Current Developments in Mathematics, 2021

Reciprocity and symmetric power functoriality

Jack A. Thorne

Abstract. Symmetric power functoriality is the one of the first interesting cases of the Langlands functoriality conjectures, which predict the existence of liftings of automorphic forms from one reductive group to another. These conjectures are closely tied, through the theory of L-functions, to questions in number theory of independent interest, most famously the Sato–Tate conjecture.

In this article we first give an introduction to these L-functions and their connection with the Langlands programme, before giving a guide to our proof, with James Newton, of symmetric power functoriality for holomorphic modular forms.

Contents

The author's work received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 714405).

© 2023 International Press

1. Introduction

Consider Ramanujan's function $\tau(n) : \mathbf{N} \to \mathbf{Z}$, defined by the formula

$$\Delta(q) = q \prod_{n \geq 1} (1 - q^n)^{24} = \sum_{n \geq 1} \tau(n) q^n.$$

This function appears in the statement of the following theorem, proved by Ramanujan in 1916 [**Ram16**]:

THEOREM 1.1. *If $n \in \mathbf{N}$, then let*

$$r_{24}(n) = \left\{ (x_1, \dots, x_{24}) \in \mathbf{Z}^{24} \mid \sum_{i=1}^{24} x_i^2 = n \right\}$$

be the number of ways of writing n as a sum of 24 squares. Then for any odd natural number n we have the formula

$$r_{24}(n) = \frac{16}{691} \sum_{d|n} d^{11} + \frac{33152}{691} \tau(n).$$

Ramanujan observed several properties of the function $\tau(n)$. For example, there is the famous congruence

$$\tau(n) \equiv \sum_{d|n} d^{11} \bmod 691.$$

He also considered the Dirichlet series $L(\Delta, s) = \sum_{n \geq 1} \tau(n) n^{-s}$, conjecturing the following properties:

- There is an Euler product expansion
$$L(\Delta, s) = \prod_{p \text{ prime}} (1 - \tau(p) p^{-s} + p^{11-2s})^{-1}.$$
- For any prime number p, factorise
$$(1 - \tau(p) p^{-s} + p^{11-2s}) = (1 - \alpha_p p^{-s})(1 - \beta_p p^{-s})$$
for complex numbers α_p, β_p. Then we have
$$|\alpha_p| = |\beta_p| = p^{11/2}.$$

Ramanujan's conjectures have turned out to be connected to some of the most beautiful aspects of algebraic number theory in the last century. In the 1960's Serre observed that many of these connections could be explained by the existence of a compatible family of p-adic Galois representations

$$\rho_{\Delta,p} : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\mathbf{Q}_p)$$

characterised by the requirement that the number $\tau(l)$ (for a prime number $l \neq p$) appear as the trace, under $\rho_{\Delta,p}$, of a Frobenius element of the Galois group $G_{\mathbf{Q}} = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ at the prime l. In the paper [**Ser69**], Serre associated to Δ the L-functions

$$L_m(s) = \prod_{p \text{ prime}} \prod_{i=0}^{m} (1 - \alpha_p^{m-i}\beta_p^i p^{-s})^{-1}$$

and conjectured that they admit an analytic continuation to the whole complex plane $\mathbf{C}$, satisfying the functional equation $\Lambda_m(s) = \Lambda_m(11m + 1 - s)$, where Λ_m is the completed L-function (a product of $L_m(s)$ and certain explicitly given Γ-factors). Deligne's construction [**Del71a**] of the representations $\rho_{\Delta,p}$ showed that these L-functions may be placed within the broader class of L-functions associated to the p-adic representations of $G_{\mathbf{Q}}$ appearing in the étale cohomology of algebraic varieties over $\mathbf{Q}$ (defined in [**Ser70**]). Indeed, Deligne's construction implies that the composite

$$\mathrm{Sym}^m \rho_{\Delta,p} : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{Q}_p) \to \mathrm{GL}_{m+1}(\mathbf{Q}_p)$$

of $\rho_{\Delta,p}$ with the m^{th} symmetric power of the standard representation of GL_2 has such a realisation, and its L-function is $L_m(s)$. The functions $L_m(s)$ are the symmetric power L-functions associated to Ramanujan's modular form Δ.

Around the same time, Langlands introduced in [**Lan70**] a class of L-functions associated to pairs (π, R) consisting of an automorphic representation π of a reductive group G over $\mathbf{Q}$ and an L-homomorphism $R : {}^L G \to \mathrm{GL}_n$ from the Langlands dual group. His fundamental functoriality conjecture predicted that these L-functions should be associated to a 'functorial lift' $R_*(\pi)$, which would be an automorphic representation of $\mathrm{GL}_n(\mathbf{A}_{\mathbf{Q}})$. In the special case where $G = \mathrm{GL}_2$, the possible choices of R are (up to twist, and restricting to irreducible representations R) precisely the symmetric powers $\mathrm{Sym}^m : \mathrm{GL}_2 \to \mathrm{GL}_{m+1}$ of the standard representation of GL_2. When $\pi = \pi_\Delta$ is the automorphic representation of $\mathrm{GL}_2(\mathbf{A}_{\mathbf{Q}})$ associated to the Ramanujan Δ function, the L-functions $L(\pi_\Delta, R, s)$ are none other than the L-functions $L_m(s)$ introduced above from the point of view of arithmetic geometry.

The goal of this article is give an introduction to the proof, by Newton and the author, of the existence of Langlands's functorial lift $\mathrm{Sym}^m_*(\pi_\Delta)$, and therefore the analytic continuation of Serre's L-function $L_m(s)$ [**NT21a, NT21b**]. (More generally, we establish the existence of the symmetric power liftings of all automorphic representations of $\mathrm{GL}_2(\mathbf{A}_{\mathbf{Q}})$ associated to holomorphic Hecke eigenforms.) We divide the exposition into two parts.

In Part I we describe more fully some of the ideas leading up to the formulation of the conjectures described in this introduction, starting with the Riemann zeta function and the first examples (due to Dirichlet and Artin) of L-functions associated to representations of Galois groups. The endpoint is a formulation of the general reciprocity conjecture for p-adic Galois representations which describes the relation they should have with automorphic representations of general linear groups, and which helps to motivate our approach to Langlands functoriality, which applies only for those automorphic representations which do have associated Galois representations. Our sketch is far from complete, both from a mathematical perspective and a historical one: for a more detailed survey covering much of the same ground, see [**Eme21**], while for a more careful approach to the reciprocity conjecture, including in particular a precise definition of automorphic representation, see [**Tay04**].

In Part II we get stuck into the details of our proof of symmetric power functoriality for holomorphic newforms. Our approach to functoriality is through the reciprocity Conjecture 2.5: we try to prove the automorphy of the symmetric powers of the 2-dimensional Galois representations associated to holomorphic newforms. As emphasised by Mazur [**Maz89**], p-adic Galois representations often come in p-adic families, which should be reflected in the existence of p-adic families of (necessarily non-classical) automorphic forms. In fact, there are many different notions of p-adic automorphic forms, corresponding to the different conditions one might impose on p-adic Galois representations, that lead to larger or smaller families. A feature of our proof is that we use several different notions of p-adic automorphic form, selecting the most appropriate one for each step in the argument. For applications to Ramanujan's modular form Δ, the most important class of p-adic automorphic forms is the class of finite slope overconvergent Hecke eigenforms, which together form the Coleman–Mazur eigencurve. For another view on the ideas appearing in our proof, see [**New22**].

2. Part I

2.1. Primes. Let us begin by playing the following game. We list the prime numbers

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61,\ 67, 71, \ldots$$

and then throw away everything but the last digit:

$$2, 3, 5, 7, 1, 3, 7, 9, 3, 9, 1, 7, 1, 3, 7, 3, 9, 1, 7, 1, \ldots$$

We can then take a census of all of these last remaining digits:

	1	3	7	9
$100 < p < 1000$	35	35	40	33
$1000 < p < 10000$	266	268	262	265
$10000 < p < 100000$	2081	2092	2103	2087

We see straight away that the last digits of primes are remarkably evenly distributed between the different residue classes. That this should be the case is a consequence of Dirichlet's theorem[1] on primes in arithmetic progressions:

THEOREM 2.1 (Dirichlet, 1837). *Let $a, N \in \mathbf{N}$ be coprime. Then*

$$\lim_{X \to \infty} \frac{\#\{p < X \text{ prime } \mid p \equiv a \bmod N\}}{\#\{p < X \text{ prime}\}} = \frac{1}{\phi(N)},$$

where $\phi(N) = \#(\mathbf{Z}/N\mathbf{Z})^\times$.

In other words, for any modulus N, the primes are distributed evenly between the different possible residue classes modulo N. By the end of the first part of this article, we will have seen sweeping generalisations of both the statement of this theorem and the methods used in the proof.

2.2. The Riemann zeta function. Before we get to the proof, let us take a step back and consider how the primes themselves are distributed. Define

$$\pi(X) = \#\{p < X \text{ prime}\},$$

the prime counting function. The asymptotic behaviour of $\pi(X)$ is the subject of the Prime Number Theorem:

THEOREM 2.2 (Hadamard, de la Vallée Poussin, 1896). *As $X \to \infty$, $\pi(X) \sim X/\log(X)$.*

The first proofs of the Prime Number Theorem used the connection between the primes and the Riemann zeta function

$$\zeta(s) = \sum_{n \geq 1} n^{-s}.$$

We consider this sum in the first instance as defined for complex numbers s such that $\mathrm{Re}(s) > 1$. In this case the sum is absolutely convergent, and $\zeta(s)$ then defines a holomorphic function in the right half-plane $\mathrm{Re}(s) > 1$. The first thing that we need to know is the Euler product expression for the Riemann zeta function:

$$\zeta(s) = \sum_{n \geq 1} n^{-s} = \prod_{p \text{ prime}} (1 - p^{-s})^{-1},$$

again valid in the region $\mathrm{Re}(s) > 1$. This can be viewed as an analytic expression of the fundamental theorem of arithmetic: we may expand the

[1] In fact, the statement given here, which computes the natural density of primes in a given residue class, is stronger than Dirichlet's original statement (often given in terms of the *Dirichlet density* of primes in a given residue class).

Euler product as a product of geometric series, and the fundamental theorem of arithmetic implies that each term n^{-s} ($n \in \mathbf{N}$) appears exactly once.

The Euler product also allows us to give an analytic proof that there are infinitely many primes. Indeed, $\zeta(s)$ diverges as s tends to 1 through values in the interval $(1, \infty)$ (because the harmonic series $\sum_{n \geq 1} n^{-1}$ diverges). On the other hand, if p is a prime number, then $(1 - p^{-s})^{-1}$ tends to a finite limit as $s \to 1$, so if there were only finitely many primes we would obtain a contradiction.

More refined information about $\zeta(s)$ leads to more refined information about the distribution of the prime numbers. For example, in order to prove the prime number theorem, it is enough to know that $\zeta(s)$ admits a meromorphic continuation to $\mathbf{C}$ which is holomorphic and non-vanishing on the line $\mathrm{Re}(s) = 1$ (except for the point $s = 1$, where there is a simple pole). Let us first explain how these properties lead to a proof of the prime number theorem. One deduction proceeds by considering the logarithmic derivative

$$\frac{\zeta'(s)}{\zeta(s)} = -\sum_{n \geq 1} \Lambda(n) n^{-s},$$

where $\Lambda(n)$ is the von Mangoldt function, taking the value $\log p$ when $n = p^k$ is a prime power and the value 0 otherwise. An elementary argument shows that the prime number theorem is equivalent to the asymptotic

$$\sum_{n < X} \Lambda(n) \sim X \text{ as } X \to \infty. \tag{2.1}$$

If $\zeta(s)$ has the claimed properties, then its logarithmic derivative is meromorphic in $\mathbf{C}$ and holomorphic in the region $\mathrm{Re}(s) \geq 1$ (excepting the point $s = 1$, where it has a simple pole). A Tauberian theorem[2] can then be applied which leads directly to the asymptotic (2.1).

How does one obtain these properties of the Riemann zeta function? The existence of the meromorphic continuation of the zeta function to the whole complex plane was known already to Riemann, who gave two proofs of its existence in his famous monograph [**Rie60**]. The proof that is of the greatest interest to us is the expression of the zeta function as an integral transform of the Jacobi theta function $\theta(\tau)$. This is our first encounter in this article with an *automorphic form*.

The theta function

$$\theta(\tau) = \sum_{n \in \mathbf{Z}} e^{\pi i n^2 \tau}$$

[2] For example, Ikehara's Tauberian theorem [**CQ15**, Theorem 3.5.2], which states that if $f(s) = \sum_{n \geq 1} a_n n^{-s}$ is a Dirichlet series with non-negative coefficients, absolutely convergent in $\mathrm{Re}(s) > 1$, and such that $f(s)$ admits a meromorphic continuation to a neighbourhood of the half-plane $\mathrm{Re}(s) \geq 1$ which is holomorphic on the line $\mathrm{Re}(s) = 1$, except for a simple pole of residue C at $s = 1$, then $\sum_{n < X} a_n \sim CX$ as $X \to \infty$. A version of this result where the a_n may be complex numbers is given in [**Lan94**, Ch. XV, §3].

is defined for τ in the complex upper half plane $\mathfrak{h} = \{\tau \in \mathbf{C} \mid \operatorname{Im}(\tau) > 0\}$. It follows quickly from the definition that $\theta(\tau)$ is holomorphic and satisfies the transformation property $\theta(\tau+2) = \theta(\tau)$. Its relation with $\zeta(s)$ is expressed by the formula (valid a priori in $\operatorname{Re}(s) > 1$):

$$\pi^{-s/2}\Gamma(s/2)\zeta(s) = \frac{1}{2}\int_{y=0}^{\infty}(\theta(iy)-1)y^{s/2}\frac{dy}{y}. \tag{2.2}$$

Let $\xi(s)$ denote the left-hand side of (2.2) (which might be called the completed ζ-function). We now use that the function $\theta(\tau)$ satisfies the additional symmetry

$$\theta(\tau) = \sqrt{\tau/i}^{\,-1}\theta(-1/\tau) \tag{2.3}$$

for all $\tau \in \mathfrak{h}$. Splitting the integral (2.2) into two pieces gives an expression and applying this gives a new expression

$$\xi(s) = \frac{1}{s(s-1)} + \frac{1}{2}\int_{y=1}^{\infty}(\theta(iy)-1)(y^{s/2}+y^{(1-s)/2})\frac{dy}{y}, \tag{2.4}$$

where the integral is now absolutely convergent for every value of $s \in \mathbf{C}$. We see that $\xi(s)$ therefore admits a meromorphic continuation to $\mathbf{C}$ (with poles only at $s = 0, 1$) and satisfies the functional equation $\xi(s) = \xi(1-s)$. The non-vanishing of $\zeta(s)$ on the line $\operatorname{Re}(s) = 1$ relies on a more subtle argument that we won't discuss here (but see e.g. [**New98**]).

The equations relating $\theta(\tau+2)$ and $\theta(-1/\tau)$ with $\theta(\tau)$ should be seen as generating a whole group of symmetries which preserve θ. The group $\mathrm{SL}_2(\mathbf{R})$ acts on the complex upper half plane $\mathfrak{h}$ by Möbius transformations:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \tau = \frac{a\tau+b}{c\tau+d}.$$

The matrices $\left(\begin{smallmatrix} 1 & 2 \\ 0 & 1 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ together generate a finite index subgroup Γ of $\mathrm{SL}_2(\mathbf{Z})$. In general, automorphic forms may be viewed as functions on homogeneous spaces of real groups satisfying both differential equations (here taken to be the Cauchy–Riemann equations, expressing the fact that θ is holomorphic) and some kind of invariance property under an arithmetic group such as Γ. They are acted upon by *automorphic representations*. We will return to this topic in §2.7.

2.3. Dirichlet L-functions. Let us now come back to the proof of Dirichlet's theorem on primes in arithmetic progressions. Fix a modulus $N \in \mathbf{N}$ and a base $a \in \mathbf{N}$ prime to N. Based on our experience with the Riemann zeta function, we might hope to prove Dirichlet's theorem by analyzing the asymptotics of the function

$$\sum_{\substack{n<X \\ n\equiv a \bmod N}} \Lambda(n), \tag{2.5}$$

perhaps by considering the logarithmic derivative of the partial Euler product

$$\prod_{\substack{p \text{ prime} \\ p \equiv a \bmod N}} \frac{1}{1-p^{-s}}.$$

Analysing this function is difficult when at the start of the proof we do not even know if the product is finite or infinite! It turns out to be more fruitful to introduce the Dirichlet L-functions $L(\chi, s)$, associated to a homomorphism $\chi : (\mathbf{Z}/N\mathbf{Z})^\times \to \mathbf{C}^\times$:

$$L(\chi, s) = \sum_{n \geq 1, (n,N)=1} \chi(n \bmod N) n^{-s} = \prod_{p \text{ prime}, p \nmid N} (1 - \chi(p)p^{-s})^{-1}.$$

Using Fourier analysis on the finite group $(\mathbf{Z}/N\mathbf{Z})^\times$ (in other words, the character theory of finite groups), we can write the indicator function of the residue class $a \bmod N$ as

$$\mathbf{1}_{a \bmod N}(n) = \frac{1}{\phi(N)} \sum_{\chi} \chi(a^{-1})\chi(n \bmod N).$$

This reduces an analysis of the function (2.5) to an analysis of the functions

$$\sum_{\substack{p^k < X \\ p \nmid N}} \chi(p \bmod N) \log(p) \tag{2.6}$$

as χ varies over the characters of $(\mathbf{Z}/N\mathbf{Z})^\times$. We already understand the case of the trivial character, which is just the prime number theorem. For non-trivial characters χ we need to show that the sum (2.6) is $o(X/\log X)$, and this can be shown to follow if $L(\chi, s)$ has a meromorphic continuation to $\mathbf{C}$ which is holomorphic and non-vanishing on the line $\mathrm{Re}(s) = 1$ (including now at the point $s = 1$). The non-vanishing at $s = 1$ is the hardest part of Dirichlet's proof.

2.4. Artin L-functions. Let us now consider the problem of generalising Dirichlet's L-functions. There are at least two reasons for doing this: first, we might hope to be able to prove statements generalising Dirichlet's theorem on primes in arithmetic progressions. Second, the functions $L(\chi, s)$ are beautiful objects in their own right; for example, their values at integer arguments are expected to have deeper arithmetic significance, as exemplified by the Dirichlet class number formula, which describes the value at $s = 1$ when χ is a quadratic character.

To see the path to such a generalisation, we first revisit the definition of the Dirichlet L-function, using results which go back to Gauss [**Gau66**]. For $N \in \mathbf{N}$, we introduce the cyclotomic field $K_N = \mathbf{Q}(e^{2\pi i/N})$. This is a Galois extension of $\mathbf{Q}$, and its Galois group is isomorphic to the group of units modulo N:

$$\begin{aligned} \mathrm{Gal}(K_N/\mathbf{Q}) &\xrightarrow{\sim} (\mathbf{Z}/N\mathbf{Z})^\times \\ \sigma_a &\mapsto a, \end{aligned} \tag{2.7}$$

where $\sigma_a(e^{2\pi i/N}) = e^{2\pi i a/N}$. In particular, if p is a prime number not dividing N then there is an associated automorphism $\sigma_p \in \mathrm{Gal}(K_N/\mathbf{Q})$ which satisfies $\sigma_p(e^{2\pi i/N}) = e^{2\pi i p/N}$. Identifying χ now with a character of $\mathrm{Gal}(K_N/\mathbf{Q})$, we have an equivalent expression

$$L(\chi, s) = \prod_{p \text{ prime}, p\nmid N} (1 - \chi(\sigma_p)p^{-s})^{-1},$$

We can now explain the perspective of Artin L-functions: replace K_N by any Galois extension $K/\mathbf{Q}$, and the character χ by any representation $\rho : \mathrm{Gal}(K/\mathbf{Q}) \to \mathrm{GL}_n(\mathbf{C})$. We then want to define

$$L(\rho, s) = \prod_{p \text{ prime}} L_p(\rho, p^{-s}),$$

where $L_p(\rho, T)$ is the corresponding Euler factor (a rational function in T). In order to define $L_p(\rho, T)$, we need to say what is the analogue of the element $\sigma_p \in \mathrm{Gal}(K_N/\mathbf{Q})$. This is the so-called Frobenius element: a misnomer, since it is in fact a conjugacy class (or even a conjugacy class of cosets) in $\mathrm{Gal}(K/\mathbf{Q})$. We first consider the case where the prime number p is unramified in K (this is the case for all but finitely many prime numbers, namely the primes which do not divide the discriminant of K). If $\mathcal{O}_K$ denotes the ring of integers of K, then there is a unique factorisation of ideals

$$p\mathcal{O}_K = \mathfrak{p}_1 \dots \mathfrak{p}_r,$$

where the $\mathfrak{p}_i$ are distinct prime ideals of $\mathcal{O}_K$. Choose any prime ideal $\mathfrak{p}$ among these: then $k(\mathfrak{p}) = \mathcal{O}_K/\mathfrak{p}$ is a finite extension of $\mathbf{Z}/p\mathbf{Z} = \mathbf{F}_p$, and as such the Galois group $\mathrm{Gal}(k(\mathfrak{p})/\mathbf{F}_p)$ has a canonical generator, the Frobenius automorphism $\phi_p : x \mapsto x^p$. The Galois group $\mathrm{Gal}(K/\mathbf{Q})$ acts on the set of prime factors $\mathfrak{p}_i$ (because it acts on $\mathcal{O}_K$ by ring automorphisms) and there is a natural map

$$\mathrm{Stab}_{\mathrm{Gal}(K/\mathbf{Q})}(\mathfrak{p}) \to \mathrm{Gal}(k(\mathfrak{p})/\mathbf{F}_p),$$

which is in fact an isomorphism – so we can lift the Frobenius ϕ_p uniquely to an element $\sigma_\mathfrak{p} \in \mathrm{Gal}(K/\mathbf{Q})$. This element depends on the choice of prime factor $\mathfrak{p}$, but $\mathrm{Gal}(K/\mathbf{Q})$ acts transitively on the factors $\mathfrak{p}_i$ and so the conjugacy class of $\sigma_\mathfrak{p} \in \mathrm{Gal}(K/\mathbf{Q})$ depends only on p, and we write σ_p for any representative of this conjugacy class.

When p is ramified, there is a factorisation

$$p\mathcal{O}_K = (\mathfrak{p}_1 \dots \mathfrak{p}_r)^e$$

for some integer $e \geq 2$. There is still a surjective homomorphism

$$\mathrm{Stab}_{\mathrm{Gal}(K/\mathbf{Q})}(\mathfrak{p}) \to \mathrm{Gal}(k(\mathfrak{p})/\mathbf{F}_p)$$

but now its kernel, the inertia group $I_\mathfrak{p}$, is non-trivial (in fact of order e). The pre-image of ϕ_p is a coset of $I_\mathfrak{p}$ whose conjugacy class in $\mathrm{Gal}(K/\mathbf{Q})$ depends only on p.

We are now ready to define the L-factor $L_p(\rho, s)$: writing V for the vector space on which ρ acts, it is given by the formula

$$L_p(\rho, T) = \det(1 - \sigma_\mathfrak{p} T : V^{I_\mathfrak{p}} \to V^{I_\mathfrak{p}}).$$

Because the characteristic polynomial of a matrix depends only on its conjugacy class, this really does depend only on the prime number p and not the choice of prime ideal $\mathfrak{p}$ lying above it. It is a simple exercise to check that when $K = K_N$, $\rho = \chi$, and χ is primitive, this recovers the local L-factor associated to the Dirichlet L-function.

This definition of the L-function $L(\rho, s)$ attached to a representation $\rho : \mathrm{Gal}(K/\mathbf{Q}) \to \mathrm{GL}_n(\mathbf{C})$ (commonly called an Artin representation) was given by Artin in 1924 [**Art24**]. One application of the Artin L-function is to prove the Chebotarev density theorem:[3]

Theorem 2.3. *Let $K/\mathbf{Q}$ be a Galois extension, and let $C \subset \mathrm{Gal}(K/\mathbf{Q})$ be a conjugacy class. Then*

$$\lim_{X \to \infty} \frac{\#\{p < X \text{ prime, unramified in } K \mid \sigma_p \in C\}}{\#\{p < X \text{ prime}\}} = \frac{\#C}{\#G}.$$

This beautiful statement is a constantly useful tool in algebraic number theory. When $K = K_N = \mathbf{Q}(e^{2\pi i/N})$, it reduces (using the isomorphism (2.7)) to Dirichlet's theorem. The proof is similar: using the character theory of finite groups, we can write the indicator function $\mathbf{1}_C : G \to \mathbf{C}$ of the conjugacy class C as a linear combination of characters of irreducible representations of G:

$$\mathbf{1}_C(g) = \sum_\rho \langle \rho, \mathbf{1}_C \rangle \cdot \mathrm{tr}\, \rho(g).$$

We may therefore reduce the proof of the Chebotarev density theorem to the problem of showing that the Artin L-function $L(\rho, s)$ of a non-trivial irreducible representation ρ of G has a meromorphic continuation to $\mathbf{C}$ which is holomorphic and non-vanishing on the line $\mathrm{Re}(s) = 1$. This can be reduced to the abelian case [**Wei74**, Ch. XIII, §12].

2.5. L-functions of p-adic representations. Artin L-functions are only the beginning of the story. We can greatly expand the class of arithmetic L-functions by enlarging the class of representations of Galois groups under consideration. In order to do this, we first replace the Galois groups of

[3] Although other approaches exist, particularly if one is willing to vary the base number field (taken here to be $\mathbf{Q}$). In particular, Chebotarev's original proof, the history of which is described in the enjoyable article [**SL96**], goes by reduction to the case of cyclotomic extensions of an arbitrary base number field.

finite extensions of $\mathbf{Q}$ by the absolute Galois group (with respect to a fixed algebraic closure $\overline{\mathbf{Q}}/\mathbf{Q}$)

$$G_{\mathbf{Q}} = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) = \varprojlim_{K \subset \overline{\mathbf{Q}} \text{ Galois}} \mathrm{Gal}(K/\mathbf{Q}),$$

a profinite group endowed with its Krull topology. Any Artin representation determines, by inflation, a representation $\rho : G_{\mathbf{Q}} \to \mathrm{GL}_n(\mathbf{C})$, with no need to mention a finite Galois extension through the Galois group of which ρ factors. In fact, one can show that any continuous representation $\rho : G_{\mathbf{Q}} \to \mathrm{GL}_n(\mathbf{C})$ necessarily factors through a finite quotient $\mathrm{Gal}(K/\mathbf{Q})$ (and in fact is continuous even when $\mathbf{C}$ is endowed with the discrete topology).

The next step therefore is to introduce replace the field $\mathbf{C}$ of coefficients by a non-archimedean local field, such as $\mathbf{Q}_p$, the completion of $\mathbf{Q}$ with respect to its p-adic absolute value (for some prime number p). The class of continuous representations

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

is much larger than the class of representations $\rho : G_{\mathbf{Q}} \to \mathrm{GL}_n(\mathbf{C})$ (and in fact too large: it is necessary to put restrictions on the representations under consideration in order to be able to do things such as write down reasonable L-functions).

The most basic example of a p-adic representation is the p-adic cyclotomic character ϵ. This can be constructed from the homomorphisms

$$\epsilon_n : G_{\mathbf{Q}} \to \mathrm{Gal}(K_{p^n}/\mathbf{Q}) \to (\mathbf{Z}/p^n\mathbf{Z})^\times.$$

These are compatible as n varies, in the sense that $\epsilon_{n+1} \bmod p^n = \epsilon_n$. Passing to the inverse limit, we obtain a continuous homomorphism (where $\mathbf{Z}_p$ denotes the ring of p-adic integers in $\mathbf{Q}_p$)

$$\epsilon : G_{\mathbf{Q}} \to \varprojlim_n (\mathbf{Z}/p^n\mathbf{Z})^\times = \mathbf{Z}_p^\times = \mathrm{GL}_1(\mathbf{Z}_p) \subset \mathrm{GL}_1(\mathbf{Q}_p).$$

One important property of ϵ that generalises is that if we evaluate it at a Frobenius element $\sigma_l \in G_{\mathbf{Q}}$ at a prime $l \neq p$, then the number $\epsilon(\sigma_l)$ is not just a p-adic number but in fact an integer, namely l.

The next examples of p-adic representations we consider are those associated to elliptic curves over $\mathbf{Q}$. Let (E, ∞) be an elliptic curve over $\mathbf{Q}$: thus E is a smooth, projective curve over $\mathbf{Q}$ of genus 1 and $\infty \in E(\mathbf{Q})$ is a marked rational point. There is a unique way to make E into a commutative algebraic group with identity ∞. For any field extension $K/\mathbf{Q}$, the set $E(K)$ of K-rational points is thus an abelian group, and $E(\overline{\mathbf{Q}})$ is even a $\mathbf{Z}[G_{\mathbf{Q}}]$-module. To construct a p-adic representation, we consider just a small part of $E(\overline{\mathbf{Q}})$, namely the subgroup of p^∞-torsion points (i.e. those which are p^n-torsion for some $n \geq 1$; this is a $\mathbf{Z}[G_{\mathbf{Q}}]$-submodule because the group law of E is defined over $\mathbf{Q}$).

We can use complex analytic considerations to describe $E(\overline{\mathbf{Q}})[p^\infty]$ as an abelian group. There is an isomorphism of topological groups $E(\mathbf{C}) \cong$

$S^1 \times S^1$, hence

$$E(\mathbf{C})[p^\infty] = (S^1 \times S^1)[p^\infty] \cong (\mathbf{Q}/\mathbf{Z} \times \mathbf{Q}/\mathbf{Z})[p^\infty] \cong \mathbf{Q}_p/\mathbf{Z}_p \times \mathbf{Q}/\mathbf{Z}_p.$$

Again using the fact that the group law of E is defined over $\mathbf{Q}$, we see that the points of $E(\mathbf{C})[p^\infty]$ are in fact defined over $\overline{\mathbf{Q}}$, and finally that $E(\overline{\mathbf{Q}})[p^\infty] \cong (\mathbf{Q}_p/\mathbf{Z}_p)^2$. If we define

$$T_pE = \mathrm{Hom}_{\mathbf{Z}}(\mathbf{Q}_p/\mathbf{Z}_p, E(\overline{\mathbf{Q}})[p^\infty]),$$

then T_pE is a free $\mathbf{Z}_p$-module of rank 2 which receives an action of $G_{\mathbf{Q}}$ which is continuous when T_pE is endowed with its p-adic topology. This is the *p-adic Tate module* of E. Writing $V_pE = T_pE \otimes_{\mathbf{Z}_p} \mathbf{Q}_p$, we see that V_pE is a 2-dimensional $\mathbf{Q}_p$-vector space and that we have constructed a continuous representation

$$\rho_{E,p} : G_{\mathbf{Q}} \to \mathrm{GL}_{\mathbf{Q}_p}(V_pE).$$

How can we construct an L-function from the representation $\rho_{E,p}$? Following the construction of Artin, we would like to define[4]

$$L_l(\rho_{E,p}, T) = \det(1 - \rho_{E,p}(\sigma_l)T : (V_pE)^{I_l} \to (V_pE)^{I_l})^{-1}.$$

However, it is not a priori clear that this makes sense, since we want this local L-factor to be an element of $\mathbf{C}(T)$, but V_pE is a vector space over $\mathbf{Q}_p$, which is not in any natural way a subfield of $\mathbf{C}$!

The representations $\rho_{E,p}$ have some remarkable properties which save the day. We first consider the case where $l \neq p$ is a prime where the elliptic curve has good reduction (the case for all but finitely many primes $l \neq p$, namely those that do not divide the minimal discriminant Δ_E of E). In this case the inertia group I_l acts trivially and $\rho_{E,p}(\sigma_l)$ is an endomorphism of V_pE, so might be represented by a matrix with entries in $\mathbf{Q}_p$ – but the characteristic polynomial of this matrix has coefficients in $\mathbf{Z}$! We in fact have the identity

$$\det(1 - \sigma_l T : V_pE \to V_pE) = 1 - a_l T + lT^2,$$

where a_l is given by the formula

$$a_l = l + 1 - \#E(\mathbf{F}_l). \tag{2.8}$$

(The set of $\mathbf{F}_l$-points of E makes sense because of our assumption that E has good reduction.) In particular, a_l is an integer. A similar formula holds when $l \neq p$ is a place of bad reduction. What about when $l = p$? The story here is more complicated. The formula $\det(1 - \sigma_p T : (V_pE)^{I_p} \to (V_pE)^{I_p})$ would

[4]Here we are using notation generalising that introduced in the last section for finite Galois extensions. For each prime number l, we can extend the embedding $\mathbf{Q} \to \mathbf{Q}_l$ to an embedding $\overline{\mathbf{Q}} \to \overline{\mathbf{Q}}_l$ of algebraic closures. This choice determines a homomorphism $G_{\mathbf{Q}_l} \to G_{\mathbf{Q}}$, where $G_{\mathbf{Q}_l}$ is the absolute Galois group of $\mathbf{Q}_l$. If $\overline{\mathbf{F}}_l$ denotes the residue field of the valued field $\overline{\mathbf{Q}}_l$ (which is indeed an algebraic closure of $\mathbf{F}_l$) and $G_{\mathbf{F}_l} = \mathrm{Gal}(\overline{\mathbf{F}}_l/\mathbf{F}_l)$, then the reduction map $G_{\mathbf{Q}_l} \to G_{\mathbf{F}_l}$ is surjective, its kernel $I_l \subset G_{\mathbf{Q}_l}$ is a closed subgroup, and it make sense to speak of the coset $\sigma_l I_l \subset G_{\mathbf{Q}_l} \subset G_{\mathbf{Q}}$ which is the pre-image of the Frobenius automorphism in $G_{\mathbf{F}_l}$. Up to conjugacy in $G_{\mathbf{Q}}$, this coset depends only on the prime l.

give the "wrong" answer. For example, if p is a prime of good reduction we would like to obtain $1 - a_pT + pT^2$, where a_p is defined as above. However, in this case we never have $V_p = V_p^{I_p}$. This reflects the fact that something special is happening at the prime p.

The subject of p-adic Hodge theory explains that $\rho_{E,p}$ contains information both about $\#E(\mathbf{F}_p)$ and about the Hodge decomposition of the de Rham cohomology of $E(\mathbf{C})$. This begins with the work of Tate [**Tat67**], who showed that for an elliptic curve (or more generally abelian variety) with good reduction over $\mathbf{Q}_p$, there is a canonical isomorphism (of $\mathbf{C}_p[G_{\mathbf{Q}_p}]$-modules, $\mathbf{C}_p$ being the completion of $\overline{\mathbf{Q}}_p$, and $\mathbf{C}_p(k)$ being the twist of $\mathbf{C}_p$ by the k^{th} power of the p-adic cyclotomic character ϵ):

$$\mathrm{Hom}_{\mathbf{Q}_p}(V_pE, \mathbf{C}_p) \cong (H^1(E, \mathcal{O}) \otimes_{\mathbf{Q}_p} \mathbf{C}_p) \oplus (H^0(E, \Omega^1) \otimes_{\mathbf{Q}_p} \mathbf{C}_p(-1)).$$

We can then recover the Hodge numbers $h^{p,q} = \dim_{\mathbf{Q}_p} H^p(E, \Omega^q_E)$ using the formulae

$$H^1(E, \mathcal{O}) \cong \mathrm{Hom}_{\mathbf{Q}_p}(V_pE, \mathbf{C}_p)^{G_{\mathbf{Q}_p}},$$
$$H^0(E, \Omega^1) \cong \mathrm{Hom}_{\mathbf{Q}_p}(V_pE, \mathbf{C}_p(1))^{G_{\mathbf{Q}_p}}.$$

If we did have $V_pE = (V_pE)^{I_p}$ then we'd have $h^{1,0} = 2$ and $h^{0,1} = 0$ contradicting the fact that both of these numbers are equal to 1 (because E is a curve of genus 1). In general, if V is a finite-dimensional $\mathbf{Q}_p[G_{\mathbf{Q}_p}]$-module then we have the inequality

$$\sum_{k \in \mathbf{Z}} \dim_{\mathbf{Q}_p} \mathrm{Hom}_{\mathbf{Q}_p}(V, \mathbf{C}_p(k))^{G_{\mathbf{Q}_p}} \leq \dim_{\mathbf{Q}_p} V,$$

which inspires the following definition:

DEFINITION 2.1. *Let V be a finite-dimensional $\mathbf{Q}_p[G_{\mathbf{Q}_p}]$-module. We say that V is Hodge–Tate if $\sum_{k \in \mathbf{Z}} \dim_{\mathbf{Q}_p} \mathrm{Hom}_{\mathbf{Q}_p}(V, \mathbf{C}_p(k))^{G_{\mathbf{Q}_p}} = \dim_{\mathbf{Q}_p} V$. If V is Hodge–Tate, we define the multiset of Hodge–Tate numbers of V to be the set of integers k such that $\dim_{\mathbf{Q}_p} \mathrm{Hom}_{\mathbf{Q}_p}(V, \mathbf{C}_p(-k))^{G_{\mathbf{Q}_p}} \neq 0$, each appearing with multiplicity equal to the dimension of this finite-dimensional $\mathbf{Q}_p$-vector space.*

These ideas were greatly extended and refined by Fontaine (see e.g. [**Fon82, Fon94a, Fon94c**]) who introduced several more interesting categories of $\mathbf{Q}_p[G_{\mathbf{Q}_p}]$-modules, each contained inside the next:

$$(\text{Hodge–Tate}) \supset (\text{de Rham}) \supset (\text{potentially semi-stable}) \supset (\text{semi-stable}) \supset (\text{crystalline}).$$

As an indication of the utility of these categories, Fontaine defined a functor D_{crys} from the category of crystalline $\mathbf{Q}_p[G_{\mathbf{Q}_p}]$-modules to the category of filtered ϕ-modules: i.e. the category of finite-dimensional $\mathbf{Q}_p$-vector spaces D endowed with a $\mathbf{Q}_p$-linear map $\phi : D \to D$ and a decreasing filtration $\mathrm{Fil}_\bullet D$. Moreover, he showed that if E is an elliptic curve with good reduction over

$\mathbf{Q}_p$, then there is a functorial isomorphism with the algebraic de Rham cohomology

$$D_{crys}(V_pE) \cong H^1_{dR}(E/\mathbf{Q}_p)$$

and an equality

$$\det(1 - \phi T : D_{crys}(V_pE) \to D_{crys}(V_pE)) = 1 - a_pT + pT^2.$$

Thus we can refine the Hodge–Tate decomposition and define the correct local L-factor in this case. More generally, Fontaine gave a recipe to extract from any potentially semi-stable $\mathbf{Q}_p[G_{\mathbf{Q}_p}]$-module a *Weil–Deligne representation* (r, N) [**Fon94b**].[5] Even without the hypothesis of good reduction at the prime p, the representation $\rho_{E,p}|_{G_{\mathbf{Q}_p}}$ is potentially semi-stable and one can show that the associated Weil–Deligne representation gives the correct local L-factor.

Having defined local L-factors at every prime, we can define

$$L(E, s) = \prod_{l \text{ prime}} L_l(\rho_{E,p}, l^{-s}).$$

How do we interpret this expression? First, we may think of it as a formal Dirichlet series with coefficients in $\mathbf{C}$, because the L-factors $L_l(\rho_{E,p}, T)$ are rational functions of T with rational coefficients. Second, we may think of it as defining a holomorphic function in the right half-plane $\mathrm{Re}(s) > 3/2$. Indeed, the Hasse–Weil theorem states that if we factorise the local L-factor $L_l(\rho_{E,p}, T)$ at a prime of good reduction as

$$L_l(\rho_{E,p}, T) = \frac{1}{1 - a_lT + lT^2} = \frac{1}{(1 - \alpha_lT)(1 - \beta_lT)},$$

then the numbers α_l, β_l (which satisfy $\alpha_l\beta_l = l$) both have absolute value $l^{1/2}$. An elementary argument then shows that the formal Dirichlet series defining $L(E, s)$ converges absolutely when $\mathrm{Re}(s) > 3/2$.

The L-function we have associated to the elliptic curve E is precisely the usual Hasse–Weil L-function, which appears in the formulation of the Birch–Swinnerton-Dyer conjecture [**Wil06**]. Some general remarks are now in order. In order to make sense of $L(\rho_{E,p}, s)$ as a function of a complex variable s, we have relied on the fact that the local L-factors $L_l(\rho_{E,p}, T)$

[5] Weil–Deligne representations are a technical but useful tool, so let us say a few words about them. By definition, the Weil group $W_{\mathbf{Q}_p}$ is the subgroup of $G_{\mathbf{Q}_p}$ consisting of automorphisms which induce an integer power of Frobenius on the residue field $\overline{\mathbf{F}}_p$. It is endowed with the topology which makes $I_{\mathbf{Q}_p}$ into an open subgroup (*not* the subspace topology of $G_{\mathbf{Q}_p}$). A Weil–Deligne representation over a field Ω (say of characteristic 0) is a pair (r, N) where $r : W_{\mathbf{Q}_p} \to \mathrm{GL}_n(\Omega)$ is a homomorphism with open kernel and $N : \Omega^n \to \Omega^n$ is a nilpotent linear map satisfying $r(\sigma)N = p^{-k}Nr(\sigma)$ for any $\sigma \in W_{\mathbf{Q}_p}$ inducing $x \mapsto x^{p^k}$ on the residue field $\overline{\mathbf{F}}_p$.

The utility of this notion is that it is independent of the topology of the base field (cf. [**Tat79**, §4.2]), and that one can associate to any potentially semi-stable representation (as we will see below, this includes all p-adic representations which arises from geometry) a Weil–Deligne representation.

are rational functions of T with integer coefficients (even though the representation $\rho_{E,p}$ is defined over $\mathbf{Q}_p$). We have also made use of the fact that $\rho_{E,p}|_{G_{\mathbf{Q}_p}}$ is potentially semi-stable, in order to be able to define the local L-factor at p. These properties cannot be expected to hold for an arbitrary p-adic representation $G_{\mathbf{Q}} \to \mathrm{GL}_n(\mathbf{Q}_p)$ (for example a non-integer power of the p-adic cyclotomic character would fail on both counts). For representations which do have these properties, we'd like to define local L-factors as for the p-adic representations attached to elliptic curves. If there is a number field $M/\mathbf{Q}$ such that all the local L-factors $L_l(\rho, T)$ have coefficients in M, then we can view $L(\rho, s) = \prod_{l \text{ prime}} L_l(\rho, l^{-s})$ as a formal Dirichlet series with coefficients in M. Given a choice of embedding $\iota : M \to \mathbf{C}$, we would then obtain a formal Dirichlet series $L(\iota\rho, s)$ with coefficients in $\mathbf{C}$. This discussion motivates the following definition:

Definition 2.2. *Let $K/\mathbf{Q}_p$ be a finite extension. We say that a continuous semi-simple representation $\rho : G_{\mathbf{Q}} \to \mathrm{GL}_n(K)$ is algebraic if $\rho|_{G_{\mathbf{Q}_p}}$ is potentially semi-stable and for all but finitely many prime numbers l, $\rho|_{G_{\mathbf{Q}_l}}$ is unramified.*

A very rich source of p-adic representations which are algebraic in this sense is the étale cohomology groups of algebraic varieties over $\mathbf{Q}$. This includes the p-adic Tate modules of elliptic curves. A basic computation in the étale cohomology of algebraic curves shows that there is an isomorphism of $\mathbf{Q}_p[G_{\mathbf{Q}}]$-modules[6]

$$H^1(\overline{E}, \mathbf{Q}_p) \cong \mathrm{Hom}_{\mathbf{Q}_p}(V_p E, \mathbf{Q}_p).$$

This computation explains a slight change in normalisations that we may as well mention now. We now write $\mathrm{Frob}_l \in G_{\mathbf{Q}_l} \subset G_{\mathbf{Q}}$ for the inverse of the element σ_l introduced previously, and call it the geometric Frobenius element. From now on the local L-factors we consider will be defined using geometric Frobenius elements as $L_l(\rho, T) = \det(1 - \mathrm{Frob}_l T : V^{I_l} \to V^{I_l})^{-1}$. The reason for making this change is that the L-function associated to e.g. the H^1 of an elliptic curve will agree with the L-function $L(E, s)$ introduced above. Geometric Frobenius elements are also more natural from the point of view of étale cohomology (where they can be compared with the action on étale cohomology induced by the geometric Frobenius endomorphism of the reduction of E modulo l).

In general, let X be a smooth proper algebraic variety over $\mathbf{Q}$, and let $\overline{X}$ be its base change to $\overline{\mathbf{Q}}$. The étale cohomology groups $H^*(\overline{X}, \mathbf{Q}_p)$ have the following properties:

- For all but finitely many primes l (including those primes $l \neq p$ where X has good reduction[7]), the action of $G_{\mathbf{Q}_l}$ on $H^*(\overline{X}, \mathbf{Q}_p)$ is unramified.

[6] Where $\overline{E}$ denotes the base change to our fixed algebraic closure $\overline{\mathbf{Q}}$ and the cohomology is p-adic étale cohomology.

[7] In the sense that X arises as the generic fibre of a smooth proper scheme over $\mathbf{Z}_l$.

- If w is an integer, then $H^w(\overline{X}, \mathbf{Q}_p)$ is pure of weight w: for any prime $l \neq p$ where X has good reduction, the eigenvalues of Frob_l on $H^w(\overline{X}, \mathbf{Q}_p)$ are algebraic numbers, all of whose complex conjugates have absolute value $l^{w/2}$.
- $H^*(\overline{X}, \mathbf{Q}_p)$ is a potentially semi-stable $\mathbf{Q}_p[G_{\mathbf{Q}_p}]$-module, in the sense of Fontaine.

The last two bullet points rely on some of the most important advances in arithmetic geometry of the 20th century. The purity of $H^w(\overline{X}, \mathbf{Q}_p)$ (which generalises Hasse's theorem for elliptic curves) is part of the Weil conjectures, proved by Deligne [**Del74**], while the potential semi-stability of $H^*(\overline{X}, \mathbf{Q}_p)$ follows from Tsuji's proof of Fontaine's conjecture C_{st} and de Jong's theory of alterations [**Tsu98, dJ96**]. The local L-factors associated to these representations were first defined (in the case $l \neq p$) by Serre [**Ser70**], who conjectured further that they have coefficients in $\mathbf{Q}$. Fontaine made the same conjecture also in the case $l = p$ (with the further expectation that $L_l(H^*(\overline{X}, \mathbf{Q}_p), T)$ is independent of p). It is known that $L_l(H^w(\overline{X}, \mathbf{Q}_p), T)$ has rational coefficients whenever X has good reduction at l. This is also known in some cases when X has bad reduction (including when $w = 1$, or X is an abelian variety, or $\dim X \leq 2$ [**Sai03**]) but not in general. In order to make unconditional statements, we can choose an isomorphism $\iota : \overline{\mathbf{Q}}_p \to \mathbf{C}$ and form the L-function $L(\iota H^w(\overline{X}, \mathbf{Q}_p), s)$; this will converge in a right half-plane, by purity. The conjectures in [**Ser70, Fon94b**] would imply that this L-function is independent of the choice of ι.

More generally, we can notice that the representations $H^w(\overline{X}, \mathbf{Q}_p)$ may be reducible, and decompose them into pieces. This leads to the following definition.

Definition 2.3. *Let $K/\mathbf{Q}_p$ be a finite extension and let $\rho : G_{\mathbf{Q}} \to \mathrm{GL}_n(K)$ be a continuous semi-simple representation. We say that ρ arises from geometry if there is a smooth proper variety X over $\mathbf{Q}$ and integers w, j such that ρ is isomorphic to a subquotient of $H^w(\overline{X}, \mathbf{Q}_p) \otimes_{\mathbf{Q}_p} K(j)$.*

It is reasonable to expect that if ρ arises from geometry then there is a number field M such that the local L-factors $L_l(\rho, T)$ (including for $l = p$) all have coefficients in M.

One special case that is worth remarking on is the case $w = 0$. Then $H^0(\overline{X}, \mathbf{Q}_p)$ may be identified with $\mathbf{Q}_p[\pi_0(\overline{X})]$, the permutation representation of $G_{\mathbf{Q}}$ on the set of connected components of $\overline{X}$. In particular, the action of $G_{\mathbf{Q}}$ factors through a finite quotient, and any representation $G_{\mathbf{Q}} \to \mathrm{GL}_n(\mathbf{Q}_p)$ with finite image appears as a submodule of $H^0(\overline{X}, \mathbf{Q}_p)$ for some X (which we may even take to be 0-dimensional). In this way, Artin representations may be viewed as the 'dimension 0' case of p-adic representations arising from geometry.

A very important perspective on the class of representations arising from geometry is given by the Fontaine–Mazur conjecture [**FM95**], which aims to give an internal classification of such representations. More precisely:

CONJECTURE 2.1. *Let $K/\mathbf{Q}_p$ be a finite extension, and let $\rho : G_{\mathbf{Q}} \to \mathrm{GL}_n(K)$ be a continuous semi-simple representation. Then ρ is algebraic if and only if it arises from geometry.*

(Both Definition 2.2 and Definition 2.2 extend to representations of G_M, where M is a number field, and the Fontaine–Mazur conjecture makes sense in this setting too.) Put another way, suppose given a continuous semi-simple representation $\rho : G_{\mathbf{Q}} \to \mathrm{GL}_n(K)$ which is unramified at all but finitely many prime numbers l. The Fontaine–Mazur conjecture asserts that the necessary condition for ρ to arise from geometry, namely that $\rho|_{G_{\mathbf{Q}_p}}$ is potentially semi-stable, is in fact also sufficient. This conjecture, and the question of how to classify the Galois representations that satisfy weakenings of this condition, have become very useful organizing principles in the arithmetic part of the Langlands programme. We will return to these ideas at the beginning of Part II of this article.

2.6. The Sato–Tate conjecture. We have motivated the importance of L-functions using their applications to results such as Dirichlet's theorem on primes in arithmetic progressions and the Chebotarev density theorem. The L-functions of p-adic representations arising from geometry can also conjecturally be applied this way. This was first suggested by Tate, in connection with the p-adic representations of self-products E^k of elliptic curves over $\mathbf{Q}$ [**Tat65**], and by Serre, who gave a beautiful statement in the generality of a pure motive over $\mathbf{Q}$ [**Ser16**].

Let us begin with the Sato–Tate conjecture. If E is an elliptic curve over $\mathbf{Q}$ and l is a prime of good reduction then purity implies, as we have already seen, that the integer

$$a_l = l + 1 - \#E(\mathbf{F}_l) = \alpha_l + \beta_l$$

is of absolute value $|a_l| \leq 2l^{1/2}$ (by the triangle inequality). Re-normalizing, we obtain a real number $a_l/2l^{1/2} \in [-1, 1]$, defined for all but finitely many prime numbers l. The Sato–Tate conjecture concerns the distribution of these numbers as l varies. To formulate it, we need to split into cases. The first case is where E has complex multiplication, in the sense that $\mathrm{End}(\overline{E})$ is strictly bigger than $\mathbf{Z}$. In this case, $\mathrm{End}(\overline{E}) \otimes_{\mathbf{Z}} \mathbf{Q} = M$ is an imaginary quadratic field, and one can show that the p-adic representation $\rho_{E,p}$ of E has the form $\mathrm{Ind}_{G_M}^{G_{\mathbf{Q}}} \chi$, for some p-adic character χ of $G_M = \mathrm{Gal}(\overline{\mathbf{Q}}/M)$. In particular, there is an additional symmetry: if the prime l is inert in M (as is the case for half of all prime numbers l, by Dirichlet's theorem!) then $\mathrm{tr}\, \rho_{E,p}(\mathrm{Frob}_l) = a_l = 0$. The remaining, generic, case is where E has no complex multiplication, in which case Serre showed that the image of

$\rho_{E,p} : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{Q}_p)$ is Zariski dense in GL_2 (and even of finite index in $\mathrm{GL}_2(\mathbf{Z}_p)$ [**Ser72**]).

The Sato–Tate conjecture concerns the case where E has no complex multiplication:

Conjecture 2.2. *Let E be an elliptic curve over $\mathbf{Q}$ without complex multiplication. Then the quantities $a_l/2l^{1/2}$ are equidistributed in $[-2,2]$ with respect to the Sato–Tate measure $\frac{2}{\pi}\sqrt{1-t^2}\,dt$: more precisely, for any continuous function $f : [-1,1] \to \mathbf{R}$, we have*

$$\lim_{X\to\infty} \frac{\sum_{l<X, l\nmid \Delta_E} f(a_l/2l^{1/2})}{\sum_{l<X, l\nmid\Delta_E} 1} = \frac{2}{\pi}\int_{t=-1}^{1} f(t)\sqrt{1-t^2}\,dt.$$

How is this analogous to the Chebotarev density theorem? That theorem states that if $\rho : G_{\mathbf{Q}} \to \mathrm{GL}_n(\mathbf{C})$ is a continuous representation, then the images $\rho(\sigma_l)$ of Frobenius elements are equidistributed within the conjugacy classes of $\rho(G_{\mathbf{Q}})$. This formulation does not quite make sense here since $\rho_{E,p}$ is a p-adic representation, but the Sato–Tate conjecture concerns the distribution of the real numbers $a_l/2l^{1/2}$. However, the polynomial $1 - a_l/l^{1/2}T + T^2$ can arise as the characteristic polynomial $\det(1-gT)$ of an element $g \in \mathrm{SU}_2(\mathbf{R})$, so we can ask if the characteristic polynomials $\det(1-\rho_{E,p}(\sigma_l)T)$ of Frobenius elements are equidistributed within the set of characteristic polynomials of elements of $\mathrm{SU}_2(\mathbf{R})$. The Sato–Tate conjecture asserts that this is indeed the case: the map

$$\mathrm{SU}_2(\mathbf{R}) \to [-1,1]$$
$$g \mapsto \frac{1}{2}\operatorname{tr} g$$

has fibres precisely the conjugacy classes of the group $\mathrm{SU}_2(\mathbf{R})$, and the Weyl integration formula for $\mathrm{SU}_2(\mathbf{R})$ implies that for any continuous function $f : [-1,1] \to \mathbf{R}$ there is an equality

$$\int_{G\in\mathrm{SU}_2(\mathbf{R})} f(\frac{1}{2}\operatorname{tr} g)\,dg = \frac{1}{2\pi}\int_{\theta=0}^{\pi} |e^{i\theta} - e^{-i\theta}|^2 f(\cos\theta)\,d\theta.$$

A change of variable now recovers the formulation in Conjecture 2.2.

As observed by Serre, this representation-theoretic formulation of the Sato–Tate conjecture also suggests a strategy for its proof. Choose, for each prime $l \nmid \Delta_E$, a representative $t_l \in \mathrm{SU}_2(\mathbf{R})$ of the conjugacy class of elements with characteristic polynomial $1 - a_l/l^{1/2}T + T^2$. In order to show that Conjecture 2.2 holds for every continuous function $f : [-1,1] \to \mathbf{R}$, it is enough (by the Peter–Weyl theorem) to show that for each irreducible representation $R : \mathrm{SU}_2(\mathbf{R}) \to \mathrm{GL}_n(\mathbf{C})$, we have

$$\lim_{X\to\infty} \frac{\sum_{l<X, l\nmid\Delta_E} (\operatorname{tr} R)(t_l)}{\sum_{l<X, l\nmid\Delta_E} 1} = \int_{g\in\mathrm{SU}_2(\mathbf{R})} (\operatorname{tr} R)(g)\,dt.$$

The right-hand side is computed by character orthogonality: it is 1 if R is the trivial representation, and 0 otherwise. The left-hand side is also easily seen to be 1 if R is the trivial representation, so we are led to the problem of showing the identity

$$\lim_{X\to\infty} \frac{\sum_{l<X, l\nmid\Delta_E} (\operatorname{tr} R)(t_l)}{\sum_{l<X, l\nmid\Delta_E} 1} = 0$$

for every non-trivial representation R of $\mathrm{SU}_2(\mathbf{R})$. Just as in the proofs of Dirichlet's theorem and the Chebotarev density theorem, this identity would follow if we could show that for each non-trivial representation R the L-function

$$L(E, R, s) = \prod_{l\nmid\Delta_E} \det(1 - R(t_l)l^{-s})^{-1},$$

a priori absolutely convergent and therefore holomorphic in the region $\mathrm{Re}(s) > 1$, admits a meromorphic continuation to some neighbourhood of the line $\mathrm{Re}(s) = 1$ which is holomorphic and non-vanishing on the line $\mathrm{Re}(s) = 1$.

This is our first point of contact with symmetric power L-functions! The non-trivial irreducible representations of $\mathrm{SU}_2(\mathbf{R})$ are easy to describe: we have the standard (or identity) representation of $\mathrm{SU}_2(\mathbf{R}) \subset \mathrm{GL}_2(\mathbf{C})$, and its symmetric powers $\mathrm{Sym}^m \mathbf{C}^2$ for each $m \geq 2$ – and each irreducible represenation of $\mathrm{SU}_2(\mathbf{R})$ is isomorphic to exactly one of these. The standard L-function $L(E, \mathbf{C}^2, s)$ (so-called because it corresponds to the standard representation) equals the shifted usual Hasse–Weil L-function $L(E, s+1/2)$, up to finitely many Euler factors. The higher symmetric power L-functions are genuinely new, but can similarly be described as shifts $L(\mathrm{Sym}^m \rho_{E,p}, s+m/2)$ of the L-functions associated to the symmetric power Galois representations

$$\mathrm{Sym}^m \rho_{E,p} : G_{\mathbf{Q}} \to \mathrm{GL}_{m+1}(\mathbf{Q}_p).$$

These representations also come from geometry: the Künneth formula implies that they can be realised inside the étale cohomology of the self-product E^m.

Serre has described a generalisation of the Sato–Tate conjecture in which the elliptic curve E may be replaced by an arbitrary smooth, proper algebraic variety X over $\mathbf{Q}$ (or more generally, pure motive over $\mathbf{Q}$). The key point is the definition of the 'Sato–Tate group' K, a compact Lie group which is the analogue of the group $\mathrm{SU}_2(\mathbf{R})$, which can be done either from the point of view of the motivic Galois group [**Ser94**] or from the point of view of p-adic representations [**Ser12**, Ch. 8]. In either case the generalisation of the Sato–Tate conjecture may be seen to follow from conjectural properties of the family of L-functions associated to a sequence of conjugacy classes of K and indexed by the set of irreducible representation of K.

2.7. Langlands L-functions. How do we actually go about proving the analytic properties of L-functions of p-adic representations needed to prove the Sato–Tate conjecture and its generalisations? The preceding discussion covers roughly the years 1924–1968 (the years spanning the publication of Artin's article [**Art24**] and Serre's book [**Ser68**]). During these years a parallel development was taking place, involving a different class of L-functions which do not obviously arise from representations of Galois groups. This eventually led to the definition, by Langlands, of families of L-functions indexed by algebraic group representations, in remarkable likeness to the L-functions appearing in Serre's criterion for the Sato–Tate conjecture to hold.

To begin this side of the story, we recall again the definition of Ramanujan's modular form

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

As described in the introduction to this article, Ramanujan made the following conjectures concerning the coefficients $\tau(n)$ of the q-expansion of Δ:

CONJECTURE 2.3. *Let* $L(\Delta, s) = \sum_{n \geq 1} \tau(n) n^{-s}$*. Then:*

(1) *There is an Euler product expansion* $L(\Delta, s) = \prod_{l \text{ prime}} (1 - \tau(l) l^{-s} + l^{11-2s})^{-1}$.
(2) *Let l be a prime number, and factorise* $1 - \tau(l) T + l^{11} T^2 = (1 - \alpha_l T)(1 - \beta_l T)$*. Then* $|\alpha_l| = |\beta_l| = l^{11/2}$.

The first part of the conjecture was proved soon afterwards by Mordell [**Mor20**], basically as a consequence of the fact that the fact that the modular form Δ is an eigenvector for the Hecke operators T_l – of which more in a moment. The second part of the conjecture remained unsolved for much longer. A strategy to prove it was suggested by Langlands as part of his functoriality conjectures [**Lan70**]. Langlands's suggestion was to think of the polynomial $1 - \tau(l) l^{-11/2} T + T^2$ as the determinant $\det(1 - t_l T)$ of a uniquely defined conjugacy class of semisimple elements $t_l \in \mathrm{SL}_2(\mathbf{C})$. For any non-trivial irreducible algebraic representation R of $\mathrm{SL}_2(\mathbf{C})$, one can then write down the L-function

$$L(\Delta, R, s) = \prod_{l \text{ prime}} \det(1 - l^{-s} R(t_l))^{-1}.$$

If each of these L-functions, a priori absolutely convergent in some right half-plane, admits an analytic continuation to the whole complex plane, then the Ramanujan conjecture $|\alpha_l| = |\beta_l| = l^{11/2}$ holds.[8] Since the non-trivial algebraic representations of $\mathrm{SL}_2(\mathbf{C})$ are precisely the symmetric powers of the standard representation $\mathbf{C}^2$, this is our second point of contact with symmetric power L-functions.

[8]This is an exercise using Landau's lemma, namely that if $\sum_{n \geq 1} a_n n^{-s}$ is a Dirichlet series with non-negative real coefficients, absolutely convergent in $\mathrm{Re}(s) > \sigma$ and which admits an analytic continuation to the half-plane $\mathrm{Re}(s) > \sigma_0$ for some $\sigma_0 < \sigma$, then the series is in fact absolutely convergent in the half-plane $\mathrm{Re}(s) > \sigma_0$.

Langlands's functoriality conjectures were in fact made in the context of automorphic forms on an arbitrary reductive group G (say over a number field M), and the good properties of these L-functions expected to follow as a consequence of the existence of functorial lifts, i.e. automorphic representations of various other reductive groups associated to homomorphisms of L-groups. For the sake of simplicity we explain some more of these ideas here just in the case where $G = \mathrm{GL}_m$ is a general linear group.

We first need to introduce the notion of automorphic representation of the group $\mathrm{GL}_m(\mathbf{A}_\mathbf{Q})$. We first describe this group a little bit more. First, the ring $\mathbf{A}_\mathbf{Q}^\infty = \prod_l' \mathbf{Q}_l$ of finite adeles is a restricted direct product, which contains $\prod_l \mathbf{Z}_l$ as an open subring. The ring of adeles is $\mathbf{A}_\mathbf{Q} = \mathbf{A}_\mathbf{Q}^\infty \times \mathbf{R}$. The group $\mathrm{GL}_m(\mathbf{A}_\mathbf{Q}^\infty) = \prod_l' \mathrm{GL}_m(\mathbf{Q}_l)$ may also be realised as a restricted direct product, with $\prod_l \mathrm{GL}_m(\mathbf{Z}_l)$ as an open subgroup, and $\mathrm{GL}_m(\mathbf{A}_\mathbf{Q}) = \mathrm{GL}_m(\mathbf{A}_\mathbf{Q}^\infty) \times \mathrm{GL}_m(\mathbf{R})$.

An automorphic representation π of $\mathrm{GL}_m(\mathbf{A}_\mathbf{Q})$ is a tensor product $\pi^\infty \otimes \pi_\infty$, where:

- π^∞ is an irreducible $\mathbf{C}[\mathrm{GL}_m(\mathbf{A}_\mathbf{Q}^\infty)]$-module which is smooth, in the sense that each vector $v \in \pi^\infty$ has open stabilizer.
- π_∞ is an irreducible admissible $(M_n(\mathbf{C}), O_n(\mathbf{R}))$-module (see e.g. [**Wal88**, Ch. 3]).
- The tensor product $\pi^\infty \otimes \pi_\infty$ is a subquotient of the space of automorphic forms on $\mathrm{GL}_m(\mathbf{A}_\mathbf{Q})$ (a subspace of the space of all functions on the quotient $\mathrm{GL}_m(\mathbf{Q})\backslash\mathrm{GL}_m(\mathbf{A}_\mathbf{Q})$, see [**BJ79**]).

Within the space of automorphic forms lies the subspace of cuspidal automorphic forms; an automorphic representation is said to be cuspidal if it is isomorphic to a subquotient of the space of cuspidal automorphic forms.

If $\pi = \pi^\infty \otimes \pi_\infty$ is an automorphic representation of $\mathrm{GL}_m(\mathbf{A}_\mathbf{Q})$, then π^∞ itself admits a restricted tensor product decomposition $\pi^\infty = \otimes_l' \pi_l$, where each π_l is an irreducible smooth representation of $\mathrm{GL}_m(\mathbf{Q}_l)$ and all but finitely many of these representations are unramified, in the sense that the space $\pi_l^{\mathrm{GL}_m(\mathbf{Z}_l)}$ is non-zero. As an example of an automorphic representation, we consider the function $\phi_\Delta : \mathrm{GL}_2(\mathbf{Q})\backslash\mathrm{GL}_2(\mathbf{A}_\mathbf{Q}) \to \mathbf{C}$ defined by the formula

$$\phi_\Delta(\gamma g^\infty g_\infty) = \Delta\left(\frac{ai+b}{ci+d}\right)(ci+d)^{-12},$$

where

$$\gamma \in \mathrm{GL}_2(\mathbf{Q}),\ g^\infty \in \mathrm{GL}_2(\widehat{\mathbf{Z}}),\ \text{and } g_\infty = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbf{R})^{\det>0}.$$

(This makes sense because any element of $\mathrm{GL}_2(\mathbf{A}_\mathbf{Q})$ can be written as a product $\gamma g^\infty g_\infty$; the fact that the value obtained is independent of the choice of expression is essentially equivalent to the fact that Δ is invariant under the classical weight 12 action of $\mathrm{SL}_2(\mathbf{Z})$.) The function ϕ_Δ is then a cuspidal automorphic form, and it generates an irreducible $\mathrm{GL}_2(\mathbf{A}_\mathbf{Q}^\infty) \times$

$(M_n(\mathbf{C}), O_n(\mathbf{R}))$-module which is a cuspidal automorphic representation π_Δ. The Hecke operators T_l which appear in the proof of the first part of Conjecture 2.3 may be seen to be a shadow of the action of the group $\mathrm{GL}_2(\mathbf{A}_\mathbf{Q}^\infty)$ on π_Δ. Although the point of view of automorphic representations of adele groups involves introducing a certain amount of new language and notation (some of which we have elided here), it has value even for studying classical modular forms, since many questions become much more transparent from the point of view of representation theory.

One example of this is the definition of the standard L-function associated to an automorphic representation π of $\mathrm{GL}_m(\mathbf{A}_\mathbf{Q})$. We can describe this in a way analogous to our definition of the L-function of a p-adic Galois representation by using the local Langlands correspondence for the groups $\mathrm{GL}_m(\mathbf{Q}_l)$. This is a bijection $\mathrm{rec}_{\mathbf{Q}_l}$ between the following two sets of objects:

- The set of isomorphism classes of irreducible smooth representations of $\mathrm{GL}_m(\mathbf{Q}_l)$.
- The set of isomorphism classes of Frobenius semi-simple Weil–Deligne representations (r, N) of $W_{\mathbf{Q}_p}$ of dimension n. (We have defined Weil–Deligne representations in footnote 5; the condition 'Frobenius semi-simple' means that the representation r is semi-simple.)

This bijection restricts to a bijection between unramified smooth representations and unramified Weil–Deligne representations. This unramified correspondence is all that was known to exist at the time Langlands's article [**Lan70**], but is enough to describe L-functions up to finitely many Euler factors. The standard L-function is given by the formula

$$L(\pi, s) = \prod_{l \text{ prime}} L_l(\mathrm{rec}_{\mathbf{Q}_l}(\pi_l), l^{-s})^{-1}.$$

An important point is that these standard L-functions are known to admit a meromorphic continuation to $\mathbf{C}$ and to satisfy a functional equation, which may be described explicitly. If $m > 1$ and π is cuspidal, then the continuation is even holomorphic everywhere. For example, the automorphic L-function $L(\pi_\Delta, s)$, which turns out to equal Ramanujan's L-function $L(\Delta, s) = \sum_{n \geq 1} \tau(n) n^{-s}$, admits a holomorphic continuation to $\mathbf{C}$ and satisfies the functional equation $\Lambda(\Delta, s) = \Lambda(\Delta, 12 - s)$, where

$$\Lambda(\Delta, s) = (2\pi)^{-s}\Gamma(s)L(\Delta, s).$$

This may be proved by using the expression

$$\Lambda(\Delta, s) = \int_{y=0}^{\infty} \Delta(iy) y^s \frac{dy}{y}$$

analogous to Riemann's expression for $\zeta(s)$ as an integral transform of the Jacobi theta function. Godement–Jacquet used analogous ideas to prove the continuation and functional equation of $L(\pi, s)$ in general [**GJ72**].

To go beyond the standard L-function, suppose given an algebraic representation $R : \mathrm{GL}_m \to \mathrm{GL}_n$ and define

$$L(\pi, R, s) = \prod_{l \text{ prime}} L_l(R \circ \mathrm{rec}_{\mathbf{Q}_l}(\pi_l), l^{-s})^{-1}.$$

Langlands's functoriality conjecture implies that these L-functions are the standard L-functions of automorphic representations of higher rank general linear groups. More precisely:

CONJECTURE 2.4. *There exists an automorphic representation* $\Pi = R_*(\pi)$ *of* $\mathrm{GL}_n(\mathbf{A}_\mathbf{Q})$ *such that for each prime number* l, $\mathrm{rec}_{\mathbf{Q}_l}(\Pi_l) = R \circ \mathrm{rec}_{\mathbf{Q}_l}(\pi_l)$.[9]

We see that, essentially by definition, we would have the equality

$$L(\pi, R, s) = L(R_*(\pi), s),$$

showing why functoriality would lead to the analytic continuation of the non-standard L-functions $L(\pi, R, s)$. In particular, taking $\pi = \pi_\Delta$ to be the automorphic representation of $\mathrm{GL}_2(\mathbf{A}_\mathbf{Q})$ associated to the Ramanujan Δ function, we see that the analytic continuation of the L-functions $L(\Delta, R, s)$ is implied by the following theorem, namely the existence of the Langlands functorial lift of π_Δ along the symmetric powers of the standard representation of GL_2:

THEOREM 2.4. *For each* $m \geq 1$, *there exists a cuspidal automorphic representation* Π *of* $\mathrm{GL}_{m+1}(\mathbf{A}_\mathbf{Q})$ *such that for every prime number* l, $\mathrm{rec}_{\mathbf{Q}_l}(\Pi_l) = \mathrm{Sym}^m \circ \mathrm{rec}_{\mathbf{Q}_l}(\pi_{\Delta,l})$.

This is a special case of the main theorem of [**NT21a**]. The most general statement we prove may be stated somewhat informally as follows:

THEOREM 2.5. *Let* f *be a holomorphic newform of weight* $k \geq 2$ *and let* π_f *be the associated cuspidal automorphic representation of* $\mathrm{GL}_2(\mathbf{A}_\mathbf{Q})$. *Suppose that* f *does not have CM (equivalently, there is no non-trivial Hecke character* $\chi : \mathbf{Q}^\times \backslash \mathbf{A}_\mathbf{Q}^\times \to \mathbf{C}^\times$ *such that* $\pi \cong \pi \otimes (\chi \circ \det)$*). Then for each* $m \geq 1$, *there exists a cuspidal automorphic representation* Π *of* $\mathrm{GL}_{m+1}(\mathbf{A}_\mathbf{Q})$ *such that for every prime number* l, $\mathrm{rec}_{\mathbf{Q}_l}(\Pi_l) = \mathrm{Sym}^m \circ \mathrm{rec}_{\mathbf{Q}_l}(\pi_l)$.

(The theorem is true, and much easier, also in the case where π_f does have CM, except the symmetric power lifts will no longer be cuspidal in this case.)

We end this section with some brief remarks concerning the general case of Langlands's functoriality conjectures, which concern liftings of automorphic representations associated to L-homomorphisms

$$^L H \to {}^L G$$

[9] One can also specify what Π_∞ should be after introducing the local Langlands correspondence $\mathrm{rec}_\mathbf{R}$ for $\mathrm{GL}_m(\mathbf{R})$ [**Lan89**], which classifies the irreducible admissible representations of $\mathrm{GL}_m(\mathbf{R})$ in terms of continuous semisimple representations of the Weil group $W_\mathbf{R} = \mathbf{C}^\times \sqcup \mathbf{C}^\times \jmath$, where $\jmath^2 = -1$ and $\jmath z \jmath^{-1} = \overline{z}$ for $z \in \mathbf{C}^\times$.

of L-groups of reductive groups H, G over a fixed number field M. As Langlands has emphasised, even the case where H is trivial and $G = \mathrm{GL}_n$ is interesting; in this case an L-homomorphism is essentially the data of a continuous representation $\rho : G_{\mathbf{Q}} \to \mathrm{GL}_n(\mathbf{C})$, and functoriality is equivalent to the strong Artin conjecture, which states that the Artin L-function $L(\rho, s)$ is in fact the standard L-function of an automorphic representation of $\mathrm{GL}_n(\mathbf{A}_{\mathbf{Q}})$. This basic case of Langlands's functoriality conjecture remains open (although some cases are known). The arguments in the second part of this article make constant use of known cases of functoriality for this broader class of L-homomorphisms (going beyond algebraic representations $\mathrm{GL}_m \to \mathrm{GL}_n$).

2.8. Reciprocity and the proof of the Sato–Tate conjecture. We have now seen two situations where families of symmetric power L-functions arise:

- Starting with an elliptic curve E over $\mathbf{Q}$ (say without CM), with associated p-adic representation $\rho_{E,p} : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{Q}_p)$, the symmetric power L-functions
$$L(\mathrm{Sym}^m \circ \rho_{E,p}, s).$$
The holomorphic continuation of these L-functions, non-vanishing on the line $\mathrm{Re}(s) = 1 + m/2$, would imply the Sato–Tate conjecture.
- Starting with a holomorphic newform f with associated cuspidal automorphic representation $\pi = \pi_f$ of $\mathrm{GL}_2(\mathbf{A}_{\mathbf{Q}})$ (say without CM), the symmetric power L-functions
$$L(\pi, \mathrm{Sym}^m, s).$$
Langlands functoriality would imply that they admit a holomorphic continuation to the entire complex plane.

In fact, one of these classes of L-functions contains the other! If f is a newform of weight 2 with rational coefficients, then there is an associated elliptic curve E_f such that
$$L(\mathrm{Sym}^m \circ \rho_{E,p}, s) = L(\pi, \mathrm{Sym}^m, s)$$
for each $m \geq 1$. This is the construction of Eichler and Shimura. More generally, Deligne [**Del71a**] showed that if f is a newform of weight $k \geq 2$ then for any isomorphism $\iota : \overline{\mathbf{Q}}_p \to \mathbf{C}$ there exists an associated p-adic representation $r_\iota(f) : G_{\mathbf{Q}} \to \mathrm{GL}_2(\overline{\mathbf{Q}}_p)$, and in fact it is the case that
$$L(\iota \,\mathrm{Sym}^m \circ r_\iota(f), s) = L(\pi_f, \mathrm{Sym}^m, s)$$
for each $m \geq 1$. We see that one path to the Sato–Tate conjecture is to follow the two steps:

- Show that if E is an elliptic curve over $\mathbf{Q}$, then there is a newform f such that $L(\pi_f, s) = L(E, s)$. This is the modularity conjecture for elliptic curves over $\mathbf{Q}$.

- Establish the existence of the symmetric power lifts $\mathrm{Sym}^m_*(\pi_f)$, as predicted by Langlands.

The modularity conjecture was proved for semistable elliptic curves over $\mathbf{Q}$ by Wiles and Taylor [**Wil95, TW95**], and in general by Breuil–Conrad–Diamond–Taylor [**BCDT01**]. The Sato–Tate conjecture was proved for most elliptic curves over $\mathbf{Q}$ by Clozel–Shepherd-Barron–Harris–Taylor in 2007 [**CHT08, Tay08, HSBT10**] and in general by Barnet-Lamb–Geraghty–Harris–Taylor in 2011 [**BLGHT11**]. However, the strategy used by these authors to prove Sato–Tate was different to the one described above: they established a weaker 'potential' version of Langlands functoriality for holomorphic newforms that suffices to establish the necessary properties of the L-functions needed for the Sato–Tate conjecture.

In order to explain this more carefully, it is helpful to first back up and describe the conjectural relation between Galois representations and automorphic representations. This combines conjectures of Langlands, Clozel, and Fontaine–Mazur, and includes the modularity conjecture for elliptic curves over $\mathbf{Q}$ (and more generally, abelian varieties over any number field) as a very special case. Which objects will participate in this correspondence? We have already singled out irreducible algebraic Galois representations $\rho : G_{\mathbf{Q}} \to \mathrm{GL}_m(\overline{\mathbf{Q}}_p)$ as the right objects on the Galois side to which to attach L-functions. We have also asserted that automorphic representations of $\mathrm{GL}_m(\mathbf{A}_{\mathbf{Q}})$ have associated L-functions. However, it is not the case that every automorphic representation should be associated to a Galois representation. The simplest examples arise in the case $m = 1$, in which case automorphic representations may be thought of equivalently as continuous Hecke characters $\chi : \mathbf{Q}^\times \backslash \mathbf{A}_{\mathbf{Q}}^\times \to \mathbf{C}^\times$. There is no sensible way to associate a p-adic Galois representation to e.g. the character $x \in \mathbf{A}_{\mathbf{Q}}^\times \mapsto \|x\|^\alpha$, for a transcendental complex number α.[10] More interesting examples are given by the automorphic representations of $\mathrm{GL}_2(\mathbf{A}_{\mathbf{Q}})$ generated by non-algebraic Maass forms; see [**Gel75**, §7] for explicit examples.

The right automorphic representations for this purpose are the algebraic ones. Algebraicity is a condition on the infinite component π_∞ of an automorphic representation π, which may be phrased using the local Langlands correspondence for $\mathrm{GL}_m(\mathbf{R})$:[11]

DEFINITION 2.4. *Let π be an automorphic representation of $\mathrm{GL}_m(\mathbf{A}_{\mathbf{Q}})$. We say that π is algebraic if $\mathrm{rec}_{\mathbf{R}}(\pi_\infty |\det|^{(1-m)/2})|_{\mathbf{C}^\times}$ is a sum of characters of the form $z \mapsto z^a \overline{z}^b$ where a, b are integers.*

[10] By contrast, when α is an integer, reciprocity for GL_1 (which is just class field theory) associates to this character an integral power of the p-adic cyclotomic character.

[11] Here we follow the definition given by Clozel [**Clo90**]. The twist by $|\cdot|^{(1-m)/2}$ is included in order to ensure that the class of algebraic representations includes those that contribute to the cohomology of congruence subgroups of $\mathrm{GL}_n(\mathbf{Z})$. In the article [**BG14**], Buzzard and Gee consider different possible notions of algebraicity, both for GL_n and other reductive groups, all of which however may ultimately be related by character twists.

A similar definition applies to automorphic representations of $\mathrm{GL}_n(\mathbf{A}_M)$ for any number field M, making it possible to formulate the following conjecture:

CONJECTURE 2.5 (Reciprocity conjecture). *Let M be a number field, let p be a prime number, and let $\iota : \overline{\mathbf{Q}}_p \to \mathbf{C}$ be an isomorphism. Then there is a unique bijection between the following two sets of objects:*

(1) *The set of irreducible algebraic representations $\rho : G_M \to \mathrm{GL}_m(\overline{\mathbf{Q}}_p)$, up to isomorphism.*
(2) *The set of algebraic cuspidal automorphic representations π of $\mathrm{GL}_m(\mathbf{A}_M)$.*

with the following property: if ρ and π correspond, then for each finite place v of M, $\iota^{-1}\mathrm{rec}_{M_v}(\pi_v|\det|^{(1-m)/2}) \cong \mathrm{WD}(\rho|_{G_{M_v}})$.[12] *In particular, if ρ and π correspond then $L(\iota\rho, s) = L(\pi, s - \frac{m-1}{2})$.*

We remark that it can be seen already at this point that there is at most one bijection with these properties. The strong multiplicity one theorem [**JS81**] implies that π^∞ determines π_∞, so there is at most one automorphic representation corresponding to any given ρ. On the other hand, the Chebotarev density theorem implies that the character of an algebraic p-adic representation ρ is determined by the characteristic polynomials $\det(1 - T\rho(\mathrm{Frob}_v))$ at unramified places v. If ρ is irreducible, then it is itself determined up to isomorphism by its character. If π and ρ are related, then we usually write $\rho = r_\iota(\pi)$.

We say that a (say irreducible, algebraic) p-adic Galois representation ρ is automorphic if it corresponds to an algebraic cuspidal automorphic representation π of $\mathrm{GL}_m(\mathbf{A}_M)$. The reciprocity conjecture suggests an alternate path to proving Langlands functoriality for algebraic automorphic representations of general linear groups. Many conjectured functorial properties of automorphic representations become transparent after transporting to the Galois side. If $R : \mathrm{GL}_m \to \mathrm{GL}_M$ is an algebraic representation, and π corresponds to ρ, then the existence of the functorial lift $R_*(\pi)$ is implied by the automorphy of the representation $R \circ \rho$. Another expected property of automorphic representations is base change, or in other words a transfer from automorphic representations of $\mathrm{GL}_m(\mathbf{A}_M)$ to automorphic representations of $\mathrm{GL}_m(\mathbf{A}_{M'})$ for any extension M'/M of number fields. On the Galois side, this corresponds simply to replacing ρ by its restriction $\rho|_{G_{M'}}$ to a finite index subgroup.

We can now explain how the Sato–Tate conjecture for elliptic curves over $\mathbf{Q}$ was proved: namely, by establishing that the symmetric power representations $\mathrm{Sym}^m \rho_{E,p}$ are potentially automorphic, in the sense that there

[12] Here WD denotes the Weil–Deligne representation (see footnote 5) associated to $\rho|_{G_{M_v}}$ using the recipe described in [**Tat79**]. Since ι is not continuous, Weil–Deligne representations are useful here as a way to describe Galois representations in a way independent of the topology of the base field. If $\rho|_{G_{M_v}}$ is unramified, then we simply have $\mathrm{WD}(\rho|_{G_{M_v}}) = \rho|_{W_{M_v}}$.

exists a finite Galois extension $M/\mathbf{Q}$ (perhaps depending on m) such that the Galois representation $\mathrm{Sym}^m \rho_{E,p}|_{G_M}$ is automorphic. Using Brauer's induction theorem, one can express the L-function $L(\iota \,\mathrm{Sym}^m \rho_{E,p}, s)$ as a ratio of automorphic L-functions associated to automorphic representations of $\mathrm{GL}_{m+1}(\mathbf{A}_{M_i})$, for some intermediate fields $M/M_i/\mathbf{Q}$. Using the known analytic continuation of these automorphic L-functions, one obtains the meromorphic continuation of $L(\iota \,\mathrm{Sym}^m \rho_{E,p}, s)$, and crucially the required properties (holomorphy and non-vanishing) on the line $\mathrm{Re}(s) = 1 + m/2$. However, one does not in this way have any control over the poles in the region $\mathrm{Re}(s) < 1 + m/2$.

2.9. Evidence for the reciprocity conjecture. Historically the reciprocity conjecture has been attacked in two stages: first, show that any automorphic representation π in a given class admits an associated p-adic representation $r_\iota(\pi)$. Then, try to show that the induced map $\pi \mapsto r_\iota(\pi)$ is surjective. The greatest progress so far has been for the class of regular algebraic objects.

Definition 2.5. *Let $K/\mathbf{Q}_p$ be a finite extension. An algebraic representation $\rho : G_{\mathbf{Q}} \to \mathrm{GL}_n(K)$ is said to be regular algebraic if the multiset of Hodge–Tate numbers of $\rho|_{G_{\mathbf{Q}_p}}$ is in fact a set, i.e. is multiplicity-free.*

This condition holds for the p-adic representations associated to elliptic curves, but not for the p-adic representations associated to $H^1(\overline{A}, \mathbf{Q}_p)$ for abelian varieties A of dimension $g > 1$ (in which case the multiset of Hodge–Tate numbers has elements 0, 1, each appearing with multiplicity g).

Definition 2.6. *An algebraic automorphic representation π of $\mathrm{GL}_n(\mathbf{A}_{\mathbf{Q}})$ is said to be regular algebraic if the restriction $\mathrm{rec}_{\mathbf{R}}(\pi_\infty |\det|^{(1-n)/2})|_{\mathbf{C}^\times}$ is multiplicity-free.*

A refined version of the reciprocity conjecture would describe the expected relation between the Hodge–Tate numbers of ρ and the composition factors of $\mathrm{rec}_{\mathbf{R}}(\pi_\infty |\det|^{(1-n)/2})|_{\mathbf{C}^\times}$, explaining why these two definitions should match up. One can generalise these definitions to a general base number field M. If $\rho : G_M \to \mathrm{GL}_n(\overline{\mathbf{Q}}_p)$ is an algebraic representation, then there is one associated multiset $H_\tau(\rho)$ of Hodge–Tate numbers for every embedding $\tau : M \to \overline{\mathbf{Q}}_p$. If $\iota : \overline{\mathbf{Q}}_p \to \mathbf{C}$ is an isomorphism, then this set would in turn be related to the Langlands parameter of π_v, where v is the infinite place of M corresponding to the embedding $\iota\tau : M \to \mathbf{C}$.

Let us make this more explicit in the case $n = 2$. The regular algebraic cuspidal automorphic representations of $\mathrm{GL}_2(\mathbf{A}_{\mathbf{Q}})$ are precisely the ones associated to holomorphic newforms of weight $k \geq 2$ (see [**Clo90**, §1.2.3]). As mentioned above, Deligne showed [**Del71a**] how to construct the p-adic representations associated to such newforms inside the étale cohomology of a p-adic local system on a modular curve, by decomposing this cohomology under the action of the adele group $\mathrm{GL}_2(\mathbf{A}^\infty_{\mathbf{Q}})$. The first major success in showing a broad class of regular algebraic p-adic representations are automorphic came in the works [**Wil95, TW95**] on the modularity conjecture for elliptic curves: they proved the first *automorphy lifting theorems*. Since the idea of automorphy lifting theorems will be very important in the second part of this article, we expand on this here.

If M is a number field and $\rho : G_M \to \mathrm{GL}_m(\overline{\mathbf{Q}}_p)$ is a continuous representation, then we can find (using the compactness of G_M) a finite extension $K/\mathbf{Q}_p$ such that, after replacing ρ by a conjugate, ρ takes values in $\mathrm{GL}_m(\mathcal{O}_K)$ (here $\mathcal{O}_K$ is the ring of integers of K). Pushing forward along the homomorphism $\mathcal{O}_K \to \overline{\mathbf{F}}_p$, we obtain a representation $G_M \to \mathrm{GL}_m(\overline{\mathbf{F}}_p)$. This representation might depend on the choice of conjugate of ρ valued in $\mathrm{GL}_m(\mathcal{O}_K)$, but its semisimplification is independent, up to isomorphism, of any choices, and we write $\overline{\rho} : G_M \to \mathrm{GL}_m(\overline{\mathbf{F}}_p)$ for this semisimplified residual representation. The archetypal automorphy lifting theorem takes the following form:

THEOREM 2.6 (Ideal). *Let M be a number field, and let $\rho, \rho' : G_M \to \mathrm{GL}_m(\overline{\mathbf{Q}}_p)$ be irreducible regular algebraic Galois representations. Suppose that the following conditions are satisfied:*

(1) *ρ' is automorphic.*
(2) *There is an isomorphism $\overline{\rho} \cong \overline{\rho}'$.*

Then ρ is automorphic.

Unconditional theorems of this type, proved under varying additional technical conditions on ρ and ρ', are the main tool we have in proving the automorphy of Galois representations. The first automorphy lifting theorems in the case $M = \mathbf{Q}$, $m = 2$ included the requirement that $\overline{\rho}$ be irreducible and ρ, ρ' have Hodge–Tate numbers 0, 1 (a necessary condition to arise from a newform of weight 2). The automorphy of the residual representation $\overline{\rho}$ was verified, in the case that ρ arises from an elliptic curve, using the fact that the image is constrained to lie in the soluble group $\mathrm{GL}_2(\mathbf{F}_3)$ – in fact this was deduced from a known case of the Artin conjecture [**Tun81**]. The difficulty in verifying this residual automorphy hypothesis in general is one of the main challenges in successfully applying automorphy lifting theorems and is the reason that the proof of the Sato–Tate conjecture given in [**BLGHT11**] goes by the route of establishing only *potential* automorphy of the symmetric power Galois representations.

Two advances have led to very strong results towards the surjectivity part of the reciprocity conjecture for regular algebraic representations of $\mathrm{GL}_2(\mathbf{A}_\mathbf{Q})$. First, the proof of Serre's modularity conjecture by Khare–Wintenberger [**KW09a, KW09b**] implies that any irreducible residual representation $\overline{\rho} : G_\mathbf{Q} \to \mathrm{GL}_2(\overline{\mathbf{F}}_p)$[13] is automorphic. Second, powerful automorphy lifting theorems have been proved by Kisin [**Kis09a**] and Emerton

[13]Satisfying the necessary condition to arise from a regular algebraic automorphic representation, namely oddness.

([**Eme**], generalised further by Pan [**Pan21**]) using the p-adic local Langlands correspondence for $\mathrm{GL}_2(\mathbf{Q}_p)$.

Going beyond the case of $\mathrm{GL}_2(\mathbf{A}_\mathbf{Q})$ presents myriad technical difficulties. The main source of Galois representations in general is the étale cohomology of Shimura varieties. We recall that, in Deligne's formulation [**Del71b**], a Shimura variety is a variety (or rather, family of varieties), defined over a number field M and attached to a *Shimura datum* (G, X) consisting of a reductive group G over $\mathbf{Q}$ and a homogeneous space X for $G(\mathbf{R})$ satisfying certain axioms. Modular curves are the Shimura varieties associated to the pair $(\mathrm{GL}_2, \mathbf{C} - \mathbf{R})$. The étale cohomology of Shimura varieties (or more generally p-adic local systems on Shimura varieties) receives an action of $G_M \times G(\mathbf{A}_\mathbf{Q}^\infty)$. The basic idea, laid out in [**Lan77**], is to analyse the cohomology by understanding the action of G_M on the $G(\mathbf{A}_\mathbf{Q}^\infty)$-isotypic pieces in terms of automorphic representations of $G(\mathbf{A}_\mathbf{Q})$, by comparing the Grothendieck–Lefschetz trace formula in étale cohomology with the Arthur–Selberg trace formula in the theory of automorphic forms.

The Shimura varieties with the best understood cohomology (beyond the case of Shimura curves) are those attached to unitary groups of CM number fields M (such as imaginary quadratic fields). One prediction of Langlands functoriality is that the automorphic representations of these unitary groups should be related to automorphic representations π of $\mathrm{GL}_m(\mathbf{A}_M)$ which are conjugate self-dual, in the sense that the contragredient of π is isomorphic to its image under the automorphism of $\mathrm{GL}_m(\mathbf{A}_M)$ induced by complex conjugation on M. After several decades of work by many mathematicians, beginning with that of Kottwitz [**Kot92**] and Clozel [**Clo91**], the Galois representations associated to regular algebraic, cuspidal, conjugate self-dual (RACSDC) automorphic representations of $\mathrm{GL}_m(\mathbf{A}_M)$ are now known to exist in complete generality and to satisfy the expected local-global compatibility at every place.[14] The case of essentially self-dual representations of $\mathrm{GL}_m(\mathbf{A}_{M'})$, where M' is a totally real field, can often be reduced to this setting using a combination of base change and patching results (such as the one exposited in [**Sor20**]). Strong automorphy lifting theorems are also available for conjugate self-dual Galois representations; see especially [**BLGGT14, PT15**] for automorphy lifting theorems and potential automorphy theorems for compatible systems for Galois representations which go far beyond what is needed to prove the Sato–Tate conjecture for elliptic curves over totally real number fields. Our proofs of symmetric power functoriality for automorphic representations of $\mathrm{GL}_2(\mathbf{A}_\mathbf{Q})$ take place in this powerful and flexible context.

Some work has also been done outside the RACSDC context, both for Galois representations which are algebraic but not regular algebraic (starting

[14] We cannot begin to survey this story here, but see [**Shi20**] for a description of these developments, which rely on many important advances in arithmetic geometry and harmonic analysis.

with [DS74] and continuing up to the recent work [BCGP21], which establishes the potential automorphy of the Galois representations associated to abelian surfaces over totally real fields) and for Galois representations over CM fields which are regular algebraic but not conjugate self-dual (as in [ACC+18], which proves the Sato–Tate conjecture for elliptic curves over CM fields). In both cases the needed Galois representations, which in general do not appear in the étale cohomology of Shimura varieties [JT20], are constructed as p-adic limits of Galois representations which *do* appear in the étale cohomology of Shimura varieties. Understanding the limits of these techniques is an interesting topic for future work but leads in a different direction to the one of interest for these notes. We invite the reader to look at the survey [Cal21] for more about these exciting recent developments.

2.10. Our main theorem. Let us state again our main theorem.

THEOREM 2.7. *Let π be a regular algebraic, cuspidal automorphic representation of $\mathrm{GL}_2(\mathbf{A}_{\mathbf{Q}})$, without CM. Then for each $m \geq 1$, there exists a regular algebraic, cuspidal automorphic representation Π of $\mathrm{GL}_{m+1}(\mathbf{A}_{\mathbf{Q}})$ such that for every prime number l, $\mathrm{rec}_{\mathbf{Q}_l}(\Pi_l) = \mathrm{Sym}^m \circ \mathrm{rec}_{\mathbf{Q}_l}(\pi_l)$.*

Here are two important corollaries:

COROLLARY 2.1.

(1) *For each $m \geq 1$, the L-function $L(\Delta, \mathrm{Sym}^m, s)$ admits an analytic continuation to $\mathbf{C}$.*
(2) *Let E be an elliptic curve over $\mathbf{Q}$. Then for each $m \geq 1$, the L-function $L(E, \mathrm{Sym}^m, s)$ admits an analytic continuation to $\mathbf{C}$.*

We emphasise that the existence of a meromorphic continuation of these L-functions (and the fact that they satisfy functional equations) is already a consequence of the potential automorphy of the associated symmetric power Galois representations. This potential automorphy also implies the Sato–Tate conjecture. Our results have applications beyond this. For example, Thorner [Tho21b] has used Theorem 2.7 to give a form of the Sato–Tate conjecture with an effective rate of convergence.

Corollary 2.1 answers a question asked more than 50 years ago. Our proof, however, relies on many ideas in the theory of p-adic automorphic forms that have been developed much more recently. We will give an overview of this proof in the second part of this article.

3. Part II

3.1. First ideas on how to prove the main theorem. The proof of Theorem 2.7 occupies the articles [NT21a, NT21b] which in turn rely on the papers [NT20, ANT20, AT20]. In this introductory section, we give an overview of some of the themes that inform the most important arguments in these papers, focusing on the case of symmetric power functoriality for level 1 modular forms (such as Δ); as we will see, this special case serves as a starting point for the proof of the general case.

Our strategy, as suggested by the reciprocity conjecture, will be to prove the automorphy of the Galois representations $\mathrm{Sym}^m r_\iota(\pi)$ (for regular algebraic cuspidal automorphic representations π of $\mathrm{GL}_2(\mathbf{A}_\mathbf{Q})$, and for some choice of prime number p and isomorphism $\iota : \overline{\mathbf{Q}}_p \to \mathbf{C}$). The main tools will be deformation theory and the p-adic variation of p-adic Galois representations. We now introduce some of the basic objects in this theory.

We fix a finite extension $K/\mathbf{Q}_p$ with ring of integers $\mathcal{O}$, uniformizer ϖ, and residue field $\mathcal{O}/(\varpi) = k$. Suppose given a number field M and a continuous representation $\overline{\rho} : G_M \to \mathrm{GL}_m(k)$. Let $\mathcal{C}_\mathcal{O}$ denote the category of complete Noetherian local $\mathcal{O}$-algebras A with residue field k (i.e. equipped with a local homomorphism $\mathcal{O} \to A$ inducing an isomorphism $\mathcal{O}/\varpi\mathcal{O} = k \cong A/\mathfrak{m}_A$).

Definition 3.1. *A lifting of $\overline{\rho}$ to an object $A \in \mathcal{C}_\mathcal{O}$ is a homomorphism $\rho_A : G_M \to \mathrm{GL}_m(A)$ such that the composite map*

$$G_M \to \mathrm{GL}_m(A) \to \mathrm{GL}_m(k)$$

induced by reduction modulo $\mathfrak{m}_A$ equals $\overline{\rho}$. A deformation of $\overline{\rho}$ is a $\ker(\mathrm{GL}_m(A) \to \mathrm{GL}_m(k))$*-conjugacy class of liftings.*

Mazur [**Maz89**] initiated Galois deformation theory by defining deformation functors of Galois representations and studying them using Galois cohomology. For example, suppose we fix a finite set S of places of M, including the p-adic places of M, and let

$$\mathrm{Def}_{\overline{\rho},S} : \mathcal{C}_\mathcal{O} \to \mathrm{Sets}$$

denote the functor which associates to any $A \in \mathcal{C}_\mathcal{O}$ the set of deformations of $\overline{\rho}$ to A which are unramified outside S. Mazur showed that if $\overline{\rho}$ is absolutely irreducible then the functor $\mathrm{Def}_{\overline{\rho},S}$ is representable. The spectrum $\mathrm{Spec}\, R_{\overline{\rho},S}$ of the representing object (or perhaps, its rigid generic fibre in the sense of Berthelot [**dJ95**, §7.1]) then deserves to be called the universal deformation space of $\overline{\rho}$. Mazur also introduced a tangent-obstruction theory for $\mathrm{Def}_{\overline{\rho},S}$: there is a canonical isomorphism[15]

$$H^1(M_S/M, \mathrm{ad}\,\overline{\rho}) \cong \mathrm{Hom}_k(\mathfrak{m}_{R_{\overline{\rho},S}}/(\varpi, \mathfrak{m}^2_{R_{\overline{\rho},S}}), k),$$

and using Tate's Euler characteristic formula in Galois cohomology one can show that there is a presentation

$$R_{\overline{\rho},S} \cong \mathcal{O}[\![X_1, \dots, X_g]\!]/(f_1, \dots, f_r), \tag{3.1}$$

where $g - r = [M : \mathbf{Q}]n^2 + 1 - \sum_{v|\infty} \dim_k H^0(M_v, \mathrm{ad}\,\overline{\rho})$. Here we use the traditional notation $\mathrm{ad}\,\overline{\rho}$ for the composition of $\overline{\rho} : G_M \to \mathrm{GL}_n(k)$ with the

[15] Here the H^1 is the continuous group cohomology of $\mathrm{Gal}(M_S/M)$, the profinite Galois group of the maximal subextension M_S of $\overline{M}/M$ which is unramified outside S. We use similar notation where M_S/M is replaced by M_v for a place v of M (and $\mathrm{Gal}(M_S/M)$ by the absolute Galois group G_{M_v}).

adjoint representation $\mathrm{GL}_n(k) \to \mathrm{End}_k(M_n(k))$. This gives in particular a lower bound on the Krull dimension of the ring $R_{\overline{\rho},S}$, which in many cases one might expect to be an equality. For example when $n = 2$, $M = \mathbf{Q}$, and $\overline{\rho}$ is odd, we obtain the formula (cf. [**Maz89**, Corollary 3])

$$\dim R_{\overline{\rho},S} \geq 4.$$

With refinements, Mazur's deformation theory has become a very effective tool for proving the automorphy of p-adic Galois representations. Let us describe two such approaches that have been applied in the case $m = 2$, $M = \mathbf{Q}$, and under the assumption that $\overline{\rho}$ is irreducible (so that $R_{\overline{\rho},S}$ exists). In the first, which is the approach of [**Wil95**], one first cuts down $R_{\overline{\rho},S}$ by imposing conditions from p-adic Hodge theory. For example, if $\overline{\rho} = \overline{\rho}_{E,p}$ for an elliptic curve E with good reduction at an odd prime p, then one can define a quotient $R^{fl}_{\overline{\rho},S}$ of $R_{\overline{\rho},S}$, which represents the functor of deformations of $\overline{\rho}$ such that $\rho|_{G_{\mathbf{Q}_p}}$ comes from a finite flat group scheme over $\mathbf{Z}_p$. This property is in particular satisfied for the lifts $G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{Z}_p)$ of $\overline{\rho}$ which come from elliptic curves with good reduction at p, or more generally those lifts $G_{\mathbf{Q}} \to \mathrm{GL}_2(\overline{\mathbf{Z}}_p)$ of $\overline{\rho}$ arising from newforms of weight 2 and level prime to p. We are led to consider a homomorphism

$$R_{\overline{\rho},S} \to \mathbf{T}(H^1(\Gamma(N_S), \mathcal{O}))_{\mathfrak{m}},$$

where $H^1(\Gamma(N_S), \mathcal{O})$ is the group cohomology of the congruence subgroup $\Gamma_1(N_S) \leq \mathrm{SL}_2(\mathbf{Z})$ of level N_S depending on S with $\mathcal{O}$-coefficients, $\mathbf{T}(H^1(\Gamma(N_S), \mathcal{O}))$ is the subalgebra of $\mathrm{End}_{\mathcal{O}}(H^1(\Gamma(N_S), \mathcal{O}))$ generated by the Hecke operators T_l for prime numbers $l \notin S$, and $\mathfrak{m} \subset \mathbf{T}(H^1(\Gamma(N_S), \mathcal{O}))$ is the maximal ideal which is the kernel of the homomorphism $\mathbf{T}(H^1(\Gamma(N_S), \mathcal{O})) \to k$ which sends T_l to $\mathrm{tr}\, \overline{\rho}(\mathrm{Frob}_l)$. The existence of this maximal ideal is equivalent to the assertion that $\overline{\rho}$ is the residual representation attached to an automorphic representation π which contributes to $H^1(\Gamma(N_S), \mathcal{O})$. The homomorphism from $R_{\overline{\rho},S}$ to the integral Hecke algebra can be constructed, using Carayol's lemma [**Car94**], by gluing together the $\overline{\mathbf{Q}}_p$-representations $r_\iota(\pi)$ for those automorphic representations π which contribute to $H^1(\Gamma(N_S), \mathcal{O})_{\mathfrak{m}}$. Local-global compatibility at the prime p for the representations $r_\iota(\pi)$ implies that this homomorphism factors through a surjective homomorphism

$$R^{fl}_{\overline{\rho},S} \to \mathbf{T}(H^1(\Gamma_1(N_S), \mathcal{O}))_{\mathfrak{m}},$$

and to prove an automorphy lifting theorem here one wants to prove that this map is an isomorphism (or at least induces a bijection on $\overline{\mathbf{Q}}_p$-points). The ring $\mathbf{T}(H^1(\Gamma(N_S), \mathcal{O}))_{\mathfrak{m}}$ is a finite flat $\mathcal{O}$-algebra; a computation analogous to the one leading to the presentation (3.1) implies that the ring $R^{fl}_{\overline{\rho},S}$ has 'expected dimension' 1, but we know rather little about it a priori. Several tools were introduced in the papers [**Wil95, TW95**] in order to prove that such maps are isomorphisms, but the one that has had the most lasting impact is the Taylor–Wiles method. The basic idea of the Taylor–Wiles method is to start with a presentation

$$\mathcal{O}[\![X_1, \dots, X_{g+1}]\!]/(f_1, \dots, f_g) \twoheadrightarrow R^{fl}_{\overline{\rho},S}$$

and consider auxiliary sets Q of prime numbers such that, allowing ramification at the primes of Q, the maps

$$R^{fl}_{\overline{\rho},S\cup Q} \to \mathbf{T}(H^1(\Gamma_1(N_{S\cup Q}), \mathcal{O}))_{\mathfrak{m}}$$

can be made to fit into a diagram

$$\begin{array}{ccccc} \mathcal{O}[\![X_1, \dots, X_{g+1}]\!] & \twoheadrightarrow & R^{fl}_{\overline{\rho},S\cup Q} & \longrightarrow & \mathbf{T}(H^1(\Gamma_1(N_{S\cup Q}), \mathcal{O}))_{\mathfrak{m}} \\ & & \downarrow & & \downarrow \\ & & R^{fl}_{\overline{\rho},S} & \longrightarrow & \mathbf{T}(H^1(\Gamma_1(N_S), \mathcal{O}))_{\mathfrak{m}}. \end{array}$$

This means in essence that the tangent space of $R^{fl}_{\overline{\rho},S\cup Q}$ is the same size as that $R^{fl}_{\overline{\rho},S}$; this is equivalent, thanks to duality theorems in Galois cohomology, to the vanishing of a certain dual Selmer group (with coefficients in the Tate dual $\operatorname{ad}\overline{\rho}(1)$), and can be arranged under some hypotheses using a clever application of the Chebotarev density theorem. Roughly speaking, the idea is that as Q varies, the rings $\mathbf{T}(H^1(\Gamma_1(N_{S\cup Q}), \mathcal{O}))_{\mathfrak{m}}$ become 'closer and closer' approximations to $\mathcal{O}[\![X_1, \dots, X_{g+1}]\!]$ – this requires a kind of lower bound on the growth of the spaces $H^1(\Gamma_1(N_{S\cup Q}), \mathcal{O})_{\mathfrak{m}}$ of modular forms as Q varies. Since $R^{fl}_{\overline{\rho},S\cup Q}$ is sandwiched between $\mathcal{O}[\![X_1, \dots, X_{g+1}]\!]$ and the Hecke algebra $\mathbf{T}(H^1(\Gamma_1(N_{S\cup Q}), \mathcal{O}))_{\mathfrak{m}}$, this should lead to useful information about the relation between the Galois deformation ring R and Hecke algebra $\mathbf{T}$.

What are the drawbacks to using this approach to prove the automorphy of representations of type $\operatorname{Sym}^m r_\iota(\pi)$? The most significant is that using the Taylor–Wiles method requires the residual representation $\overline{\rho}$ to be quite non-degenerate, not only so that the ring $R_{\overline{\rho},S}$ exists but also so that we can 'kill the dual Selmer group'. This will force us to restrict to primes $p \geq m+1$ (as otherwise the residual representation is forced by to reducible, because Sym^m is a reducible representation of GL_2 in characteristic $p < m+1$). Then there is the separate issue of establishing the residual automorphy of these residual representations $G_{\mathbf{Q}} \to \mathrm{GL}_{m+1}(k)$, which generally have insoluble image – which seems hard, if not impossible. We would rather be able to take $p = 2$, $\pi = \pi_\Delta$, in which case the residual representation $\operatorname{Sym}^m r_\iota(\pi_\Delta)$ is trivial, but this takes us well outside of the realm in which one can apply the Taylor–Wiles method.

In another approach, taken by Emerton in his work on the Fontaine–Mazur conjecture [**Eme**], we keep the ring $R_{\overline{\rho},S}$ associated to a (supposed absolutely irreducible) representation $\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_2(k)$, and rather enlarge the Hecke algebra $\mathbf{T}$ under consideration, by allowing non-classical (and therefore purely p-adic) systems of Hecke eigenvalues. The approach taken in

[**Eme**] to define this enlarged Hecke algebra uses the completed cohomology of modular curves. We define

$$\widetilde{H}^1(\Gamma_1(N_S), \mathcal{O}) = \varprojlim_s \varinjlim_r H^1(\Gamma_1(N_S p^r), \mathcal{O}/\varpi^s \mathcal{O}).$$

Then one can show that there is a surjective homomorphism

$$R_{\overline{\rho},S} \to \mathbf{T}(\widetilde{H}^1(\Gamma_1(N_S), \mathcal{O}))_{\mathfrak{m}}.$$

The idea now is to use softer techniques to show that this map is an isomorphism (for example, if we know in advance that $R_{\overline{\rho},S} \cong \mathcal{O}[\![X_1, X_2, X_3]\!]$ and that $\mathbf{T}(\widetilde{H}^1(\Gamma_1(N_S), \mathcal{O}))_{\mathfrak{m}}$ has Krull dimension at least 4, then the map is forced to be an isomorphism for dimension reasons). If we are given a homomorphism $R_{\overline{\rho},S} \to \mathcal{O}$ associated to an algebraic deformation ρ of $\overline{\rho}$, then it comes from a homomorphism $\mathbf{T}(\widetilde{H}^1(\Gamma_1(N_S), \mathcal{O}))_{\mathfrak{m}} \to \mathcal{O}$, and one might hope to be able to prove a classicality result asserting that the algebraicity[16] of ρ implies that this homomorphism factors through the homomorphism

$$\mathbf{T}(\widetilde{H}^1(\Gamma_1(N_S), \mathcal{O}))_{\mathfrak{m}} \to \mathbf{T}(H^1(\Gamma_1(N_S), \mathcal{O}))_{\mathfrak{m}}$$

to the classical Hecke algebra, and therefore corresponds to the Hecke eigenvalues of an algebraic automorphic representation. This is exactly what Emerton does, using the p-adic local Langlands correspondence for $\mathrm{GL}_2(\mathbf{Q}_p)$.

One could object that this approach also relies on the irreducibility of $\overline{\rho}$, in order to have the existence of $R_{\overline{\rho},S}$. However, we can circumvent this difficulty using the theory of pseudocharacters or determinants, which are a substitute for representations with many good properties. They were first applied in number theory by Wiles in order to construct the p-adic Galois representations associated to p-ordinary Hilbert modular forms [**Wil88**]. A theory that works well even in small characteristics was given by Chenevier using results in geometric invariant theory [**Che14**]. A further generalisation of the notion of pseudocharacter, in which GL_n may be replaced by an arbitrary reductive group G, was given by V. Lafforgue in his work on the Langlands correspondence over global fields of positive characteristic [**Laf18**]. Emerson showed that when $G = \mathrm{GL}_n$, this notion is equivalent to Chenevier's notion [**Eme18**]. Let us recall Lafforgue's definition of pseudocharacter for GL_n here:

Definition 3.2. *Let Γ be a group and let A be a ring. A pseudocharacter of Γ of dimension n over a ring A is a collection $\Theta = (\Theta_k)_{k \geq 1}$ of algebra homomorphisms $\Theta_k : \mathbf{Z}[\mathrm{GL}_n^k]^{\mathrm{GL}_n} \to \mathrm{Map}(\Gamma^k, A)$ satisfying the following conditions:*

(1) *For all $k, l \geq 1$ and for each map $\zeta : \{1, \dots, k\} \to \{1, \dots, l\}$, each $f \in \mathbf{Z}[\mathrm{GL}_n^k]^{\mathrm{GL}_n}$, and for each $\gamma_1, \dots, \gamma_l \in \Gamma$, we have*

$$\Theta_l(f^\zeta)(\gamma_1, \dots, \gamma_l) = \Theta_k(f)(\gamma_{\zeta(1)}, \dots, \gamma_{\zeta(k)}),$$

[16] In other words, the potential semi-stability of $\rho|_{G_{\mathbf{Q}_p}}$

where $f^\zeta \in \mathbf{Z}[\mathrm{GL}_n^l]^{\mathrm{GL}_n}$ *is defined by* $f^\zeta(g_1, \ldots, g_l) = f(g_{\zeta(1)}, \ldots, g_{\zeta(k)})$.

(2) *For each* $k \geq 1$, *for each* $\gamma_1, \ldots, \gamma_{k+1} \in \Gamma$, *and for each* $f \in \mathbf{Z}[\mathrm{GL}_n^k]^{\mathrm{GL}_n}$, *we have*

$$\Theta_{k+1}(\hat{f})(\gamma_1, \ldots, \gamma_{k+1}) = \Theta_k(f)(\gamma_1, \ldots, \gamma_{k-1}, \gamma_k\gamma_{k+1},$$

where $\hat{f} \in \mathbf{Z}[\mathrm{GL}_n^{k+1}]^{\mathrm{GL}_n}$ *is defined by* $\hat{f}(g_1, \ldots, g_{k+1}) = f(g_1, \ldots, g_{k-1}, g_k g_{k+1})$.

If Γ *is a profinite group and* A *is a topological ring, we say that* Θ *is continuous if for each* $k \geq 1$, Θ_k *takes values in* $\mathrm{Map}_{cts}(\Gamma^k, A)$.

It follows from the definition that if $\rho : \Gamma \to \mathrm{GL}_n(A)$ is a (continuous) representation, then $\mathrm{tr}\,\rho = (\Theta_k)_{k\geq 1}$,[17] defined by $\Theta_k(f)(\gamma_1, \ldots, \gamma_k) = f(\rho(\gamma_1), \ldots, \rho(\gamma_k)))$, is a pseudocharacter. Pseudocharacters have good functorial properties, and the following theorem gives some evidence as to why they are a good proxy for representations:

THEOREM 3.1. *Let* Γ *be a profinite group and let* $F = \overline{\mathbf{F}}_p$ *or* $\overline{\mathbf{Q}}_p$. *Then the map* $\rho \mapsto \mathrm{tr}\,\rho$ *sets up a bijection between the set of conjugacy classes of semisimple continuous representations* $\rho : \Gamma \to \mathrm{GL}_n(F)$ *and the set of continuous pseudocharacters* Θ *of* Γ *over* F *of rank* n.

Coming back to the situation of a number field M equipped with a finite set of finite places S, we can consider the problem of deforming a continuous pseudocharacter $\bar{t}$ of $G_{M,S}$ over k. More precisely, we can define a functor

$$\mathrm{PDef}_{\bar{t},S} : \mathcal{C}_\mathcal{O} \to \mathrm{Sets}$$

by assigning to $A \in \mathcal{C}_\mathcal{O}$ the set of pseudocharacters t_A of $G_{M,S}$ over A such that $t_A \bmod \mathfrak{m}_A$ equals $\bar{t}$. ('PDef' stands for 'pseudodeformation', i.e. a deformation of a pseudocharacter.)

THEOREM 3.2.

(1) *The functor* $\mathrm{PDef}_{\bar{t},S}$ *is representable. We write* $P_{\bar{t},S} \in \mathcal{C}_\mathcal{O}$ *for the representing object.*
(2) *Suppose that* $\bar{t} = \mathrm{tr}\,\bar{\rho}$ *for a continuous representation* $\bar{\rho} : G_{M,S} \to \mathrm{GL}_n(k)$. *Then the map* $\rho_A \mapsto \mathrm{tr}\,\rho_A$ *determines a natural transformation* $\mathrm{Def}_{\bar{\rho},S} \to \mathrm{PDef}_{\bar{t},S}$. *If* $\bar{\rho}$ *is absolutely irreducible, then this is a natural isomorphism.*

When $\bar{\rho}$ is not absolutely irreducible, the ring $P_{\bar{t},S}$ can thus stand in for the universal deformation ring, although it is harder to get a handle on that $R_{\bar{\rho},S}$. For example, the paper [**Che14**] gives only a weak upper bound for the dimension of the Zariski tangent space of $P_{\bar{t},S}$. Wang-Erickson [**WE20**] has

[17] This is an abuse of notation, since $\mathrm{tr}\,\rho$ already denotes the usual character of the representation ρ (which is a function $\Gamma \to A$). Note however that Θ_1 is equivalent to the data, for each $\gamma \in \Gamma$, of the characteristic polynomial of $\rho(\gamma)$, and that Θ_1 determines Θ_k for each $k \geq 1$.

given a refined tangent-obstruction theory for residually multiplicity-free for $P_{\bar{t},S}$, but it remains to be seen whether this theory will have applications to e.g. proving automorphy lifting theorems.

Nevertheless, we can describe $P_{\bar{t},S}$ in some situations. For example, suppose again that $M = \mathbf{Q}$, $p = 2$, and $S = \{2\}$, and that $\bar{t}$ is the trivial pseudocharacter of rank 2 (thus associated to the reduction modulo 2 of the 2-adic Galois representation attached to Ramanujan's modular form Δ). In this case, Chenevier [**Che14**, Theorem 5.1] has shown that the quotient of $P_{\bar{t},S}$ corresponding to odd pseudodeformations of $\bar{t}$ is isomorphic to $\mathcal{O}[\![X_1, X_2, X_3]\!]$ – a beautifully simple form for the universal pseudodeformation ring that might lead us to hope we could apply an argument in the style of [**Eme**] to the problem of symmetric power functoriality for the modular form Δ. Here however there is a critical missing ingredient – currently the p-adic Langlands correspondence is known to exist only for the group $\mathrm{GL}_2(\mathbf{Q}_p)$, and not for higher rank groups, and in particular there are no known classicality theorems for the completed cohomology of GL_n when $n > 2$.

We are thus led, Goldilocks-like, to look for a theory of p-adic automorphic forms that admits a non-trivial geometry (unlike the case of classical automorphic forms, for which the spectrum of the Hecke algebra is a finite set of points) and which admits useful classicality results (so we can show that p-adic automorphic forms with associated Galois representations which are algebraic are in fact classical, which will be a necessary step if we are to establish functoriality). The theory of overconvergent p-adic modular forms of finite slope has these properties, and the geometry is that of the Coleman–Mazur eigencurve $\mathcal{E}_p$, to which we now turn.

3.2. The eigencurve. The eigencurve $\mathcal{E}_p$ is a p-adic rigid analytic space (in the sense of Tate) which can be thought of as a moduli space for families of newforms (equivalently, regular algebraic automorphic automorphic representations) of finite slope. We review one of the two descriptions of the eigencurve given by Coleman–Mazur [**CM98**] in the case of tame level 1 modular forms.[18]

The space $S_k(\Gamma_1(p^r))$ of cuspidal modular forms of weight k and level $\Gamma_1(p^r)$ (for some $r \geq 1$, and say with $\overline{\mathbf{Q}}_p$ coefficients) admits an action of the Hecke operators T_l (for prime numbers $l \neq p$) and U_p. The eigencurve $\mathcal{E}_p$ admits a Zariski dense set of classical points associated to normalised Hecke eigenforms (i.e. simultaneous eigenvectors of all these Hecke operators) $f = 1 + \sum_{n \geq 2} a_p(f) q^n \in S_k(\Gamma_1(p^r))$ of finite slope. Let us describe these classical points a bit more. Given a Hecke eigenform $f \in S_k(\Gamma_1(p^r))$, we can associate

[18] Coleman–Mazur also impose the condition $p \neq 2$. Buzzard extended the definition of eigencurve to include the case $p = 2$ and also to the case of general tame level [**Buz07**]. Note as well that we consider only the cuspidal part of the eigencurve here (i.e. we excise the contribution of the ordinary Eisenstein series). Many equivalent definitions of the eigencurve now exist, some of which will be discussed in greater detail below.

various quantities: first, the slope $s(f) = v_p(a_p(f))$, where $v_p : \overline{\mathbf{Q}}_p \to \mathbf{Q} \cup \{\infty\}$ is the p-adic valuation, normalised to have $v_p(p) = 1$. We say that f has finite slope if $s(f) < \infty$, or equivalently if $a_p(f) \neq 0$.

We note that f need not be a newform. For example, the modular form $\Delta \in S_{12}(\mathrm{SL}_2(\mathbf{Z}))$ is not an eigenvector for the Hecke operator $U_p \in S_k(\Gamma_1(p^r))$ when $p \geq 1$, but generates a 2-dimensional subspace with a basis of U_p-eigenvectors

$$\Delta(q) - \alpha_p \Delta(q^p),\ \Delta(q) - \beta_p \Delta(q^p), \tag{3.2}$$

corresponding to the factors of the Hecke polynomial $1 - \tau(p)T + p^{11}T = (1 - \alpha_p T)(1 - \beta_p T)$. This phenomenon can be explained a bit more transparently by passing to the viewpoint of automorphic representations. If $f \in S_k(\Gamma_1(p^r))$ is a Hecke eigenform, then the lift of ιf to an automorphic form on $\mathrm{GL}_2(\mathbf{A_Q})$ generates a cuspidal automorphic representation π of $\mathrm{GL}_2(\mathbf{A_Q})$ such that π_∞ is 'of weight k', in the sense that its infinitesimal character coincides with that of the representation $(\mathrm{Sym}^{k-2}\,\mathbf{C}^2)^\vee$, and such that π^∞ is unramified away from the prime p. When f has finite slope, π_p is isomorphic to a subrepresentation of the normalized induction $i_{B_2(\mathbf{Q}_p)}^{\mathrm{GL}_2(\mathbf{Q}_p)} \chi_1 \otimes \chi_2$, where[19]

$$\chi_1 \otimes \chi_2 : T_2(\mathbf{Q}_p) = \mathbf{Q}_p^\times \times \mathbf{Q}_p^\times \to \mathbf{C}^\times$$

is a smooth character such that χ_1 is unramified. We can further choose χ_1 so that $\iota a_p(f) = p^{1/2}\chi_1(p)$. A choice of character $\chi = \chi_1 \otimes \chi_2$ of $T_2(\mathbf{Q}_p)$ such that π_p occurs as a subrepresentation of $i_{B_2(\mathbf{Q}_p)}^{\mathrm{GL}_2(\mathbf{Q}_p)} \chi_1 \otimes \chi_2$ is called an accessible refinement of π_p. The association $f \leftrightarrow (\pi, \chi)$ sets up a bijection between Hecke eigenforms in $S_k(\Gamma_1(p^r))$ of finite slope (for varying $r \geq 1$) and cuspidal automorphic representations of weight k, unramified outside p, and which are equipped with an accessible refinement as above. The paper [**CM98**] uses the language of Hecke eigenforms, while later works (particularly those setting up the theory of eigenvarieties for higher rank groups) use the language of accessible refinements; both are useful. The set of classical points embeds into the set of $\overline{\mathbf{Q}}_p$-points of the eigencurve $\mathcal{E}_p$.

We can now give one description of the eigencurve $\mathcal{E}_p$. It is a disjoint union of pieces corresponding to the possible residual representations $\overline{r_\iota(\pi)}$ associated to pairs (π, χ) as above. We fix a choice of residual representation $\overline{\rho}$ which arises in this fashion. Let $\mathcal{Z}(\overline{\rho})$ denote the set of pairs (π, χ) as above such that $\overline{r_\iota(\pi)} \cong \overline{\rho}$. Let $P_{\mathrm{tr}\,\overline{\rho},\{p\}}$ denote the universal pseudodeformation ring of $\mathrm{tr}\,\overline{\rho}$, and let $\mathcal{X}$ denote its rigid generic fibre. Let $\mathcal{T} = \mathrm{Hom}(T_2(\mathbf{Q}_p), \mathbf{G}_m)$ denote the character variety of $\mathbf{Q}_p^\times \times \mathbf{Q}_p^\times$. Thus $\mathcal{T}(K)$ is the set of continuous characters $\mathbf{Q}_p^\times \times \mathbf{Q}_p^\times \to K^\times$. If $(\pi, \chi) \in \mathcal{Z}(\overline{\rho})$ then we can write down a point $x_{\pi,\chi} \in (\mathcal{X} \times \mathcal{T})(\overline{\mathbf{Q}}_p)$ as follows:

[19]Here we introduce the notation B_n for the upper-triangular Borel subgroup in GL_n, and T_n for the diagonal maximal torus of GL_n.

- The $\mathcal{X}$-component corresponds to the pseudocharacter $\operatorname{tr} r_\iota(\pi)$.
- The $\mathcal{T}$-component is the character $\delta(t_1, t_2) = t_2^{2-k}\iota^{-1}\chi_1(t_1)\chi_2(t_2)|t_1/t_2|_p^{-1/2}$.

We can define the eigencurve $\mathcal{E}_p$ as the Zariski closure of the set $\{x_{\pi,\chi} \mid (\pi, \chi) \in \mathcal{Z}(\overline{\rho})\}$ of classical points (i.e. the smallest rigid analytic closed subvariety of $\mathcal{X} \times \mathcal{T}$ containing all of these points). Of course, there is no reason at all a priori why this space should have reasonable properties! The fundamental result is proved in [**CM98**]:

THEOREM 3.3. *Let $\mathcal{W} = \operatorname{Hom}(\mathbf{Z}_p^\times, \mathbf{G}_m)$, and $\kappa : \mathcal{E}_p \to \mathcal{W}$ be the map $\kappa(\rho, \delta) = \delta|_{1\times\mathbf{Z}_p^\times}^{-1}$*[20]*. Then κ is, locally on $\mathcal{E}_p$, finite flat: each point of $\mathcal{E}_p$ admits an affinoid neighbourhood $\mathcal{U}$ such that $\kappa|_{\mathcal{U}}$ is finite flat onto the image $\kappa(\mathcal{U}) \subset \mathcal{W}$. In particular, $\mathcal{E}_p$ is equidimensional of dimension 1.*

This theorem gives a sense to the idea that the points $x_{\pi,\chi}$ lie in p-adic families indexed by the weight $\kappa(x_{\pi,\chi}) \in \mathcal{W}$. In order to prove the theorem, it is necessary to actually construct these families, and Coleman–Mazur do this by describing all of the points of $\mathcal{E}_p$ (not just the classical points) in terms of overconvergent modular forms of finite slope. Overconvergent modular forms of a given weight $w \in \mathcal{W}$ lie in a $\mathbf{Q}_p$-Banach space $\mathcal{S}_w$, which receives an action of the Hecke operators T_l and a compact action of the operator U_p, which therefore has a discrete spectrum with finite-dimensional eigenspaces. Since the work of Coleman–Mazur, many alternative constructions of the eigencurve have been given, using varying notions of p-adic modular form (see e.g. [**AS97, Che04, Eme06, Buz07, Urb11, AIP15, Han17**]). What is important for us is not so much the precise construction but the following properties of the eigencurve, which are very useful in applications:

- Classicality criterion: we can define a continuous map $s : \mathcal{E}_p(\overline{\mathbf{Q}}_p) \to \mathbf{Q}$ by $s(\rho, \delta) = v_p(\delta(p, 1))$. If $x_{\pi,\chi}$ is the classical point associated to a Hecke eigenform f of finite slope $s(f) = v_p(a_p(f))$, then $s(x_{\pi,\chi}) = s(f)$. If f has weight k, then $s(f) \in [0, k-1]$. Coleman's classicality criterion states that if $x \in \mathcal{E}_p(\overline{\mathbf{Q}}_p)$ and the character $\kappa(x)$ agrees with $t \mapsto t^{k-2}$ on an open subgroup of $\mathbf{Z}_p^\times$, and moreover $s(x) \in [0, k-1)$, then x is in fact a classical point.
- Accumulation property of classical points: classical points of the eigencurve $\mathcal{E}_p$ satisfy a density property stronger than Zariski density in the ambient space $\mathcal{X} \times \mathcal{T}$. More precisely, if $x \in \mathcal{E}_p(\overline{\mathbf{Q}}_p)$ is a classical point, then for any affinoid neighbourhood $\mathcal{U}$ of x in $\mathcal{E}_p$, the classical points of $\mathcal{U}$ are Zariski dense in $\mathcal{U}$.

Although the eigencurve is a curve, in the sense of being a 1-dimensional rigid space, it is very far from being an algebraic curve, and its geometry remains rather mysterious. Some results on the local geometry are available (see e.g. [**DL16**]) but it is not known in any case whether, for example, the set

[20]The sign here is to assist comparison with classical normalisations.

of irreducible components is finite or infinite. One very intriguing question concerns the geometry 'close to the boundary of weight space' – we will come back to this very soon. Since the eigencurve is constructed as a finite cover of the spectral variety of the U_p operator, such questions are closely tied up with questions about slopes of p-adic modular forms. Many mysterious phenomena have been observed here which are yet to be explained – see for example the fascinating conjectures in the articles [**Buz05**, **BG16**, **BP19**].

3.3. Analytic continuation of functoriality. We now come to the first key principle that animates our proof: the idea that the existence of the symmetric power lifting can be analytically continued along irreducible components of the eigencurve $\mathcal{E}_p$. (Following [**Con99**], we define irreducible components as corresponding to connected components of the normalization of $\mathcal{E}_p$.) The following theorem, a special case of the results proved in [**NT21a**], makes this precise:

THEOREM 3.4. *Fix $m \geq 1$. Let $x_{\pi,\chi}$, $x_{\pi',\chi'}$ be classical points of the eigencurve $\mathcal{E}_p$ which lie on a common irreducible component. Suppose that the following conditions hold:*

(1) *Neither of π, π' has CM, and neither of π, π' is ι-ordinary.*
(2) *The refinement $\chi' = \chi'_1 \otimes \chi'_2$ is $(m+1)$-regular ([*NT21a*, Definition 2.23]). This means that for each $i = 1, \ldots, m$ we have $(\chi'_1)^i \neq (\chi'_2)^i$.*
(3) *$\mathrm{Sym}^m r_\iota(\pi)$ is automorphic.*

Then $\mathrm{Sym}^m r_\iota(\pi')$ is automorphic.

We will sketch the proof of this theorem in the next section. First we describe how it may be applied to obtain the following corollary:

COROLLARY 3.1. *Fix $m \geq 1$. Suppose there exists an everywhere unramified, regular algebraic cuspidal automorphic representation π of $\mathrm{GL}_2(\mathbf{A}_\mathbf{Q})$ such that $\mathrm{Sym}^m_*(\pi)$ exists. Then $\mathrm{Sym}^m_*(\pi)$ exists for every such automorphic representation of $\mathrm{GL}_2(\mathbf{A}_\mathbf{Q})$.*

One ideal scenario would be that the eigencurve $\mathcal{E}_p$ was irreducible (for some prime p). In general this cannot be true since there are at least as many components of $\mathcal{E}_p$ as there are odd, semisimple residual representations $\overline{\rho} : G_{\mathbf{Q},\{p\}} \to \mathrm{GL}_2(\overline{\mathbf{F}}_p)$. As we have already remarked, there is a unique such representation if we stick to the case $p = 2$. However, even in this case we don't know how to say anything about the set of irreducible components of $\mathcal{E}_2$.

Remarkably, however, Buzzard–Kilford [**BK05**] were able to compute explicitly a large part of $\mathcal{E}_2$, namely the part 'at the boundary of weight space'. In order to say what this means, let us make our description of weight space $\mathcal{W} = \mathrm{Hom}(\mathbf{Z}_p^\times, \mathbf{G}_m)$ slightly more explicit. When $p = 2$, we can decompose $\mathbf{Z}_p^\times = \{\pm 1\} \times (1 + 4\mathbf{Z}_2)$. The eigencurve $\mathcal{E}_2$ is supported above the connected component $\mathcal{W}_+$ of $\mathcal{W}$ corresponding to characters which are trivial on $-1 on \mathbf{Z}_p^\times$ (essentially because there are no level 1 modular forms of odd weight). The group $1+4\mathbf{Z}_2$ is procyclic, so we may identify $\mathcal{W}_+$ with the 2-adic open unit disc $\{|w|<1\}$ by sending a character δ to $w(\delta)=\delta(5)-1$. Here is the result of Buzzard–Kilford:

THEOREM 3.5. *Let $\mathcal{A}=\{w \mid |8|<|w|<1\}$. Then there is a decomposition $\kappa^{-1}(\mathcal{A})=\sqcup_{i=1}^{\infty}X_i$ as a countable disjoint union of admissible open subspaces. The components X_i have the following properties:*

(1) *For each $i\geq 1$, the restriction $\kappa|_{X_i}:X_i\to\mathcal{A}$ is an isomorphism.*
(2) *If $x\in X_i(\overline{\mathbf{Q}}_p)$, then $s(x)=iv_p(w(\kappa(x)))$.*

More informally, the pre-image of the boundary annulus $\mathcal{A}$ of weight space $\mathcal{W}_+$ in $\mathcal{E}_2$ decomposes as an infinite disjoint union of copies of $\mathcal{A}$, each of which maps isomorphically to $\mathcal{A}$. Moreover, on each copy the slope is a simple function of the weight.

In order to prove Corollary 3.1, we want to be able to reach any point from any other point by analytic continuation. This is too much to ask. However, another move is available to us. This can be seen most transparently for the classical points associated to everywhere unramified regular algebraic cuspidal automorphic representations π. Such a representation will admit two accessible refinements χ, χ' (cf. the two p-stabilizations of Δ written down in (3.2)) which determine two distinct classical points $x_{\pi,\chi}$, $x_{\pi,\chi'}$ on the eigencurve $\mathcal{E}_p$. Although these points have the same image under the map κ, they (usually) have distinct slopes! If π has weight k then the slopes are related by the formula

$$s(x_{\pi,\chi})+s(x_{\pi,\chi'})=k-1.$$

Since the existence (or otherwise) of $\mathrm{Sym}^m_*(\pi)$ depends only on π, and not on the choice of refinement, swapping refinements allows us to move between different components X_i of $\kappa^{-1}(\mathcal{A})$ while preserving the property of the existence of $\mathrm{Sym}^m_*(\pi)$ on that component[21]. Moreover, every irreducible component of $\mathcal{E}_2$ meets $\kappa^{-1}(\mathcal{A})$ because the image of any irreducible component of $\mathcal{E}_2$ under κ has finite complement in $\mathcal{W}_+$, by [**CM98**, Theorem B]. Therefore every irreducible component of $\mathcal{E}_2$ contains some X_i. These moves taken together are enough to establish Corollary 3.1.

3.4. Proof of the analytic continuation principle. We now sketch the proof of Theorem 3.4, which is based on ideas going back to Kisin's work [**Kis03**] on the eigencurve and the Fontaine–Mazur conjecture. Recall that we have defined the (tame level 1) eigencurve $\mathcal{E}_p$ as the Zariski closure inside $\mathcal{X}\times\mathcal{T}$ of the set of classical points $x_{\pi,\chi}$. One can ask how close the closed

[21]There is a slight wrinkle here: the classical points corresponding to level 1 forms have even weights, and so do not lie above $\mathcal{A}$ (cf. the discussion after [**NT21a**, Lemma 3.3]). One can also define a 'swapping refinements' map for classical points which are ramified at p by introducing a character twist, in order to ensure that the condition 'χ_1 unramified' is preserved. Ultimately we find that if $x_{\pi,\chi}$ is a classical point lying in some X_i, and $x_{\pi',\chi'}$ is its image under this swapping map, then $x_{\pi',\chi'}$ lies in $X_{i'}$, where i, i' are related by the formula $i+i'=(k-1)/v_p(w(x_{\pi,\chi}))$.

immersion $\mathcal{E}_p \subset \mathcal{X} \times \mathcal{T}$ is to being an isomorphism. The answer is: rather far, since $\dim \mathcal{X} = 3$ and $\dim \mathcal{T} = 2$, while $\dim \mathcal{E}_p = 1$. Even ignoring the factor of $\mathcal{T}$, it is still the case that $\mathcal{X}$ contains points corresponding to many more Galois representations $\rho : G_{\mathbf{Q},\{p\}} \to \mathrm{GL}_2(\overline{\mathbf{Q}}_p)$ than the ones associated to finite slope overconvergent Hecke eigenforms. We need to impose a more stringent condition on Galois representations appearing in $\mathcal{X}$.

In [**Kis03**], Kisin proposed such a condition. Roughly speaking, he defined a closed subvariety $\mathcal{X}_{fs} \subset \mathcal{X} \times \mathcal{T}$ (where fs stands for 'finite slope') with the following properties:

- For 'generic' points $(t, \delta) \in \mathcal{X} \times \mathcal{T}$ with $t = \mathrm{tr}\,\rho$, $\rho : G_{\mathbf{Q}} \to \mathrm{GL}(V)$, we have $(t, \delta) \in \mathcal{X}_{fs}$ if and only if there is a $G_{\mathbf{Q}_p}$-equivariant homomorphism

$$V \hookrightarrow (B^+_{cris} \otimes_{\mathbf{Q}_p} K)^{\varphi = \delta(p,1)}. \tag{3.3}$$

- There is an inclusion $\mathcal{E}_p \subset \mathcal{X}_{fs}$.
- At 'generic' classical points $x_{\pi,\chi} \in \mathcal{E}_p(K)$, the inclusion $\mathcal{E}_p \subset \mathcal{X}_{fs}$ is a local isomorphism provided that the Bloch–Kato Selmer group $H^1_f(\mathbf{Q}, \mathrm{ad}\, r_\iota(\pi))$ is zero.

(We will recall the definition of the Bloch–Kato Selmer group in the next section. Here we use the word generic in the loose sense that there are certain technical conditions which are satisfied in many examples.) The ring B^+_{cris} is the positive part of Fontaine's ring of crystalline periods. Condition (3.3) arises since it is classified for level 1 classical points (for which the space of $G_{\mathbf{Q}_p}$-equivariant homomorphisms $V \to B_{cris} \otimes_{\mathbf{Q}_p} K$ is in fact two-dimensional, because $r_\iota(\pi)|_{G_{\mathbf{Q}_p}}$ is crystalline) and behaves well in families. Optimistically, one might ask if the inclusion $\mathcal{E}_p \subset \mathcal{X}_{fs}$ is in fact an equality, and this has been proved in many cases [**Eme**, Theorem 1.2.4].

Kisin's condition was reformulated by Colmez [**Col08**], using a different set of ideas from p-adic Hodge theory. Let $\mathcal{R}$ be the Robba ring, i.e. the ring of formal series $f(X) = \sum_{n \in \mathbf{Z}} a_n X^n$ ($a_n \in \mathbf{Q}_p$) which converge in the annulus $(\rho, 1)$ for some $0 < \rho(f) < 1$. It is equipped with a Frobenius endomorphism $\varphi : \mathcal{R} \to \mathcal{R}$ and a commuting action of the group $\Gamma = \mathbf{Z}_p^\times$. By definition, a (φ, Γ)-module of rank n is a finite free $\mathcal{R}$-module D equipped with commuting semilinear actions of φ and Γ such that $\varphi(D)$ generates D. The category of (φ, Γ)-modules is not abelian, but Kedlaya proved that its objects have a canonical slope decomposition [**Ked04**]. The fundamental theorem relating (φ, Γ)-modules to p-adic representations is as follows (cf. [**CC98**, **Ked04**]):

Theorem 3.6. *There are mutually inverse equivalences of categories*

$$\left(\substack{\textit{Finite dimensional} \\ \textit{continuous } \mathbf{Q}_p[G_{\mathbf{Q}_p}]\textit{-modules}} \right) \leftrightarrow ((\varphi, \Gamma)\textit{-modules over } \mathcal{R} \textit{ of slope } 0).$$

Colmez observed that for a 2-dimensional representation V of $G_{\mathbf{Q}_p}$, Kisin's condition on the existence of a crystalline period can be replaced by the condition that its associated (φ, Γ)-module is reducible in the category of all (φ, Γ)-modules over $\mathcal{R}$, or in other words an extension of 1-dimensional objects in this category. The (φ, Γ)-modules of rank 1 over $\mathcal{R}$ may be described explicitly: they correspond to the continuous characters $\delta : \mathbf{Q}_p^\times \to K^\times$.[22]

DEFINITION 3.3. *Let $\rho : G_{\mathbf{Q}_p} \to \mathrm{GL}_d(K)$ be a continuous representation and let $\delta = \delta_1 \otimes \cdots \otimes \delta_n : T_n(\mathbf{Q}_p) \to K^\times$ be a continuous character. We say that ρ is trianguline of parameter δ if $D = D_{rig}(V)$ admits a filtration*

$$0 = \mathrm{Fil}^0 D \subset \mathrm{Fil}^1 D \subset \cdots \subset \mathrm{Fil}_n D = D$$

by (φ, Γ)-submodules such that for each $i = 1, \ldots, n$, $\mathrm{Fil}^i D$ is a (φ, Γ)-submodule, direct summand as $\mathcal{R}$-module, of rank i, and there is an isomorphism $\mathrm{Fil}^i D / \mathrm{Fil}^{i-1} D \cong \mathcal{R}(\delta_i)$. If ρ is trianguline of some parameter, we simply say that it is trianguline.

This leads to another candidate definition for the 'finite slope subspace' $\mathcal{X}_{fs} \subset \mathcal{X} \times \mathcal{T}$: as the Zariski closure of the set of points (t, δ), where t is the pseudocharacter of a continuous representation $\rho : G_{\mathbf{Q},\{p\}} \to \mathrm{GL}_2(\overline{\mathbf{Q}}_p)$ which is trianguline of parameter δ. Since (φ, Γ)-modules over $\mathcal{R}$ behave well in families (thanks to the work of Kedlaya–Pottharst–Xiao [**KPX14**] and Liu [**Liu15**]), this space $\mathcal{X}_{fs}$ is well-behaved. In particular, all of its points correspond to trianguline Galois representations.[23]

We can now describe the role these ideas play with respect to symmetric power functoriality. The notion of triangulation plays well with respect to tensor operations on group representations. In particular, if $\rho : G_{\mathbf{Q}_p} \to \mathrm{GL}_2(K)$ is trianguline of parameter $\delta = (\delta_1, \delta_2)$, then for any $m \geq 1$ the representation $\mathrm{Sym}^m \rho : G_{\mathbf{Q}_p} \to \mathrm{GL}_{m+1}(K)$ will be trianguline of parameter $\mathrm{Sym}^m \delta := (\delta_1^m, \delta_1^{m-1}\delta_2, \ldots, \delta_2^m)$. This construction makes sense in families, and we obtain a diagram

$$\begin{array}{ccc} \mathcal{E}_p & & \\ \downarrow & & \\ (\mathcal{X} \times \mathcal{T})^{tri} & \xrightarrow{\mathrm{Sym}^m} & (\mathcal{X}_{m+1} \times \mathcal{T}_{m+1})^{tri}, \end{array}$$

where $\mathcal{X}_{m+1}$ is the pseudodeformation space of $\mathrm{tr}\,\mathrm{Sym}^m \overline{\rho}$ and we define $\mathcal{T}_{m+1} = \mathrm{Hom}(T_{m+1}(\mathbf{Q}_p), \mathbf{G}_m)$.

[22] The associated (φ, Γ)-module is $\mathcal{R}(\delta)$, which denotes $\mathcal{R}$ equipped with the actions of φ, Γ satisfying $\varphi(1) = \delta(p)$ and $\gamma \cdot 1 = \delta(\gamma)$.

[23] We are ignoring some technicalities here. One minor point is that the 'Zariski closure' results of [**KPX14**] require the points taken to be 'strictly trianguline', a condition that is satisfied for the most points on the eigencurve. A more significant problem is that the results of [**KPX14**] require us to have a true family of Galois representations, not just a family of pseudocharacters. In practice, we apply the results of [**KPX14**] only in an affinoid neighbourhood of a point (t, δ) where the representation ρ is irreducible, in which case the pseudocharacter can be lifted to a family of Galois representations in this neighbourhood in an essentially unique way, by [**Che14**, Proposition G].

To go further, we need to introduce another eigenvariety (i.e. a higher rank analogue of the eigencurve) whose points will give rise to Galois representations of rank $m+1$ (and which will fit into the empty spot in the top right of the above diagram). Here it is technically convenient to pass to a soluble CM extension $M/\mathbf{Q}$, and to replace $\mathrm{GL}_2(\mathbf{A_Q})$ by $G_2(\mathbf{A}_{M^+})$, where G_2 is a definite unitary group in two variables over M^+, and $\mathrm{GL}_{m+1}(\mathbf{A_Q})$ by $G_{m+1}(\mathbf{A}_{M^+})$, where G_{m+1} is a definite unitary group in $m+1$ variables over M^+. This is valid since if π is a regular algebraic cuspidal automorphic representation of $\mathrm{GL}_2(\mathbf{A_Q})$, then we can find an algebraic Hecke character $\chi : M^\times \backslash \mathbf{A}_M^\times \to \mathbf{C}^\times$ such that the character twist $\mathrm{BC}_{M/\mathbf{Q}}(\pi) \otimes \chi$ of the base change of π to M is RACSDC[24], hence descends to the group G_2. (The Galois-theoretic manifestation of this is that the representation $\rho = r_\iota(\pi)|_{G_M} \otimes r_\iota(\chi)$ satisfies $\rho^c \cong \rho^\vee \otimes \epsilon^{-1}$.)

We lose nothing in doing this, since soluble base change and descent between $\mathrm{GL}_n(\mathbf{A_Q})$ and $\mathrm{GL}_n(\mathbf{A}_M)$ is well understood [**AC89**][25] and base change and descent between conjugate self-dual, regular algebraic automorphic representations of $\mathrm{GL}_{m+1}(\mathbf{A}_M)$ and automorphic representations of $G_{m+1}(\mathbf{A}_{M^+})$ is also reasonably well-understood (using e.g. the theorems proved in [**Lab11**]) provided we choose our auxiliary data carefully (for example, we want the p-adic places of M^+ to split in M so that $G(M^+ \otimes_{\mathbf{Q}} \mathbf{Q}_p)$ looks like a product of general linear groups).[26]

On the other hand, we gain in that the basic properties of the eigenvarieties of definite unitary groups have been very well-studied (in particular, there are strong classicality theorems that give Galois-theoretic conditions under which a point is in fact a classical point, i.e. arising from an automorphic representation of $G(\mathbf{A}_{M^+})$ and an accessible refinement of said representation – see [**NT21a**, Lemma 2.30], which combines ideas from [**Che11**] and [**BHS17**]). There are also genuine technical simplifications: one expects that when $m > 1$ the eigenvariety of $\mathrm{GL}_{m+1}(\mathbf{A_Q})$ does not admit a dense set of classical points and is not (locally on the source) finite and surjective over weight space, two properties of the eigencurve $\mathcal{E}_p$ that *are* enjoyed by the eigenvarieties of definite unitary groups (see e.g. [**Che04**, Théorème 6.3.6]).

Having done so, we arrive at a diagram

$$\begin{array}{ccc} \mathcal{E}_{G_2} & \overset{?}{\dashrightarrow} & \mathcal{E}_{G_{m+1}} \\ \downarrow{\scriptstyle i_2} & & \downarrow{\scriptstyle i_{m+1}} \\ (\mathcal{X}_{2,M} \times \mathcal{T}_{2,M})^{tri} & \xrightarrow{\mathrm{Sym}^m} & (\mathcal{X}_{m+1,M} \times \mathcal{T}_{m+1,M})^{tri}, \end{array}$$

[24] Recall this stands for regular algebraic, conjugate self-dual, cuspidal – see §2.9 of this article. The base change automorphic representation is the one whose associated Galois representation is $r_\iota(\pi)|_{G_M}$.

[25] At least, when we stick to regular algebraic automorphic representations π such that $r_\iota(\pi)|_{G_M}$ is irreducible – see [**BLGHT11**, Lemma 1.4] for a positive statement.

[26] The work [**KMSW14**], the second half of which has yet to appear, will give a complete description of the discrete automorphic representations of any unitary group.

where:

- $\mathcal{E}_{G_2}$, $\mathcal{E}_{G_{m+1}}$ are eigenvarieties for the definite unitary groups G_2, G_{m+1}, respectively. They each contain dense sets of classical points associated to pairs (π, χ) consisting of an automorphic representation and accessible refinement, and admit maps $\kappa_2 : \mathcal{E}_{G_2} \to \mathcal{T}_{2,M}$, $\kappa_{m+1} : \mathcal{E}_{G_{m+1}} \to \mathcal{T}_{m+1,M}$, where e.g. $\mathcal{T}_{M,2}$ is the character variety of a maximal torus in $G_2(M^+ \otimes_{\mathbf{Q}} \mathbf{Q}_p)$.
- $\mathcal{X}_{2,M}$ is the rigid space associated parameterizing conjugate self-dual pseudocharacters lifting the residual pseudocharacter of $r_\iota(\pi)|_{G_M} \otimes r_\iota(\chi)$.
- $\mathcal{X}_{m+1,M}$ is the rigid space associated parameterizing conjugate self-dual pseudocharacters lifting the residual pseudocharacter of $\mathrm{Sym}^m(r_\iota(\pi)|_{G_M} \otimes r_\iota(\chi))$.
- The bottom horizontal arrow is 'passage to symmetric power'.
- The superscript 'tri' indicates in each case that we have passed to the closed subspace which is the Zariski closure of the set of trianguline points.
- The vertical arrows are given by 'passage to associated Galois representation' – they are closed immersions.

The arrow we would ultimately like to construct is the broken horizontal arrow, which would be a kind of overconvergent symmetric power functoriality, and from which we would hope to be able to deduce the symmetric power functoriality in the usual sense (i.e. for classical automorphic representations) by applying a classicality theorem to points of $\mathcal{E}_{G_{m+1}}$ which are images of classical points of $\mathcal{E}_{G_2}$.[27]

This is asking too much. What we are able to do (and this is enough to be able to deduce Theorem 3.4) is show that if $\mathcal{C} \subset \mathcal{E}_{G_2}$ is an irreducible component containing a non-critical classical point $x_{\pi,\chi}$ such that $\mathrm{Sym}^m_*(\pi)$ exists, then $\mathrm{Sym}^m(i_2(\mathcal{C}))$ is contained in the image of i_{m+1}. Here we use the useful fact that if $\mathcal{Z}$ is an irreducible rigid analytic space and $\mathcal{W} \subset \mathcal{Z}$ is a Zariski closed subspace which contains a non-empty affinoid open of $\mathcal{Z}$, then $\mathcal{W} = \mathcal{Z}$. This means we need only show that the image of i_{m+1} contains the image under Sym^m of an open affinoid neighbourhood of $x_{\pi,\chi}$. Ultimately we need only work with affinoid local version of the above diagram where the associated pseudocharacters are irreducible; this resolves one of the technical problems alluded to in footnote 23.

[27]The classicality theorem we need is somewhat stronger than the 'numerically non-critical implies classical' theorem which is the naive generalisation of Coleman's criterion $v_p(a_p) < k - 1$. This is because the symmetric powers of a 2-dimensional Galois representation of small slope are no longer necessarily of small slope. However, they will (as observed by Chenevier [**Che11**, Example 3.26] – this requires the $(m+1)$-regular condition in the statement of Theorem 3.4) satisfy the weaker condition that each triangulation is non-critical, cf. [**NT21a**, Lemma 2.7]. The criticality (or otherwises) of classical points is closely tied to the local geometry of the eigenvariety around such points and has been studied in great detail by Breuil–Hellmann–Schraen [**BHS17, BHS19**].

We now come back to the result proved by Kisin [**Kis03**] for the eigencurve $\mathcal{E}_p$: namely that the closed immersion $\mathcal{E}_p \to \mathcal{X}_{fs}$ (his analogue of our map i_2) is a local isomorphism in an affinoid neighbourhood of a classical point $x_{\pi,\chi}$ such that the adjoint Bloch–Kato Selmer group $H^1_f(\mathbf{Q}, \operatorname{ad} r_\iota(\pi))$ is 0 (and satisfying some other mild technical conditions). This holds in our context, as can be seen using a tangent space calculation first done in ths unitary group case by Bellaïche–Chenevier [**BC09**, Corollary 2.6.1]. The essential point is that the local surjectivity of the eigenvariety over weight space gives a lower bound for the dimension of the eigenvariety. If the adjoint Bloch–Kato Selmer group vanishes and the character δ is sufficiently generic, one can show that an upper bound for the tangent space of $(\mathcal{X}_{2,M} \times \mathcal{T}_{2,M})^{tri}$ that is equal to this lower bound, showing that the eigenvariety is smooth and that the map i_2 is an isomorphism in a suitable affinoid neighbourhood. Exactly the same argument is valid for $\mathcal{E}_{G_{m+1}}$. Thus the proof is complete, *provided* we can show that

$$H^1_f(M^+, \operatorname{ad} \operatorname{Sym}^m r_\iota(\pi)|_{G_M}) = 0.$$

This is non-trivial and requires a different set of ideas, which we address in the next section.

3.5. Vanishing of the adjoint Bloch–Kato Selmer group. Let M be a CM field, let π be a RACSDC automorphic representation of $\mathrm{GL}_n(\mathbf{A}_M)$, and let $\rho = r_\iota(\pi)$. Let us assume (as is conjecturally always the case) that ρ is irreducible. There is an isomorphism $\rho^c \cong \rho^\vee \otimes \epsilon^{1-n}$. Fixing a choice of complex conjugation $c \in G_{M^+}$, this implies the existence of a non-degenerate bilnear form $\langle \cdot, \cdot \rangle : \overline{\mathbf{Q}}_p^n \times \overline{\mathbf{Q}}_p^n \to \overline{\mathbf{Q}}_p$ satisfying the identity

$$\langle \rho(g^c)v, \rho(g)w \rangle = \epsilon^{1-n}(g)\langle v, w \rangle$$

for all $g \in G_M$, $v, w \in \overline{\mathbf{Q}}_p^n$. The adjoint representation $\operatorname{ad} \rho = M_n(\overline{\mathbf{Q}}_p)$ is a $\overline{\mathbf{Q}}_p[G_M]$-module; we can extend the action of G_M to an action of G_{M^+} by defining $c \cdot X = -X^*$, where $*$ denotes adjoint with respect to $\langle \cdot, \cdot \rangle$.

The Galois cohomology group $H^1(M^+, \operatorname{ad} \rho)$ parameterizes deformations of ρ to $\overline{\mathbf{Q}}_p[\epsilon]$ which are conjugate self-dual (a variant of the computation of the tangent space to the universal deformation ring of an absolutely irreducible residual representation, already discussed above). The Bloch–Kato Selmer group is defined to be

$$H^1_f(M^+, \operatorname{ad} \rho) = \ker(H^1(M^+, \operatorname{ad} \rho) \to \prod_v H^1(M^+_v, \operatorname{ad} \rho)/H^1_f(M^+_v, \operatorname{ad} \rho)),$$

where the product runs over the set of all finite places v of M^+, and we define

$$H^1_f(M^+_v, \operatorname{ad} \rho) = H^1_{ur}(M^+_v, \operatorname{ad} \rho)$$

to be the unramified subgroup if $v \nmid p$ and

$$H^1_f(M^+_v, \operatorname{ad} \rho) = \ker(H^1(M^+_v, \operatorname{ad} \rho) \to H^1(M^+_v, \operatorname{ad} \rho \otimes_{\mathbf{Q}_p} B_{crys}))$$

if $v|p$. Its elements correspond to those deformations of ρ which are potentially semistable. Its vanishing is predicted by the Bloch–Kato conjecture[28] and, as we have seen in the previous section, would have consequences for the geometry of unitary group eigenvarieties.

The vanishing of $H^1_f(M^+, \operatorname{ad}\rho)$ is in some cases a natural byproduct of the proof of automorphy lifting theorems. Recall that one approach to proving an automorphy lifting theorem is to try to prove that a map $R \to \mathbf{T}$ from a Galois deformation ring (say with conditions from p-adic Hodge theory imposed) to a Hecke algebra (say acting on the cuspidal part of the cohomology of an arithmetic group) is an isomorphism. If our representation ρ is classified by a homomorphism $f : R \to \overline{\mathbf{Q}}_p$, then the Bloch–Kato Selmer group $H^1_f(M^+, \operatorname{ad}\rho)$ can be identified with the dual of the Zariski tangent space[29] $(\ker(f)/\ker(f)^2)[1/p]$. On the other hand, if we can show that $R \to \mathbf{T}$ is an isomorphism, then this Zariski tangent space will be forced to be zero, since $\mathbf{T}[1/p]$ will be an étale algebra (essentially because the space of cuspidal automorphic forms is unitary and admissible as a representation of the adele group, and therefore in particular semisimple). Such vanishing results appear already in [**Wil95**].

Recalling the discussion in §3.1, we might expect to need strong conditions on the residual representation $\overline{\rho}$ (in particular, absolute irreducibility). However, in our applications we need to be able to treat the case where $\overline{\rho}$ is a trivial representation in arbitrary degree (as this is required in order to be able to prove the existence of $\mathrm{Sym}^m_*(\pi_\Delta)$ using the 2-adic eigenvariety). A slightly different tack is therefore required.[30]

In the papers [**NT20, Tho21a**], we prove the following vanishing theorem.

Theorem 3.7. *Let M be a CM field and let π be a RACSDC automorphic representation of $\mathrm{GL}_n(\mathbf{A}_M)$. Suppose that $r_\iota(\pi)|_{G_{M(\zeta_{p^\infty})}}$ is irreducible.*[31] *Then we have*

$$H^1_f(M^+, \operatorname{ad} r_\iota(\pi)) = 0.$$

---

[28] We don't have space here to discuss further the Bloch–Kato conjecture, which predicts the order of vanishing at the point $s = 1$ of the L-function $L(\iota V, s)$ of an algebraic Galois representation $V = 0$. When V is non-trivial but of weight 0 (such as in the case $V = \operatorname{ad}\rho$) then the L-function is expected to be non-vanishing at $s = 1$ (in line with the non-vanishing results for automorphic L-functions on the line $\mathrm{Re}(s) = 1$ which may be used to prove the Sato–Tate conjecture). See Fontaine's survey [**Fon92**] for more details.

[29] This is unsurprising, although not quite a formal consequence of the definitions. See [**Kis09c**, §2.3] for a justification of a similar statement.

[30] Another approach is taken in Kisin's paper [**Kis04**], which proves the vanishing of $H^1_f(\mathbf{Q}, \operatorname{ad} r_\iota(\pi))$ in many cases for regular algebriac cuspidal automorphic representations of $\mathrm{GL}_2(\mathbf{A}_\mathbf{Q})$. This uses the realisation of these p-adic Galois representations in the étale cohomology of modular curves in a way which seems difficult to generalise to the higher rank situation.

[31] Here $M(\zeta_{p^\infty})$ denotes the subfield of $\overline{M}$ obtained by adjoining all p-power roots of unity to M.

This is rather general. Indeed, there is no condition at all on the residual representation or on the prime p. If π arises by symmetric power lifting and base change of a regular algebraic cuspidal automorphic representation of $\mathrm{GL}_2(\mathbf{A}_\mathbf{Q})$ which does not have CM, then $r_\iota(\pi)|_{G_{M(\zeta_{p^\infty})}}$ is irreducible, so we obtain the required vanishing. (This follows from the stronger fact that the Zariski closure of the image of $r_\iota(\pi)$ in GL_{m+1} contains a conjugate of the the image of the unique irreducible representation $\mathrm{SL}_2 \to \mathrm{GL}_{m+1}$, see [**NT20**, Example 2.34].)

In proving this theorem we make use of some new ideas in the theory of pseudocharacters and in our understanding of the Taylor–Wiles method. We briefly sketch the main ones. Instead of considering a map $R \to \mathbf{T}$, we consider a map $P \to \mathbf{T}$ in the category $\mathcal{C}_\mathcal{O}$, where P is a universal pseudodeformation ring of $\mathrm{tr}\,\overline{\rho}$ and $\mathbf{T}$ is a Hecke algebra acting on a space of algebraic modular forms, with integral structure, on the definite unitary group G_n. We need to impose p-adic Hodge theoretic conditions on the pseudocharacters that appear in P. It is far from obvious how to do this, since a pseudocharacter by its very nature does not give rise to a module with a Galois action on which conditions can be imposed. This problem was solved by Wake–Wang-Erickson [**WWE19**], who show that one can get around this by using the notion of Cayley–Hamilton representation of a pseudocharacter defined by Chenevier [**Che14**]. Another result we need concerns deformations of pseudocharacters with integral structure. More precisely, we prove the following theorem ([**NT20**, Proposition 2.9]):

THEOREM 3.8. *Let Γ be a profinite group and let $\rho : \Gamma \to \mathrm{GL}_n(\mathcal{O})$ be a continuous homomorphism such that $\rho \otimes_\mathcal{O} K$ is absolutely irreducible. Let $\Theta = \mathrm{tr}\,\rho$. Consider the map $\tilde{\rho} \mapsto \mathrm{tr}\,\tilde{\rho}$ from the set of continuous homomorphisms $\rho' : \Gamma \to \mathrm{GL}_n(\mathcal{O} \oplus \epsilon K/\mathcal{O})$ lifting ρ to the set of continuous pseudocharacters $\Theta' : \Gamma$ over $\mathcal{O} \oplus \epsilon K/\mathcal{O}$ lifting Θ. Then there is a constant $C \in \mathbf{Z}_{\geq 0}$, depending only on $\rho(\Gamma)$, such that this map is 'bijective up to ϖ^C-torsion', in the following sense:*

(1) *For any Θ', there exists a representation ρ' such that $\mathrm{tr}\,\rho' = \alpha_C \circ \theta$, where α_C is the ring endomorphism of $\mathcal{O} \oplus \epsilon K/\mathcal{O}$ which sends $x + \epsilon y$ to $x + \epsilon\varpi^C y$.*
(2) *For any pair ρ', ρ'' of representations such that $\mathrm{tr}\,\rho' = \mathrm{tr}\,\rho''$, $\alpha_C \circ \rho'$ and $\alpha_C \circ \rho''$ are conjugate under the group $1 + \epsilon M_n(K/\mathcal{O})$.*

The proof of this theorem uses the definition of pseudocharacter introduced by Lafforgue in [**Laf18**], and is in some sense an integral refinement of the proof there that pseudocharacters biject with semisimple representations over an algebraically closed field. It is interesting to note that the different but equivalent definitions of pseudocharacter given in the papers [**Che14, Laf18**] both have a role to play.

To proceed, we consider the map $f : P \to \overline{\mathbf{Q}}_p$ associated to $r_\iota(\pi)$ (which factors through $\mathbf{T}$, essentially by definition), and try to show that the map

$P_{\ker f} \to \mathbf{T}_{\ker f}$ induced after localisation and completion at $\ker f$ is an isomorphism. This will suffice to prove the vanishing of the adjoint Bloch–Kato Selmer group, for the same reasons as outlined above: the (completed and localised) Hecke algebra is étale (and even a field in this case) while the Zariski tangent space of $P_{\ker f}$ may be identified with the Bloch–Kato Selmer group (here we crucially use the results of [**Liu07**] to match up the Bloch–Kato Selmer condition with the one arising from the p-adic Hodge theoretic condition imposed in [**WWE19**], cf. [**WWE19**, Remark 5.2.2]).

To show that $P_{\ker f} \to \mathbf{T}_{\ker f}$ is an isomorphism, we use a variant of the Taylor–Wiles method, which requires patching the action of certain auxiliary rings P_Q on auxiliary spaces H_Q of automorphic forms on the definite unitary group. In the classical Taylor–Wiles method, Q is a set of auxiliary places of the base number field at which additional ramification is allowed. There is an finite abelian group Δ_Q (a quotient of the product of the inertia groups at the places of Q) and a homomorphism $\mathcal{O}[\Delta_Q] \to P_Q$ such that $P_Q \otimes_{\mathcal{O}[\Delta_Q]} \mathcal{O}$ may be naturally identified with P. The space of H_Q of automorphic forms with additional level at Q is supposed to be a P_Q-module which is finite free as $\mathcal{O}[\Delta_Q]$-module, and which satisfies the analogous property $H_Q \otimes_{\mathcal{O}[\Delta_Q]} \mathcal{O} \cong H$, where H is the space of automorphic forms on which the Hecke algebra $\mathbf{T}$ at minimal level acts faithfully. (This freeness of H_Q over the ring $\mathcal{O}[\Delta_Q]$ is what provides the lower bound on the growth of the Q-ramified Hecke algebra, mentioned in our sketch of the Taylor–Wiles method in §3.1.)

In order to be able to define H_Q (and in particular, construct an isomorphism $H_Q \otimes_{\mathcal{O}[\Delta_Q]} \mathcal{O} \cong H$), one typically needs the residual representation $\overline{\rho}$ to be sufficiently non-degenerate; the places $v \in Q$ are selected so that e.g. $\overline{\rho}$ is unramified at v and $\overline{\rho}(\mathrm{Frob}_v)$ has no repeated eigenvalues. In our situation, where $\overline{\rho}$ can even be trivial, this is no longer possible, and we need lose control over the auxiliary spaces H_Q. In particular, although they are $\mathcal{O}[\Delta_Q]$-modules, they may no longer be of bounded rank. Since the Taylor–Wiles method traditionally used a kind of diagonalization or compactness argument to construct the patched objects, this looks like a serious problem!

In his work on the Fontaine–Mazur conjecture for residually reducible representations $\rho : G_{\mathbf{Q}} \to \mathrm{GL}_2(\overline{\mathbf{Q}}_p)$, Pan [**Pan21**] contended with exactly this difficulty. He showed how to use ultrafilters to construct very large 'ultra-patched' modules. This may not have reasonable properties but it turns out that after localization and completion at the prime ideal corresponding to ρ, enough finiteness properties hold that one can still make the Taylor–Wiles method work in a neighbourhood of ρ. The same technique applies equally well in our case. It seems an interesting problem to clarify the role played by ultrafilters in this argument.

3.6. Seed points. Let $m \geq 1$. The work described in the last few sections shows that, if we want to establish the existence of $\mathrm{Sym}^m_*(\pi)$ for every everywhere unramified regular algebraic cuspidal automorphic representation of $\mathrm{GL}_2(\mathbf{A}_{\mathbf{Q}})$, then it is enough to show this existence for a single such representation π. In order to do show the existence of such a π, we use an argument in which many of the same themes appear (p-adic automorphic forms, residually reducible Galois representations) but in slightly different form, in a spirit closer to our previous work with Clozel on symmetric power functoriality in small degrees [**CT14, CT15, CT17**].

To explain the idea, suppose given a regular algebraic cuspidal automorphic representation π of $\mathrm{GL}_2(\mathbf{A}_\mathbf{Q})$ with the following properties:

- π does not have CM.
- There exists a prime number p and an isomorphism $\iota : \overline{\mathbf{Q}}_p \to \mathbf{C}$ such that $\overline{r_\iota(\pi)} \cong \mathrm{Ind}_{G_K}^{G_\mathbf{Q}} \overline{\chi}$, for some imaginary quadratic extension $K/\mathbf{Q}$ and continuous character $\overline{\chi} : G_K \to \overline{\mathbf{F}}_p^\times$.

It is easy to construct examples of such automorphic representations. For example, we could start with an elliptic curve E over $\mathbf{Q}$ which does have CM, and is therefore associated to a CM regular algebraic cuspidal automorphic representation π_E of $\mathrm{GL}_2(\mathbf{A}_\mathbf{Q})$. Then choose a prime number p such that $\overline{\rho}_{E,p}$ is irreducible and a prime number $l \neq p$ such that E has good reduction at l and $\overline{\rho}_{E,p}|_{G_{\mathbf{Q}_l}}$ is the trivial representation. Ribet's level-raising theorem then implies the existence of another regular algebraic automorphic representation π such that $\overline{r_\iota(\pi)} \cong \overline{r_\iota(\pi_E)}$ and π_l is the Steinberg representation of $\mathrm{GL}_2(\mathbf{Q}_l)$ (i.e. the unique irreducible admissible representation of trivial central character and conductor l). The local condition at l implies that π does not have CM.

We now consider the problem of establishing the existence of the m^{th} symmetric power lifting of π. Restricting to K, we see that the residual representation of $\mathrm{Sym}^m r_\iota(\pi)$ has the simple form

$$\mathrm{Sym}^m \overline{r_\iota(\pi)}|_{G_K} \cong \bigoplus_{i=0}^m \overline{\chi}^{m-i} \overline{\chi}^{c,i-1}.$$

Although $\mathrm{Sym}^m r_\iota(\pi)$ is irreducible, its residual representation is highly reducible. Automorphy lifting theorems are one of the most powerful tools that exist to prove Galois representations are automorphic, but establishing the residual automorphy of the residual representation is often a sticking point. Here we might hope that the relatively simple form of the residual representation makes it easier to check the residual automorphy, and that an automorphy lifting theorem might be available which applies in this residually reducible situation.

Such an automorphy lifting theorem is proved in [**ANT20**], generalising the one proved in [**Tho15**] that was applied to the symmetric power functoriality problem in the papers [**CT14, CT15, CT17**]. Here is the statement.

THEOREM 3.9. *Let M be a CM number field, let $\iota : \overline{\mathbf{Q}}_p \to \mathbf{C}$ be an isomorphism, and let $\rho : G_M \to \mathrm{GL}_n(\overline{\mathbf{Q}}_p)$ be a conjugate self-dual Galois representation which satisfies the following conditions:*

(1) *There are continuous characters* $\overline{\chi}_1, \ldots, \overline{\chi}_n : G_M \to \overline{\mathbf{F}}_p^\times$ *such that there is an isomorphism*

$$\overline{\rho} \cong \oplus_{i=1}^n \overline{\chi}_i,$$

and moreover for each $i = 1, \ldots, n$, $\overline{\chi}_i \overline{\chi}_i^c = \epsilon^{1-n}$, *and for each* $1 \leq i < j \leq n$, *the character* $\overline{\chi}_i / \overline{\chi}_j |_{G_{M(\zeta_p)}}$ *has order strictly greater than* $2n$.

(2) $n \geq 3$ *and* $p > n$.

(3) ρ *is unramified at all but finitely many places of* M *and there is* $\lambda = (\lambda_{\tau,i})^{\mathrm{Hom}(M,\overline{\mathbf{Q}}_p)}$ *such that for each* $\tau \in \mathrm{Hom}(M, \overline{\mathbf{Q}}_p)$, λ_τ *is dominant in the sense that* $\lambda_{\tau,1} \geq \lambda_{\tau,2} \geq \cdots \geq \lambda_{\tau,n}$, *and* ρ *is ordinary of weight* λ, *in the sense that for each place* $v|p$ *of* M *there is an isomorphism*

$$\rho|_{G_{F_v}} \cong \begin{pmatrix} \psi_{v,1} & * & * & * \\ 0 & \psi_{v,n} & * & * \\ \vdots & \ddots & \ddots & * \\ 0 & \cdots & 0 & \psi_{v,n} \end{pmatrix},$$

where for each $i = 1, \ldots, n$, $\psi_{v,i} : G_{F_v} \to \overline{\mathbf{Q}}_p^\times$ *is a continuous character such that for some open subgroup* $U_{v,i} \subset \mathcal{O}_{F_v}^\times$ *and for all* $u \in U_{v,i}$, *we have* $\psi_{v,i} \circ \mathrm{Art}_{F_v}(u) = \prod_{\tau : M_v \to \overline{\mathbf{Q}}_p} \tau(u)^{-\lambda_{\tau,n-i+1}+i-1}$.

(4) *There is a RACSDC automorphic representation* π *of* $\mathrm{GL}_n(\mathbf{A}_M)$ *such that* $\overline{r_\iota(\pi)} \cong \overline{\rho}$ *and* π *is* ι*-ordinary.*

(5) $[F(\zeta_p) : F] = p - 1$.

(6) *There exists a place* $v_0 \nmid p$ *of* M *such that* π_{v_0} *is an unramified twist of the Steinberg representation and there is an unramified character* $\psi : G_{M_{v_0}} \to \overline{\mathbf{Q}}_p^\times$ *such that* $\rho|_{G_{M_{v_0}}}^{ss} \cong \oplus_{i=1}^n \psi \epsilon^{1-i}$.

Then ρ *is ordinarily automorphic, in the sense that there is an* ι*-ordinary RACSDC automorphic representation* Π *of* $\mathrm{GL}_n(\mathbf{A}_M)$ *such that* $\rho \cong r_\iota(\Pi)$.

We do not define the ι-ordinary condition on π here[32], but only note that it generalies the well-known 'slope of U_p-eigenvalue equals 0' condition on classical Hecke eigenforms and corresponds, under local-global compatibility at the p-adic places of M, to the requirement that $r_\iota(\pi)$ is ordinary of weight corresponding to that of π.

The proof of this theorem relies on yet another class of p-adic automorphic forms, namely the ordinary p-adic automorphic forms, which may be thought of as overconvergent modular forms of slope 0. Starting in the 1980's with the paper [**Hid86**], Hida developed a powerful theory of ordinary p-adic modular forms, using it to construct families of Galois representations with values in Hecke algebras which are finite covers of the Iwasawa algebra Λ whose rigid generic fibre is the usual weight space. Geraghty [**Ger19**]

[32] But see [**Ger19**, Definition 5.3].

developed Hida theory for definite unitary groups and used it to prove automorphy lifting theorems for ordinary, conjugate self-dual Galois representations in arbitrary rank (although with a residual irreducibility condition). Skinner–Wiles used Hida theory for classical modular forms (and more generally, p-adic Hilbert modular forms) to prove an automorphy lifting theorem for residually reducible ordinary Galois representations of dimension 2. (Ordinary modular forms can be viewed as a precursor of overconvergent modular forms of finite slope, of which they are a special case. They remain useful since they generally have easy-to-understand integral structures.)

The starting point in [**ANT20**] for the proof of the above theorem is the idea of applying the Skinner–Wiles strategy in Geraghty's context of Hida theory for definite unitary groups. We start with a diagram

$$R^{ord} \leftarrow P^{ord} \rightarrow \mathbf{T}^{ord},$$

where $\mathbf{T}^{ord}$ is the 'big ordinary Hecke algebra' (a finite Λ-algebra), P^{ord} is a pseudodeformation ring of $\operatorname{tr}\overline{\rho}$ with an 'ordinary of variable weight' condition imposed at the p-adic places of M, and R^{ord} is the corresponding Galois deformation ring. (A pleasant feature of this situation is that although $\overline{\rho}$ is reducible, R^{ord} exists, because $\overline{\rho}$ is irreducible *as a representation valued in the Langlands dual group of the unitary group* G_n – in other words, is Schur in the sense of [**CHT08**]. This is in contrast to the situation considered in [**SW99**], where one needs to consider one Galois deformation ring for every possible indecomposable residual representation with the same pseudocharacter as $\overline{\rho}$.) The gain in using modular forms of variable weight is that $\mathbf{T}^{ord}$ now has Krull dimension $\dim \Lambda = 1 + n[M^+ : \mathbf{Q}]$, and one can find plentiful 'generic' prime ideals $\mathfrak{p} \subset \mathbf{T}^{ord}$ such that the specialisation of the universal pseudocharacter to $\operatorname{Frac} \mathbf{T}^{ord}/\mathfrak{p}$ is absolutely irreducible, and suitable for application of a modified version of the Taylor–Wiles method. This makes it possible to prove that $P^{ord}_{\mathfrak{p}} \rightarrow \mathbf{T}^{ord}_{\mathfrak{p}}$ is an isomorphism.

Thinking geometrically, this implies that any irreducible component of $\operatorname{Spec} P^{ord}$ which contains $\mathfrak{p}$ is in fact contained in the closed subspace $\operatorname{Spec} \mathbf{T}^{ord}$. The wonderful idea introduced in [**SW99**] is to use the fact that for any two irreducible components Z_0, Z_1 of $\operatorname{Spec} P^{ord}$, one can find a chain $W_1, W_2, \dots, W_d$ of irreducible components of $\operatorname{Spec} P^{ord}$ such that all of the intersections

$$Z_0 \cap W_1, W_1 \cap W_2, W_2 \cap W_3, \dots, W_{d-1} \cap W_d, W_d \cap Z_1$$

have a dimension which is 'not too small'.[33] Using that the dimension is not too small, we can find generic prime ideals in each of these intersections, and

[33]The lower bound dimension of these intersections is the so-called 'connectedness dimension' of the ring P^{ord} and is more easily computed for the ring R^{ord}. In the set-up of this section, the relationship between R^{ord} and P^{ord} is rather simple (in fact, [**ANT20**, Proposition 3.2] identifies P^{ord} as the subring of invariants of R^{ord} under a finite abelian group). If we have a presentation $R^{ord} \cong \mathcal{O}[\![X_1, \dots, X_g]\!]/(f_1, \dots, f_r)$, then the connectedness dimension of R^{ord} is at least $g - r$. In practice this can be made as large as desired by first replacing the field M by a soluble CM extension.

so propagate the automorphy of any irreducible component of Spec P^{ord} to any other component.

The key step in this argument of finding 'generic' primes in the intersections $W_i \cap W_{i+1}$ is possible only if the locus of reducible pseudocharacters in P^{ord} is 'not too large'. Unfortunately, it can definitely be too large in practice. An important form of functoriality relevant for the unitary group G_n is given by the theory of endoscopy, which implies (roughly speaking) the existence of functorial liftings from groups $G_k \times G_{n-k}$ to G_n. Since the dimension of the ordinary Hecke algebra for G_k is $k[M^+ : \mathbf{Q}]$, we would expect to see components in Spec $\mathbf{T}^{ord}$ of dimension $k[M^+ : \mathbf{Q}] + (n-k)[M^+ : \mathbf{Q}] = n[M^+ : \mathbf{Q}] = \dim \mathbf{T}^{ord}$ arising by endoscopic transfer (and therefore corresponding to Galois representations which are reducible in GL_n). We therefore need to impose additional conditions to rule out the existence of entire components of Spec $\mathbf{T}^{ord}$ corresponding to reducible Galois representations. (This is one point where we depart from the context of [**SW99**], where there are no endoscopic forms and the arguments needed are substantially different.)

This explains the presence of the final condition (6) in the statement of Theorem 3.9 – it is a local condition that can be used effectively to eliminate the contribution of endoscopic lifts, since the Steinberg representation (corresponding as it does under the local Langlands correspondence to an indecomposable Weil–Deligne representation) cannot be the local component of a lift from an endoscopic group of G_n. We have dwelled on this point because it ties into the problem of verifying the residual automorphy of the residual representation $\mathrm{Sym}^m \overline{r_\iota(\pi)}|_{G_K}$ (or rather a character twist of this representation which is conjugate self-dual). We can hope to verify the residual automorphy by, informally speaking, taking an endoscopic lift from the group $U(1)^n$, where automorphic representations can be written down 'by hand' in terms of conjugate self-dual Hecke characters. However, this automorphic representation of G_n (or its base change to $\mathrm{GL}_n(\mathbf{A}_M)$) is not suitable for application of Theorem 3.9 – we need to modify it so that its local component at the place v_0 is an unramified twist of the Steinberg representation. In other words, we need to raise the level by finding a congruence to another automorphic representation of G_n which has the same residual representation but this specified local behaviour at the place v_0.

This level-raising argument is one of the most technically intricate part of the paper [**NT21a**], where it is carried out using an inductive argument, which treats the case of even m using the theory of types and the case of odd m by reference to the paper [**AT20**], which establishes a version of Ihara's lemma for unitary groups of the form $U(2, D)$ (D a division algebra). The forthcoming paper [**Tho22**] will give a direct approach (not based on induction on m).

3.7. The case of level $N > 1$. Let π be a regular algebraic cuspidal automorphic representation of $\mathrm{GL}_2(\mathbf{A}_{\mathbf{Q}})$, without CM. The arguments described so far suffice to prove the existence of the symmetric power liftings of π in every degree provided that π is everywhere unramified (therefore lifted from a newform of level 1), or more generally provided that π determines a point of the 2-adic, tame level 1 Coleman–Mazur eigencurve $\mathcal{E}_2$. We can interpret this latter condition in terms of the local component π_2: it means that π_2 has an accessible refinement.[34] We can phrase this equivalently as the condition that π_2 has non-trivial Jacquet module, or that π_2 is not supercuspidal, or that the Weil–Deligne representation $\mathrm{rec}_{\mathbf{Q}_p}(\pi_2)$ is reducible.

In order to establish the existence of the symmetric power liftings of the automorphic representation π without any conditions on its ramification, we essentially induct on the level. This kind of strategy is familiar from other situations in the literature. For example, in the original work of Wiles [**Wil95**], a modularity lifting theorem is proved first at 'minimal level' (relative to the ramification of the residual representation $\overline{\rho}$). This minimal level modularity lifting theorem is then used as a stepping stone to an $R = \mathbf{T}$ theorem where additional ramification is allowed on both the Galois and automorphic sides. Another example which is particularly relevant to us is the proof of Serre's modularity conjecture [**Ser87**] by Khare–Wintenberger [**KW09a, KW09b**]. Recall that Serre's conjecture asserts that any representation $\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_2(\overline{\mathbf{F}}_p)$ which is of 'S-type', in the sense that it is is absolutely irreducible and odd, arises as the residual representation of a newform f. Serre associated to any such representation $\overline{\rho}$ a weight $k(\overline{\rho})$ and tame conductor $N(\overline{\rho})$, and conjectured moreover that it should be possible to choose f to be of weight $k(\overline{\rho})$ and level $N(\overline{\rho})$. The approach taken in [**KW09a**] is essentially to induct on the number of primes dividing $N(\overline{\rho})$, the base case where $N(\overline{\rho}) = 1$ (in other words, where $\overline{\rho}$ is ramified only above the prime p) having been treated already in [**Kha06**]. We will come to the technique of 'killing ramification', introduced in [**KW09a**], and which makes this induction argument possible, below.

In our case it is natural to divide the proof up slightly differently into two parts, as follows. Let π again be a regular algebraic cuspidal automorphic representation of $\mathrm{GL}_2(\mathbf{A}_{\mathbf{Q}})$, without CM.

- First, we establish the existence of the symmetric power liftings of π in every degree under the assumption that π has no supercuspidal local components; in other words, that for every prime p, π_p admits an accessible refinement. In this part of the proof, we induct on the

[34] In fact, we require that π_2 admits an accessible refinement $\chi = \chi_1 \otimes \chi_2$ such that χ_1 is unramified. However, if π_2 admits an accessible refinement χ we can always twist π so that χ_1 is unramified, and twisting is easily seen to preserve the existence (or otherwise) of the symmetric power lifting in any given degree.

number of primes p such that π_p is ramified, the base case where π is everywhere unramified having been taken care of already.
- Then, we establish the existence of the symmetric power liftings of π in complete generality. In this part of the proof, we induct on the number of primes p such that π_p is supercuspidal, the base case where there are no such primes being the content of the previous bullet point.

The first bullet point takes us to the end of the paper [**NT21a**], and represents the natural reach of the methods that have been introduced so far in this article. The second bullet point is covered by the follow-up paper [**NT21b**] and requires a new idea, that of the 'functoriality lifting theorem'. We address these in turn.

3.7.1. *The everywhere locally refineable case.* Suppose then that π is a regular algebraic cuspidal automorphic representation of $\mathrm{GL}_2(\mathbf{A_Q})$, without CM, and that π_p is ramified but admits an accessible refinement χ. Let $N = Mp^r$ be the conductor of π, where $(p, M) = 1$. Then (π, χ) defines a classical point of the tame level M, p-adic eigencurve $\mathcal{E}_p(M)$ (say as defined in [**Buz07**]). Using the accumulation property of the eigencurve, it is possible to find an affinoid neighbourhood $\mathcal{U}$ of $x_{\pi,\chi}$ in $\mathcal{E}_p(M)$ and a dense set of classical points $\mathcal{E}_p(M)$ in $\mathcal{U}$ corresponding to automorphic representations which are *unramified at* p, and for which the symmetric power liftings are therefore known to exist, by our inductive hypothesis. To conclude, we invoke our 'analytic continuation of functoriality' principle. In our first statement of this above (as Theorem 3.4), we restricted to the case of the tame level 1 eigencurve $\mathcal{E}_p = \mathcal{E}_p(1)$, but in fact it is valid without restriction on the tame level (and even for automorphic representations of unitary groups over CM fields – the main technical ingredient is the vanishing of the adjoint Bloch–Kato Selmer group, which is indeed proved for RACSDC automorphic representations over an arbitrary CM field). Here is such a statement.

Theorem 3.10. *Fix $m \geq 1$. Let p be a prime number and let (π, χ) and (π', χ') be pairs consisting of regular algebraic automorphic representations, without CM, and equipped with accessible refinements of the components at p. Suppose that π, π' have prime-to-p conductors dividing M, and that the following conditions are satisfied:*

(1) *The refinement χ has small slope.*
(2) *Both refinements χ, χ' are $(m+1)$-regular.*
(3) $\mathrm{Sym}^m_*(\pi)$ *exists.*

If the classical points $x_{\pi,\chi}$, $x_{\pi',\chi'}$ of $\mathcal{E}_p(M)$ lie on a common irreducible component of $\mathcal{E}_p(M)$, then $\mathrm{Sym}^m_*(\pi)$ *exists too.*

One wrinkle is the $(m+1)$-regularity condition on the refinements. This condition does not trouble us in the level 1 case as one can show that any for any everywhere unramified regular algebraic cuspidal automorphic representation π of $\mathrm{GL}_2(\mathbf{A_Q})$, every refinement of π_2 is $(m+1)$-regular for every

$m \geq 1$ [**NT21a**, Lemma 3.4]. In general, this condition can fail (for example, take the newform corresponding to an elliptic curve E and any refinement at a prime $p \geq 5$ at which E has good supersingular reduction).

The result that can be proved most naturally by induction using Theorem 3.10 is the following one:

PROPOSITION 3.1. *Fix $m \geq 1$. If π is a regular algebraic cuspidal automorphic representation of $\mathrm{GL}_2(\mathbf{A}_{\mathbf{Q}})$ without CM such that for every prime number p such that π_p is ramified, π_p admits an accessible refinement χ_p which is $(m+1)$-regular, then $\mathrm{Sym}_*^m(\pi)$ exists.*

In order to remove the $(m+1)$-regularity condition, we use a trick based on the Taylor–Wiles method. This appears to be a novel application of the Taylor–Wiles method, which we apply to prove a kind of density statement for the local components of automorphic representations in a given congruence class. One can draw an analogy with equidistribution theorems such as those proved in [**Ser97, Shi12**], which prove that the set of local components of automorphic representations of varying weight and level is equidistributed with respect to Plancherel measure; we obtain a weaker statement, namely Zariski density of the corresponding points in the spectrum of a local lifting ring, but with the crucial difference that we are considering the set of automorphic representations with a given residual representation with respect to a fixed isomorphism $\iota : \overline{\mathbf{Q}}_p \to \mathbf{C}$. The precise statement we prove is as follows.

THEOREM 3.11. *Fix $m \geq 1$, and let π be a cuspidal automorphic representation of $\mathrm{GL}_2(\mathbf{A}_{\mathbf{Q}})$ of weight k, without CM. Suppose that for each prime number l, π_l admits an accessible refinement. Then we can find a prime number p, an isomorphism $\iota : \overline{\mathbf{Q}}_p \to \mathbf{C}$, and another cuspidal automorphic representation π' of weight k with the following properties:*

(1) *$\mathrm{Sym}_*^m(\pi)$ exists if and only if $\mathrm{Sym}_*^m(\pi')$ does.*
(2) *For each prime number l, π'_l admits an accessible refinement. If moreover π'_l is ramified, then this refinement is $(m+1)$-regular.*

This theorem, combined with Proposition 3.1, suffices to prove the existence of $\mathrm{Sym}_*^m(\pi)$ when π has no supercuspidal local components. We sketch the proof. First, we will choose the prime p to be large relative to m, k, and π[35]. We will construct π' to have the same weight k, the same residual representation $\overline{r_\iota(\pi)} \cong \overline{r_\iota(\pi')}$ with respect to some isomorphism $\iota : \overline{\mathbf{Q}}_p \to \mathbf{C}$, and also to be unramified at p. We may then apply [**BLGGT14**, Theorem 4.2.1] to conclude that the first point in the theorem statement is true (in the

[35]The needed conditions are: $p > \max(2(m+2), mk)$, $\overline{r_\iota(\pi)}$ contains a conjugate of $\mathrm{SL}_2(\mathbf{F}_p)$, and π_p is unramified. This implies in particular that for any isomorphism $\iota : \overline{\mathbf{Q}}_p \to \mathbf{C}$, $\mathrm{Sym}^m r_\iota(\pi)$ is potentially diagonalisable and we are in a position to apply the very general automorphy lifting theorems for $(m+1)$-dimensional Galois representations established in the paper [**BLGGT14**].

equivalent form that the Galois representation $\mathrm{Sym}^m r_\iota(\pi)$ is automorphic if and only if $\mathrm{Sym}^m r_\iota(\pi')$ is).

To construct π', we use Kisin's refinement of the Taylor–Wiles method. We need to introduce some relevant lifting and deformation rings. Let $\psi = \det r_\iota(\pi)$, and let S denote the set of primes at which $r_\iota(\pi)$ is ramified. Fixing a large enough coefficient field K, we let $R_{\overline{\rho},S,\psi} \in \mathcal{C}_\mathcal{O}$ denote the object representing the functor of deformations of $\overline{\rho}$ of determinant ψ, unramified outside S, and Fontaine–Laffaille of Hodge–Tate weights $\{0, k-1\}$ on the decomposition group $G_{\mathbf{Q}_p}$. If $l \in S$, we write $R_{\overline{\rho},l,\psi} \in \mathcal{C}_\mathcal{O}$ for the object representing the functor of *liftings* of $\overline{\rho}|_{G_{\mathbf{Q}_l}}$ of determinant ψ (which are moreover Fontaine–Laffaille of Hodge–Tate weights $\{0, k-1\}$ if $l = p$). Choosing p sufficiently large relative to p, we can assume that each of the rings $R_{\overline{\rho},l,\psi}$ is formally smooth over $\mathcal{O}$ of relative dimension 3.

In order to compare the global deformation ring with the local lifting rings, we need to introduce the device of framed liftings. Accordingly, we let $R^{\square_S}_{\overline{\rho},S,\psi} \in \mathcal{C}_\mathcal{O}$ denote the object representing the functor of orbits of pairs $(\rho, \{\alpha_l\}_l \in S)$, where ρ is a lifting of type $R_{\overline{\rho},S,\psi}$ and for each $l \in S$, $\alpha_l \in \ker(\mathrm{GL}_2(A) \to \mathrm{GL}_2(k))$. The orbits are taken under the action of the group $\ker(\mathrm{GL}_2(A) \to \mathrm{GL}_2(k))$, which acts by the formula $g \cdot (\rho, \{\alpha_l\}_l \in S) = (g\rho g^{-1}, \{g\alpha_l\}_{l\in S})$. There is a natural ring homomorphism $R_{\overline{\rho},S,\psi} \to R^{\square_S}_{\overline{\rho},S,\psi}$, which corresponds (under the Yoneda lemma) to the functor of 'forgetting the matrices α_l'. This ring homomorphism is formally smooth and in fact has a section, so that $R^{\square_S}_{\overline{\rho},S,\psi}$ is non-canonically isomorphic to a power series ring over $R_{\overline{\rho},S,\psi}$.

Having introduced the framed version of the global deformation ring, we introduce also the completed tensor product $R^{loc}_{\overline{\rho},S,\psi} = \widehat{\otimes}_{l\in S} R_{\overline{\rho},l,\psi} \in \mathcal{C}_\mathcal{O}$, which represents the functor of tuples $(\{\rho_l\}_{l\in S})$ of liftings of the local representations $\overline{\rho}|_{G_{\mathbf{Q}_l}}$ (these liftings supposed to be all of determinant $\psi|_{G_{\mathbf{Q}_l}}$). A tangent space calculation generalising the one described in §3.1 implies that there is a presentation

$$R^{loc}_{\overline{\rho},S,\psi}[\![X_1, \ldots, X_g]\!] \twoheadrightarrow R^{\square_S}_{\overline{\rho},S,\psi}.$$

Introducing sets Q of Taylor–Wiles primes (valid since $\overline{r_\iota(\pi)}$ is absolutely irreducible, although its symmetric power might not be), we obtain auxiliary diagrams

$$R^{loc}_{\overline{\rho},S,\psi}[\![X_1, \ldots, X_g]\!] \twoheadrightarrow R^{\square_S}_{\overline{\rho},S\cup Q,\psi} \twoheadrightarrow R^{\square_S}_{\overline{\rho},S,\psi}.$$

We may introduce spaces of automorphic forms that these deformation rings act on (through a map to an appropriate Hecke algebra). The Taylor–Wiles–Kisin patching argument can then be used to 'patch' these together to obtain a 'patched module' of automorphic forms H_∞, which is a faithful $R^{loc}_{\overline{\rho},S,\psi}[\![X_1, \ldots, X_g]\!]$-module. Typically at this point in the argument, one would 'return to level 1' to find that $R^{\square_S}_{\overline{\rho},S,\psi}$ acts faithfully (or at least has nilpotent annihilator) on a space of automorphic forms at base level, which in turn implies an automorphy lifting theorem.

Here however we want to do something slightly different. Recalling the statement of Theorem 3.11, let us suppose for contradiction that for every set Q of Taylor–Wiles primes for $\overline{\rho}$ there is no automorphic representation π' of weight k which is unramified away from $S \cup Q - \{p\}$, satisfies $\overline{r_\iota(\pi')} \cong \overline{r_\iota(\pi)}$, and with the property that π'_l is $(m+1)$-regular for each $l \in S$. This implies that for every such π', the local representations $r_\iota(\pi')|_{G_{\mathbf{Q}_l}}$ $(l \in S)$ must lie in the 'non-$(m+1)$-regular' locus of $\operatorname{Spec} R^{loc}_{\overline{\rho},S,\psi}$, a proper Zariski closed subset. This leads to a contradiction because this property is inherited by the patched module H_∞, implying in particular that H_∞ cannot be a faithful $R^{loc}_{\overline{\rho},S,\psi}[\![X_1,\ldots,X_g]\!]$-module. We deduce that there must be at least one set Q of Taylor–Wiles primes and one automorphic representation π' with the desired $(m+1)$-regular refinements at primes $l \in S$. We're now done: we just need to check that if π'_q is ramified for some $q \in Q$ then π'_q is also $(m+1)$-regular, but this is automatic since the image $r_\iota(\pi')(I_{\mathbf{Q}_q})$ is generated by a non-scalar matrix of p-power order and p is large relative to m.

3.7.2. *The case of supercuspidal local components.* We now treat the general case of our main Theorem 2.7. To reduce this case to the previous one, we use the the following 'functoriality lifting theorem':

THEOREM 3.12. *Let π, π' be cuspidal automorphic representations of $\mathrm{GL}_2(\mathbf{A}_\mathbf{Q})$ of weight 2. Let $m \geq 1$, let p be a prime number, and let $\iota : \overline{\mathbf{Q}}_p \to \mathbf{C}$ be an isomorphism such that the following conditions are satisfied:*

1. *There is an isomorphism $\overline{r_\iota(\pi)} \cong \overline{r_\iota(\pi')}$.*
2. *Neither π nor π' is ι-ordinary.*
3. *For each prime $l \neq p$, π_l is a twist of the Steinberg representation if and only if π'_l is.*
4. *There exists $a \geq 1$ such that $p^a > \max(5, 2m+1)$ and there is a sandwich*

$$\mathrm{PSL}_2(\mathbf{F}_{p^a}) \subset \operatorname{Proj} \overline{r_\iota(\pi)}(G_\mathbf{Q}) \subset \mathrm{PGL}_2(\mathbf{F}_{p^a})$$

up to conjugacy in $\mathrm{PGL}_2(\overline{\mathbf{F}}_p)$. (Here $\operatorname{Proj} \overline{r_\iota(\pi)}$ denotes the projective representation associated to $\overline{r_\iota(\pi)}$.)

Then $\mathrm{Sym}^m_*(\pi)$ exists if and only if $\mathrm{Sym}^m_*(\pi')$ does.

Let us explain why we call this a functoriality lifting theorem. It could be viewed as a kind of automorphy lifting theorem: it says that the automorphy of the Galois representation $\mathrm{Sym}^m r_\iota(\pi)$ implies the automorphy of the Galois representation $\mathrm{Sym}^m r_\iota(\pi')$, which has the same associated residual representation. However, because we impose more stringent conditions, namely that these Galois representations arise as functorial transfers of 2-dimensional Galois representations which are known already to be automorphic, we are able to obtain a more refined statement that we would otherwise. In particular, the above theorem is valid for any prime p (including the case $p = 2$!) and when p is small relative to m (in which case the residual representation $\mathrm{Sym}^m \, r_\iota(\pi)$ could have arbitrarily many Jordan–Hölder factors). This is also reflected in the proof, where instead of performing Taylor–Wiles patching on a single (pseudo-)deformation ring, we patch two (pseudo-)deformation rings (corresponding to deformations of $\overline{r_\iota(\pi)}$ and $\mathrm{tr}\,\mathrm{Sym}^m \, \overline{r_\iota(\pi)}$ respectively) and the map between them induced by the representation $\mathrm{Sym}^m : \mathrm{GL}_2 \to \mathrm{GL}_{m+1}$.

3.7.3. *Getting to the end.* Before getting on to the proof of Theorem 3.12, let's sketch how it leads to the completion of the proof of symmetric power functoriality for regular algebraic cuspidal automorphic representations of $\mathrm{GL}_2(\mathbf{A}_\mathbf{Q})$. Let π be a cuspidal automorphic representation of $\mathrm{GL}_2(\mathbf{A}_\mathbf{Q})$ of weight k and without CM. We induct on the number $sc(\pi)$ of primes l such that π_l is supercuspidal, the base case $sc(\pi) = 0$ having been already established.

The first reduction is that we can assume $k = 2$. This is familiar from the proof of Serre's conjecture: we can choose a prime p, large enough with respect to π so that for any isomorphism $\iota : \overline{\mathbf{Q}}_p \to \mathbf{C}$, $\overline{r_\iota(\pi)}(G_\mathbf{Q})$ contains a conjugate of $\mathrm{SL}_2(\mathbf{F}_p)$, π_p is unramified, and another cuspidal automorphic representation π' of $\mathrm{GL}_2(\mathbf{A}_\mathbf{Q})$ of weight 2 such that at primes $l \neq p$, π'_l and π_l are unramified twists of each other, but such that π'_p is now ramified (of principal series type, say). Then $r_\iota(\pi')$ will be potentially Barsotti–Tate and so potentially diagonalisable, in the sense of [**BLGGT14**]. Here we are using a result from [**GK14**], itself making use of Kisin's description of the irreducible components of the local lifting rings of 2-dimensional Barsotti–Tate representations [**Kis09c**]. The automorphy lifting theorems established in [**BLGGT14**] then show that $\mathrm{Sym}^m_*(\pi)$ exists if and only if $\mathrm{Sym}^m_*(\pi')$ does.

The next reduction is that, selecting a prime p such that π_p is supercuspidal, we can assume that there is an isomorphism $\iota : \overline{\mathbf{Q}}_p \to \mathbf{C}$ such that the residual representation $\overline{r_\iota(\pi)}$ satisfies the 'large image' hypothesis (4) of Theorem 3.12. The idea that one can force large image (and in particular, non-dihedral image) of the mod p Galois representation by adding in additional primes of ramification is important in the proof of [**KW09a**], where 'good dihedral primes' are introduced that force the residual image $\overline{r_\iota(\pi)}$ to be large for all primes p up to a specified bound. Using a similar idea, we can modify π so that it is 'seasoned'[36], in the sense of [**NT21b**, Definition 3.6], and therefore has large residual image with respect to at least one isomorphism $\iota : \overline{\mathbf{Q}}_p \to \mathbf{C}$ for each prime number p such that π_p is supercuspidal.

[36]This definition requires 3 prime numbers to be in a good position with respect to π. The word is supposed to suggest that, having been sprinkled with a few well-chosen primes, the automorphic representation π is especially palatable. An additional subtlety over the case considered in [**KW09a**] is that we need control of the residual representation of π including at the prime q where additional ramification is imposed.

We are now ready to carry out the induction step. We choose $\iota : \overline{\mathbf{Q}}_p \to \mathbf{C}$ such that $\overline{r_\iota(\pi)}$ has large image and π_p is supercuspidal. Using well-known principles (in particular, the theory of types as in [**Kha01**]), we can find another cuspidal automorphic representation π' of weight 2 such that $\overline{r_\iota(\pi)} \cong \overline{r_\iota(\pi')}$, such that for each prime $l \neq p$, π_l, π'_l have the same inertial type, and (crucially) such that π'_p is not supercuspidal. Thus $\mathrm{Sym}^m_*(\pi')$ exists, by induction, and Theorem 3.12 implies that $\mathrm{Sym}^m_*(\pi)$ exists too.

3.7.4. *Proving Theorem 3.12.* Let us now describe the modification of the Taylor–Wiles–Kisin method that leads to the proof of Theorem 3.12. The case $p = 2$ requires additional complications, analogous to those appearing in the article [**Kis09b**] relative to [**Kis09c**], so we concentrate on the case where p is odd. In fact, we adopt the following simplifying hypotheses:

- After base change to a soluble totally real extension $F/\mathbf{Q}$, $r_\iota(\pi)|_{G_F}$ is unramified away from p but crystalline (hence Barsotti–Tate) at the p-adic places of F.
- $\det r_\iota(\pi) = \epsilon^{-1}$.
- $[F : \mathbf{Q}]$ is even.
- Let S_p denote the set of p-adic places of F. If $v \in S_p$, then $\overline{r_\iota(\pi)}|_{G_{F_v}}$ is the trivial representation.

Let $\rho = r_\iota(\pi)|_{G_F}$, let S_p denote the set of p-adic places of F, and let $R_{\overline{\rho},S_p,\epsilon^{-1}} \in \mathcal{C}_\mathcal{O}$ denote the object representing the functor of deformations of $\overline{\rho}$ which are unramified outside S_p, of determinant ϵ^{-1}, and Barsotti–Tate non-ordinary in the sense of [**Kis09c**]. The hypothesis that $[F : \mathbf{Q}]$ is even implies that there is a quaternion algebra D over F which is split at finite places and definite at the infinite places; the deformation ring $R_{\overline{\rho},S_p,\epsilon^{-1}}$ acts on a module H_D of non-ordinary algebraic modular forms with integral structure on D.

The Taylor–Wiles–Kisin method, as exhibited in [**Kis09c**] and sketched in the previous section, makes it possible to construct an object $S_\infty \in \mathcal{C}_\mathcal{O}$, an S_∞-algebra R_∞ and an R_∞-module $H_{D,\infty}$ with the following properties:

- There is a number q such that $S_\infty \cong \mathcal{O}[\![S_1, \dots, S_q]\!]$.
- There is an isomorphism $R_\infty \otimes_{S_\infty} \mathcal{O} \cong R_{\overline{\rho},S_p,\epsilon^{-1}}$ and a compatible isomorphism $H_{D,\infty} \otimes_{S_\infty} \mathcal{O} \cong H_D$.
- $H_{D,\infty}$ is finite free as S_∞-module and R_∞ is an integral domain of the same Krull dimension as S_∞.

These facts together imply that R_∞ acts faithfully on $H_{D,\infty}$, hence that the annihilator of H_D as $R_{\overline{\rho},S_p,\epsilon^{-1}}$-module is nilpotent – and this directly implies that any deformation of $\overline{\rho}$ corresponding to a homomorphism $R_{\overline{\rho},S_p,\epsilon^{-1}} \to \mathcal{O}$ is automorphic.

We now introduce the structures associated to $\mathrm{Sym}^m \overline{\rho}$. If p is larger than m then this representation will be reducible, so we need to use pseudodeformation rings. Accordingly, we introduce a quadratic CM extension

M/F and let $\overline{t} = \operatorname{tr} \operatorname{Sym}^m \overline{\rho}|_{G_M}$, and write $P \in \mathcal{C}_{\mathcal{O}}$ for the object classifying conjugate self-dual pseudodeformations of $\overline{t}$ which are semistable with Hodge–Tate weights in the interval $[0, m]$; the same ring whose existence we have exploited already in §3.5. We introduce the definite unitary group G_{m+1} over F and write H_G for a space of algebraic modular forms on G_{m+1} with integral structure, on which the ring P acts. We observe that there is a tautological ring homomorphism

$$P \to R_{\overline{\rho}, S_p, \epsilon^{-1}},$$

which classifies the natural transformation which associates to a deformation ρ' of $\overline{\rho}$ the pseudocharacter $\operatorname{tr} \operatorname{Sym}^m \rho'|_{G_M}$. After introducing Taylor–Wiles primes and carrying out a patching argument, we can construct, in addition to the objects R_∞, $H_{D,\infty}$ considered above, an S_∞-algebra P_∞ and a P_∞-module $H_{G,\infty}$ with the following properties:

- There is an isomorphism $P_\infty \otimes_{S_\infty} \mathcal{O} \cong P$ and a compatible isomorphism $H_{G,\infty} \otimes_{S_\infty} \mathcal{O} \cong H_G$.
- $H_{G,\infty}$ is finite free as S_∞-module.
- There is a ring homomorphism $P_\infty \to R_\infty$ making the diagram

$$\begin{array}{ccc} P_\infty & \longrightarrow & R_\infty \\ \downarrow & & \downarrow \\ P & \longrightarrow & R \end{array}$$

  commute.

What we do *not* have is any control on the Krull dimension of the set of irreducible components of P_∞: getting such information for the ring R_∞ relies on being able to compute the tangent space in Galois cohomological terms and to choose Taylor–Wiles primes which kill the dual Selmer group, which is not possible for P_∞ because of the potentially degenerate form of the residual representation $\operatorname{Sym}^m \overline{\rho}|_{G_M}$.

As a substitute for this, we will use the vanishing of the adjoint Bloch–Kato Selmer group $H^1_f(F, \operatorname{ad} \operatorname{Sym}^m \rho)$. Let $\mathfrak{p}, \mathfrak{p}' \subset P$ denote the prime ideals which are the kernels of the maps $P \to \mathcal{O}$ associated to the two representations $\operatorname{Sym}^m r_\iota(\pi)|_{G_M}$ and $\operatorname{Sym}^m r_\iota(\pi')|_{G_M}$, and let $\mathfrak{p}_\infty, \mathfrak{p}'_\infty \subset P_\infty$ denote the pullbacks of these prime ideals to P_∞. The vanishing of the adjoint Bloch–Kato Selmer group implies that the Zariski tangent space of the ring $P_\mathfrak{p}$ is 0, and therefore that the Zariski tangent space of the ring $P_{\infty,\mathfrak{p}_\infty}$ has dimension at most $\dim S_\infty[1/p]$. On the other hand, if we assume that $\operatorname{Sym}^m_*(\pi)$ exists then $\mathfrak{p}_\infty$ is in the support of $H_{G,\infty}$, which is a free S_∞-module, implying that the Krull dimension of $P_{\infty,\mathfrak{p}_\infty}$ is at least $\dim S_\infty[1/p]$. It follows that $P_{\infty,(\mathfrak{p}_\infty)}$ is a regular local ring and that there is a unique irreducible component Z of $\operatorname{Spec} P_\infty$ containing $\mathfrak{p}_\infty$, which is necessarily in the support of $H_{G,\infty}$ as P_∞-module.

We now consider the map $\operatorname{Spec} R_\infty \to \operatorname{Spec} P_\infty$. The map $P_\infty \to R_\infty$ is finite, so this map has closed irreducible image of dimension $\dim S_\infty$ and containing both the points $\mathfrak{p}_\infty$, $\mathfrak{p}'_\infty$. We conclude that this image equals Z and that $\mathfrak{p}'_\infty$ is in the support of $H_{G,\infty}$ as P_∞-module. Passing back to P, we find that $\mathfrak{p}'$ is in the support of H_G as P-module, or in other words that $\mathrm{Sym}^m r_\iota(\pi')|_{G_M}$ is automorphic and $\mathrm{Sym}^m_*(\pi')$ exists, as required.

3.8. The defects of our proof. We conclude this article with a reflection on the reach of our methods. There is no doubt that reciprocity is an effective tool for studying Langlands functoriality for those automorphic representations which have associated Galois representations. After the theorems discussed in this article, the next goal would be to establish symmetric power functoriality for all automorphic representations associated to Hilbert modular forms (in other words, regular algebraic cuspidal automorphic representations of $\mathrm{GL}_2(\mathbf{A}_F)$, without CM, where F is an arbitrary totally real field), beyond the case $F = \mathbf{Q}$. Many of our techniques apply equally well in this case: in particular, the 'analytic continuation of functoriality' principle on the eigenvariety, the vanishing of the adjoint Bloch–Kato Selmer group for Galois representations associated to RACSDC automorphic representations, and the 'functoriality lifting theorem' discussed in the previous section (which is in fact stated in [**NT21b**] over an arbitrary totally real base field).

However, there is one critical ingredient which is missing in general, namely the beautifully simple structure of the 2-adic tame level 1 eigencurve. Although the rough structure of the eigencurve at the boundary of weight space may be expected to generalise (and this is even proved in many cases, as in [**LWX17**]), it is not reasonable to expect such a clean statement to hold over base fields other than $\mathbf{Q}$. One can draw an analogy here with the proof of Serre's conjecture. As we have alluded to above, this may be proved by reducing to the case of S-type representations $\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_2(\overline{\mathbf{F}}_p)$ which are unramified outside p. The proof in this case in turn is by induction on the prime p, beginning with Tate's theorem [**Tat94**] that the conjecture holds vacuously when $p = 2$ (because the set of S-type representations in this case is empty).

Serre's modularity conjecture is expected to hold equally over any totally real number field (or even over any number field at all, in the correct formulation) [**BDJ10**]. However, we cannot expect the above strategy to work without modification because when the base number field F is large enough, there are plentiful low weight S-type representations which are ramified only at the p-adic places. (For example, Dembélé [**Dem09**] constructs a surjective representation $G_F \to \mathrm{SL}_2(\mathbf{F}_{2^8})$ with $F = \mathbf{Q}(\zeta_{32}) \cap \mathbf{R}$, ramified only at the 2-adic places.) There is no base case for the induction!

In the case of symmetric power functoriality, we can point to two possible ways out. The first is to try to follow the conjectural programme outlined in [**CT14**], which aims to establish the existence of the symmetric power lifting $\mathrm{Sym}^m_*(\pi)$ by induction on m. Holding us back is the hypothesis that one already has access to sufficiently many cases of $\mathrm{GL}_2 \times \mathrm{GL}_r \to \mathrm{GL}_{2r}$ tensor product functoriality – itself a non-trivial problem. More speculatively, non-abelian (or non-soluble) descent for algebraic automorphic representations would allow one to upgrade potential automorphy, as established in great generality in the papers [**BLGGT14, PT15**] to true automorphy. We look forward to finding out what the future holds.

References

[AC89] James Arthur and Laurent Clozel. *Simple algebras, base change, and the advanced theory of the trace formula*, volume 120 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 1989. MR 1007299

[ACC+18] Patrick B. Allen, Frank Calegari, Ana Caraiani, Toby Gee, David Helm, Bao V. Le Hung, James Newton, Peter Scholze, Richard Taylor, and Jack A. Thorne. Potential automorphy over CM fields, 2018.

[AIP15] Fabrizio Andreatta, Adrian Iovita, and Vincent Pilloni. p-adic families of Siegel modular cuspforms. *Ann. of Math. (2)*, 181(2):623–697, 2015. MR 3275848

[ANT20] Patrick B. Allen, James Newton, and Jack A. Thorne. Automorphy lifting for residually reducible l-adic Galois representations, II. *Compos. Math.*, 156(11):2399–2422, 2020. MR 4190048

[Art24] E. Artin. Über eine neue art von L-Reihen. *Abh. Math. Sem. Univ. Hamburg*, 3(1):89–108, 1924. MR 3069421

[AS97] Avner Ash and Glenn Stevens. p-adic deformations of cohomology classes of subgroups of $\mathrm{GL}(n, \mathbf{Z})$. volume 48, pages 1–30. 1997. Journées Arithmétiques (Barcelona, 1995). MR 1464013

[AT20] Christos Anastassiades and Jack A. Thorne. Raising the level of automorphic representations of GL_{2n} of unitary type, 2020. MR 4454347

[BC09] Joël Bellaïche and Gaëtan Chenevier. Families of Galois representations and Selmer groups. *Astérisque*, (324):xii+314, 2009. MR 2656025

[BCDT01] Christophe Breuil, Brian Conrad, Fred Diamond, and Richard Taylor. On the modularity of elliptic curves over $\mathbf{Q}$: wild 3-adic exercises. *J. Amer. Math. Soc.*, 14(4):843–939, 2001. MR 1839918

[BCGP21] George Boxer, Frank Calegari, Toby Gee, and Vincent Pilloni. Abelian surfaces over totally real fields are potentially modular. *Publ. Math. Inst. Hautes Études Sci.*, 134:153–501, 2021. MR 4349242

[BDJ10] Kevin Buzzard, Fred Diamond, and Frazer Jarvis. On Serre's conjecture for mod l Galois representations over totally real fields. *Duke Math. J.*, 155(1):105–161, 2010. MR 2730374

[BG14] Kevin Buzzard and Toby Gee. The conjectural connections between automorphic representations and Galois representations. In *Automorphic forms and Galois representations. Vol. 1*, volume 414 of *London Math. Soc. Lecture Note Ser.*, pages 135–187. Cambridge Univ. Press, Cambridge, 2014. MR 3444225

[BG16] Kevin Buzzard and Toby Gee. Slopes of modular forms. In *Families of automorphic forms and the trace formula*, Simons Symp., pages 93–109. Springer, [Cham], 2016. MR 3675164

[BHS17] Christophe Breuil, Eugen Hellmann, and Benjamin Schraen. Une interprétation modulaire de la variété trianguline. *Math. Ann.*, 367(3-4):1587–1645, 2017. MR 3623233

[BHS19] Christophe Breuil, Eugen Hellmann, and Benjamin Schraen. A local model for the trianguline variety and applications. *Publ. Math. Inst. Hautes Études Sci.*, 130:299–412, 2019. MR 4028517

[BJ79] A. Borel and H. Jacquet. Automorphic forms and automorphic representations. In *Automorphic forms, representations and L-functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 1*, Proc. Sympos. Pure Math., XXXIII, pages 189–207. Amer. Math. Soc., Providence, R.I., 1979. With a supplement "On the notion of an automorphic representation" by R. P. Langlands. MR 0546598

[BK05] Kevin Buzzard and L. J. P. Kilford. The 2-adic eigencurve at the boundary of weight space. *Compos. Math.*, 141(3):605–619, 2005. MR 2135280

[BLGGT14] Thomas Barnet-Lamb, Toby Gee, David Geraghty, and Richard Taylor. Potential automorphy and change of weight. *Ann. of Math. (2)*, 179(2):501–609, 2014. MR 3152941

[BLGHT11] 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(1):29–98, 2011. MR 2827723

[BP19] John Bergdall and Robert Pollack. Slopes of modular forms and the ghost conjecture. *Int. Math. Res. Not. IMRN*, (4):1125–1144, 2019. MR 3915298

[Buz05] Kevin Buzzard. Questions about slopes of modular forms. Number 298, pages 1–15. 2005. Automorphic forms. I. MR 2141701

[Buz07] Kevin Buzzard. Eigenvarieties. In *L-functions and Galois representations*, volume 320 of *London Math. Soc. Lecture Note Ser.*, pages 59–120. Cambridge Univ. Press, Cambridge, 2007. MR 2392353

[Cal21] Frank Calegari. Reciprocity in the Langlands program since Fermat's Last Theorem, 2021.

[Car94] Henri Carayol. Formes modulaires et représentations galoisiennes à valeurs dans un anneau local complet. In *p-adic monodromy and the Birch and Swinnerton-Dyer conjecture (Boston, MA, 1991)*, volume 165 of *Contemp. Math.*, pages 213–237. Amer. Math. Soc., Providence, RI, 1994. MR 1279611

[CC98] F. Cherbonnier and P. Colmez. Représentations p-adiques surconvergentes. *Invent. Math.*, 133(3):581–611, 1998. MR 1645070

[Che04] Gaëtan Chenevier. Familles p-adiques de formes automorphes pour GL_n. *J. Reine Angew. Math.*, 570:143–217, 2004. MR 2075765

[Che11] Gaëtan Chenevier. On the infinite fern of Galois representations of unitary type. *Ann. Sci. Éc. Norm. Supér. (4)*, 44(6):963–1019, 2011. MR 2919688

[Che14] Gaëtan Chenevier. The p-adic analytic space of pseudocharacters of a profinite group and pseudorepresentations over arbitrary rings. In *Automorphic forms and Galois representations. Vol. 1*, volume 414 of *London Math. Soc. Lecture Note Ser.*, pages 221–285. Cambridge Univ. Press, Cambridge, 2014. MR 3444227

[CHT08] 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):1–181, 2008. With Appendix A, summarizing unpublished work of Russ Mann, and Appendix B by Marie-France Vignéras. MR 2470687

[Clo90] Laurent Clozel. Motifs et formes automorphes: applications du principe de fonctorialité. In *Automorphic forms, Shimura varieties, and L-functions, Vol. I (Ann Arbor, MI, 1988)*, volume 10 of *Perspect. Math.*, pages 77–159. Academic Press, Boston, MA, 1990. MR 1044819

[Clo91] Laurent Clozel. Représentations galoisiennes associées aux représentations automorphes autoduales de $\mathrm{GL}(n)$. *Inst. Hautes Études Sci. Publ. Math.*, (73):97–145, 1991. MR 1114211

[CM98] R. Coleman and B. Mazur. The eigencurve. In *Galois representations in arithmetic algebraic geometry (Durham, 1996)*, volume 254 of *London Math. Soc.*

Lecture Note Ser., pages 1–113. Cambridge Univ. Press, Cambridge, 1998. MR 1696469

[Col08] Pierre Colmez. Représentations triangulines de dimension 2. Number 319, pages 213–258. 2008. Représentations p-adiques de groupes p-adiques. I. Représentations galoisiennes et (ϕ, Γ)-modules. MR 2493219

[Con99] Brian Conrad. Irreducible components of rigid spaces. *Ann. Inst. Fourier (Grenoble)*, 49(2):473–541, 1999. MR 1697371

[CQ15] D. Choimet and H. Queffélec. *Twelve landmarks of twentieth-century analysis*. Cambridge University Press, New York, 2015. Illustrated by Michaël Monerau, Translated from the 2009 French original by Danièle Gibbons and Greg Gibbons, With a foreword by Gilles Godefroy. MR 3445361

[CT14] Laurent Clozel and Jack A. Thorne. Level-raising and symmetric power functoriality, I. *Compos. Math.*, 150(5):729–748, 2014. MR 3209793

[CT15] Laurent Clozel and Jack A. Thorne. Level raising and symmetric power functoriality, II. *Ann. of Math. (2)*, 181(1):303–359, 2015. MR 3272927

[CT17] Laurent Clozel and Jack A. Thorne. Level-raising and symmetric power functoriality, III. *Duke Math. J.*, 166(2):325–402, 2017. MR 3600753

[Del71a] Pierre Deligne. Formes modulaires et représentations l-adiques. In *Séminaire Bourbaki. Vol. 1968/69: Exposés 347–363*, volume 175 of *Lecture Notes in Math.*, pages Exp. No. 355, 139–172. Springer, Berlin, 1971. MR 3077124

[Del71b] Pierre Deligne. Travaux de Shimura. In *Séminaire Bourbaki, 23ème année (1970/71), Exp. No. 389*, pages 123–165. Lecture Notes in Math., Vol. 244. 1971. MR 0498581

[Del74] Pierre Deligne. La conjecture de Weil. I. *Inst. Hautes Études Sci. Publ. Math.*, (43):273–307, 1974. MR 0340258

[Dem09] Lassina Dembélé. A non-solvable Galois extension of $\mathbf{Q}$ ramified at 2 only. *C. R. Math. Acad. Sci. Paris*, 347(3-4):111–116, 2009. MR 2538094

[dJ95] A. J. de Jong. Crystalline Dieudonné module theory via formal and rigid geometry. *Inst. Hautes Études Sci. Publ. Math.*, (82):5–96 (1996), 1995. MR 1383213

[dJ96] A. J. de Jong. Smoothness, semi-stability and alterations. *Inst. Hautes Études Sci. Publ. Math.*, (83):51–93, 1996. MR 1423020

[DL16] Hansheng Diao and Ruochuan Liu. The eigencurve is proper. *Duke Math. J.*, 165(7):1381–1395, 2016. MR 3498869

[DS74] Pierre Deligne and Jean-Pierre Serre. Formes modulaires de poids 1. *Ann. Sci. École Norm. Sup. (4)*, 7:507–530 (1975), 1974. MR 0379379

[Eme] Local-global compatibility in the p-adic Langlands programme for $\mathrm{GL}_2/\mathbf{Q}$.

[Eme06] Matthew Emerton. On the interpolation of systems of eigenvalues attached to automorphic Hecke eigenforms. *Invent. Math.*, 164(1):1–84, 2006. MR 2207783

[Eme18] Kathleen Emerson. *Comparison of different definitions of pseudocharacter*. PhD thesis, Princeton University, November 2018.

[Eme21] Matthew Emerton. Langlands reciprocity: L-functions, automorphic forms, and Diophantine equations. In *The genesis of the Langlands Program*, volume 467 of *London Math. Soc. Lecture Note Ser.*, pages 301–386. Cambridge Univ. Press, Cambridge, 2021. MR 4274535

[FM95] Jean-Marc Fontaine and Barry Mazur. Geometric Galois representations. In *Elliptic curves, modular forms, & Fermat's last theorem (Hong Kong, 1993)*, Ser. Number Theory, I, pages 41–78. Int. Press, Cambridge, MA, 1995. MR 1363495

[Fon82] Jean-Marc Fontaine. Sur certains types de représentations p-adiques du groupe de Galois d'un corps local; construction d'un anneau de Barsotti-Tate. *Ann. of Math. (2)*, 115(3):529–577, 1982. MR 0657238

[Fon92] Jean-Marc Fontaine. Valeurs spéciales des fonctions L des motifs. Number 206, pages Exp. No. 751, 4, 205–249. 1992. Séminaire Bourbaki, Vol. 1991/92. MR 1206069

[Fon94a] Jean-Marc Fontaine. Le corps des périodes p-adiques. Number 223, pages 59–111. 1994. With an appendix by Pierre Colmez, Périodes p-adiques (Bures-sur-Yvette, 1988). MR 1293971

[Fon94b] Jean-Marc Fontaine. Représentations l-adiques potentiellement semi-stables. Number 223, pages 321–347. 1994. Périodes p-adiques (Bures-sur-Yvette, 1988). MR 1293977

[Fon94c] Jean-Marc Fontaine. Représentations p-adiques semi-stables. Number 223, pages 113–184. 1994. With an appendix by Pierre Colmez, Périodes p-adiques (Bures-sur-Yvette, 1988). MR 1293972

[Gau66] Carl Friedrich Gauss. *Disquisitiones arithmeticae*. Yale University Press, New Haven, Conn.-London, 1966. Translated into English by Arthur A. Clarke, S. J. MR 0197380

[Gel75] Stephen S. Gelbart. *Automorphic forms on adèle groups*. Annals of Mathematics Studies, No. 83. Princeton University Press, Princeton, N.J.; University of Tokyo Press, Tokyo, 1975. MR 0379375

[Ger19] David Geraghty. Modularity lifting theorems for ordinary Galois representations. *Math. Ann.*, 373(3-4):1341–1427, 2019. MR 3953131

[GJ72] Roger Godement and Hervé Jacquet. *Zeta functions of simple algebras*. Lecture Notes in Mathematics, Vol. 260. Springer-Verlag, Berlin-New York, 1972. MR 0342495

[GK14] Toby Gee and Mark Kisin. The Breuil-Mézard conjecture for potentially Barsotti-Tate representations. *Forum Math. Pi*, 2:e1, 56, 2014. MR 3292675

[Han17] David Hansen. Universal eigenvarieties, trianguline Galois representations, and p-adic Langlands functoriality. *J. Reine Angew. Math.*, 730:1–64, 2017. With an appendix by James Newton. MR 3692014

[Hid86] Haruzo Hida. Galois representations into $\mathrm{GL}_2(\mathbf{Z}_p[[X]])$ attached to ordinary cusp forms. *Invent. Math.*, 85(3):545–613, 1986. MR 0848685

[HSBT10] Michael Harris, Nick Shepherd-Barron, and Richard Taylor. A family of Calabi-Yau varieties and potential automorphy. *Ann. of Math. (2)*, 171(2):779–813, 2010. MR 2630056

[JS81] H. Jacquet and J. A. Shalika. On Euler products and the classification of automorphic representations. I. *Amer. J. Math.*, 103(3):499–558, 1981. MR 0618323

[JT20] Christian Johansson and Jack A. Thorne. *On subquotients of the étale cohomology of Shimura varieties*, page 306–334. London Mathematical Society Lecture Note Series. Cambridge University Press, 2020. MR 4439215

[Ked04] Kiran S. Kedlaya. A p-adic local monodromy theorem. *Ann. of Math. (2)*, 160(1):93–184, 2004. MR 2119719

[Kha01] Chandrashekhar Khare. A local analysis of congruences in the (p,p) case. II. *Invent. Math.*, 143(1):129–155, 2001. MR 1802794

[Kha06] Chandrashekhar Khare. Serre's modularity conjecture: the level one case. *Duke Math. J.*, 134(3):557–589, 2006. MR 2254626

[Kis03] Mark Kisin. Overconvergent modular forms and the Fontaine-Mazur conjecture. *Invent. Math.*, 153(2):373–454, 2003. MR 1992017

[Kis04] Mark Kisin. Geometric deformations of modular Galois representations. *Invent. Math.*, 157(2):275–328, 2004. MR 2076924

[Kis09a] Mark Kisin. The Fontaine-Mazur conjecture for GL_2. *J. Amer. Math. Soc.*, 22(3):641–690, 2009. MR 2505297

[Kis09b] Mark Kisin. Modularity of 2-adic Barsotti-Tate representations. *Invent. Math.*, 178(3):587–634, 2009. MR 2551765

[Kis09c] Mark Kisin. Moduli of finite flat group schemes, and modularity. *Ann. of Math. (2)*, 170(3):1085–1180, 2009. MR 2600871

[KMSW14] Tasho Kaletha, Alberto Minguez, Sug Woo Shin, and Paul-James White. Endoscopic classification of representations: Inner forms of unitary groups, 2014.

[Kot92] Robert E. Kottwitz. On the λ-adic representations associated to some simple Shimura varieties. *Invent. Math.*, 108(3):653–665, 1992. MR 1163241

[KPX14] Kiran S. Kedlaya, Jonathan Pottharst, and Liang Xiao. Cohomology of arithmetic families of (φ, Γ)-modules. *J. Amer. Math. Soc.*, 27(4):1043–1115, 2014. MR 3230818

[KW09a] Chandrashekhar Khare and Jean-Pierre Wintenberger. Serre's modularity conjecture. I. *Invent. Math.*, 178(3):485–504, 2009. MR 2551763

[KW09b] Chandrashekhar Khare and Jean-Pierre Wintenberger. Serre's modularity conjecture. II. *Invent. Math.*, 178(3):505–586, 2009. MR 2551764

[Lab11] J.-P. Labesse. Changement de base CM et séries discrètes. In *On the stabilization of the trace formula*, volume 1 of *Stab. Trace Formula Shimura Var. Arith. Appl.*, pages 429–470. Int. Press, Somerville, MA, 2011. MR 2856380

[Laf18] Vincent Lafforgue. Chtoucas pour les groupes réductifs et paramétrisation de Langlands globale. *J. Amer. Math. Soc.*, 31(3):719–891, 2018. MR 3787407

[Lan70] R. P. Langlands. Problems in the theory of automorphic forms. In *Lectures in modern analysis and applications, III*, pages 18–61. Lecture Notes in Math., Vol. 170. 1970. MR 0302614

[Lan77] R. P. Langlands. Shimura varieties and the Selberg trace formula. *Canadian J. Math.*, 29(6):1292–1299, 1977. MR 0498400

[Lan89] R. P. Langlands. On the classification of irreducible representations of real algebraic groups. In *Representation theory and harmonic analysis on semisimple Lie groups*, volume 31 of *Math. Surveys Monogr.*, pages 101–170. Amer. Math. Soc., Providence, RI, 1989. MR 1011897

[Lan94] Serge Lang. *Algebraic number theory*, volume 110 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1994. MR 1282723

[Liu07] Tong Liu. Torsion p-adic Galois representations and a conjecture of Fontaine. *Ann. Sci. École Norm. Sup. (4)*, 40(4):633–674, 2007. MR 2191528

[Liu15] R. Liu. Triangulation of refined families. *Comment. Math. Helv.*, 90(4):831–904, 2015. MR 3433281

[LWX17] Ruochuan Liu, Daqing Wan, and Liang Xiao. The eigencurve over the boundary of weight space. *Duke Math. J.*, 166(9):1739–1787, 2017. MR 3662443

[Maz89] B. Mazur. Deforming Galois representations. In *Galois groups over* $\mathbf{Q}$ *(Berkeley, CA, 1987)*, volume 16 of *Math. Sci. Res. Inst. Publ.*, pages 385–437. Springer, New York, 1989. MR 1012172

[Mor20] L. J. Mordell. On Mr Ramanujan's empirical expansions of modular functions. *Proceedings of the Cambridge Philosophical Society*, 19:117–124, 1920.

[New98] Donald J. Newman. *Analytic number theory*, volume 177 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1998. MR 1488421

[New22] James Newton. Modularity of Galois representations and Langlands functoriality, 2022.

[NT20] James Newton and Jack A. Thorne. Adjoint Selmer groups of automorphic Galois representations of unitary type, 2020.

[NT21a] James Newton and Jack A. Thorne. Symmetric power functoriality for holomorphic modular forms. *Publ. Math. Inst. Hautes Études Sci.*, 134:1–116, 2021. MR 4349240

[NT21b] James Newton and Jack A. Thorne. Symmetric power functoriality for holomorphic modular forms, II. *Publ. Math. Inst. Hautes Études Sci.*, 134:117–152, 2021. MR 4349241

[Pan21] Lue Pan. The Fontaine-Mazur conjecture in the residually reducible case, 2021. MR 4467307

[PT15] Stefan Patrikis and Richard Taylor. Automorphy and irreducibility of some l-adic representations. *Compos. Math.*, 151(2):207–229, 2015. MR 3314824

[Ram16] S. Ramanujan. On certain arithmetical functions. *Transactions of the Cambridge Philosophical Society*, 22(9):159–184, 1916. MR 4055699

[Rie60] B. Riemann. Über die Anzahl der Primzahlen unter einer gegebenen größe. *Monatsberichte der Berliner Akademie 1859*, pages 671–680, 1860.

[Sai03] Takeshi Saito. Weight spectral sequences and independence of l. *J. Inst. Math. Jussieu*, 2(4):583–634, 2003. MR 2006800

[Ser68] Jean-Pierre Serre. *Abelian l-adic representations and elliptic curves*. W. A. Benjamin, Inc., New York-Amsterdam, 1968. McGill University lecture notes written with the collaboration of Willem Kuyk and John Labute. MR 0263823

[Ser69] Jean-Pierre Serre. Une interprétation des congruences relatives à la fonction τ de Ramanujan. In *Séminaire Delange-Pisot-Poitou: 1967/68, Théorie des Nombres, Fasc. 1, Exp. 14*, page 17. Secrétariat mathématique, Paris, 1969. MR 0244147

[Ser70] Jean-Pierre Serre. Facteurs locaux des fonctions zêta des varietés algébriques (définitions et conjectures). In *Séminaire Delange-Pisot-Poitou. 11e année: 1969/70. Théorie des nombres. Fasc. 1: Exposés 1 à 15; Fasc. 2: Exposés 16 à 24*, page 15. Secrétariat Math., Paris, 1970. MR 3618526

[Ser72] Jean-Pierre Serre. Propriétés galoisiennes des points d'ordre fini des courbes elliptiques. *Invent. Math.*, 15(4):259–331, 1972. MR 0387283

[Ser87] Jean-Pierre Serre. Sur les représentations modulaires de degré 2 de $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. *Duke Math. J.*, 54(1):179–230, 1987. MR 0885783

[Ser94] Jean-Pierre Serre. Propriétés conjecturales des groupes de Galois motiviques et des représentations l-adiques. In *Motives (Seattle, WA, 1991)*, volume 55 of *Proc. Sympos. Pure Math.*, pages 377–400. Amer. Math. Soc., Providence, RI, 1994. MR 1265537

[Ser97] Jean-Pierre Serre. Répartition asymptotique des valeurs propres de l'opérateur de Hecke T_p. *J. Amer. Math. Soc.*, 10(1):75–102, 1997. MR 1396897

[Ser12] Jean-Pierre Serre. *Lectures on $N_X(p)$*, volume 11 of *Chapman & Hall/CRC Research Notes in Mathematics*. CRC Press, Boca Raton, FL, 2012. MR 2920749

[Ser16] Jean-Pierre Serre. Lettre á Armand Borel. In *Frobenius distributions: Lang-Trotter and Sato-Tate conjectures*, volume 663 of *Contemp. Math.*, pages 1–9. Amer. Math. Soc., Providence, RI, 2016. MR 3502936

[Shi12] Sug Woo Shin. Automorphic Plancherel density theorem. *Israel J. Math.*, 192(1):83–120, 2012. MR 3004076

[Shi20] Sug Woo Shin. *Construction of automorphic Galois representations: the self-dual case*, page 209–250. London Mathematical Society Lecture Note Series. Cambridge University Press, 2020. MR 4439211

[SL96] P. Stevenhagen and H. W. Lenstra, Jr. Chebotarëv and his density theorem. *Math. Intelligencer*, 18(2):26–37, 1996. MR 1395088

[Sor20] Claus M. Sorensen. *A Patching Lemma*, page 297–305. London Mathematical Society Lecture Note Series. Cambridge University Press, 2020. MR 4439214

[SW99] C. M. Skinner and A. J. Wiles. Residually reducible representations and modular forms. *Inst. Hautes Études Sci. Publ. Math.*, (89):5–126 (2000), 1999. MR 1793414

[Tat65] John T. Tate. Algebraic cycles and poles of zeta functions. In *Arithmetical Algebraic Geometry (Proc. Conf. Purdue Univ., 1963)*, pages 93–110. Harper & Row, New York, 1965. MR 0225778

[Tat67] J. T. Tate. p-divisible groups. In *Proc. Conf. Local Fields (Driebergen, 1966)*, pages 158–183. Springer, Berlin, 1967. MR 0231827

[Tat79] J. Tate. Number theoretic background. In *Automorphic forms, representations and L-functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 2*, Proc. Sympos. Pure Math., XXXIII, pages 3–26. Amer. Math. Soc., Providence, R.I., 1979. MR 0546607

[Tat94] John Tate. The non-existence of certain Galois extensions of $\mathbf{Q}$ unramified outside 2. In *Arithmetic geometry (Tempe, AZ, 1993)*, volume 174 of *Contemp. Math.*, pages 153–156. Amer. Math. Soc., Providence, RI, 1994. MR 1299740

[Tay04] Richard Taylor. Galois representations. *Ann. Fac. Sci. Toulouse Math. (6)*, 13(1):73–119, 2004. MR 2060030

[Tay08] Richard Taylor. Automorphy for some l-adic lifts of automorphic mod l Galois representations. II. *Publ. Math. Inst. Hautes Études Sci.*, (108):183–239, 2008. MR 2470688

[Tho15] Jack A. Thorne. Automorphy lifting for residually reducible l-adic Galois representations. *J. Amer. Math. Soc.*, 28(3):785–870, 2015. MR 3327536

[Tho21a] Jack A. Thorne. On the vanishing of adjoint Bloch–Kato Selmer groups of irreducible automorphic Galois representations, 2021. MR 4538044

[Tho21b] Jesse Thorner. Effective forms of the Sato–Tate conjecture. *Research in the Mathematical Sciences*, 8(1), Jan 2021. MR 4198758

[Tho22] Jack A. Thorne. A p-adic approach to level-raising for GL_n, 2022.

[Tsu98] Takeshi Tsuji. p-adic Hodge theory in the semi-stable reduction case. In *Proceedings of the International Congress of Mathematicians, Vol. II (Berlin, 1998)*, number Extra Vol. II, pages 207–216, 1998. MR 1648071

[Tun81] Jerrold Tunnell. Artin's conjecture for representations of octahedral type. *Bull. Amer. Math. Soc. (N.S.)*, 5(2):173–175, 1981. MR 0621884

[TW95] Richard Taylor and Andrew Wiles. Ring-theoretic properties of certain Hecke algebras. *Ann. of Math. (2)*, 141(3):553–572, 1995. MR 1333036

[Urb11] Eric Urban. Eigenvarieties for reductive groups. *Ann. of Math. (2)*, 174(3):1685–1784, 2011. MR 2846490

[Wal88] Nolan R. Wallach. *Real reductive groups. I*, volume 132 of *Pure and Applied Mathematics*. Academic Press, Inc., Boston, MA, 1988. MR 0929683

[WE20] Carl Wang-Erickson. Presentations of non-commutative deformation rings via A_∞-algebras and applications to deformations of Galois representations and pseudorepresentations, 2020.

[Wei74] André Weil. *Basic number theory*. Die Grundlehren der mathematischen Wissenschaften, Band 144. Springer-Verlag, New York-Berlin, third edition, 1974. MR 0427267

[Wil88] A. Wiles. On ordinary λ-adic representations associated to modular forms. *Invent. Math.*, 94(3):529–573, 1988. MR 0969243

[Wil95] Andrew Wiles. Modular elliptic curves and Fermat's last theorem. *Ann. of Math. (2)*, 141(3):443–551, 1995. MR 1333035

[Wil06] Andrew Wiles. The Birch and Swinnerton-Dyer conjecture. In *The millennium prize problems*, pages 31–41. Clay Math. Inst., Cambridge, MA, 2006. MR 2238272

[WWE19] Preston Wake and Carl Wang-Erickson. Deformation conditions for pseudorepresentations. *Forum Math. Sigma*, 7:Paper No. e20, 44, 2019. MR 3987305

DPMMS, Wilberforce Road, Cambridge, CB3 0WB, UK
Email address: thorne@dpmms.cam.ac.uk