ (in that order), and the slopes of the Hodge polygon are k_1, k_2 (in that order). Since $\mathbf{D}_{\text{pst}}(\rho)$ is weakly admissible, the endpoints of the Hodge and Newton polygons have to agree, i.e.

$$k_1 + k_2 = 2v_p(\varphi_1) + 1.$$

This implies $v_p(\varphi_1) < k_2$, hence (see [NT2021, Lemma 2.7, 2.8]) the triangulation we have constructed (the one given by the subobject spanned by e_1) is numerically non-critical of the claimed parameter.

Now we consider the situation where ρ is potentially cristalline, i.e. $N = 0$, and irreducible. In this situation, $\mathbf{D}_{\text{pst}}$ has exactly two subobjects[14], namely the two lines on which φ acts by $\chi_1(\mathrm{Art}_{\mathbf{Q}}(p)), \chi_2(\mathrm{Art}_{\mathbf{Q}}(p))$. As before, call these two lines $\overline{\mathbf{Q}}_p e_1, \overline{\mathbf{Q}}_p e_2$. If ρ is irreducible, then neither one of these subobjects get to be weakly admissible (else ρ would have a subobject). Since they are one-dimensional, not being weakly admissible is the same as having Newton strictly BELOW Hodge, i.e.

$$v_p(\varphi_i) < k_i' \leq k_2, \qquad i = 1, 2,$$

[13] I won't make it explicit ever again, but this stuff comes from the definition of $\mathrm{WD}(\rho)$ in terms of $\mathbf{D}_{\text{pst}}$.

[14] Here we are using the fact from the elementary theory of Weil–Deligne representations, namely that being Frobenius semisimple can be checked on a single lift of Frobenius; if we wanted we could have used this last time as well.

where k_i' is the Hodge–Tate weight of $\overline{\mathbf{Q}}_p e_i$. In other words, both of the triangulations are numerically non-critical, and so they have exactly the claimed parameters.

If ρ is potentially cristalline and reducible, but not indecomposable, the same type of analysis works, but we need to include an extra step to rule out the possibility that one of the two triangulations is critical (which of course is necessary to even know what the parameter is, since we need to know what order the Hodge–Tate weights come in). The key point is that exactly one of the subobjects $\overline{\mathbf{Q}}_p e_1$ and $\overline{\mathbf{Q}}_p e_2$ are weakly admissible (if both were, then ρ would be the direct sum of two subobjects, and if none were, then ρ would be irreducible). Without loss of generality, suppose that $\overline{\mathbf{Q}}_p e_1$ is the weakly admissible one. Then the same argument as before shows that the triangulation $\overline{\mathbf{Q}}_p e_2$ is noncritical. In other words,

$$\mathrm{Fil}^{k_1+1}\mathbf{D}_{\mathrm{pst}}(\rho) \cap \overline{\mathbf{Q}}_p e_2 = 0.$$

To prove that $\overline{\mathbf{Q}}_p e_1$ is noncritical, we need to prove that

$$\mathrm{Fil}^{k_1+1}\mathbf{D}_{\mathrm{pst}}(\rho) \cap \overline{\mathbf{Q}}_p e_1 = 0.$$

But if this intersection were nonzero, then we would have

$$\mathrm{Fil}^{k_1+1}\mathbf{D}_{\mathrm{pst}}(\rho) = \overline{\mathbf{Q}}_p e_1,$$

and thus

$$\mathbf{D}_{\mathrm{pst}}(\rho) = \overline{\mathbf{Q}}_p e_1 \oplus \overline{\mathbf{Q}}_p e_2$$

as an object in the category of filtered $(\varphi, N, G_{\mathbf{Q}_p})$-modules. But since $\mathbf{D}_{\mathrm{pst}}(\rho)$ is weakly admissible, and $\overline{\mathbf{Q}}_p e_1$ is weakly admissible, the same is true of $\overline{\mathbf{Q}}_p e_2$, a contradiction (see [Con1999a, Proposition 8.2.10]). We conclude that both triangulations are noncritical and of the desired form.

In the last case (when ρ is decomposable), we have the same two triangulations, and since $\mathbf{D}_{\mathrm{pst}}$ (including its Hodge filtration) is compatible with direct sums, its clear that $\overline{\mathbf{Q}}_p e_1$ is the triangulation with Hodge–Tate weight k_1 (i.e. the noncritical one) and $\overline{\mathbf{Q}}_p e_2$ is the one with Hodge–Tate weight k_2 (i.e. the critical one). □

The point of Lemma 3.2.8 is that we will use it to show that all triangulations of the symmetric power are noncritical (this property is preserved under taking symmetric powers by [Che2011, Example 3.26]) and therefore be able to apply the classicality result Proposition 2.2.5. Anyhow, we now finally restate and prove Theorem 3.0.1

Theorem 3.2.9 ([NT2021], Theorem 2.33)**.** *Fix an isomorphism $\iota : \overline{\mathbf{Q}}_p \cong \mathbf{C}$. Let $(\pi_0, \chi_0), (\pi_0', \chi_0')$ be accessibly refined automorphic representations of $GL_2(\mathbf{A}_{\mathbf{Q}})$, and let $n \geq 2$. Let z_0, z_0' be the corresponding classical points on $\mathscr{E}_0$. Suppose the following are true:*

1. *χ_0 is noncritical and n-regular.*

2. *χ_0' is n-regular.*

3. *The Zariski closures of $r_{\pi_0,\iota}(G_{\mathbf{Q}_p})$ and $r_{\pi_0,\iota}(G_{\mathbf{Q}_p})$ contain SL_2.*

4. *z_0, z_0' lie on the same irreducible component of $\mathscr{E}_0 \times_{\mathbf{Q}_p} \mathbf{C}_p$.*

Then automorphicity of $\mathrm{Sym}^{n-1} r_{\pi_0,\iota}$ implies automorphicity of $\mathrm{Sym}^{n-1} r_{\pi_0',\iota}$.

Proof. The point is to use the diagram of rigid spaces that we have constructed above,

$$\begin{tikzcd}
\mathfrak{X}_0 \times \widetilde{\mathcal{T}}_0 \arrow[r, "b"] & \mathfrak{X}_2 \times \mathcal{T}_2 \\
\mathscr{E}_0 \times_{\mathcal{T}_0} \widetilde{\mathcal{T}}_0 \arrow[u, "\tilde{i}:=i\times\mathrm{id}"] & \mathscr{E}_2 \arrow[u, "i_2"]
\end{tikzcd}$$

The two points z_0, z_0' start out on the same irreducible component of the bottom-left corner (technically we need to be careful about lifting to an irreducible component of the base-change of the irreducible component to $\widetilde{\mathcal{T}}_0$ but this kind of thing is taken care of by Brian Conrad's general theory of irreducible components of rigid spaces). We want to apply [NT2021, Corollary 2.28] (the one that makes the analytic continuation work for symmetric powers from rank-2 definite unitary group to rank-2). To do this, we need a guarantee that $b \circ \tilde{i}(z_0), b \circ \tilde{i}(z_0)$ lie on the same irreducible component of $\mathscr{E}_2$. This is immediate by hypothesis (4), as long as the irreducible component $\mathscr{C} \subset \mathscr{E}_0 \times_{\mathcal{T}_0} \widetilde{\mathcal{T}}_0$ containing z_0, z_0' is mapped to $i_2(\mathscr{E}_2)$ under $b \circ \tilde{i}$. This is in turn is essentially base change (twisted by $\psi_{\kappa(z)}$) from $\mathrm{GL}_2(\mathbf{A}_{\mathbf{Q}})$ to $\mathrm{GL}_2(\mathbf{A}_F)$, followed by descent from $\mathrm{GL}_2(\mathbf{A}_F)$ to $G_2(\mathbf{A}_{F^+})$, where G_2 is the rank-2 definite unitary group that comes from the data we defined above. There are two wrinkles in this that must be smoothed over:

- This base change/descent argument only works when given an actual automorphic representation, i.e., only on classical points on $\mathcal{C}$. But this is okay because the classical points Z_0 are Zariski dense: so if we show that $\tilde{i} \circ b(Z_0) \subset i_2(\mathscr{E}_2)$, we know the same is true of $\tilde{i} \circ b(\mathcal{C})$ (i_2 is a closed embedding).

- Let $z \in Z_0$ come from an accessibly refined automorphic representation (π, χ). The base-change step
$$\pi \rightsquigarrow \pi_F \otimes \iota\psi_z^{-1}$$
is okay because $F/\mathbf{Q}$ is solvable. To do the descent step, however, we need an assurance that π_F is cuspidal, i.e. that $r_{\pi,\iota}|_{G_F}$ is irreducible. This is immediate from hypothesis (3). Also, the construction of ψ ensures that $\pi_F \otimes \iota\psi_z^{-1}$ is conjugate-self dual, which of course is necessary to apply descent (the theorem being cited is [Lab2011, Théorème 5.4]).

The inclusion $b \circ \tilde{i}(\mathcal{C}) \subset i_2(\mathscr{E}_2)$ having now been established, and $b \circ \tilde{i}(z_0), b \circ \tilde{i}(z_0')$ lying on the same irreducible component of $\mathscr{E}_2$, we can use [NT2021, Corollary 2.28] (the analytic continuation result for definite unitary groups, which uses the vanishing of the adjoint Selmer group and the smoothness argument I explained a few weeks ago), in conjunction with Lemma 3.2.8 (which shows explicitly that all triangulations of $r_{\pi_0,\iota}$ are noncritical, thanks to hypothesis (3) which guarantees that the last case in Lemma 3.2.8 cannot happen) to see that π_0' has a symmetric power lift π_n', an automorphic representation for $G_n(\mathbf{A}_{F^+})$. In fact, π_n' and any finite base change thereof is again cuspidal, thanks again to the irreducibility of the symmetric power Galois representation ensured by hypothesis (3). It follows that we can base change π_n' to $\mathrm{GL}_n(\mathbf{A}_F)$ to get a cuspidal automorphic representation, thus being able use soluble descent (the exact meaning of "soluble descent" to be made more precise in the next section) to $\mathrm{GL}_n(\mathbf{A}_\mathbf{Q})$. After untwisting by the appropriate power of $\psi_{z'}$, we obtain the desired automorphicity. $\square$

3.3 | Bonus details on base-change

In the previous section, there are some potentially confusing points regarding descent and base change, which we clear up here.

- The brand-new result of Clozel and Rajan [CR2021] is not actually used — only the cyclic prime order case, which is already correct in [AC1989].
- If we follow to the letter the arguments of [BLGHT2011], the proof uses the fact that all 2-dimensional Galois representations are essentially self-dual. At first this may seem to be an obstacle to extending the techniques of Newton–Thorne to the case of $\mathrm{GL}(3)$, but in fact the added irreducibility hypotheses that apply in our setting make this unnecessary (of course it forces us to rely on the omnipresent "big image" hypotheses of [NT2021]).

Still, we prove the second bullet point since it is an interesting fact.

Lemma 3.3.1. *Let K be an arbitrary number field. Then for any continuous representation*

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

ρ is essentially self-dual. In fact,

$$\rho \cong \rho^\vee \otimes \det \rho.$$

Proof. This is thanks to the identity

$$\begin{pmatrix} & 1 \\ -1 & \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} & 1 \\ -1 & \end{pmatrix}^{-1} = (ad - bc) \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{\top,-1},$$

which shows that the change of basis

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

provides the desired isomorphism. Indeed, the result is true for any 2-dimensional representation of any group with coefficients in any field. $\square$

Let ι be a fixed choice of isomorphism $\iota : \overline{\mathbf{Q}}_p \to \mathbf{C}$, and let $F/\mathbf{Q}$ be a solvable CM extension. At the end of the proof of Theorem 3.0.1, we were armed with the following data:

- A regular algebraic cuspidal automorphic representation π_0 of $\mathrm{GL}_2(\mathbf{A}_\mathbf{Q})$, satisfying the property that $\mathrm{Sym}^{n-1} r_{\pi_0,\iota}|_{G_F}$ is irreducible.
- A regular algebraic cuspidal conjugate self-dual automorphic representation $\widetilde{\pi_n}$ of $\mathrm{GL}_n(\mathbf{A}_F)$ (obtained by a bunch of steps of descent and base change between GL_m and rank-m definite unitary group, $m \in \{0, n\}$) satisfying the property that

$$r_{\widetilde{\pi_n},\iota} \cong \mathrm{Sym}^{n-1} r_{\pi_0,\iota}|_{G_F},$$

i.e. $\mathrm{Sym}^{n-1} r_{\pi_0,\iota}|_{G_F}$ is automorphic.

It is ultimately our goal to show that $\mathrm{Sym}^{n-1} r_{\pi_0,\iota}$ (an extension of $\mathrm{Sym}^{n-1} r_{\pi_0,\iota}|_{G_F}$ from the finite-index subgroup $G_F \subset G_\mathbf{Q}$) is automorphic, and this is where the lemmas of [BLGHT2011, §1] come in. The first step is to show that $\mathrm{Sym}^{n-1} r_{\pi_0,\iota}|_{G_{F^+}}$ is automorphic, and this is the role that is played by [BLGHT2011, Lemma 1.5] (essentially just the solvable descent theorem of [AC1989] in the cyclic degree 2 case plus some details).

Lemma 3.3.2. $\mathrm{Sym}^{n-1} r_{\pi_0,\iota}|_{G_{F^+}}$ *is automorphic, and in particular corresponds to a regular algebraic cuspidal essentially self-dual automorphic representation of* $\mathrm{GL}_n(\mathbf{A}_{F^+})$.

Proof. The assumptions on the automorphic representations involved, plus Lemma 3.3.1 (and the easy consequence of it for a symmetric power of a 2-dimensional representation), plus the fact that $\mathrm{Sym}^{n-1} r_{\pi_0,\iota}|_{G_{F^+}}$ comes from restricting a Galois representation of domain $G_\mathbf{Q}$ and therefore its determinant has values not depending on the choice of complex conjugation (as it is $\mathrm{Gal}(F^+/F)$-invariant), imply that the hypotheses of [BLGHT2011, Lemma 1.5] hold.

For the purpose of making sure I know the full detail, I will now plagiarize the proof of [BLGHT2011, Lemma 1.5], adapted for this setting (i.e. with the various variables replaced with what you plug in).

In order to apply the cyclic-degree-2 case of the Arthur–Clozel descent theorem [AC1989, Theorem 4.2], we need to check that $\widetilde{\pi_n} \cong \widetilde{\pi_n}^c$. It suffices to check that the corresponding Galois representation $r_{\widetilde{\pi_n},\iota} : G_F \to \mathrm{GL}_n(\overline{\mathbf{Q}}_p)$ satisfies the same property. But this is immediate from the fact that $r_{\widetilde{\pi_n},\iota}$ has the privilege of extending all the way to a representation with domain $G_{\mathbf{Q}}$, namely $\mathrm{Sym}^{n-1} r_{\pi_0,\iota}$ (or even G_{F^+} would have been okay). This way, we apply Arthur–Clozel and produce the descent π_{F^+} of $\widetilde{\pi_n}$ to $\mathrm{GL}_n(\mathbf{A}_{F^+})$, which a priori gets to be regular algebraic cuspidal.

However, this is not enough: we know that $r_{\pi_{F^+}}$ extends $r_{\widetilde{\pi_n},\iota}$, but we do not know that it agrees with $\mathrm{Sym}^{n-1} r_{\pi_0,\iota}|_{G_{F^+}}$ (indeed, we will need to twist by an appropriate character to get this to be true), nor do we know that it is essentially self-dual. First, we check that the descent π_{F^+} is essentially self-dual (this obviously remains true after a character twist so it is okay to do this verification right away). Since $\tilde{\pi}_n$ is conjugate-self-dual, we have

$$r^{\vee}_{\widetilde{\pi_n},\iota} = r^{c}_{\widetilde{\pi_n},\iota}\epsilon^{n-1} = r_{\widetilde{\pi_n},\iota}\epsilon^{n-1}$$

where ϵ denotes the p-adic cyclotomic character (this is part of the statement of how the Galois representations corresponding to automorphic representations work — see [BLGHT2011, Theorem 1.2]).

The Galois representation $r_{\pi_{F^+},\iota}$ is an extension of $r_{\tilde{\pi_n},\iota}$ to G_{F^+}, so we conclude that the two representations of G_{F^+}

$$r^{\vee}_{\pi_{F^+},\iota}, \qquad r_{\pi_{F^+},\iota}\epsilon^{n-1}$$

agree on $G_F \subset G_{F^+}$, and hence differ[15] by a $\overline{\mathbf{Q}}_p$-valued character of $\mathrm{Gal}(F/F^+)$, i.e. π_{F^+} is essentially self-dual (the condition that the Hecke character corresponding to ψ has the same value on -1 in each of the archimedean places of F^+ is obvious from class field theory and the fact that F is an imaginary quadratic extension of the totally real field F^+)[16]. The exact same argument (using the same sublemma proved in the footnote) also shows that since the representation $r_{\pi_{F^+},\iota}$ of G_{F^+} agrees with $\mathrm{Sym}^{n-1} r_{\pi_0,\iota}|_{G_{F^+}}$ on G_F, they differ by a character of $\mathrm{Gal}(F/F^+)$, and hence by twisting π_{F^+} by the corresponding Hecke character, we obtain the desired essentially self-dual automorphic representation of $\mathrm{GL}_n(\mathbf{A}_{F^+})$ that corresponds to $\mathrm{Sym}^{n-1} r_{\pi_0,\iota}|_{G_{F^+}}$. □

[15] Let ρ_1, ρ_2 be two irreducible finite-dimensional representations of a group G valued in the same finite-dimensional $k = \overline{k}$-vector space V, and suppose that they agree on a normal subgroup $H \subset G$ such that G/H is abelian. Suppose furthermore that they remain irreducible when restricted to H. For any $g \in G$, consider the linear operator $\varphi_g : V \to V$ given by $\rho_1(g)\rho_2(g)^{-1}$. For $h \in H$ and $v \in V$, we have

$$\rho_1(h)\varphi_g v = \rho_1(hg)\rho_2(g)^{-1} v = \varphi_{hg}\rho_2(h)v.$$

In fact, $\varphi_{hg} = \varphi_g$ since φ_g only depends on the coset gH and G/H is abelian, so we conclude that φ_g is intertwining between irreducibles $\rho_1|_H$ and $\rho_2|H$. By Schur's lemma we conclude that φ_g acts by a scalar, and hence that ρ_1, ρ_2 differ by a character of G/H. This applies in our situation because all of our automorphic representations are cuspidal, and hence all of our Galois representations are irreducible

[16] Here our argument is allowed to differ slightly from that of [BLGHT2011, Lemma 1.5] — in that lemma, there is no guarantee that the representation "r" (in our case that role is played by $\mathrm{Sym}^{n-1} r_{\pi_0,\iota}|_{G_{F^+}}$) is irreducible, so an added assumption that r is self-dual up to character twist must be added. For this reason, I thought that it would be necessary to use Lemma 3.3.1, but now I think it should be okay.

Finally, the exact same type of argument as in Lemma 3.3.2 allows us to go all the way down to $\mathbf{Q}$ (this time it is [BLGHT2011, Lemma 1.3] that we will follow).

Lemma 3.3.3. $\mathrm{Sym}^{n-1} r_{\pi_0,\iota}$ *is automorphic, corresponding to an essentially self-dual representation of* $\mathrm{GL}_n(\mathbf{A}_{\mathbf{Q}})$.

Proof. In Lemma 3.3.2, we produced a regular algebraic cuspidal essentially self-dual automorphic representation π_{F^+} [17] of $\mathrm{GL}_n(\mathbf{A}_{F^+})$ such that

$$\mathrm{Sym}^{n-1} r_{\pi_0,\iota}|_{G_{F^+}} = r_{\pi_{F^+},\iota}.$$

Taking a subextension $\mathbf{Q} \subset E \subset F^+$ such that F/E is cyclic of prime degree (always possible as $F^+/\mathbf{Q}$ is solvable), wewill prove that $\mathrm{Sym}^{n-1} r_{\pi_0,\iota}|_{G_E}$ is automorphic corresponding to a regular algebraic cuspidal essentially self-dual automorphic representation of $\mathrm{GL}_n(\mathbf{A}_E)$. By induction this suffices to prove the desired statement.

Now the rest of the argument goes down exactly the same as in Lemma 3.3.2 (where we didn't need an induction as F/F^+ is already cyclic of degree 2).

Since $r_{\pi_{F^+},\iota}$ comes from restricting a representation of G_E (in fact $G_{\mathbf{Q}}$), it is invariant under $\mathrm{Gal}(F^+/E)$. Therefore, so is π_{F^+}, so by [AC1989, Theorem 4.2], and all the irreducibility we have for all of our Galois representations, it descends to an automorphic representation π_E of $\mathrm{GL}_n(\mathbf{A}_E)$ which is regular algebraic cuspidal. It is essentially self-dual, thanks to the fact that $r_{\pi_{F^+}}$ is irreducible and essentially self-dual (same argument as in the footnote of Lemma 3.3.2). Again by the same argument as in Lemma 3.3.2, we can twist π_E by a Hecke character to obtain a regular algberaic cuspidal essentially self-dual automorphic representation of $\mathrm{GL}_n(\mathbf{A}_E)$ whose corresponding Galois representation is exactly $\mathrm{Sym}^{n-1} r_{\pi_0,\iota}|G_E$. $\square$

[17] technically we had to twist by a character, but let us just keep this name for it

Chapter 4

Ping–Pong

"The game's afoot."

King Henry V, Shakespeare's *Henry V*

We now fill in the remaining details of the proof of Theorem 1.1.6. This is what [NT2021] call "ping-pong." Other than the actual table tennis game which is played between various components of the eigencurve over a boudnary annulus of weight space, it mostly consists of verifying that level 1 forms, as well as the auxilliary forms that appear in ping-pong, are going to satisfy the hypotheses required in the general machinery Theorem 3.0.1 of analytic continuation of symmetric power lifts.

The most "fun" part is this curious lemma that boils down to explicit analysis of mod p modular forms for some small values of p. It would be interesting to see whether this lemma can be generalized to higher-rank groups, or perhaps to Hilbert modular forms.

Lemma 4.0.1 (Lemma 3.5 of [NT2021])**.** *Let f be a level 1 cuspidal eigenform of weight $k \geq 2$. Then every accessible refinement of π at the prime $p = 2$ is numerically non-critical and n-regular for every $n \geq 2$.*

Proof. A refinement at 2 is just the data of an ordering of the two Frobenius eigenvalues associated to f at 2, i.e. a choice of root of

$$X^2 - a_2X + 2^{k-1}$$

where a_2 is the T_2-eigenvalue of f. In other words, it is a choice of eigenform in the space of oldforms for $\Gamma_0(2)$ generated by $f(z)$ and $f(2z)$ (conjecturally there are always two such choices). Let α, β be the two U_p-eigenvalues, i.e. the two roots of $X^2 - a_2X + 2^{k-1}$ in $\overline{\mathbf{Q}}_2$. Without loss of generality (there is definitely at least one eigenvector), suppose that there is an eigenform g in our oldform space coming from f with U_2-eigenvalue α.

Assume for the sake of contradiction that $v_2(\alpha) = 0$, i.e. that g is ordinary. Let $\tilde{g}$ be an element of the Hida family passing through g, where $\tilde{g}$ is an ordinary 2-adic overconvergent eigenform for $\Gamma_0(2)$ of classical weight 2. By Coleman's classicality criterion (alternatively I think this is also part of the statement of Hida theory), $\tilde{g}$ is classical (here we use the fact that it is in weight 2 and not weight 1, as 2 minus 1 is positive, hence bigger than the slope of $\tilde{g}$ which is zero). Alternatively, we could have used [Ser1973, Théorème 11] to arive here. But since the only weight-2 modular form for $\Gamma_0(2)$ is the Eisenstein series[1], and we have restricted to the cuspidal locus, this is a contradiction, hence g was not ordinary to begin with.

This implies that every accessible refinement of π at 2 is numerically noncritical. Indeed:

- Suppose that U_2 has at least two eigenvectors in the space of oldforms coming from f. Then by what we just said (applied to α as well as β, since they both have a corresponding eigenform g we can look at), $v_2(\alpha), v_2(\beta) > 0$. But since α, β are the two roots of $X^2 - a_2 X + 2^{k-1}$, we also know $v_2(\alpha) + v_2(\beta) = k - 1$, so we have

$$v_2(\alpha), v_2(\beta) < k - 1,$$

  i.e., both accessible refinements of π are numerically non-critical.

- Suppose that U_2 has just one eigenvector in the space of oldforms coming from f, WLOG with eigenvalue α. Of course, since we have assumed U_2 is not diagonalizable, we have $\alpha = \beta$, and hence both α and β have positive 2-adic valuation, which implies

$$v_2(\alpha) < k - 1,$$

  i.e. the single accessible refinement of π at 2, namely the one coming from the eigenvector for α, is numerically non-critical.

The above was "full detail" for the single sentence of [NT2021] that goes "Numerical non-criticality of every refinement is immediate from the fact that there are no cusp forms of level 1 that are ordinary at 2."

Now it is time to prove the regularity claim, which is the fun part. First, I just remark that n-regularity of every accessible refinement of π_2 for all $n \geq 2$ is implied by the claim that α/β is not a root of unity[2], which we now go ahead and prove. Assume for the sake of contradiction that α/β is a root of unity.

[1] To be completely honest I just checked the dimensions explicitly using SAGE, which returns 1 after both queries `dimension_modular_forms(Gamma0(2),2)` and `dimension_eis(Gamma0(11),2)`.

[2] Here is the full detail: The actual definition of an n-regular accessible 2-adic refinement $\chi = \chi_1 \otimes \chi_2$ of π is that the corresponding triangulation of $\mathbf{D}^{\dagger}_{\mathrm{rig}}(r_{\pi,\iota_2})$ has parameter (δ_1, δ_2) satisfying $\delta_1^n/\delta_2^n \notin x^{\mathbf{N}}$ for any n (so that all the symmetric powers of the triangulation are regular in the usual sense and we can apply the tools of deformation theory to study the image of the eigenvariety under the symmetric power map). To check the claim I made requires us to do two things: to understand explicitly the relationship between the accessible refinement χ and the two Frobenius eigenvalues α, β; and to understand explicitly the relationship between χ and the parameters

As we all know, α, β are the Frobenius eigenvalues in the 2-adic Weil–Deligne module $\mathrm{rec}^{\mathrm{T}}_{\mathbf{Q}_2}(\iota_2^{-1}\pi_2)$. By definition of the global Galois representation π_{r_ℓ,ι_ℓ} (in particular the definition of what the characteristic polynomial of Frobenii are supposed to be), these eigenvalues map under $\iota_\ell^{-1} \circ \iota_2$ to the Frobenius eigenvalues of the 2-dimensional ℓ-adic Galois representation $r_{\pi,\iota_\ell}|_{G_{\mathbf{Q}_2}}$, for any rational prime ℓ. First set $\ell = 3$. In that case, the global Galois representation r_{π,ι_3} has the property that its mod-3 semisimplification has Frob_2-eigenvalues adding up to $a_2 \mod 3 \in \mathbf{F}_3 \subset \overline{\mathbf{F}}_3$, of course. But we further know (by [Ser1975, Théorème 3], the proof of which was first published in [Joc1982, Theorem 4.1] using trace formula techniques) that the mod-3 system of Hecke eigenvalues Φ of f is equal to $\theta^{\circ\nu}(\Psi)$ for some system of Hecke eigenvalues Ψ occuring in the mod-3 modular forms of level 1 in weight between 2 and $3 + 1 = 4$. Of course, by the explicit structure theory of mod-3 modular forms in level 1 (see [SD1973, Theorem 3] and especially the first sentence of [SD1973, p. 19]), the only such system of Hecke eigenvalues is the one associated with $E_4 \mod 3 = 1 \in \mathbf{F}_3[[q]]$, i.e. it is the Ψ given by

$$\Psi(T_n) = \sigma_3(n) \mod 3$$

(note that 1 is not a normalized eigenform, since the q-coefficient is zero and not 1, so the computation of Ψ is either done by applying the formula based on q-expansions, or by lifting to E_4, dividing by 240, and reading off Fourier coefficients — there is no harm in doing this thanks to [AS1986a, Proposition 1.2.3]). Note that even though $E_4 \mod 3 = 1$, the associated system of Hecke eigenvalues Ψ is not trivial. Anyway, putting this all together, we conclude that[3]

$$a_2 \mod 3 = \Phi(T_2) = (\theta^{\circ\nu}\Psi)(T_2) = 2^\nu \sigma_3(2) = 2^\nu(1+8) = 0 \in \overline{\mathbf{F}}_3.$$

Since the Frob_2-eigenvalues of r_{π,ι_3} are none other than $\iota_3^{-1}\iota_2(\alpha), \iota_3^{-1}\iota_2(\beta)$, which we just saw have to add up to zero when reduced modulo 3, and since α/β is a root of unity (and hence $\iota_3^{-1}\iota_2(\alpha/\beta) \in \mathcal{O}_{\overline{\mathbf{Q}}_3}$), we conclude that

$$\omega_3 := \iota_3^{-1}\iota_2\left(\frac{\alpha}{\beta}\right) \in \mu_{\overline{\mathbf{Q}}_3} \cap (-1 + \mathfrak{m}_{\mathcal{O}_{\overline{\mathbf{Q}}_3}}).$$

In particular, $-\omega_3$ is a root of unity in $\overline{\mathbf{Q}}_3$ which is congruent to 1 modulo the maximal ideal.

---

of the triangulation that comes from it. The first thing is accomplished by [NT2021, Lemma 2.18], which says that the accessible refinement χ induces an increasing filtration of the Weil–Deligne module $\mathrm{rec}^{\mathrm{T}}_{\mathbf{Q}_2}(\iota_2^{-1}\pi_2)$ with graded pieces given by $\chi_i|\cdot|^{-1/2} \circ \mathrm{Art}_{\mathbf{Q}_p}$, $i = 1, 2$ — this proves that the Frobenius eigenvalues α, β are given in terms of χ by $\chi_i(p)p^{1/2}$, $i = 1, 2$ (see [BC2009a, Proposition 2.4.1]). The second thing is accomplished by [NT2021, Lemma 2.8], which gives us a triangulation with parameter $\delta_i(x) = \chi_i(x)|x|^{-1/2}x^{k_i}$. Now we conclude: α/β not a root of unity implies that $\chi_1(p)/\chi_2(p)$ is not a root of unity, which implies that $(\chi_1/\chi_2)^n \neq 1$ for all $n \geq 1$, which implies that $\delta_1/\delta_2 \notin x^{\mathbf{N}}$, as desired.

[3] In reality for this particular step, using Serre–Tate was a bit overkill — we could have just thought directly about the systems of Hecke eigenvalues mod 3 associated to powers of Δ thanks to [SD1973, Theorem 3].

This implies that $-\omega_3$ is a 3-power root of unity[4], and hence that α/β is -1 times a 3-power root of unity.

Now we repeat the same game except all of the systems of eigenvalues and modular forms will be modulo 5 instead of modulo 3 (we are still interested in the Frobenius at 2 though). This is slightly harder, since $a_2 \mod \ell$ no longer needs to be 0 now that $\ell = 5$. In particular, we need to use more information about exactly what the mod-ℓ Galois representations associated to cusp forms congruent to Eisenstein series look like. Luckily this is standard knowledge from [SD1973]. Again, let Φ be the mod-5 system of Hecke eigenvalues associated to f. By [Ser1975, Théorème 3] (though now that $\ell > 3$ [AS1986b, Theorem 1.3] would also do the trick), there exists a system of Hecke eigenvalues Ψ appearing in the mod-5 modular forms of level 1 in weight $2, 4, or 6 = 5 + 1$ such that $\Phi = \theta^{\circ\nu}(\Psi)$ for some $\nu \geq 0$. Since the only such systems are those associated to E_4 and E_6 (these are the only modular forms of level one in $\mathbf{Z}[[q]]$ of weight at most 6), there are only two possibilities for Ψ, namely $\Psi(T_n) = \sigma_3(n)$ or $\Psi(T_n) = \sigma_5(n)$. Note that none of these will have $\Psi(T_2) = 0$, as $1 + 2^3 = 9$ and $1 + 2^5 = 33$ are both not zero modulo 5. So to figure out the ratio of the Frob_2-eigenvalues, we work a little bit harder by using the fact that the mod-5 Galois representation coming from a cusp form with system of Hecke eigenvalues congruent to that of E_k mod 5 is always conjugate to

$$\begin{pmatrix} 1 & * \\ & \epsilon_5^{k-1} \end{pmatrix}$$

(of course the idea of thinking about it this way comes straight from [SD1973], but none of the nontrivial calculations of that paper are necessary here – once the traces of Frobenius are what you want, which is obvious thanks to q-expansions, you are done by Chebotarev and Brauer–Nesbitt as usual). Here ϵ is the 5-adic cyclotomic character, which takes Frob_2 to 2. Therefore, the Frob_2 eigenvalues of Ψ are either $\{1, 2^{4-1}\}$ or $\{1, 2^{6-1}\}$. We conclude (since $2^3, 2^5 \equiv \pm 2 \mod 5$) that the root of unity

$$\iota_5 \iota_2^{-1}\left(\frac{\alpha}{\beta}\right) = (\pm i) \cdot (\text{a 5-power root of unity})$$

(the argument is the same as last time, using the fact that $\pm i$ are lifts of ± 2 from $\mathbf{F}_5$ to $\overline{\mathbf{Q}}_5$). Since ι_ℓ are all isomorphisms of fields, we conclude that if α/β is a root of unity, then it is at the same time $-$(a 3-power root of unity) and $(\pm i)$(a 5-power root of unity). This is impossible, for example because the first thing implies after applying ι_2^{-1} that it is of the form $e^{2\pi i\left(\frac{1}{2}+\frac{a}{3^x}\right)} \in \mathbf{C}$

[4]Indeed, write $-\omega_3 = \zeta_{3^a}^x \zeta_M^y$ with $(3, M) = 1$. The only 3-power root of unity in $\overline{\mathbf{F}}_3$ is 1, so $\zeta_{3^a}^x \equiv 1 \mod \mathfrak{m}$, which implies by assumption that $\zeta_M^y \equiv 1 \mod \mathfrak{m}$ is a M-th root of unity in $\overline{\mathbf{Q}}_3$. By Hensel's lemma (here we use the fact that $(3, M) = 1$), the only such thing is 1, which implies that $-\omega_3 = \zeta_{3^a}^x$ is a 3-power root of unity.

and the second thing implies that it is of the form $e^{2\pi i\left(\pm\frac{1}{4}+\frac{b}{5^y}\right)}$, but we can never have

$$\frac{1}{2}+\frac{a}{3^x}\equiv\pm\frac{1}{4}+\frac{b}{5^y}\mod \mathbf{Z}$$

for $a, b, x, y \in \mathbf{N}$.

□

Remark 4.0.2. Obviously it is the work of Ash–Stevens that we would have to use in order to have any hope of generalization. However, we remark that the proof here could have been written down much earlier than Ash–Stevens came out in 1986, since the result about everything coming from weight $\leq \ell+1$ up to twist was known by Serre and Tate [Ser1975] in level 1 at least as early as 1975, and the explicit analysis of mod-ℓ modular forms in level 1 had by then certainly already been carried out by Swinnerton-Dyer [SD1973]. On the other hand, if we use the approach of Ash–Stevens, the condition $\ell > 3$ is needed to make the theory go through (this is assumed throughout [AS1986b] due to the hypotheses in [AS1986a, Theorem 1.3.5]). Although they fail to make explicit what they are using, Newton–Thorne probably thought about this using Ash–Stevens, which explains why they used the primes 5 and 7; and Richard Taylor in his course [Ye2021] probably thought about it the way explained above with the primes 3 and 5[5]).

Remark 4.0.3. Since α and β are roots of a polynomial with coefficients in $\mathbf{Q}$, note that the huge number of choices of isomorphisms $\iota_2, \iota_3, \iota_5$ is not really relevant here, up to possibly switching α, β.

Lemma 4.0.1 gives us most of the hypotheses we need to do analytic continuation for level 1 forms, but we are still missing something that tells us that the image of the 2-adic Galois representations corresponding to the overconvergent modular forms that we will care about will have image containing SL_2 when restricted to $G_{\mathbf{Q}_2}$. For the purposes of ping-pong, we will need to know this about some things which are not of level 1. First, a lemma about irreducibility is required (it will be very nice to know that things are irreducible before trying to prove they have big image). Before doing this, we need to understand the Galois representations associated to ordinary modular forms a bit better. Technically all we need is contained in the original papers of Mazur–Wiles [MW1986] and Wiles [Wil1988], but Newton–Thorne chose to do everything using the general theory for GL_n, stated in the general language of accessible refinements as in [Tho2015, Ger2019]. Here I essentially follow them, but note that their way of thinking about this is necessarily different (and less complicated than) Wiles and Mazur–Wiles, the point being that in the 1980s when those papers were written, they did not have access to the theorem of

[5] Indeed, Lynnelle cites [Ser1987] in the notes — this paper is about observations made by Serre long before the existence of Ash–Stevens, and for the reference regarding the fact that everything is up to twist something in weight at most $\ell+1$, he cites [Ser1975], where the fact is stated in level 1 without any restriction on ℓ; Serre's proof is unpublished but appears in more generality in [Joc1982], which indeed uses trace formula techniques as claimed by [Ser1975]

Colmez–Fontaine, and therefore had a harder time proving that things were reducible[6]. The following lemma is our expanded version of [NT2021, Lemma 3.5(1)].

Lemma 4.0.4. *Let π be a cuspidal automorphic representation of $\mathrm{GL}_2(\mathbf{A}_\mathbf{Q})$ of weight $k \geq 2$, p an arbitrary rational prime, and $\iota : \overline{\mathbf{Q}}_p \to \mathbf{C}$ an arbitrary isomorphism of fields. Then $r_{\pi,\iota}|_{G_{\mathbf{Q}_p}}$ is reducible if and only if π is ι-ordinary.*

Proof. One direction, the fact that ι-ordinary implies reducible, follows from the fact (from [MW1986, Wil1988]) that the Galois representation associated to an ordinary (at p) modular form is upper-triangular when restricted to $G_{\mathbf{Q}_p}$. Let me deduce this for myself, using the easier arguments of [Tho2015, Ger2019] that utilize the more advanced p-adic Hodge theory now available thanks to Colmez–Fontaine. The fact that π is ι-ordinary just means that π has an accessible refinement $\chi_1 \otimes \chi_2 : T_2(\mathbf{Q}_p) \to \overline{\mathbf{Q}}_p^\times$ with the property that

$$v_p(\chi_1(p)p^{1/2}) = 0, \qquad v_p(\chi_2(p)p^{1/2}) = k-1$$

(since by [NT2021, Lemma 2.18] such a refinement provides a filtration of the Weil–Deligne with associated graded $\chi_1|\cdot|^{-1/2} \circ \mathrm{Art}_{\mathbf{Q}_p}^{-1} \oplus \chi_2|\cdot|^{-1/2} \circ \mathrm{Art}_{\mathbf{Q}_p}$[7] and this is what we read off the Frobenius eigenvalues from; note that this extra $1/2$ added to the valuations ensures that the convention of [Tho2015] for ordinary automorphic representations agrees with the usual thing for modular forms). Recall that the notion of $\chi = \chi_1 \otimes \chi_2$ being an accessible refinement depends on ι: it says that the complex representation π_p admits an embedding into the normalized induction $i_B^{\mathrm{GL}_2}(\iota\chi)$, so all this stuff on the automorphic side is using ι just as much as the Galois representation $r_{\pi,\iota}$ is. In any event, the filtration of the Frobenius-semisimple Weil-Deligne module $\mathrm{WD}(r_{\pi,\iota}|_{G_{\mathbf{Q}_p}})^{F-ss} = \mathrm{rec}^{\mathrm{Tate}}(\iota^{-1}\pi_p)$ produced from χ induces a filtration of $\mathbf{D}_{\mathrm{pst}}(r_{\pi,\iota}|_{G_{\mathbf{Q}_p}})$ with associated gradeds having Frobenii $\chi_1(p)p^{1/2}$ and $\chi_2(p)p^{1/2}$ (direct consequence of definition of the WD-module from the $\mathbf{D}_{\mathrm{pst}}$). To prove that $r_{\pi,\iota}|_{G_{\mathbf{Q}_p}}$ is reducible, we just need to show that the first step in the filtration is weakly admissible in the sense of Colmez–Fontaine (so that it corresponds to a bona fide subobject of the local Galois representation we are interested in). Since $\mathbf{D}_{\mathrm{pst}}(r_{\pi,\iota}|_{G_{\mathbf{Q}_p}})$ is weakly admissible, we just need to prove that the subobject $D' \subset \mathbf{D}_{\mathrm{pst}}(r_{\pi,\iota}|_{G_{\mathbf{Q}_p}})$ with $\varphi = \chi_1(p)p^{1/2}$ has Newton-polygon-endpoint at most the Hodge-polygon-endpoint (the opposite inequality is taken care of by the weak admissibility of the big object). But this is obvious, since $v_p(\chi_1(p)p^{1/2}) = 0$ and the Hodge–Tate weights $0, k-1$ are both nonnegative.

[6] Eknath Ghate explained to me the idea of their argument, which was by analytic continuation from a très Zariski dense subset of a Hida family, where the forms in this dense subset are of weight 2 and arbitrarily deep level at p and can therefore be linked to abelian varieties (Jacobian of modular curve, say $X_1(p^n)$) where the usual analyses can be made.

[7] N.B. there is a typo in the statement of [NT2021, Lemma 2.18], at least according to the convention for the direction of the local Artin reciprocity map given in the introduction.

For the converse, the key point is the classification of de Rham (i.e., potentially semistable) p-adic characters of $G_{\mathbf{Q}_p}$: they are, up to finite order, of the form $\psi \cdot \epsilon^{-k}$, where $\psi : G_{\mathbf{Q}_p} \to \overline{\mathbf{Z}}_p^\times$ has $|\psi(I)| < \infty$ and k is the Hodge–Tate weight. See for example [FM1995, §10] or [BC2009b, Proposition 8.3.4] (apply it to the restriction of the character to a finite-index subgroup over which it becomes semistable) — the point is just the classify the weakly admissible 1-dimensional filtered $(\varphi, N, \mathrm{Gal}(K/\mathbf{Q}_p))$-modules. If the 2-dimensional p-adic Galois representation $r_{\pi,\iota}|_{G_{\mathbf{Q}_p}}$ is reducible, then there are exactly two Jordan–Holder factors (i.e. projection to the diagonal in suitable basis), and they are de Rham as $r_{\pi,\iota}|_{G_{\mathbf{Q}_p}}$ is Hodge–Tate. As a result, since the Hodge–Tate weights are 0 and $k-1$ (so the same is true of the two Jordan–Holder factors by looking at $\mathbf{D}_{\mathrm{HT}}$ in exact sequences) the Frobenius-semisimple guy $\mathrm{WD}(r_{\pi,\iota}|G_{\mathbf{Q}_p})^{F-ss}$ is just $\psi_1 \oplus \psi_2\epsilon^{1-k}$ for some ψ_1, ψ_2 which are $\overline{\mathbf{Z}}_p^\times$-valued characters. Applying the local Langlands correspondence (Weil–Deligne modules being built from irreducibles the same way that admissible smooth representations are build from supercuspidals) with the Tate normalization, we see that π_p is a subquotient of the normalized parabolic induction of $(\psi_1 \circ \mathrm{Art}_{\mathbf{Q}_p})| \cdot |^{1/2} \otimes (\psi_2 \circ \mathrm{Art}_{\mathbf{Q}_p})| \cdot |^{3/2-k}$, and is therefore ordinary (as this refinement provides the Frobenius eigenvalue $\psi_1 \circ \mathrm{Art}_{\mathbf{Q}_p}(p)$ via the recipe of [NT2021, Lemma 2.18], which is what we just applied in reverse, and this is necessarily a p-adic unit thanks to where ψ_1 is valued). □

Thanks to Lemma 4.0.4 and the fact that the ordinary locus is excluded from the cuspidal version of the eigencurve in the $p = 2$, $N = 1$ case that we are using, the following lemma will tend to apply for all the classical forms we are interested in:

Lemma 4.0.5. *Let $\rho : G_{\mathbf{Q}_p} \to \mathrm{GL}_2(\overline{\mathbf{Q}}_p)$ be an irreducible representation which is Hodge–Tate with distinct Hodge–Tate weights. Then one of the following must be true:*

1. *The Zariski closure of $\rho(G_{\mathbf{Q}_p})$ contains $\mathrm{SL}_2(\overline{\mathbf{Q}}_p)$.*
2. *The representation ρ is induced from a character of an index-2 closed subgroup of $G_{\mathbf{Q}_p}$.*

Proof. Let $H \subset \mathrm{GL}_2$ be the Zariski closure of the image of ρ. This is an algebraic subgroup of GL_2. Since ρ is irreducible, H is moreover a reductive algebraic group: indeed, any unipotent normal subgroup $H' \subset H$ is trivial (the definition of unipotent is that every nonzero representation has a nonzero fixed vector, so the H'-fixed vectors of V, being a nontrivial submodule [since H' is normal] of the simple module V, must be all of V), which is the definition of reductive.

Also, since the Hodge–Tate weights of ρ are distinct, the Sen operator $\Phi \in \mathfrak{gl}_2 \otimes \mathbf{C}_p$ is diagonalizable with two distinct eigenvalues in $\mathbf{Z}$. By [Sen1973, Theorem 1], applied to the group $G_{\mathbf{Q}_p^{\mathrm{ur}}} \subset G_{\mathbf{Q}_p}$ (the hypothesis of algebraically closed residue field is required in [Sen1973] — thanks to Tongmu He for pointing this out to me), the Lie algebra of H, when base-changed to $\mathbf{C}_p$, contains Φ. In fact, this implies that the $\overline{\mathbf{Q}}_p$-linear subspace $\mathrm{Lie}(H) \subset \mathfrak{gl}_2$ contains an element Φ' which is diagonalizable over $\overline{\mathbf{Q}}_p$ with distinct eigencalues. One way to prove this is as follows: consider the commutative diagram

$$\begin{array}{ccc} \mathrm{Lie}(H) \otimes \mathbf{C}_p & \xrightarrow{\text{charpoly}} & \mathbf{C}_p^2 \\ \uparrow & & \uparrow \\ \mathrm{Lie}(H) & \xrightarrow{\text{charpoly}} & \overline{\mathbf{Q}}_p^2 \end{array}$$

where the vertical maps are the obvious inclusions (which are dense because a $\overline{\mathbf{Q}}_p$ is dense in $\mathbf{C}_p$ and $\mathrm{Lie}(H)$ is a finite-dimensional $\overline{\mathbf{Q}}_p$-vector space, equipped with, say, the sup norm according to some basis), and the $\mathrm{charpoly}$ map is just $(-\mathrm{Tr}, \det)$ (i.e. it takes a matrix to the full information of the coefficients of its charpoly). Let $V_{\mathbf{C}_p} \subset \mathbf{C}_p^2$ be the set of points of the form $(-2a, a^2)$ for $a \in \mathbf{C}_p$. This is a closed set, since $\mathbf{C}_p$ is a topological ring and $V_{\mathbf{C}_p}$ is the zero locus of the continuous function $(x, y) \mapsto y - x^2/4$. Define $V_{\mathbf{Q}_p} \subset \overline{\mathbf{Q}}_p^2$ as being the set of points of the form $(-2a, a^2)$ for $a \in \overline{\mathbf{Q}}_p$. It is obviously contained in $V_{\mathbf{C}_p}$ (in fact it is equal to $V_{\mathbf{C}_p} \cap \overline{\mathbf{C}}_p^2$ but we will not use this). The reason for defining these closed subsets is that it expresses exactly the condition for a 2×2 matrix to have charpoly with exactly one root over the algebraically closed field in question. If $\mathrm{Lie}(H)$ contains no element which is diagonalizable with distinct eigenvalues (i.e. has charpoly with two distinct roots), that is the same as saying that $\mathrm{charpoly}(\mathrm{Lie}(H)) \subset V_{\mathbf{Q}_p}$. Since $V_{\mathbf{Q}_p} \subset V_{\mathbf{C}_p}$, which is a closed subset of $\mathbf{C}_p^2$, and $\mathrm{Lie}(H) \subset \mathrm{Lie}(H) \otimes \mathbf{C}_p$ is dense, we conclude that

$$\mathrm{charpoly}(\mathrm{Lie}(H) \otimes \mathbf{C}_p) \subset V_{\mathbf{C}_p},$$

which contradicts the fact that $\Phi \in \mathrm{Lie}(H) \otimes \mathbf{C}_p$ has distinct eigenvalues. We conclude the existence of an element $\Phi' \in \mathrm{Lie}(H)$ which is diagonalizable over $\overline{\mathbf{Q}}_p$ with distinct eigenvalues. It might have been possible to avoid this step by simply base-changing everything to $\mathbf{C}_p$, but I felt it would be easier to just do this rather than check that everything we care about is preserved by this base change.

By conjugation by an element of $\mathrm{GL}_2(\overline{\mathbf{Q}}_p)$ (which obviously doesn't affect the hypotheses since this is just a change of basis that will not affect the isomorphism class of the representation), we can therefore assume that

$$\begin{pmatrix} k_1 & 0 \\ 0 & k_2 \end{pmatrix} \in \mathrm{Lie}(H),$$

where $k_1 \neq k_2$ are the Hodge–Tate weights of ρ. According to [Bor1991, §7.3(2)] (which is really part of the basic theory of diagonalizable group schemes, as in [Bor1991, §8]), this implies that

$$H \supseteq \left\{ \begin{pmatrix} t_1 & \\ & t_2 \end{pmatrix} : t_1, t_2 \in \overline{\mathbf{Q}}_p, t_1^p t_2^q = 1 \right\},$$

where $p, q \in \mathbf{Z}$ are coprime such that k_1/k_2 in lowest terms is $-q/p$. In particular, H contains an element $\mathrm{diag}(t_1, t_2)$ where $t_1 \neq t_2$, and in fact contains the torus

$$\lambda : \mathbf{G}_m \hookrightarrow H \subset \mathrm{GL}_2$$

given by

$$\lambda(t) = \begin{pmatrix} t^{-q} & \\ & t^p \end{pmatrix}.$$

Let $T_H \subset H$ be a maximal torus of the reductive group H such that T_H contains λ. Such a maximal torus exists, for example by [Con2022a, Lemma 2.2]. For the same reason, we can take a maximal torus T of $\mathrm{GL}_2(\overline{\mathbf{Q}}_p)$ such that

$$\lambda \subset T_H \subset T.$$

By [Con2022b, Theorem 23.2.2], there exists $g \in \mathrm{GL}_2(\overline{\mathbf{Q}}_p)$ such that gTg^{-1} is the diagonal torus of $\mathrm{GL}_2(\overline{\mathbf{Q}}_p)$. Since

$$g \begin{pmatrix} t_1 & \\ & t_2 \end{pmatrix} g^{-1} \in \begin{pmatrix} * & \\ & * \end{pmatrix}$$

implies that

$$g \begin{pmatrix} t_1 & \\ & t_2 \end{pmatrix} g^{-1} = \begin{pmatrix} t_1 & \\ & t_2 \end{pmatrix}$$

when $t_1 \neq t_2$, we can assume (without changing our definition of λ, and especially the fact that λ contains a diagonal element with distinct entries) that T is the diagonal torus. Now there are two possibilities:

1. Suppose that H° is a torus, which forces $T_H = H^\circ$ (as tori are connected). Then (as H always normalizes H° since conjugation by a fixed element is a morphism of algebraic groups $H \to H$) we know that H normalizes T_H. Moreover, we proved above that $T_H \subset T$ contains a diagonal element of $\mathrm{GL}_2(\overline{\mathbf{Q}}_p)$ with distinct entries, and hence that the centralizer of T_H in $\mathrm{GL}_2(\overline{\mathbf{Q}}_p)$ is exactly the diagonal torus T (here I have used that $T_H \subset T$ and done a computation with matrices). As a formal consequence of the fact that H normalizes T_H, we know that H normalizes the centralizer in $\mathrm{GL}_2(\overline{\mathbf{Q}}_p)$ of T_H, and hence that H normalizes the diagonal torus of $\mathrm{GL}_2(\mathbf{Q}_p)$. Since ρ is irreducible, we know H is not contained in the diagonal torus. The normalizer of the maximal torus being

$$N_{\mathrm{GL}_2(\overline{\mathbf{Q}}_p)}(T) = \begin{pmatrix} * & \\ & * \end{pmatrix} \sqcup \begin{pmatrix} & * \\ * & \end{pmatrix} = T \sqcup wT,$$

where w is the unique nontrivial Weyl group element, we conclude that $\rho^{-1}(H \cap T)$ is an index-2 closed subgroup G_K of $G_{\mathbf{Q}_p}$ (K a quadratic extension of $\mathbf{Q}_p$), and hence (for example by explicit definition of induction) that ρ is induced from the character χ_1 of G_K given by

$$\rho|G_K = \begin{pmatrix} \chi_1 & \\ & \chi_2 \end{pmatrix}.$$

Indeed, if $\gamma \in G_{\mathbf{Q}_p} \setminus G_K$, then χ_2 and χ_1 are conjugate to each other under γ, since for any $g \in G_K$,

$$\rho(\gamma g \gamma^{-1}) = \begin{pmatrix} \chi_1^\gamma(g) & \\ & \chi_2^\gamma(g) \end{pmatrix}$$

is the same as

$$\rho(\gamma g \gamma^{-1}) = \begin{pmatrix} \chi_2(g) & \\ & \chi_1(g) \end{pmatrix}$$

(explicit matrix computation using the fact that $\rho(\gamma) \in wT$).

2. In the case that H° is not a torus, by the weight space decomposition (see e.g. [Mil2018, §18.15] — we have no problems as H° is a connected reductive group over $\overline{\mathbf{Q}}_p$), there must be a nontrivial root with respect to the maximal torus T_H, i.e. $\Phi(H^\circ, T_H) \neq \emptyset$. In that case, if we take the root groups corresponding to a root and its inverse (which always exists since root data for reductive groups are closed under negation, as it can be written as a special case of reflection — see [Con2022c]), we end up with two copies of $\mathbf{G}_a$ living in GL_2. In fact, these two copies of $\mathbf{G}_a$ are automatically root groups for GL_2 (the T_H-root spaces they come from in $\mathrm{Lie}(H) \subset \mathfrak{gl}_2$ are automatically T-root spaces, and all the root spaces are 1-dimensional and the root groups are always just $\mathbf{G}_a$ so they can't get bigger on the way upstairs to GL_2). By [Con2022c, example 1.4], these two copies of $\mathbf{G}_a$ in GL_2 are just the upstairs and downstairs unipotent radical, which together generate SL_2. Since these subgroups started their life as root groups coming from $\mathrm{Lie}(H)$, we have $\mathrm{SL}_2 \subset H$, as desired.

The two possibilities above account for both possibilities in the claim. □

In the $p = 2$, $N = 1$ case (where the ordinary forms are automatically excluded), the first option in Lemma 4.0.5 will always be satisfied for classical level 1 forms of weight $k \geq 2$, as I now explain

Lemma 4.0.6. *Let π be the everywhere unramified cuspidal automorphic representation of $\mathrm{GL}_2(\mathbf{A}_\mathbf{Q})$ associated to a level 1 cusp form f of weight $k \geq 2$. Let $p = 2$. Then the Zariski closure of the image of $r_{\pi,\iota}|_{G_{\mathbf{Q}_p}}$ contains SL_2.*

Proof. In the proof of Lemma 4.0.1, we argued that f cannot be ordinary. It follows from Lemma 4.0.4 that $r_{\pi,\iota}|_{G_{\mathbf{Q}_p}}$ is irreducible. We also know that it has distinct Hodge–Tate weights $0, k-1$ (this is where we use $k \geq 2$). Therefore, by Lemma 4.0.5 , we just need to rule out the possibility that $r_{\pi,\iota}|_{G_{\mathbf{Q}_p}}$ is induced by a character $\psi : G_K \to \overline{\mathbf{Q}}_p^\times$, for K a quadratic extension of $\mathbf{Q}_p$. Suppose that this is the case. Since f is of level one, the representation $r_{\pi,\iota}|_{G_{\mathbf{Q}_p}}$ is unramified, and hence ψ and $K/\mathbf{Q}_p$ are unramified. This hypothesis is absolutely crucial, because it implies that ψ has an extension to $G_{\mathbf{Q}_p}$ (by local class field theory — note that technically we only used the fact that ψ was unramified, as the CFT argument works even if $K/\mathbf{Q}_p$ is ramified). By Mackey theory (or just by explicit analysis of the definition of the induced representation), the fact that ψ has an extension to $G_{\mathbf{Q}_p}$ implies that $r_{\pi,\iota}|_{G_{\mathbf{Q}_p}} = \mathrm{Ind}_{G_K}^{G_{\mathbf{Q}_p}} \psi$ is actually reducible. This contradicts the fact (deduced above from Lemma 4.0.4) that it is actually irreducible. □

Without any assumptions on the level or the choice of prime p, the exact same argument can be made to work, as long as we add in a regularity hypothesis.

Lemma 4.0.7. *Let π be the cuspidal automorphic representation of $\mathrm{GL}_2(\mathbf{A}_\mathbf{Q})$ associated to a cusp eigenform f of weight $k \geq 2$. Suppose also that $r_{\pi,\iota}|_{G_{\mathbf{Q}_p}}$ is irreducible and that π_p admits a 3-regular refinement. Then the Zariski closure of the image of $r_{\pi,\iota}|_{G_{\mathbf{Q}_p}}$ contains $\mathrm{SL}_2(\overline{\mathbf{Q}}_p)$.*

Proof. Thanks to the irreducibility hypothesis, and the fact that $k \geq 2$ (so the Hodge–Tate weights are distinct), by Lemma 4.0.5, we just need to rule out the case where $r_{\pi,\iota}|_{G_{\mathbf{Q}_p}}$ is induced from a character ψ of G_K, $K/\mathbf{Q}_p$ a quadratic extension. So far this proof is the same as that of Lemma 4.0.6, but now we need to use the hypothesis about the refinement instead of the level-1 hypothesis. By [NT2021, Lemma 2.18] and local-global compatibility as usual, the refinement $\chi = \chi_1 \otimes \chi_2 : T_2(\mathbf{Q}_p) \to \overline{\mathbf{Q}}_p^\times$ provides a filtration of $\mathrm{WD}(r_{\pi,\iota}|_{G_{\mathbf{Q}_p}})$ with graded pieces $\chi_1|\cdot|^{-1/2} \circ \mathrm{Art}_{\mathbf{Q}_p}, \chi_2|\cdot|^{-1/2} \circ \mathrm{Art}_{\mathbf{Q}_p}$. Moreover, the assumption that $r_{\pi,\iota}|_{G_{\mathbf{Q}_p}}$ is induced from ψ implies (by compatibility of $\mathbf{D}_{\mathrm{pst}}$ with induction) that $\mathrm{WD}(r_{\pi,\iota}|_{G_{\mathbf{Q}_p}}) = \mathrm{Ind}_{W_K}^{W_{\mathbf{Q}_p}} \psi|_{W_K}$. Since

$$\left(\mathrm{Ind}_{W_K}^{W_{\mathbf{Q}_p}} (\psi|_{W_K})\right)\Big|_{W_K} = \psi|_{W_K} \oplus \psi|_{W_K}^\gamma \tag{4.1}$$

(where $\gamma \in W_{\mathbf{Q}_p} \setminus W_K$), we conclude (without loss of generality choosing the first direct summand for where $\chi_1|\cdot|^{-1/2}$ lands) e.g. by Schur's lemma that

$$\psi|_{W_K} = (\chi_1|\cdot|^{-1/2} \circ \mathrm{Art}_{\mathbf{Q}_p})|_{W_K}.$$

But this implies (since extensions from Weil group to Galois group are unique) that $\psi : G_K \to \overline{\mathbf{Q}}_p^\times$ actually extends to a character of $G_{\mathbf{Q}_p}$, namely $\chi_1|\cdot|^{-1/2} \circ \mathrm{Art}_{\mathbf{Q}_p}$. Therefore, $\psi|_{W_K} = \psi|_{W_K}^\gamma$, and thus by Equation (4.1) and the fact that the second graded piece of the filtration of this induced module is $\chi_2|\cdot|^{-1/2} \circ \mathrm{Art}_{\mathbf{Q}_p}$, we also have

$$\psi|_{W_K} = (\chi_2|\cdot|^{-1/2} \circ \mathrm{Art}_{\mathbf{Q}_p})|_{W_K}.$$

We conclude that χ_1 and χ_2 agree on the index-2 subgroup $G_K \subset G_{\mathbf{Q}_p}$, and therefore that $\chi_1^2 = \chi_2^2$. This contradicts the assumption that χ is 3-regular (see [NT2021, Definition 2.23]), thus ruling out the case that $r_{\pi,\iota}|_{G_{\mathbf{Q}_p}}$ is induced and implying that the Zariski closure of its image contains SL_2 thanks to Lemma 4.0.5 as argued above. □

There is no advantage to stating Lemma 4.0.7 for $p = 2, N = 1$, but NB the whole ping-pong argument takes place on the eigencurve for $p = 2, N = 1$ and therefore we only need Lemma 4.0.7 in that case.

Now to finish things off with the actual ping-pong argument, which is taken directly from Newton–Thorne (all the details are essentially there).

Let $p = 2$ and $N = 1$. In this situation, [BK2005] implies that $\mathscr{E}_0$, restricted to the boundary annulus $|8| < |w| < 1$ of weight space (note that only the connected component of $\mathcal{W}$ matters, as the Nebentypus of a modular form of level 2 is always going to have $\chi(-1) = 1$), becomes a disjoint union of annuli, that is,

$$\kappa^{-1}(|8| < |w| < 1) = \bigsqcup_{i=1}^{\infty} X_i,$$

where $\kappa : X_i \to \{|8| < |w| < 1\}$ is an isomorphism of $\mathbf{Q}_p$-rigid spaces. Moreover, the slopes of points in $X_i(\overline{\mathbf{Q}}_p)$ are known explicitly: if $z \in X_i(\overline{\mathbf{Q}}_p)$, then the slope $s(z)$ is equal to $iv_p(w(\kappa(z)))$ (here w denotes the isomorphism between the connected component of weight space and the open unit disc, which on points just takes a character χ of $\mathbf{Z}_p^\times$ and sends it to $\chi(5) - 1$ — at least this is the convention used in [BK2005]). In the proof of Lemma 4.0.1, we also proved that all the points of $\mathscr{E}_0$ are NOT ordinary, and hence thanks to Lemma 4.0.4 the local irreducibility we need in Lemma 4.0.7 will always be satisfied. This is very good, because it implies that all we really need to do analytic continuation of symmetric power functoriality on $\mathscr{E}_0$ is the regularity hypothesis. Managing this hypothesis is the reason we like level 1 forms, and it is the reason why ping-pong must be done. It is also the fundamental reason why the modularity lifting stuff in the second half of [NT2021] is necessary.

Anyhow, time for ping-pong:

Theorem 4.0.8. *Let f, g be cusp eigenforms of level 1 and weight ≥ 2, and let $n \geq 3$. Suppose that $\mathrm{Sym}^{n-1} f$ exists. Then so does $\mathrm{Sym}^{n-1} g$.*

Proof. The key point is that every irreducible component $\mathcal{C}$ of $\mathscr{E}_0$ has the property that $\kappa(\mathcal{C})$ is Zariski-open in weight space, and therefore $\mathcal{C}$ meets at least one of the X_i's. The reason for this is that (as we mentioned at the beginning of Chapter 3) $\mathscr{E}_0$, and any irreducible component thereof (by [Con1999b, Theorem 4.3.2]), is a finite cover of a Fredholm variety: a particular fiber over weight space being empty is equivalent to all the nontrivial coefficients of a particular characteristic series being zero simultaneously, which defines a proper Zariski-closed subset of weight space.

The first step is to analytically continue symmetric power functoriality from f to a conveniently-chosen point in X_i, where X_i is (as "justified" by the previous paragraph) one of the boundary annuli described above. Since a point of X_i is uniquely determined by its image under κ, we can define such a point by prescribing the weight. Without bothering to deal with nebentypus (it won't be necessary), let us just take a very large integer k (the point of it being large is so that the form we get will be classical and we can apply all the nice things we just proved) and consider the weight-character χ in the connected component of $\mathcal{W}$ that is defined by $\chi(5) = 5^{k-2}$. To figure out how large k needs to be to guarantee that the point $z_{f,k} \in X_i$ such that $\kappa(z_{f,k}) = \chi$ is classical, we first need to figure out what the slope is. In order to even have χ be in the boundary annulus $|8| < |w(\chi)| < 1$ on which the result of [BK2005] works (and therefore even have $z_{f,k}$ in the first place), we must have

$$v_2(5^{k-2} - 1) = 2,$$

which is guaranteed as long as k is odd. In this situation, since $z_{f,k} \in X_i$, the slope of $z_{f,k}$ is equal to $iv_2(5^{k-2} - 1) = 2i$. For Coleman [Col1996] to guarantee us the classicality of $z_{f,k}$, we would like to have k large enough that

$$k - 1 > 2i.$$

We will also want to take the twin form, for which the usual condition $2i \neq (k-1)/2, (k-2)/2$ is useful. This condition is also useful because it guarantees that the slopes of the two accessible refinements are different, and hence (via the relationship between the slope of a refinement and the value at p that I explained explicitly earlier[8]) that every accessible refinment is n-regular for all n. For this reason, we might as well ask that k is large enough that

$$\frac{k-2}{2} > 2i.$$

As discussed above, Coleman's classicality criterion tells us that $z_{f,k}$ is classical, coming from an admissibly refined cuspidal automorphic representation (π, χ) unramified away from 2. By what we just said, this refinement (and any other) is n-regular for all n. It is also numerically noncritical (but this is just because nothing is ordinary, in the same way as in the beginning of the proof of Lemma 4.0.1). By Lemma 4.0.7, the Zariski closure of the image of $r_{\pi,\iota}|_{G_{\mathbf{Q}_p}}$ contains SL_2. By Lemma 4.0.1 and Lemma 4.0.6, the same is true for f and the automorphic representation it corresponds to. Therefore, the analytic continuation machinery mostly explained in my previous letter says that existence of $\mathrm{Sym}^{n-1} f$ implies the same thing for the modular form coming from $z_{f,k}$.

Now take the twin $z'_{f,k}$. The corresponding Galois representation is the same except possibly twisted by a character, so we don't need to do any work to check that the Zariski closure of the image of $G_{\mathbf{Q}_p}$ contains SL_2. Similarly, the construction (again the paragraph at the bottom of [NT2021, p. 56]) of the twin form implies that $z'_{f,k}$ is classical of weight $k - 1 - 2i$ and that

[8] The ratio of elements of $\overline{\mathbf{Q}}_p^\times$ with different valuations cannot be a root of unity.

the two slopes are distinct and hence all refinements are n-regular for all n. So (for the exact same reason as above), $z'_{f,k}$ also has a symmetric power lift. But $z_{f,k}$ lives in a totally different boundary annulus than $z_{f,k}$! Since it is of weight χ just like $z_{f,k}$, and of slope $k-1-2i$, we know that

$$z'_{f,k} \in X_{(k-1-2i)/2}.$$

Note also that everything applies in reverse (even though there are slight asymmetries in the statement of the analytic continuation machine, we have verified all the hypotheses for all the forms involved so there is no issue). So if we know the automorphy of $\mathrm{Sym}^{n-1} z'_{f,k}$, then we can also conclude the automorphy of f. We apply the reverse-version of what we just did to g: letting j be such that $g \in X_j$, the argument above tells us that for all sufficiently large odd numbers k' (how large depends only on j), there is a

$$z'_{g,k} \in X_{(k-1-2j)/2}$$

such that automorphy of $\mathrm{Sym}^{n-1} z'_{g,k}$ implies that of g. Since $z'_{g,k}$ and $z'_{f,k}$ both satisfy all the hypotheses of the analytic continuation machine, and since the annuli X_s are irreducible, all we need to do is prove that we can choose k and k' so that $z'_{g,k}, z'_{f,k}$ live in the same X_s (so that automorphy of $\mathrm{Sym}^{n-1} z'_{g,k}$ is implied by that of $\mathrm{Sym}^{n-1} z'_{f,k}$ and we are done). But this is straightforward, as we just need the following three things:

1. $k \gg_i 0$,
2. $k' \gg_j 0$, and
3. $(k-1-2i)/2 = (k'-1-2j)/2$.

The third condition is equivalent to $k = k' + 2(i-j)$, so it is clear that the sought-after k, k' exist. $\square$

Bibliography

[Abb2010] Ahmed Abbes, *Éléments de géométrie rigide. Volume I*, Progress in Mathematics, vol. 286, Birkhäuser/Springer Basel AG, Basel, 2010. Construction et étude géométrique des espaces rigides. [Construction and geometric study of rigid spaces], With a preface by Michel Raynaud. MR2815110

[AC1989] James Arthur and Laurent Clozel, *Simple algebras, base change, and the advanced theory of the trace formula*, Annals of Mathematics Studies, vol. 120, Princeton University Press, Princeton, NJ, 1989. MR1007299

[AIS2014] Fabrizio Andreatta, Adrian Iovita, and Glenn Stevens, *Overconvergent modular sheaves and modular forms for* $\mathbf{GL}_{2/F}$, Israel J. Math. **201** (2014), no. 1, 299–359. MR3265287

[AM2004] Ahmed Abbes and Abdellah Mokrane, *Sous-groupes canoniques et cycles évanescents p-adiques pour les variétés abéliennes*, Publ. Math. Inst. Hautes Études Sci. **99** (2004), 117–162. MR2075884

[AS1986a] Avner Ash and Glenn Stevens, *Cohomology of arithmetic groups and congruences between systems of Hecke eigenvalues*, J. Reine Angew. Math. **365** (1986), 192–220. MR826158

[AS1986b] ———, *Modular forms in characteristic l and special values of their L-functions*, Duke Math. J. **53** (1986), no. 3, 849–868. MR860675

[AT2009] Emil Artin and John Tate, *Class field theory*, AMS Chelsea Publishing, Providence, RI, 2009. Reprinted with corrections from the 1967 original. MR2467155

[Böc2001] Gebhard Böckle, *On the density of modular points in universal deformation spaces*, Amer. J. Math. **123** (2001), no. 5, 985–1007. MR1854117

[BC2009a] Joël Bellaïche and Gaëtan Chenevier, *Families of Galois representations and Selmer groups*, 2009. MR2656025

[BC2009b] Olivier Brinon and Brian Conrad, *CMI summer school notes on p-adic Hodge theory (preliminary version)*, 2009. Lecture notes. Clay Mathematical Institute Summer School, Honolulu, Hawaii. https://math.stanford.edu/~conrad/papers/notes.pdf.

[Bel2019] Joël Bellaiche, *Global applications of eigenvarieties*, 2019. Mathoverflow question https://mathoverflow.net/questions/75144/global-applications-of-eigenvarieties.

[Bel2021] Joël Bellaïche, *The eigenbook—eigenvarieties, families of Galois representations, p-adic L-functions*, Pathways in Mathematics, Birkhäuser/Springer, Cham, 2021. MR4306639

[Ber2008] Laurent Berger, *Équations différentielles p-adiques et (ϕ, N)-modules filtrés*, 2008, pp. 13–38. Représentations p-adiques de groupes p-adiques. I. Représentations galoisiennes et (ϕ, Γ)-modules. MR2493215

[BGR1984] S. Bosch, U. Güntzer, and R. Remmert, *Non-Archimedean analysis*, Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 261, Springer-Verlag, Berlin, 1984. A systematic approach to rigid analytic geometry. MR746961

[BHS2017] Christophe Breuil, Eugen Hellmann, and Benjamin Schraen, *Une interprétation modulaire de la variété trianguline*, Math. Ann. **367** (2017), no. 3-4, 1587–1645. MR3623233

[BK2005] Kevin Buzzard and L. J. P. Kilford, *The 2-adic eigencurve at the boundary of weight space*, Compos. Math. **141** (2005), no. 3, 605–619. MR2135280

[BKKN1967] R. Berger, R. Kiehl, E. Kunz, and Hans-Joachim Nastold, *Differentialrechnung in der analytischen Geometrie*, Lecture Notes in Mathematics, No. 38, Springer-Verlag, Berlin-New York, 1967. MR0224870

[BL1993] Siegfried Bosch and Werner Lütkebohmert, *Formal and rigid geometry. I. Rigid spaces*, Math. Ann. **295** (1993), no. 2, 291–317. MR1202394

[BLGGT2014] Thomas Barnet-Lamb, Toby Gee, David Geraghty, and Richard Taylor, *Potential automorphy and change of weight*, Ann. of Math. (2) **179** (2014), no. 2, 501–609. MR3152941

[BLGHT2011] Tom Barnet-Lamb, David Geraghty, Michael Harris, and Richard Taylor, *A family of Calabi-Yau varieties and potential automorphy II*, Publ. Res. Inst. Math. Sci. **47** (2011), no. 1, 29–98. MR2827723

[BM2002] Christophe Breuil and Ariane Mézard, *Multiplicités modulaires et représentations de* $\mathrm{GL}_2(\mathbf{Z}_p)$ *et de* $\mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ *en* $l = p$, Duke Mathematical Journal **115** (2002), no. 2, 205–310. With an appendix by Guy Henniart. MR1944572

[Bor1979] A. Borel, *Automorphic L-functions*, Automorphic forms, representations and L-functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 2, 1979, pp. 27–61. MR546608

[Bor1991] Armand Borel, *Linear algebraic groups*, Second, Graduate Texts in Mathematics, vol. 126, Springer-Verlag, New York, 1991. MR1102012

[Bre2015] Christophe Breuil, *Vers le socle localement analytique pour* GL_n *II*, Math. Ann. **361** (2015), no. 3-4, 741–785. MR3319547

[Buz2004] Kevin Buzzard, *On p-adic families of automorphic forms*, Modular curves and abelian varieties (Progress in Mathematics, vol. 224), 2004, pp. 23–44. MR2058640

[Buz2007] ______, *Eigenvarieties*, L-functions and Galois representations (London Mathematical Society Lecture Note Series, vol. 320), 2007, pp. 59–120. MR2392353

[BZ1976] I. N. Bernšteĭn and A. V. Zelevinskiĭ, *Representations of the group $GL(n, F)$, where F is a local non-Archimedean field*, Uspehi Mat. Nauk **31** (1976), no. 3(189), 5–70. MR0425030

[BZ1977] I. N. Bernstein and A. V. Zelevinsky, *Induced representations of reductive $\mathfrak{p}$-adic groups. I*, Ann. Sci. École Norm. Sup. (4) **10** (1977), no. 4, 441–472. MR579172

[Cal2013] Frank Calegari, *Congruences between modular forms*, 2013. Lecture notes from AWS course.

[Car2012] Ana Caraiani, *Local-global compatibility and the action of monodromy on nearby cycles*, Duke Math. J. **161** (2012), no. 12, 2311–2413. MR2972460

[Cas2001] Bill Casselman, *The L-group*, Class field theory—its centenary and prospect (Tokyo, 1998), 2001, pp. 217–258. MR1846460

[CH2013] Gaëtan Chenevier and Michael Harris, *Construction of automorphic Galois representations, II*, Camb. J. Math. **1** (2013), no. 1, 53–73. MR3272052

[Che1951] Claude Chevalley, *Deux théorèmes d'arithmétique*, Journal of the Mathematical Society of Japan **3** (1951), 36–44. MR44570

[Che2004] Gaëtan Chenevier, *Familles p-adiques de formes automorphes pour* GL_n, J. Reine Angew. Math. **570** (2004), 143–217. MR2075765

[Che2005] ———, *Une correspondance de Jacquet-Langlands p-adique*, Duke Math. J. **126** (2005), no. 1, 161–194. MR2111512

[Che2010] ———, *The infinite fern and families of quaternionic modular forms*, 2010. Lecture notes from course at IHP Galois trimester. http://gaetan.chenevier.perso.math.cnrs.fr/coursihp.html.

[Che2011] ———, *On the infinite fern of Galois representations of unitary type*, Ann. Sci. Éc. Norm. Supér. (4) **44** (2011), no. 6, 963–1019. MR2919688

[Che2013] ———, *Représentations galoisiennes automorphes et conséquences arithmétiques des conjectures de langlands et arthur*, 2013. Mémoire d'habilitation à diriger des recherches (Université Paris XI).

[Che2014] ———, *The p-adic analytic space of pseudocharacters of a profinite group and pseudorepresentations over arbitrary rings*, Automorphic forms and Galois representations. Vol. 1, 2014, pp. 221–285. MR3444227

[CHT2008] Laurent Clozel, Michael Harris, and Richard Taylor, *Automorphy for some l-adic lifts of automorphic mod l Galois representations*, Publ. Math. Inst. Hautes Études Sci. **108** (2008), 1–181. With Appendix A, summarizing unpublished work of Russ Mann, and Appendix B by Marie-France Vignéras. MR2470687

[CM1998] R. Coleman and B. Mazur, *The eigencurve*, Galois representations in arithmetic algebraic geometry (Durham, 1996), 1998, pp. 1–113. MR1696469

[CM2009] Frank Calegari and Barry Mazur, *Nearly ordinary Galois deformations over arbitrary number fields*, J. Inst. Math. Jussieu **8** (2009), no. 1, 99–177. MR2461903

[Col1989] Robert F. Coleman, *Reciprocity laws on curves*, Compositio Math. **72** (1989), no. 2, 205–235. MR1030142

[Col1996] ———, *Classical and overconvergent modular forms*, Invent. Math. **124** (1996), no. 1-3, 215–241. MR1369416

[Col1997a] ———, *Classical and overconvergent modular forms of higher level*, J. Théor. Nombres Bordeaux **9** (1997), no. 2, 395–403. MR1617406

[Col1997b] ———, *p-adic Banach spaces and families of modular forms*, Invent. Math. **127** (1997), no. 3, 417–479. MR1431135

[Col2005] ———, *The canonical subgroup of E is* $\operatorname{Spec} R[x]/(x^p + \frac{p}{E_{p-1}(E,\omega)}x)$, Asian J. Math. **9** (2005), no. 2, 257–260. MR2176608

[Con1999a] Brian Conrad, *Irreducible components of rigid spaces*, Ann. Inst. Fourier (Grenoble) **49** (1999), no. 2, 473–541. MR1697371

[Con1999b] ———, *Irreducible components of rigid spaces*, Ann. Inst. Fourier (Grenoble) **49** (1999), no. 2, 473–541. MR1697371

[Con2006] ———, *Relative ampleness in rigid geometry*, Ann. Inst. Fourier (Grenoble) **56** (2006), no. 4, 1049–1126. MR2266885

[Con2022a] ———, *Algebraic groups I. grothendieck's theorem on tori*, 2022. Lecture notes from Math 252, Stanford University. Based on notes by Sam Lichtenstein. http://virtualmath1.stanford.edu/~conrad/252Page/handouts/grthm.pdf.

[Con2022b] ———, *Linear algebraic groups I*, 2022. Lecture notes from Math 252, Stanford University. Based on notes by Sam Lichtenstein.http://virtualmath1.stanford.edu/~conrad/252Page/handouts/alggroups.pdf.

[Con2022c] ———, *Math 249b. root datum for split reductive groups*, 2022. Lecture notes from Math 249B, Stanford University. http://virtualmath1.stanford.edu/~conrad/249BW16Page/handouts/rootdatum.pdf.

[Con] ———, *Modular forms, cohomology, and the Ramanujan conjecture*. Unpublished draft book (typically known under the name "`brian.pdf`").

[CR2021] Laurent Clozel and Conjeeveram S. Rajan, *Solvable base change*, J. Reine Angew. Math. **772** (2021), 147–174. MR4227590

[Del1971] Pierre Deligne, *Formes modulaires et représentations l-adiques*, Séminaire Bourbaki. Vol. 1968/69: Exposés 347–363, 1971, pp. Exp. No. 355, 139–172. MR3077124

[dJ1998] A. J. de Jong, *Erratum to: "Crystalline Dieudonné module theory via formal and rigid geometry" [Inst. Hautes Études Sci. Publ. Math. No.* **82** *(1995), 5–96 (1996); MR1383213 (97f:14047)]*, Inst. Hautes Études Sci. Publ. Math. **87** (1998), 175. MR1659266

[DL2016] Hansheng Diao and Ruochuan Liu, *The eigencurve is proper*, Duke Math. J. **165** (2016), no. 7, 1381–1395. MR3498869

[DS1974] Pierre Deligne and Jean-Pierre Serre, *Formes modulaires de poids* 1, Ann. Sci. École Norm. Sup. (4) **7** (1974), 507–530 (1975). MR379379

[DT1994] Fred Diamond and Richard Taylor, *Nonoptimal levels of mod l modular representations*, Invent. Math. **115** (1994), no. 3, 435–462. MR1262939

[Dwo1962] Bernard Dwork, *On the zeta function of a hypersurface*, Inst. Hautes Études Sci. Publ. Math. **12** (1962), 5–68. MR159823

[Elk2019] Noam D. Elkies, *Math 229x: Introduction to analytic number theory*, 2019. Lecture notes from Harvard graduate course. https://people.math.harvard.edu/~elkies/M229.18/index.html.

[Eme2006a] Matthew Emerton, *Jacquet modules of locally analytic representations of p-adic reductive groups. I. Construction and first properties*, Ann. Sci. École Norm. Sup. (4) **39** (2006), no. 5, 775–839. MR2292633

[Eme2006b] ———, *On the interpolation of systems of eigenvalues attached to automorphic Hecke eigenforms*, Invent. Math. **164** (2006), no. 1, 1–84. MR2207783

[Eme2017] ———, *Locally analytic vectors in representations of locally p-adic analytic groups*, Mem. Amer. Math. Soc. **248** (2017), no. 1175, iv+158. MR3685952

[Eme2021] ———, *Langlands reciprocity: L-functions, automorphic forms, and Diophantine equations*, The genesis of the Langlands Program, 2021, pp. 301–386. MR4274535

[FM1995] Jean-Marc Fontaine and Barry Mazur, *Geometric Galois representations*, Elliptic curves, modular forms, & Fermat's last theorem (Hong Kong, 1993), 1995, pp. 41–78. MR1363495

[FvdP2004] Jean Fresnel and Marius van der Put, *Rigid analytic geometry and its applications*, Progress in Mathematics, vol. 218, Birkhäuser Boston, Inc., Boston, MA, 2004. MR2014891

[Ger2019] David Geraghty, *Modularity lifting theorems for ordinary Galois representations*, Math. Ann. **373** (2019), no. 3-4, 1341–1427. MR3953131

[GfGPS1990] I. M. Gel′ fand, M. I. Graev, and I. I. Pyatetskii-Shapiro, *Representation theory and automorphic functions*, Generalized Functions, vol. 6, Academic Press, Inc., Boston, MA, 1990. Translated from the Russian by K. A. Hirsch, Reprint of the 1969 edition. MR1071179

[GJ1972] Roger Godement and Hervé Jacquet, *Zeta functions of simple algebras*, Lecture Notes in Mathematics, Vol. 260, Springer-Verlag, Berlin-New York, 1972. MR0342495

[GJ1979] Stephen Gelbart and Hervé Jacquet, *Forms of $\mathrm{GL}(2)$ from the analytic point of view*, Automorphic forms, representations and L-functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 1, 1979, pp. 213–251. MR546600

[GM1998] Fernando Q. Gouvêa and Barry Mazur, *On the density of modular representations*, Computational perspectives on number theory (Chicago, IL, 1995), 1998, pp. 127–142. MR1486834

[Gou1988] Fernando Q. Gouvêa, *Arithmetic of p-adic modular forms*, Lecture Notes in Mathematics, vol. 1304, Springer-Verlag, Berlin, 1988. MR1027593

[HC1968] Harish-Chandra, *Automorphic forms on semisimple Lie groups*, Lecture Notes in Mathematics, No. 62, Springer-Verlag, Berlin-New York, 1968. Notes by J. G. M. Mars. MR0232893

[HSBT2010] Michael Harris, Nick Shepherd-Barron, and Richard Taylor, *A family of Calabi-Yau varieties and potential automorphy*, Ann. of Math. (2) **171** (2010), no. 2, 779–813. MR2630056

[HT2001] Michael Harris and Richard Taylor, *The geometry and cohomology of some simple Shimura varieties*, Annals of Mathematics Studies, vol. 151, Princeton University Press, Princeton, NJ, 2001. With an appendix by Vladimir G. Berkovich. MR1876802

[Hum2008] James E. Humphreys, *Representations of semisimple Lie algebras in the BGG category $\mathcal{O}$*, Graduate Studies in Mathematics, vol. 94, American Mathematical Society, Providence, RI, 2008. MR2428237

[JL1970] H. Jacquet and R. P. Langlands, *Automorphic forms on* $\mathrm{GL}(2)$, Lecture Notes in Mathematics, Vol. 114, Springer-Verlag, Berlin-New York, 1970. MR0401654

[JN2019] Christian Johansson and James Newton, *Parallel weight 2 points on Hilbert modular eigenvarieties and the parity conjecture*, Forum Math. Sigma **7** (2019), Paper No. e27, 36. MR4010559

[Joc1982] Naomi Jochnowitz, *Congruences between systems of eigenvalues of modular forms*, Trans. Amer. Math. Soc. **270** (1982), no. 1, 269–285. MR642341

[JS1976] Hervé Jacquet and Joseph A. Shalika, *A non-vanishing theorem for zeta functions of* GL_n, Inventiones Mathematicae **38** (1976/77), no. 1, 1–16. MR432596

[Kat1973] Nicholas M. Katz, *p-adic properties of modular schemes and modular forms*, Modular functions of one variable, III (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972), 1973, pp. 69–190. MR0447119

[Kis2003] Mark Kisin, *Overconvergent modular forms and the Fontaine-Mazur conjecture*, Invent. Math. **153** (2003), no. 2, 373–454. MR1992017

[Kis2019] M. Kisin, *Memories of Jean-Marc and his mathematics*, Gaz. Math. **162** (2019), 18–20. MR3965745

[KL2005] Mark Kisin and King Fai Lai, *Overconvergent Hilbert modular forms*, Amer. J. Math. **127** (2005), no. 4, 735–783. MR2154369

[KM1985] Nicholas M. Katz and Barry Mazur, *Arithmetic moduli of elliptic curves*, Annals of Mathematics Studies, vol. 108, Princeton University Press, Princeton, NJ, 1985. MR772569

[KPX2014] Kiran S. Kedlaya, Jonathan Pottharst, and Liang Xiao, *Cohomology of arithmetic families of (φ, Γ)-modules*, J. Amer. Math. Soc. **27** (2014), no. 4, 1043–1115. MR3230818

[Lab2011] J.-P. Labesse, *Changement de base CM et séries discrètes*, On the stabilization of the trace formula, 2011, pp. 429–470. MR2856380

[Lan1970] R. P. Langlands, *Problems in the theory of automorphic forms*, Lectures in modern analysis and applications, III, 1970, pp. 18–61. Lecture Notes in Math., Vol. 170. MR0302614

[Lan1994] Serge Lang, *Algebraic number theory*, 2nd edition, Graduate Texts in Mathematics, vol. 110, Springer-Verlag, New York, 1994. MR1282723

[LO1977] Jeffrey Lagarias and Andrew Odlyzko, *Effective versions of the Chebotarev density theorem*, Algebraic number fields: L-functions and Galois properties (Proceedings of a Symposium, University of Durham, Durham, 1975), 1977, pp. 409–464. MR0447191

[Loe2011] David Loeffler, *Overconvergent algebraic automorphic forms*, Proc. Lond. Math. Soc. (3) **102** (2011), no. 2, 193–228. MR2769113

[LW2012] David Loeffler and Jared Weinstein, *On the computation of local components of a newform*, Math. Comp. **81** (2012), no. 278, 1179–1200. MR2869056

[LWX2017] Ruochuan Liu, Daqing Wan, and Liang Xiao, *The eigencurve over the boundary of weight space*, Duke Math. J. **166** (2017), no. 9, 1739–1787. MR3662443

[Maz1997] Barry Mazur, *An "infinite fern" in the universal deformation space of Galois representations*, 1997, pp. 155–193. Journées Arithmétiques (Barcelona, 1995). MR1464022

[Mil2018] James Milne, *Reductive groups*, 2018. Online notes. https://www.jmilne.org/math/CourseNotes/RG.pdf.

[MW1986] B. Mazur and A. Wiles, *On p-adic analytic families of Galois representations*, Compositio Math. **59** (1986), no. 2, 231–264. MR860140

[NT2020] James Newton and Jack Thorne, *Adjoint selmer groups of automorphic galois representations of unitary type*, To appear in J. Eur. Math. Soc. (2020).

[NT2021] ———, *Symmetric power functoriality for holomorphic modular forms*, Publ. Math. Inst. Hautes Études Sci. **134** (2021), 1–116. MR4349240

[Ogg1969] Andrew P. Ogg, *On the eigenvalues of Hecke operators*, Math. Ann. **179** (1969), 101–108. MR269597

[OS2015] Sascha Orlik and Matthias Strauch, *On Jordan-Hölder series of some locally analytic representations*, J. Amer. Math. Soc. **28** (2015), no. 1, 99–157. MR3264764

[Pan2020] Lue Pan, *A note on overconvergence of hecke action*, arXiv preprint arXiv:2012.11845 (2020).

[Pil2013] Vincent Pilloni, *Overconvergent modular forms*, Ann. Inst. Fourier (Grenoble) **63** (2013), no. 1, 219–239. MR3097946

[Ray1967] M. Raynaud, *Passage au quotient par une relation d'équivalence plate*, Proc. Conf. Local Fields (Driebergen, 1966), 1967, pp. 78–85. MR0232781

[Ray1970] Michel Raynaud, *Anneaux locaux henséliens*, Lecture notes in mathematics, vol. 169, Springer, 1970.

[Ray1974] ———, *Géométrie analytique rigide d'après Tate, Kiehl,$\cdots$*, Table Ronde d'Analyse non archimédienne (Paris, 1972), 1974, pp. 319–327. Bull. Soc. Math. France, Mém. No. 39–40. MR0470254

[Sai1997] Takeshi Saito, *Modular forms and p-adic Hodge theory*, Invent. Math. **129** (1997), no. 3, 607–620. MR1465337

[Sat1963] Ichirô Satake, *Theory of spherical functions on reductive algebraic groups over $\mathfrak{p}$-adic fields*, Inst. Hautes Études Sci. Publ. Math. **18** (1963), 5–69. MR195863

[Sch2011] Peter Scholze, *The Langlands-Kottwitz approach for the modular curve*, Int. Math. Res. Not. IMRN **15** (2011), 3368–3425. MR2822177

[Sch2013] ———, *The local Langlands correspondence for GL_n over p-adic fields*, Invent. Math. **192** (2013), no. 3, 663–715. MR3049932

[SD1973] H. P. F. Swinnerton-Dyer, *On l-adic representations and congruences for coefficients of modular forms*, Modular functions of one variable, III (Proc. Internat. Summer School, Univ. Antwerp, 1972), 1973, pp. 1–55. Lecture Notes in Math., Vol. 350. MR0406931

[Sen1973] Shankar Sen, *Lie algebras of Galois groups arising from Hodge-Tate modules*, Ann. of Math. (2) **97** (1973), 160–170. MR314853

[Ser1962] Jean-Pierre Serre, *Endomorphismes complètement continus des espaces de Banach p-adiques*, Inst. Hautes Études Sci. Publ. Math. **12** (1962), 69–85. MR144186

[Ser1967] ———, *Local class field theory*, Algebraic Number Theory (Proceedings of an Instructional Conference, Brighton, 1965), 1967, pp. 128–161. MR0220701

[Ser1973] ———, *Formes modulaires et fonctions zêta p-adiques*, Modular functions of one variable, III (Proc. Internat. Summer School, Univ. Antwerp, 1972), 1973, pp. 191–268. Lecture Notes in Math., Vol. 350. MR0404145

[Ser1975] ———, *Valeurs propres des opérateurs de Hecke modulo l*, Journées Arithmétiques de Bordeaux (Conf., Univ. Bordeaux, 1974), 1975, pp. 109–117. Astérisque, Nos. 24–25. MR0382173

[Ser1981] ———, *Quelques applications du théorème de densité de Chebotarev*, Publications Mathématiques de l'Institut des Hautes Études Scientifiques **54** (1981), 323–401. MR644559

[Ser1987] ———, *Sur les représentations modulaires de degré 2 de $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$*, Duke Math. J. **54** (1987), no. 1, 179–230. MR885783

[Ser1989] ———, *Abelian l-adic representations and elliptic curves*, 2nd edition, Advanced Book Classics, Addison–Wesley Publishing Company, Advanced Book Program, Redwood City, CA, 1989. McGill University lecture notes written with the collaboration of Willem Kuyk and John Labute. MR1043865

[Shi1994] Goro Shimura, *Introduction to the arithmetic theory of automorphic functions*, Publications of the Mathematical Society of Japan, vol. 11, Princeton University Press, Princeton, NJ, 1994. Reprint of the 1971 original, Kanô Memorial Lectures, 1. MR1291394

[Sil1994] Joseph H. Silverman, *Advanced topics in the arithmetic of elliptic curves*, Graduate Texts in Mathematics, vol. 151, Springer-Verlag, New York, 1994. MR1312368

[Sil2009] ______, *The arithmetic of elliptic curves*, 2nd edition, Graduate Texts in Mathematics, vol. 106, Springer, Dordrecht, 2009. MR2514094

[SS2003] Elias M. Stein and Rami Shakarchi, *Complex analysis*, Princeton Lectures in Analysis, vol. 2, Princeton University Press, Princeton, NJ, 2003. MR1976398

[Taï2016] Olivier Taïbi, *Eigenvarieties for classical groups and complex conjugations in Galois representations*, Math. Res. Lett. **23** (2016), no. 4, 1167–1220. MR3554506

[Tat1967] John Tate, *p-divisible groups*, Proceedings of a Conference on Local Fields (Driebergen, 1966), 1967, pp. 158–183. MR0231827

[Tho2015] Jack A. Thorne, *Automorphy lifting for residually reducible l-adic Galois representations*, J. Amer. Math. Soc. **28** (2015), no. 3, 785–870. MR3327536

[TO1970] John Tate and Frans Oort, *Group schemes of prime order*, Ann. Sci. École Norm. Sup. (4) **3** (1970), 1–21. MR265368

[Urb2011] Eric Urban, *Eigenvarieties for reductive groups*, Ann. of Math. (2) **174** (2011), no. 3, 1685–1784. MR2846490

[Wei1956] André Weil, *On a certain type of characters of the idèle-class group of an algebraic number-field*, Proceedings of the international symposium on algebraic number theory, Tokyo & Nikko, 1955, 1956, pp. 1–7. MR0083523

[Wil1988] A. Wiles, *On ordinary λ-adic representations associated to modular forms*, Invent. Math. **94** (1988), no. 3, 529–573. MR969243

[WXZ2017] Daqing Wan, Liang Xiao, and Jun Zhang, *Slopes of eigencurves over boundary disks*, Math. Ann. **369** (2017), no. 1-2, 487–537. MR3694653

[Ye2019] Lynnelle Ye, *Slopes in Eigenvarieties for Definite Unitary Groups*, ProQuest LLC, Ann Arbor, MI, 2019. Thesis (Ph.D.)–Harvard University. MR4197772

[Ye2020] ______, *A modular proof of the properness of the coleman–mazur eigencurve*, arXiv preprint arXiv:2010.10705 (2020).

[Ye2021] ______, *Math 249b, winter 2021: Newton-thorne symmetric power automorphy lifting*, 2021. Lecture notes from course taught by Richard Taylor, Stanford University.

[EGA] Alexander and Jean Dieudonné Grothendieck, *Éléments de géométrie algébrique*, Publications Mathématiques de l'IHÉS. 1960-1967.

[SGA3.I] Philippe and Polo Gille Patrick (ed.), *Schémas en groupes (SGA 3). Tome I. Propriétés générales des schémas en groupes*, Documents Mathématiques (Paris), vol. 7, Société Mathématique de France, Paris, 2011. Séminaire de Géométrie Algébrique du Bois Marie 1962–64. A seminar directed by M. Demazure and A. Grothendieck with the collaboration of M. Artin, J.-E. Bertin, P. Gabriel, M. Raynaud and J-P. Serre. Revised and annotated edition of the 1970 French original.