Character sums, automorphic forms, equidistribution, and Ramanujan graphs [1]

Ching-Li Chai[2] and Wen-Ching Winnie Li[3]

Version 3.1, January 4, 2001

Abstract

According to the Langlands program, L-functions arising from families of exponential sums over global function fields are automorphic L-functions. We illustrate this general principle in examples related to Kloosterman sums and Ramanujan graphs.

Part I. The Kloosterman sum conjecture over function fields

§1. Introduction

Let K be a function field of one variable with the field of constants $\mathbb{F}$ being a finite field with q elements. Denote by p the characteristic of $\mathbb{F}$. Let E be an elliptic curve defined over K. For the sake of simplicity, assume that p is at least 5. Then E can be defined by an equation $y^2 = g(x)$ for some polynomial g over K of degree three. At almost all places v of K, E has a good reduction and the resulting elliptic curve E_v has zeta function

$$Z(s, E_v) = \frac{1 - a_v Nv^{-s} + Nv^{1-2s}}{(1 - Nv^{-s})(1 - Nv^{1-s})}.$$

Here Nv, the norm of v, is the cardinality of the residue field $\mathbb{F}_v$ of K at v. The Hasse-Weil L-function attached to E is

$$L(s, E) = \prod_{v \quad \text{good}} \frac{1}{1 - a_v Nv^{-s} + Nv^{1-2s}} \prod_{v \quad \text{bad}} \frac{1}{1 - a_v Nv^{-s}}.$$

At a good place v the coefficient a_v is equal to $1 + Nv$ minus the number of points of E_v over $\mathbb{F}_v$, which in turn can be expressed as the character sum

$$a_v = -\sum_{x \in \mathbb{F}_v} \chi \circ \mathrm{Nm}_{\mathbb{F}_v/\mathbb{F}}(g_v(x)).$$

Here χ is the unique multiplicative character of $\mathbb{F}^\times$ of order 2 and g_v denotes $g \mod v$, which is a polynomial over $\mathbb{F}_v$. Due to the work of Grothendieck and Deligne [5], one can apply the converse theorem for GL_2 proved by Jacquet and Langlands [14] to conclude that $L(s, E)$ is an automorphic L-function for GL_2 over K. Notice that each Euler factor $1 - a_v Nv^{-s} + Nv^{1-2s}$ is in fact the L-function of the idele class character η_v attached to the quadratic extension of the rational function field $\mathbb{F}_v(t)$ defined by $y^2 = g_v(x)$. Furthermore, Yoshida [36] has shown that the Sato-Tate conjecture holds for E when the j-invariant of E is not in $\mathbb{F}$. In this case, we get examples of automorphic forms for GL_2 over K for which the Sato-Tate conjecture holds.

[1] 2000 MSC: 11G40, 11T23, 22E55, 05C99

[2] partially supported by grant DMS 9800609 from the National Science Foundation

[3] partially supported by grant DMS96-22938 and DMS 99-70651 from the National Science Foundation

The above theme holds more generally for character sums over the rational points of a variety defined over a finite field. Given a family of character sums, often one can find a morphism $\pi : X \to B$ of schemes over $\mathbb{F}$ and a (complex of) $\overline{\mathbb{Q}_\ell}$-sheaf $\mathcal{L}$ on X, such that a typical sum in the given family is the sum of the traces of the action on (the geometric fiber of) $\mathcal{L}$ of the Frobenii of all $\mathbb{F}$-rational points of X lying over a fixed rational point $b \in B(\mathbb{F})$; similarly for finite extension fields of $\mathbb{F}$. We assume that the sheaf $\mathcal{L}$ is a *mixed* sheaf in the sense of [7], as this is the case for most applications. For a family of character sums as above, one can apply the method of ℓ-adic cohomology to obtain information about the character sums. We recall some basic facts.

(1) The cohomology sheaves $\mathrm{R}^i\pi_!(\mathcal{L})$ are mixed $\overline{\mathbb{Q}_\ell}$-sheaves by Deligne [7]: especially it is a successive extension of smooth pure $\overline{\mathbb{Q}_\ell}$-sheaves over a dense open subscheme U of B. The weights of $\mathrm{R}^i\pi_!(\mathcal{L})$ provides sharp estimates of the character sums in question.

(2) By Grothendieck, the L-function attached to this family of character sums is a rational function (in q^{-s}); Poincaré duality yields a functional equation for the L-function.

(3) The local constants for the functional equation coincides with the local constants used in the theory of automorphic representations by Deligne [5] and Laumon [21].

(4) In examples one often can compute the vanishing cycles of the cohomology sheaves $\mathrm{R}^i\pi_!(\mathcal{L})$ at the "bad" places of the base scheme B. That is, the local L-factors at the bad places often can be determined explicitly.

(5) The traces of Frobenii of closed points of the base scheme B on the cohomology sheaves $\mathrm{R}^i\pi_!(\mathcal{L})$ is equidistributed; the distribution law is governed by the geometric monodromy group of $\mathrm{R}^i\pi_!(\mathcal{L})$, i.e. the Zariski closure of the action of the geometric fundamental group of U. The vanishing cycle sheaves often help one determine the geometric monodromy group.

The L-function attached to $\mathrm{R}^i\pi_!(\mathcal{L})$ is, up to finitely many Euler factors, a finite product of L-functions $L(\rho_j, s)$ attached to irreducible $\overline{\mathbb{Q}_\ell}$-representations ρ_j of the Galois group of the function field $\mathbb{F}(B)$ of B, where the ρ_j's are the irreducible subquotients of the action of the Galois group of $\mathbb{F}(B)$ on the geometric generic fiber of $\mathrm{R}^i\pi_!(\mathcal{L})$. For each irreducible $\overline{\mathbb{Q}_\ell}$-representation ρ of the Galois group of $\mathbb{F}(B)$, the L-function attached to $\rho_j \otimes \rho$ satisfies the "correct" functional equation by Grothendieck, Deligne and Laumon, just as stated in (2), (3) above. One deduces that $L(\rho_j, s)$ is an automorphic L-function attached to $\mathrm{GL}(n_j)$ if the degree n_j of ρ_j is two by the converse theorem for $GL(2)$ proved by Hecke- Weil-Jacquet-Langlands [14], and if $n_j = 3$ by the converse theorem for $GL(3)$ established by Jacquet, Piatetski-Shapiro and Shalika [15]. The general converse theorem for $\mathrm{GL}(n)$ has been established by Cogdell and Piatetski-Shapiro [3], [2]. This combined with the global Langlands correspondence for GL(2) over function fields proved by Drinfel'd [9] implies that $L(\rho_j, s)$ is an automorphic L-function attached to GL(4) if the degree of ρ_j is 4. Recently Lafforgue announced the proof of the Langlands correspondence over function fields for $\mathrm{GL}(n)$, see [18] and [19]. So the L function attached to a family of character sums in the above setting is a quotient of finite products of automorphic L-functions.

In this two-part paper we illustrate the above theme by exhibiting several examples, with character sums arising from eigenvalues of Ramanujan graphs. More precisely, in Part I we consider Kloosterman sums, and in Part II the eigenvalues of Terras graphs. Part I of this paper is organized as follows. In §2 we give two "prototype theorems", Theorem 2.2 and Theorem 2.3, where the base scheme B is a smooth algebraic curve C over $\mathbb{F}$, $X = C \times \mathbb{P}^1$ and f is the projection $C \times \mathbb{P}^1 \to C$. In §3 we prove the Kloosterman sum conjecture over a function field K, which asserts the existence of automorphic forms of $GL(2)$ over K whose Fourier coefficients are Kloosterman sums. It is interesting to note that recent computations by A. Booker [1] based on an idea of Sarnak seem to suggest that the analogous statement over $\mathbb{Q}$, as questioned by Katz in [17], does not hold. More precisely, he showed that if an automorphic form for GL_2 over $\mathbb{Q}$ with prescribed Fourier coefficients given by Kloosterman sums were to exist, then either the level or the eigenvalue at infinity of the form would be very large. The sharp contrast between the number field case and the function field case of this problem is very interesting. We also show the equidistribution of Kloosterman sums by computing the geometric monodromy group. The connection with a family of Ramanujan graphs, called *norm graphs*, is discussed in §4, where the base curve C is specified to $\mathbb{P}^1$. For this case one has detailed information about the vanishing cycles and local factors at bad places, which in turn enable us to show that the automorphic forms occurring in the Kloosterman sum conjecture actually live on a quaternion group. In other words, they can be viewed as functions on certain Ramanujan graphs constructed by Morgenstern, arising from lifting functions on norm graphs.

§2. Two prototype families of automorphic L-functions

Observe that an idele class character of a function field over a finite field $\mathbb{F}$ with q elements, if not principal, has the associated L-function a polynomial in q^{-s} with coefficients being character sums. On the other hand, an automorphic L-function $L(s)$ of GL_n over K has an Euler product over the places of K with the factor at almost all places v being the reciprocal of a polynomial in Nv^{-s} with degree n. Given an additive or multiplicative character of $\mathbb{F}$ and a rational function $f \in K$, we construct in this section an automorphic L-functions $L(s)$ by taking a product of appropriate L-functions of GL_1, that is, L-functions attached to idele class characters of function fields.

(2.1) Let $f(x)$ be a rational function with coefficients in K. Only finitely many places of K occur as zeros or poles of the coefficients of f. These are the 'possibly bad' places, and the remaining places are called 'good' places of f. At a good place v, the coefficients of f are units in the completion of K at v. By passing to the residue field $\mathbb{F}_v$, we get a rational function $f_v := f \mod v$ with coefficients in $\mathbb{F}_v$.

To facilitate our statement, we use the notation $L_1 \sim L_2$ to mean that two Euler products L_1 and L_2 over places of K have the same Euler factors at almost all places. Since the L-functions attached to automorphic representations of GL_n over K satisfy the multiplicity one theorem, and all L-functions considered in this paper come from representations, it suffices to know the Euler factors for almost all places. Moreover, an L-function $L(s, \rho)$ attached to a representation ρ is said to *behave nicely* if it is a rational function in q^{-s} and it satisfies a functional equation

$$L(s, \rho) = \varepsilon(s) L(1 - s, \check{\rho}),$$

where $\check{\rho}$ is the contragredient of ρ and $\varepsilon(s)$ is a constant times a positive power of q^{-s}. An automorphic L-function for GL_n over K is known to behave nicely, as proved by Hecke [12] [13] and Tate [33] for $n = 1$, Jacquet and Langlands [14] for $n = 2$ and Godement and Jacquet [11] for $n \geq 3$. Moreover, the L-functions attached to cuspidal representations are in fact polynomials in q^{-s}.

We fix a choice of a nontrivial additive character of a finite field as follows. For the prime field $\mathbb{Z}/p\mathbb{Z}$, we choose ψ_p to be $\psi_p(x) = e^{2\pi i x/p}$ and for any finite field $\mathbb{F}$ of characteristic p, choose $\psi = \psi_p \circ \mathrm{Tr}_{\mathbb{F}/(\mathbb{Z}/p\mathbb{Z})}$.

(2.2) Theorem A *Let $f(x) \in K(x)$ be a non-constant rational function over K.*

(A1) *At each good place v of f, there exists an idele class character η_{ψ,f_v} of $\mathbb{F}_v(t)$ such that at the places w of $\mathbb{F}_v(t)$ where η_{ψ,f_v} is unramified, its value at a uniformizer π_w is given by the character sum*

$$\eta_{\psi,f_v}(\pi_w) = \psi \circ \mathrm{Tr}_{\mathbb{F}_v/\mathbb{F}}(\sum_{\alpha} f_v(\alpha)),$$

where α runs through all roots of π_w in an algebraic closure $\bar{\mathbb{F}}$ of $\mathbb{F}$.

(A2) *Suppose that $f(x)$ is not of the form $h_1(x)^p - h_1(x) + h_2$ for any $h_1(x) \in K(x)$ and $h_2 \in K$. Then the L-function $L(s, \eta_{\psi,f_v})$ attached to η_{ψ,f_v} is a polynomial in Nv^{-s} of degree n equal to the degree of the conductor of η_{ψ,f_v} minus 2, and n is the same for almost all v.*

(A3) *With the same assumption as in (A2), there exists a compatible family of ℓ-adic representations of $\mathrm{Gal}(K^{sep}/K)$, depending on ψ and f, such that its associated L-function $L(s, \psi, f)$ satisfies*

$$L(s, \psi, f) \sim \prod_{v \ \ good} L(s, \eta_{\psi,f_v})^{-1}.$$

Further $L(s, \psi, f)$ and its twists by idele class characters of K and by all ℓ-adic representations of $\mathrm{Gal}(K^{sep}/K)$ behave nicely.

(2.3) Theorem B *Let $g(x) \in K(x)$ be a non-constant rational function over K and let χ be a nontrivial character of $\mathbb{F}^\times$ of order d.*

(B1) *At a good place v of g, there exists an idele class character η_{χ,g_v} of $\mathbb{F}_v(t)$ such that at the places w of $\mathbb{F}_v(t)$ where η_{χ,g_v} is unramified, its value is given by the character sum*

$$\eta_{\chi,g_v}(\pi_w) = \chi \circ \mathrm{Nm}_{\mathbb{F}_v/\mathbb{F}}(\prod_{\alpha} g_v(\alpha)),$$

where α runs through all roots of π_w in $\bar{\mathbb{F}}$.

(B2) *Suppose that $g(x)$ is not of the form $h_1 h_2(x)^d$ for any $h_2(x) \in K(x)$ and any $h_1 \in K$. Then the L-function $L(s, \eta_{\chi,g_v})$ attached to η_{χ,g_v} is a polynomial in Nv^{-s} of degree n equal to the degree of the conductor of η_{χ,g_v} minus 2, and n is the same for almost all places v.*

(B3) *Under the same assumption as in (B2), there exists a compatible family of ℓ-adic representations of $\mathrm{Gal}(K^{sep}/K))$, depending on χ and g, such that the associated L-function $L(s, \chi, g)$ satisfies*

$$L(s, \chi, g) \sim \prod_{v \ \ good} L(s, \eta_{\chi,g_v})^{-1}.$$

Further $L(s, \chi, g)$ and its twists by idele class characters of K and by all ℓ-adic representations of $Gal(K^{sep}/K)$ behave nicely.

(2.3.1) Remarks (1) From the motivic point of view, the L-functions in (A3) and (B3) arise from the first cohomology group. Here we normalize them by taking the inverse so that formally they appear like automorphic L-functions. In other words, we forget the degree of the cohomology and use only the Galois representation as the Langlands parameter of the automorphic representation.

(2) The function $L(s, \psi, f)$ in (A3) is a factor of the Hasse-Weil L-function attached to the Artin-Schreier curve defined by

$$y^p - y = f(x),$$

while $L(s, \chi, g)$ in (B3) is a factor of the Hasse-Weil L-function attached to the curve defined by

$$y^d = g(x).$$

In particular, the Hasse-Weil L-function $L(s, E)$ attached to the elliptic curve E defined by $y^2 = g(x)$ at the beginning of the Introduction is equal to $L(s, \chi, g)$ with χ being the quadratic character of $\mathbb{F}^\times$.

(3) Given m rational functions $f_1(x), ..., f_m(x) \in K(x)$ as in (A2), and/or n rational functions $g_1(x), ..., g_n(x) \in K(x)$ as in (B2) and n nontrivial multiplicative characters $\chi_1, ..., \chi_n$ of $\mathbb{F}^\times$, quite often there exists a compatible family of ℓ-adic representations of $\mathrm{Gal}(K^{sep}/K)$, depending on $f_1(x), ..., f_m(x), \psi$, and/or $g_1(x), \chi_1, ..., g_n(x), \chi_n$, such that its associated L-function has its local factor at a good place v equal to the reciprocal of the L-function attached to the product of the idele class characters $\eta_{\psi, f_{1_v}}, ..., \eta_{\psi, f_{m_v}}$ and/or $\eta_{\chi_1, g_{1_v}}, ..., \eta_{\chi_n, g_{n_v}}$. Example 2.5.2 below falls in this case.

As explained in the Introduction, the two “prototype Theorems” 2.2 and 2.3 are consequences of Grothendieck's geometric interpretation of L-functions as envisioned by Weil, and Deligne's spectacular theorems in [7].

The next statement is a special case of the Langlands correspondence for $\mathrm{GL}(n)$ over global function fields, proved by Lafforgue.

(2.4) Theorem *The L-functions $L(s, \psi, f)$ and $L(s, \chi, g)$ in (A3) and (B3) respectively are automorphic L-functions for GL_n over the function field K. Moreover, these L-functions satisfy the Ramanujan-Petersson conjecture.*

(2.5) Exhibited below are specific examples of the automorphic L-functions of $GL(n)$ constructed above; some are also automorphic L-functions on quaternion groups and verify the Sato-Tate conjecture, cf. §4 of this part and §4 of Part II.

(2.5.1) Example Let $f(x) = x + a/x$ with any nonzero element a in K. We obtain an automorphic L-function $L(s, \psi, f)$ for GL_2 over K whose local factor at a good place v is

$$L(s, \eta_{\psi, f_v}) = 1 + Kl(\mathbb{F}_v; a) Nv^{-s} + Nv^{1-2s},$$

where $Kl(\mathbb{F}_v; a)$ is the Kloosterman sum

$$Kl(\mathbb{F}_v; a) = \sum_{x \in \mathbb{F}_v} \psi_p \circ \mathrm{Tr}_{\mathbb{F}_v/(\mathbb{Z}/p\mathbb{Z})}(x + a/x).$$

As we shall see in §4 that this form actually lives on a quaternion group, and it has close ties to eigenvalues of norm graphs.

(2.5.2) Example Suppose p is odd. Let $g_1(x) = (x-1)^2 + ax$ for $a \in K^\times$, χ_1 the quadratic character of $\mathbb{F}^\times$, $g_2(x) = x$ and χ_2 a nontrivial character of $\mathbb{F}^\times$. Then the product over good places v of K of the reciprocal of $L(s, \eta_{\chi_1, g_{1v}} \eta_{\chi_2, g_{2v}})$ yields an automorphic L-function for GL_2 over K; see Theorem 4.2 of Part II for the precise statement. The relation to Terras graphs is also explained in §4 of Part II.

(2.5.3) Example Let $f_1(x) = x^2 + \frac{a}{x}$ and $f_2(x) = x + \frac{a}{x^2}$ for a nonzero element a in K. If p is not 2, then $L(s, \psi, f_1)$ and $L(s, \psi, f_2)$ are automorphic L-functions for GL_3 over K.

(2.5.4) Example $L(s, \psi, f)$ is an automorphic L-function for GL_4 over K if $f(x) = x^3 + \frac{a}{x}$ and p is not equal to 3, or $f(x) = x^2 + \frac{a}{x^2}$ and p is not equal to 2. Here a is any nonzero element in K as before.

Since the Hasse-Weil zeta function of the smooth projective curve arising from the affine equation $y^p - y = f(x)$ (resp. $y^d = g(x)$) is a product of L-functions of type $L(s, \psi, f)$ (resp. $L(s, \chi, g)$), hence it is also an automorphic L-function. In other words, we have shown

(2.5.5) Corollary *The Hasse-Weil L-function attached to a smooth projective curve of genus r defined over K with affine equation $y^p - y = f(x)$ (resp. $y^d = g(x)$) is an automorphic L-function for $GL(2r)$ over K.*

The first two statements of Theorems A and B are proved in [23], Chap. 6, Theorem 4, and in [26]. They can also be seen geometrically. Since the third statement of each theorem is geometrical, using results of Grothendieck and Deligne, we shall give a unified geometric argument for all statements. The basic geometric settings below mostly follow Deligne and Katz in [5], [7], [17], [16].

(2.6) The basic method for the proof of Theorems A and B have been outlined in the Introduction. In the remainder of this section we will sketch the proof of Theorem A; the proof of Theorem B is similar.

(2.6.1) Described below is the geometric set-up for Theorem 2.2, following the notation of §1.

- The base scheme B is the smooth geometrically irreducible algebraic curve over $\mathbb{F}_q$ such that K is the field of rational functions on B.
- The source scheme X is $B \times_{\mathrm{Spec}\,\mathbb{F}_q} \mathbb{P}^1$, and $\pi : X \to B$ is the projection to the first factor,
- Let $U \subset X$ be the largest open subscheme of X such that the rational function $f(x) \in K(x)$ defines an $\mathbb{F}_q$-morphism from U to $\mathbb{G}_a$.
- Let $\mathcal{L}_\psi$ be the smooth rank-one $\overline{\mathbb{Q}_\ell}$-sheaf on $\mathbb{G}_{a/\mathrm{Spec}\,\mathbb{F}_q}$, the push-forward of the fiber $\mathbb{F}_q$ of the Lang torsor $\mathrm{Id} - \mathrm{Fr}_q : \mathbb{G}_a \to \mathbb{G}_a$ by $\psi : \mathbb{F}_q \to \overline{\mathbb{Q}_\ell}^\times$.
- The sheaf $\mathcal{L} = \mathcal{L}_{\psi, f}$ on X is $(U \hookrightarrow X)_!\,(f^*\mathcal{L}_\psi)$, the extension-by-zero of the smooth rank-one $\overline{\mathbb{Q}_\ell}$-sheaf $f^*\mathcal{L}_\psi$ on U.

(2.6.2) The hypothesis in (A2) ensures that $\mathrm{R}^i\pi_*\mathcal{L} = 0$ unless $i = 1$. The assertion about the degree in (A2) now follows from Grothendieck's Euler-Poincaré formula for constructible $\overline{\mathbb{Q}_\ell}$-sheaves over possibly open smooth curves. The rest of the statements of Theorem A all follow from the machinery of ℓ-adic cohomology. This concludes the sketch of the proof of Theorem A. The proof of Theorem B is similar, therefore omitted.

§3. The Kloosterman sum conjecture over function fields

(3.1) In this section we discuss a special occasion where Kloosterman sums arise as eigenvalues of the Hecke operators on certain automorphic forms over GL_2. To fix our notation, for each prime p we choose the additive character ψ_p of $\mathbb{Z}/p\mathbb{Z}$ to be $\psi_p(x) = e^{2\pi ix/p}$. The Kloosterman sum attached to a finite field k of characteristic p and a nonzero element b in k is defined as

$$Kl(k;b) = \sum_{x\in k^\times} \psi_p \circ \mathrm{Tr}_{k/(\mathbb{Z}/p\mathbb{Z})}(x + b/x).$$

Let K be a function field with the field of constants $\mathbb{F}$ of characteristic p as in §1 and let a be a nonzero element in K. Let f be the rational function $f(x) = x + a/x$ and choose the nontrivial additive character ψ of $\mathbb{F}$ to be $\psi_p \circ \mathrm{Tr}_{\mathbb{F}/(\mathbb{Z}/p\mathbb{Z})}$. At a place v which is neither a zero nor a pole of a, the function $f_v(x) \in \mathbb{F}_v(x)$ is regular everywhere except for a simple pole at 0 and a simple pole at ∞, hence the associated idele class character η_{ψ,f_v} of $\mathbb{F}_v(x)$ has conductor equal to $2\cdot 0 + 2\cdot\infty$ and the L-function of η_{ψ,f_v} has degree 2 in Nv^{-s} given by

$$L(s,\eta_{\psi,f_v}) = 1 + Kl(\mathbb{F}_v;a)Nv^{-s} + Nv^{1-2s}$$

(cf. [26]). Here, by passing to the residue field of K at v, we regard a as a nonzero element in $\mathbb{F}_v$. The coefficient of Nv^{-s} is obvious. We proceed to compute the coefficient of Nv^{-2s}, which we denote by b. It follows from the definition of $L(s,\eta_{\psi,f_v})$ that

$$b = \sum_{v_1,v_2} \psi(\pi_{v_1})\psi(\pi_{v_2}) + \sum_{w} \psi(\pi_w),$$

where v_1, v_2 run through places of $\mathbb{F}_v(x)$ of degree 1 other than 0 and ∞, and w runs through the places of degree 2. In view of the explicit description of η_{ψ,f_v} given in Theorem A, (A1), the first sum is over nonzero elements c, d in $\mathbb{F}_v$ of $\psi(c + a/c)\psi(d + a/d) = \psi((c + d)(1 + \frac{a}{cd}))$, while the second sum is over monic irreducible polynomials π_w of degree 2 over $\mathbb{F}_v$ of $\psi(z + a/z + z' + a/z') = \psi((z + z')(1 + \frac{a}{zz'}))$ with z, z' being the two roots of π_w. Viewed this way, the two parts of b have a uniform expression as sums over split and nonsplit degree two polynomials $x^2 - Ax + B$ over $\mathbb{F}_v$ with B nonzero, respectively. In other words, we may rewrite b as

$$b = \sum_{A\in\mathbb{F}_v} \sum_{B\in\mathbb{F}_v^\times} \psi(A(1 + \frac{a}{B})).$$

By separating the term with $A = 0$, which sums to $Nv - 1$, and then summing over B with a fixed nonzero A, which yields $-\psi(A)$, we arrive at

$$b = Nv - 1 - \sum_{A\in\mathbb{F}_v^\times} \psi(A) = Nv,$$

as asserted.

As a consequence of Theorem A and the converse theorem for GL_2, we know that $L(s, \psi, f)$ is an automorphic L-function for GL_2 over K. Alternatively, one can simply quote [34] 4.5, since the Kloosterman L-function $L(s, \psi, f)$ comes from a compatible system of two-dimensional ℓ-adic representations of $\mathrm{Gal}(K^{\mathrm{alg}}/K)$, as shown in §2. We summarize this result in

(3.2) Theorem (Kloosterman sum conjecture over a function field) *Let K be a function field with the field of constants $\mathbb{F}$ a finite field. Given a nonzero element $a \in K$, there exists an automorphic form f of GL_2 over K which is an eigenfunction of the Hecke operator T_v at all places v of K, which is neither a zero nor a pole of a, with eigenvalue $-Kl(\mathbb{F}_v; a)$. Here $\mathbb{F}_v$ denotes the residue field of the completion of K at v. In other words,*

$$L(s, f) \sim \prod_{v \quad good} \frac{1}{1 + Kl(\mathbb{F}_v; a)Nv^{-s} + Nv^{1-2s}}.$$

(3.2.1) Remark When $K = \mathbb{F}(t)$ is a rational function field, the conjecture for the case $p = 2$ and $a = t$ was proved by C. Moreno [30]. The case of $p = 2$ and 3 was proved in [25] with some help from J. Teitelbaum.

It is interesting to compare this with its counterpart over the field $\mathbb{Q}$, which is a question raised by Katz in [17].

(3.3) Question (Kloosterman sum question over $\mathbb{Q}$) Given a nonzero integer b, does there exist a Maass wave form f which is an eigenfunction of the Hecke operator T_p for all $p \nmid b$ with eigenvalue $-Kl(\mathbb{Z}/p\mathbb{Z}; b)$, that is, with

$$L(s, f) \sim \prod_{p \nmid b} \frac{1}{1 + Kl(\mathbb{Z}/p\mathbb{Z}; b)p^{-s} + p^{1-2s}} \ ?$$

(3.3.1) Remark Based on an idea of Sarnak, A. Booker [1] did some computations showing that if f existed, then either its level or its eigenvalue at infinity would be very large. He also explained the impossibility to give a counter-example by numerical computations because there always exists a modular form for $SL_2(\mathbb{Z})$ whose Fourier coefficients at the finitely many prescribed places are arbitrarily close to the prescribed numbers (satisfying the Ramanujan-Peterson conjecture) provided that either the weight (for holomorphic forms) or the eigenvalue at infinity (for Maass wave forms) is allowed to grow unboundedly.

Weil [35] showed that the Kloosterman sum $Kl(k; b)$ has absolute value bounded by $2\sqrt{|k|}$. On the other hand, each Kloosterman sum is in fact real. An interesting question is to study the distribution of a family of normalized sums $Kl(k; b)/\sqrt{|k|}$, which lie in the interval [-2, 2]. It follows from the work of Deligne [7] and our proof of the above theorem that the Sato-Tate conjecture holds for the automorphic form f in Theorem 3.2 associated to a non-constant a since f in this case does not arise from a quadratic extension of K.

(3.4) Corollary *If the element a in the statement of Theorem 3.2 is not a constant in K, then the normalized Kloosterman sums $Kl(\mathbb{F}_v; a)/\sqrt{|\mathbb{F}_v|}$ are uniformly distributed with respect to the Sato-Tate measure*

$$\mu_{ST}(x) = \frac{1}{\pi}\sqrt{1 - \frac{x^2}{4}}dx$$

as Nv tends to infinity.

PROOF. The equidistribution statement is a special case of the general equidistribution theorem in [7] 3.5.3 for $\overline{\mathbb{Q}_\ell}$-representations of the Weil group of K attached to a smooth Weil $\overline{\mathbb{Q}_\ell}$-sheaf pure of some weight on an open subset of C. See also [17] 3.6 for several forms of equidistribution statements and a self-contained exposition.

In our case the sheaf in question is the restriction to U of the sheaf $\mathrm{R}^1\,\mathrm{pr}_{1,U\,!}\mathcal{L}_{\psi,f}$, where

- U is the complement in B of the zeroes and poles of a,
- $\mathrm{pr}_{1,U} : U \times \mathbb{G}_\mathrm{m} \to U$ is the first projection,
- $f = x + \frac{a}{x}$, regarded as an $\mathbb{F}_q$-morphism from $U \times \mathbb{G}_\mathrm{m} \to \mathbb{A}_1$,
- $\mathcal{L}_{\psi,f}$ is the pull-back by f of the smooth rank-one $\overline{\mathbb{Q}_\ell}$-sheaf $\mathcal{L}_\psi$ on $\mathbb{G}_a$.

This sheaf is smooth of rank two on U and is pure of weight one, for it is well known that the product of the two eigenvalues of Fr_v is equal to q_v. We know that an equidistribution law is satisfied, and it is governed by a (possibly disconnected) semisimple group, the Zariski closure of the image of the geometric monodromy group of $\mathrm{R}^1\,\mathrm{pr}_{1,U\,!}\mathcal{L}_{\psi,f}$.

The cup product gives a perfect pairing

$$\mathrm{R}^1\,\mathrm{pr}_{1,U\,!}\mathcal{L}_{\psi,f} \times \mathrm{R}^1\,\mathrm{pr}_{1,U\,!}\mathcal{L}_{\psi,-f} \longrightarrow \mathrm{R}^2\,\mathrm{pr}_{1,U\,!}\overline{\mathbb{Q}_\ell} \xrightarrow{\sim} \overline{\mathbb{Q}_\ell}(-1)\,.$$

On the other hand, the involution $x \mapsto -x$ on $\mathbb{A}^1$ sends $\mathcal{L}_{\psi,f}$ to $\mathcal{L}_{\psi,-f}$, and induces a natural isomorphism

$$\mathrm{R}^1\,\mathrm{pr}_{1,U\,!}\mathcal{L}_{\psi,f} \xrightarrow{\sim} \mathrm{R}^1\,\mathrm{pr}_{1,U\,!}\mathcal{L}_{\psi,f}\,.$$

The duality pairing above then becomes an isomorphism $\det\bigl(\mathrm{R}^1\,\mathrm{pr}_{1,U\,!}\mathcal{L}_{\psi,f}\bigr) \xrightarrow{\sim} \overline{\mathbb{Q}_\ell}(-1)$.

Since $\det\bigl(\mathrm{R}^1\,\mathrm{pr}_{1,U\,!}\mathcal{L}_{\psi,f}\bigr)= \overline{\mathbb{Q}_\ell}(-1)$, the geometric monodromy group of $\mathrm{R}^1\,\mathrm{pr}_{1,U\,!}\mathcal{L}_{\psi,f}$ is a subgroup of $\mathrm{SL}(2)$, hence either it is equal to $\mathrm{SL}(2)$ or it is finite. We will be done if we rule out the second possibility, because the Sato-Tate measure given in the statement is exactly the push-forward of the Haar measure of the compact form of $\mathrm{SL}(2)$ to its conjugacy classes. The case that it is final is forbidden by the assumption that a is not constant. Otherwise it would mean that the sheaf $\mathrm{R}^1\,\mathrm{pr}_{1,U\,!}\mathcal{L}_{\psi,f}$ is geometrically trivial after a finite base change, or more concretely that the two eigenvalues of Fr_v are equal up to a root of unity of bounded order. That will violate the elementary fact that Kloosterman sums are the inverse Fourier transform of Gauss sums. One can also quote [6], which says among other things that the sheaf $\mathrm{R}^1\,\mathrm{pr}_{1,U\,!}\mathcal{L}_{\psi,f}$ is tamely ramified at 0, and is unipotent with exactly one Jordan block there. This of course rules out the possibility that the geometric monodromy group is finite. ∎

Let K be the rational function field $\mathbb{F}(t)$ and a be the element t. The same argument as in the proof of Corollary 3.4 in [25] combined with Corollary 3.4 gives

(3.5) Corollary (Katz [17]) *To each finite field k attach the measure*

$$\mu_k = \frac{1}{|k|-1}\sum_{b\in k^\times} \delta_{Kl(k;b)/\sqrt{|k|}},$$

where δ_x denotes the Dirac measure supported at x. Then as $|k|$ tends to infinity, the measures μ_k converge weakly to the Sato-Tate measure μ_{ST}.

(3.6) We examine the special case $K = \mathbb{F}(t)$ and $f(x) = x + \frac{t}{x}$. By Theorem 3.2, $L(s, \psi, f)$ is automorphic for GL_2 over $\mathbb{F}(t)$. Call π the associated automorphic representation. The only bad places for f are at 0 and ∞. As discussed above, the curve C defined by

$$y^p - y = f(x)$$

has semistable reduction at the place 0, hence the local component of π at 0 is a special representation (arising from inducing unramified characters) with the local L-factor equal to $1/(1-q^{-s})$, q being the cardinality of $\mathbb{F}$. At ∞, the Jacobian of C has potentially good reduction, so the local L-factor of the component of π at ∞ is 1. It follows from Remark 3.7.3 (i) that the exponent of the conductor of π at ∞ is 3, therefore local components of π are discrete series at exactly two places of $\mathbb{F}(t)$, namely, 0 and ∞, and the conductor of π, written additively, is $0 + 3\infty$. By a result of Gelbart and Jacquet [11], for any quaternion algebra H over $\mathbb{F}(t)$ ramified at 0 and ∞, there is an automorphic representation σ of $H^\times$ over $\mathbb{F}(t)$ corresponding to π locally and globally according to Langlands philosophy. In particular, the component of σ at 0 is the trivial representation, and σ and π have the same L-functions. We record this result in

(3.6.1) Theorem *Let $f(x) = x + \frac{t}{x}$ be as above and let H be a quaternion algebra over $\mathbb{F}(t)$ ramified at 0 and ∞. Then there is an automorphic representation σ of $H^\times$ over $\mathbb{F}(t)$ whose L-function is*

$$L(s, \sigma) = L(s, \psi, f) = \frac{1}{1 - q^{-s}} \prod_{v \neq 0, \infty} \frac{1}{1 + Kl(\mathbb{F}_v; t) Nv^{-s} + Nv^{1-2s}}.$$

Here q is the cardinality of $\mathbb{F}$.

(3.7) Appendix: Kloosterman curves and their Jacobians

(3.7.1) First we consider a Kloosterman curve over a field k of characteristic $p > 2$. Such a curve has an affine part given by the equation

$$y^p - y = x + \frac{a}{x}, \quad a \in k^\times$$

and we denote its projective smooth model by C_a. One quickly finds that the one-form $\frac{dx}{x}$ is a regular differential on C_a, with exactly two zeroes at $(x = 0, y = \infty)$ and $(x = \infty, y = \infty)$, both of order $p - 2$. So we conclude that C_a has genus $p - 1$.

Our curve C_a is a hyperelliptic curve; "the projection to y" makes C_a a double cover of $\mathbb{P}^1$, ramified at two points. The hyperelliptic involution is given by $(x, y) \mapsto (\frac{a}{x}, y)$. Besides the hyperelliptic involution, C_a has two automorphisms, defined by $\sigma : (x, y) \mapsto (x, y + 1)$, $\tau : (x, y) \mapsto (-x, -y)$ respectively. The two automorphisms σ and τ generate a dihedral group D_{2p} with $2p$ elements. Its integral group ring therefore operates on the Jacobian J_a of C_a. (Of course the hyperelliptic involution operates as $[-1]$ on J_a.) So we have a homomorphism $\mathbb{Q}[D_{2p}] \to \mathrm{End}^0(J_a) := \mathrm{End}(J_a) \otimes_{\mathbb{Z}} \mathbb{Q}$.

We study the semi-simple algebra $\mathbb{Q}[D_{2p}]$. From elementary representation theory for the dihedral group, one knows that it decomposes as a product

$$\mathbb{Q}[D_{2p}] \cong \mathbb{Q} \times \mathbb{Q} \times B$$

where B is a quaternion algebra over $F := \mathbb{Q}(\mu_p)^+$, the maximal totally real subfield of the cyclotomic field $\mathbb{Q}(\mu_p)$. Actually the quaternion algebra B spits: $B \cong \mathrm{M}_2(F)$. One only has to write down the following representation of D_{2p} over F:

$$\sigma \mapsto \begin{pmatrix} \cos(\frac{2\pi}{p}) & -\sin(\frac{2\pi}{p}) \\ \sin(\frac{2\pi}{p}) & \cos(\frac{2\pi}{p}) \end{pmatrix}, \qquad \tau \mapsto \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

One verifies without difficulty that the image of $\mathbb{Z}[D_{2p}]$ in $\mathrm{M}_2(F)$ generated by these two elements is maximal away from the primes above 2, but not maximal there.

From the structure of $\mathbb{Q}[D_{2p}]$ above we conclude that J_a is isogenous to a product $X_1 \times X_2$ of two abelian varieties over k, each of dimension $\frac{p-1}{2}$. For convenience we take X_1 to be the image of $1+\tau$, and X_2 the image of $1-\tau$. Both X_1 and X_2 have multiplication by $\mathcal{O}_F$, and they are isogenous to each other.

We can say a little bit about the slope condition of J_a, namely J_a is an ordinary abelian variety. To check this, first recall the general fact that the Cartier operator on $\mathrm{H}^0(C_a, \Omega^1)$ is dual to the Frobenius action on $\mathrm{H}_1(C_a, \mathcal{O}_{C_a})$. It is immediate from the definition that the inverse Cartier operator sends the regular one-form $\frac{dx}{x}$ to itself. Therefore the p-rank of J_a is strictly positive. Now the ring $\mathbb{Z}[\mu_p] \otimes_{\mathbb{Z}} \mathbb{Z}_p$ operates on the p-adic Tate module of the maximal étale quotient of the Barsotti-Tate group of J_a. This module is non-zero since the p-rank is positive, and note that $\mathbb{Q}(\mu_p)$ is totally ramified above p. This forces the p-adic Tate module to have rank $p-1$, meaning that J_a is ordinary.

Let us consider the family $\{C_t\}$ of Kloosterman curves over $\mathbb{G}_m$, where t is the coordinate function of $\mathbb{G}_m = \operatorname{Spec} k[t, t^{-1}]$. We would like to know about the behavior of this family as $t \to 0$ and $t \to \infty$. By elementary calculation, one sees that the family $\{J_t\}$ has multiplicative reduction. In fact the family of curves $\{C_t\}$ has semistable reduction at $t = 0$, the fiber at $t = 0$ being two smooth projective lines intersecting each other at p points with normal crossings. The calculation at $t = \infty$ is messy. All we want to say is that the family $\{J_t\}$ has bad reduction at $t = \infty$, as one can check easily. On the other hand it has potentially good reduction at $t = \infty$. One quick way to see this is by Raynaud's trick, see [32] Th. 5, p. 62. Assume otherwise that $\{J_t\}$ does not have potentially good reduction. We recall that $X_{1,t}$ and $X_{2,t}$ has multiplication by $\mathcal{O}_{\mathbb{Q}(\mu_p)^+}$. This forces $\{J_t\}$ to have totally multiplicative reduction above $t = \infty$ after a finite base change. Now we remember that C_t is ordinary for $t \in \mathbb{G}_m$, and it has semistable totally degenerate reduction above $t = 0$. Taking the p-torsion subgroup of the semiabelian reduction of the family $\{J_t\}$ over a finite cover of $\mathbb{P}^1$, we obtain a trivialization of the Hodge line bundle $\underline{\omega}$ of the semiabelian scheme. By the positivity of the Hodge line bundle, we deduce that the family $\{J_t\}$ is isotrivial, a contradiction. This proves the assertion. The same argument also shows that the good reduction fibers above $t = \infty$ of the potentially good reduction must have strictly positive slopes.

There is a mildly interesting question in this aspect, namely one would like to know the slopes of the reduction of $\{J_t\}$ above $t = \infty$ after realizing the potentially good reduction by making a finite base change. From the consideration above, one deduces that either there are two distinct positive slopes with equal multiplicity, or that $\frac{1}{2}$ is the only slope and the reduction is supersingular. The authors have not determined what the answer is, although it is quite conceivable that this is already known in the literature.

Of course, all of the above are compatible with the basic properties of Kloosterman sums given in [6], Th.7.8, p. 221. Each nontrivial irreducible character $\psi : \mathbb{Q}(\mu_p) \hookrightarrow \overline{\mathbb{Q}_\ell}$ cuts out a rank-two piece of the $\overline{\mathbb{Q}_\ell}$-representation given by the ℓ-power torsion points of J_t. This establishes the connection between our elementary consideration above and Deligne's more sophisticated point of view. There one sees that the monodromy is unipotent with one Jordan block at $t = 0$, while wildly ramified at $t = \infty$.

We summarize the above discussion in

(3.7.2) Proposition *Let k be a field of characteristic $p > 2$. For each $a \in k^\times$, let C_a be the projective smooth curve over k with affine model given by the equation $y^p - y = x + \frac{a}{x}$.*

1. *The curve C_a is a hyperelliptic curve of genus $p-1$.*
2. *There is a dihedral group D_{2p} acting on C_a.*
3. *The Jacobian J_a of C_a is isogenous to a product $X_1 \times X_2$ of two abelian varieties of dimension $\frac{p-1}{2}$. The two abelian varieties X_1 and X_2 are isogenous to each other. Each X_i has endomorphisms by the ring of integers of the maximal totally real subfield $\mathbb{Q}(\mu_p)^+$ in the cyclotomic field $\mathbb{Q}(\mu_p)$.*
4. *The Jacobian J_a is an ordinary abelian variety.*
5. *Suppose that the base scheme is $\mathbb{G}_m$, and consider the family $\{C_t\}$ of Kloosterman curves. This family has totally degenerate reduction at $t = 0$, the closed fiber being two copies of smooth projective lines intersecting transversally at p distinct points.*
6. *At $t = \infty$ the family $\{J_t\}$ has potentially good reduction. The fiber of the good reduction either has two distinct slopes, both strictly positive, or is supersingular.*

(3.7.3) Remarks (i) Suppose that the base scheme is $\mathbb{G}_m$ as above. At $t = 0$, both the Swan conductor and the Artin conductor of the first cohomology of $\{J_t\}$ is zero. This follows from (5) above and remains true after base change, or one can simply cite [6], p. 222, Thm. 7.8 (iii). When transformed into a representation of the Weil-Deligne group, the Artin conductor at $t = 0$ is 1 due to the nontrivial nilpotent action N. At $t = \infty$, the Swan conductor of $\{J_t\}$ is $p - 1$, while the Artin conductor is $3(p - 1)$. More precisely, each nontrivial additive character ϕ of $\mathbb{Z}/p\mathbb{Z}$ cuts out from the first cohomology of $\{J_t\}$ a rank-two smooth sheaf $\mathcal{F}_\phi$ over $\mathbb{G}_m$; and $\mathcal{F}_\phi$ is isomorphic to $\mathcal{F}_{\bar{\phi}}$. Each $\mathcal{F}_\phi$ has Swan conductor 1 and Artin conductor 3 at $t = \infty$; see [6], p. 222, Thm. 7.8 (iv).

(ii) When $p = 2$, the curve C_a is an elliptic curve. Statements (3) – (6) of Proposition 3.7.2 remain true, with the same proof, and Remark (i) above was proved in [30].

§4. Connections with Ramanujan graphs

A Ramanujan graph is a connected k-regular graph whose eigenvalues other than $\pm k$ (called nontrivial eigenvalues) are majorized by $2\sqrt{k-1}$. Such graphs are close to being random regular graphs, and hence are good communication networks. We refer to [24] and papers therein for more details concerning such graphs. Explicit constructions of infinite families of Ramanujan graphs with valency k equal to 1 plus a prime power are initiated by Margulis [29] and, independently, by Lubotzky, Phillips and Sarnak [28] based on quaternion groups over $\mathbb{Q}$, and later followed by Morgenstern [31] based on quaternion groups over function fields. In this construction the nontrivial eigenvalues of graphs are in fact eigenvalues of Hecke operators, which are known to satisfy the Ramanujan-Petersson conjecture, giving rise to the desired eigenvalue estimate. We review the essential ingredients relevant to our concern.

Denote by K the rational function field $\mathbb{F}(t)$ with $|\mathbb{F}| = q$, an odd integer. Fix a non-square $\delta \in \mathbb{F}$. Let H be the quaternion algebra over K with basis $1, i, j, ij$ satisfying $i^2 = \delta, j^2 = t - 1, ij = -ji$. Then H ramifies only at 1 and ∞, and it has class number one. Denote by D the algebraic group $H^\times$ mod its center and write $\mathcal{O}_v$ for the ring of integers of the completion of K at the place v. The vertices of a Morgenstern graph $X_\mathcal{K}$ are double cosets of adelic points of D depending on the choice of an open subgroup $\mathcal{K}$ of $\prod_{v \neq \infty} D(\mathcal{O}_v)$. More precisely,

$$X_\mathcal{K} = D(K) \setminus D(A_K)/D(K_\infty)\mathcal{K}.$$

To get the graph structure on $X_\mathcal{K}$, we find double coset representatives of $X_\mathcal{K}$ at the place 0 of K, namely,

$$X_\mathcal{K} \approx \Gamma(\mathcal{K}) \setminus D(K_0)/D(\mathcal{O}_0) = \Gamma(\mathcal{K}) \setminus PGL_2(K_0)/PGL_2(\mathcal{O}_0),$$

where $\Gamma(\mathcal{K}) = D(K) \cap D(K_\infty)\mathcal{K}$. Since $PGL_2(K_0)/PGL_2(\mathcal{O}_0)$ has a natural structure as $(q+1)$-regular infinite tree, this interpretation shows that $X_\mathcal{K}$ is a quotient graph of the tree, and hence is a $(q+1)$-regular finite graph. The first expression of $X_\mathcal{K}$ shows that the functions on the vertices of $X_\mathcal{K}$ are in fact automorphic functions of the quaternion group $D(A_K)$, and the second expression of $X_\mathcal{K}$ implies that the action of the adjacency matrix of the graph $X_\mathcal{K}$ is nothing but the Hecke operator T_0 at the place 0 acting on automorphic forms, and the nontrivial eigenvalues of $X_\mathcal{K}$ are the eigenvalues of T_0 at the nonconstant automorphic forms on $X_\mathcal{K}$. Drinfel'd [10] has shown that the nontrivial eigenvalues have absolute value majorized by $2\sqrt{q}$. Hence $X_\mathcal{K}$ is a Ramanujan graph.

On the other hand, for a finite group G and a symmetric subset S (that is, s^{-1} lies in S whenever s does) of generators of G, the Cayley graph $Cay(G, S)$ is a graph with vertices being the elements in G and the neighbors of a vertex x being xs, $s \in S$. Denote by $\mathbb{F}'$ a quadratic extension of $\mathbb{F}$, and for $a \in \mathbb{F}^\times$ let S_a consist of elements in $\mathbb{F}'$ whose norm to $\mathbb{F}$ is a. It was shown in [22] that for any $a \in \mathbb{F}^\times$ the Cayley graph $Cay(\mathbb{F}', S_a)$ is a Ramanujan graph, called a norm graph. Different choices of a result in isomorphic norm graphs. More precisely, the nontrivial eigenvalues of $Cay(\mathbb{F}', S_a)$ are the generalized Kloosterman sums

$$kl(\mathbb{F}; b) = \sum_{x \in S_b} \psi_p \circ \mathrm{Tr}_{\mathbb{F}'/(\mathbb{Z}/p\mathbb{Z})}(x)$$

for all $b \in \mathbb{F}^\times$, each occurring with multiplicity $q + 1$. Deligne in [6] showed that the generalized Kloosterman sum $kl(\mathbb{F}; b)$ is the opposite of the Kloosterman sum $Kl(\mathbb{F}; b)$, which Weil [35] has shown to have absolute value majorized by $2\sqrt{q}$.

These norm graphs are related to the Morgenstern graphs as follows.

(4.1) Proposition *(Proposition 4.1[27]) The Cayley graph $Cay(\mathbb{F}', S_{-1})$ is isomorphic to the Morgenstern graph $X_{\mathcal{K}_0}$ with $\mathcal{K}_0 = (1 + \mathcal{P}_1^2)\prod_{v \neq 1, \infty} D(\mathcal{O}_v)$. Here $\mathcal{P}_1$ denotes the maximal ideal of the integral elements $\mathcal{O}_1$ in the quaternion algebra $H(K_1)$.*

Consequently, for each $a \in \mathbb{F}^\times$, there exist $q+1$ linearly independent automorphic forms on $X_{\mathcal{K}_0} = D(K) \setminus D(A_K)/D(K_\infty)(1+\mathcal{P}_1^2)\prod_{v\neq 1,\infty} D(\mathcal{O}_v)$ which are eigenfunctions of the Hecke operator T_0 at 0 with eigenvalue $kl(\mathbb{F};a) = -Kl(\mathbb{F};a)$. We proceed to describe these forms.

The graph $X_{\mathcal{K}_0}$ has q^2 vertices. Instead of representing them using elements in $H(K_0)$ as in Morgenstern graph, we choose double coset representatives from $H(K_1)$. More precisely, by applying the strong approximation theory to the adelic double coset expression of $X_{\mathcal{K}_0}$, we see that automorphic functions on $X_{\mathcal{K}_0}$ can be identified with functions on $G = \mathcal{O}_1^\times/(1+\mathcal{P}_1^2)$ which are left invariant by the nonzero elements in the quadratic extension $\mathbb{F}(i)$ of $\mathbb{F}$ since j and $\mathbb{F}(i)^\times$, viewed as K-rational elements, are integral and invertible at all places other than 1 and j is a uniformizing element in $H(K_1)$. As $\mathbb{F}(i)^\times \setminus G$ may be represented by $1+bj$ with $b \in \mathbb{F}(i)$ viewed as elements in $H(K_1)$, these elements represent the vertices of $X_{\mathcal{K}_0}$. The space V of automorphic functions has a basis f_ϕ defined by

$$f_\phi(1+bj) = \phi(b)$$

as ϕ runs through all additive characters of $\mathbb{F}(i)$. As before, denote by ψ the nontrivial additive character $\psi_p \circ \mathrm{Tr}_{\mathbb{F}(i)/(\mathbb{Z}/p\mathbb{Z})}$, thus all characters of $\mathbb{F}(i)$ are given by ψ_x, $x \in \mathbb{F}(i)$, which maps b to $\psi(bx)$. The group G acts on V by right translations. More precisely, $a(1+cj) \in G$ sends f_{ψ_x} to $\psi(xc)f_{\psi_{xa'}}$, where $a' = \bar{a}/a$ with $\bar{a}$ being the image of a under the Frobenius automorphism of $\mathbb{F}(i)$ over $\mathbb{F}$. Clearly, when $x=0$, we obtain the one dimensional irreducible space V_0 generated by f_{ψ_0}. For a nonzero x, the orbit of f_{ψ_x} under G is a $(q+1)$-dimensional space spanned by $f_{\psi_{xu}}$ as u runs through all elements in S_1. Thus this space depends only on the norm a of x, which we denote by V_a. It is an irreducible G-space by Clifford's theorem. We have decomposed V into a direct sum of q irreducible subspaces V_a, parametrized by elements of $\mathbb{F}$, and no two of them are equivalent. The action of G on V_a extends to an irreducible representation of $D(K_1)$ on the same space via right translations ($j \in D(K_1)$ sends f_{ψ_x} to $f_{\psi_{\bar{x}}}$), which is the component at the place 1 of the global admissible irreducible automorphic representation σ_a of $D(A_K)$ such that V_a, viewed as a space of functions on $X_{\mathcal{K}_0}$, consists of the automorphic forms in the space of σ_a invariant by the largest congruence subgroup.

Next we look at the action of the Hecke operators T_v on V_a for $v \neq 1, \infty$. Here $a \neq 0$ so that the representation is nontrivial. Given such a place v, choose the uniformizer at K_v to be the irreducible polynomial π_v in $\mathbb{F}[t]$ such that when expressed in term of a polynomial in $t-1$ the constant term is 1. As the quaternion algebra H splits at v, the set of elements in $H(K_v)^\times$ with reduced norm in $\pi_v\mathcal{O}_v^\times$ is a disjoint union of $Nv+1$ right cosets $\gamma_{v,k}H(\mathcal{O}_v)^\times$. The Hecke operator T_v acts on V by sending a function f in V to the linear combination of right translations of f by $\gamma_{v,k}$. This action is independent of the choice of the coset representatives $\gamma_{v,j}$ since functions in V are right invariant by $H(\mathcal{O}_v)^\times$, and it preserves each space V_a. As the double coset space $X_{\mathcal{K}_0}$ is represented by $\mathbb{F}(i) \setminus G$, there are K-rational elements γ_k in $H(K)$ so that at the place v they form coset representatives defining the Hecke operator T_v and at other places $w \neq v, 1, \infty$ they fall in $H(\mathcal{O}_w)^\times$. Hence γ_k's are integral everywhere except possibly at 1 and ∞. Multiplying γ_k by a suitable power of j if necessary, we may assume that all $\gamma_k = f_k + g_k j$ with f_k, g_k polynomials in $t-1$ over $\mathbb{F}(i)$ and at least one of f_k, g_k has nonzero constant term. This implies that the reduced norm of γ_k, denoted by $\mathrm{Nm}_{\mathrm{rd}}\,\gamma_k$, is π_v times a nonzero element in $\mathbb{F}$. Since the norm map from $\mathbb{F}(i)$ to $\mathbb{F}$ is surjective, multiplying by a suitable element in $\mathbb{F}(i)$ if necessary, we may assume that

$$\mathrm{Nm}_{\mathrm{rd}}\,\gamma_k = f_k\bar{f}_k - (t-1)g_k\bar{g}_k = \pi_v.$$

Here $\bar{f} = \sum_m \bar{a}_m(t-1)^m$ if $f = \sum_m a_m(t-1)^m$ is a polynomial over $\mathbb{F}(i)$. The constant term of π_v is 1 by our choice, hence the constant term of f_k is nonzero and it has norm 1. We may further assume that the constant term of f_k is one for all k. To proceed, observe the following

(4.2) Lemma *Let f, g be two polynomials in $t-1$ over $\mathbb{F}(i)$ such that $g \neq 0$ and the element $\gamma = f + gj$ has reduced norm π_v. Let u be an element in S_1 and $\delta = f + ugj$. Then $\gamma H(\mathcal{O}_v)^\times = \delta H(\mathcal{O}_v)^\times$ if and only if $u = 1$.*

Proof. Sufficiency is obvious. To see the converse, suppose $\gamma H(\mathcal{O}_v)^\times = \delta H(\mathcal{O}_v)^\times$. Then $\gamma^{-1}\delta$ lies in $H(\mathcal{O}_v)^\times$. Since $\gamma^{-1} = \pi_v^{-1}(\bar{f} - gj)$, the above condition implies $u\bar{f}g \equiv g\bar{f} \pmod{\pi_v}$. Thus if we can show that f and g are both coprime to π_v over $\mathbb{F}(i)[t-1]$, then we conclude $u = 1$. To see this, start with

$$\mathrm{Nm}_{\mathrm{rd}}\,\gamma = f\bar{f} - g\bar{g}(t-1) = \pi_v.$$

If f and π_v are coprime, then so are $\bar{f}$ and π_v, and hence so are g and π_v. If $h = gcd(f, \pi_v)$ is not a constant, then h is irreducible over $\mathbb{F}(i)$, either $\pi_v = h$ or $\pi_v = h\bar{h}$. If $\pi_v = h$, then π_v also divides $\bar{f}$, which in turn implies that π_v divides both g and $\bar{g}$. Consequently, π_v^2 divides $\mathrm{Nm}_{\mathrm{rd}}\,\gamma$, a contradiction. If $\pi_v = h\bar{h}$, then $\bar{h}$ divides $\bar{f}$, and hence π_v divides $f\bar{f}$. This in turn implies that π_v also divides $g\bar{g}$. Consequently, $\mathrm{Nm}_{\mathrm{rd}}\,\gamma/\pi_v = f\bar{f}/\pi_v - (t-1)g\bar{g}/\pi_v = 1$, which is possible only if $f\bar{f} = \pi_v$ and $g = 0$, contradicting the assumption that $g \neq 0$. ∎

Lemma 4.2 shows that if $\gamma_k = f_k + g_k j$ with $g_k \neq 0$, then for all $u \in S_1$, the $q+1$ elements $f_k + ug_k j$, viewed as elements in $H(\mathcal{O}_v)$, have reduced norm π_v and represent different $H(\mathcal{O}_v)^\times$-cosets, hence they can be chosen in the definition of the Hecke operator T_v. We shall assume this is the case. For a global element $x \in H(K)$, denote by x_w its component at the place w.

We are now ready to study the action of T_v on $f \in V$. Let y be any element in G. We have

$$T_v f(y) = \sum_{1 \le k \le Nv+1} f(y(\gamma_k)_v) = \sum_{1 \le k \le Nv+1} f((\gamma_k^{-1})_1 y)$$

since f is right invariant by $\prod_{w \neq 1,v} D(\mathcal{O}_w)$ and left invariant by $D(K)$. Further, since f is right invariant by $1 + \mathcal{P}_1^2$, each term in the above sum depends only on γ_k modulo $t-1$, which has the form $1 + d_k j$ for some element $d_k \in \mathbb{F}(i)$ by our choice of γ_k. As f is invariant by the center of $H(A_K)^\times$, we may thus replace γ_k^{-1} by $1 - d_k j$ so that

$$T_v f(y) = \sum_{1 \le k \le Nv+1} f((1 - d_k j)_1 y).$$

Now take f to be f_{ψ_x} with $x \in \mathbb{F}(i)$. Given $y = a_1(1+bj)_1 \in G$ with $a, b \in \mathbb{F}(i)$ and $a \neq 0$, we have

$$\begin{aligned} T_v f_{\psi_x}(y) &= \sum_{1 \le k \le Nv+1} f_{\psi_x}((1 - d_k j)_1 a_1 (1+bj)_1) \\ &= \sum_{1 \le k \le Nv+1} f_{\psi_x}(a_1(1 + (b - d_k a')j)_1) = \sum_{1 \le k \le Nv+1} \psi_x(-d_k a') f_{\psi_x}(y), \end{aligned}$$

where $a' = \bar{a}/a$ is an element in S_1. Because of our choice of γ_k, d_k occurs if and only if all of its multiples by elements in S_1 occur, the summation $\sum_{1 \le k \le Nv+1} \psi_x(-d_k a')$ is independent of a and depends only on the norm of x. This implies that f_{ψ_x} is an eigenfunction of T_v for all $x \in \mathbb{F}(i)$, and those lie in the space V_a have the same eigenvalue with respect to T_v. Therefore V_a is a common eigenspace for all Hecke operators T_v with $v \neq 1$.

To see the eigenvalues explicitly, we have to find explicit representatives occurring in T_v as described above. Let v be a place of degree one other than 1 and ∞. Then $\pi_v = c(t-1)+1$ for some $c \in \mathbb{F}^\times$. The representatives occurring in T_v are $1 + uj$ with $u \in S_{-c}$, and the eigenvalue at T_v for $f_{\psi_x} \in V_a$ is

$$\sum_{u \in S_{-c}} \psi_x(-u) = \sum_{u \in S_{-ac}} \psi(u) = -Kl(\mathbb{F}; -ac) = -Kl(\mathbb{F}_v; a/t-1)$$

since $1/t - 1 \equiv -c \pmod{\pi_v}$.

Finally we show that no two eigenspaces V_a, $a \in \mathbb{F}$, have the same eigenvalues with respect to all T_v with degree v equal to 1 and v not the place at 1. This is obvious for V_0 since all eigenvalues are equal to $q+1$, while eigenvalues for V_a with $a \neq 0$ have absolute value at most $2\sqrt{q}$. Suppose a and b are two nonzero elements in $\mathbb{F}$ such that V_a and V_b have the same eigenvalues. Then we have $Kl(\mathbb{F}, ac) = Kl(\mathbb{F}, bc)$ for all $c \in \mathbb{F}^\times$. Let χ be a character of $\mathbb{F}^\times$ such that $\chi(a) \neq \chi(b)$. Consider

$$\begin{aligned}\sum_{c \in \mathbb{F}^\times} \chi(c) Kl(\mathbb{F}, ac) &= \sum_{x \in \mathbb{F}^\times} \psi(x) \sum_{c \in \mathbb{F}^\times} \chi(c)\psi(ac/x) = g(\chi, \psi) \sum_{x \in \mathbb{F}^\times} \psi(x)\chi(x/a) \\ &= g(\chi,\psi)^2 \chi(1/a),\end{aligned}$$

which is then equal to $\sum_{c \in \mathbb{F}^\times} \chi(c) Kl(\mathbb{F}, bc) = g(\chi,\psi)^2\chi(1/b)$ by assumption on Kloosterman sums. Here $g(\chi, \psi)$ denotes the Gauss sum which has absolute value $\sqrt{q}$. Since $\chi(a) \neq \chi(b)$, we have a contradiction.

We have shown that the eigenspaces V_a with $a \neq 0$ are distinguished by the eigenvalues $-Kl(\mathbb{F}_v, a/t-1)$ of T_v for places v of degree 1, $v \neq 1, \infty$. This combined with Theorem 3.2 and 3.6.1 yields the conclusion that for any $v \neq 1, \infty$, the eigenvalue of T_v on V_a for $a \neq 0$ is $-Kl(\mathbb{F}_v, a/t-1)$. We summarize the above discussion in

(4.3) Theorem *For each $a \in \mathbb{F}^\times$, let V_a be the $(q+1)$-dimensional subspace of automorphic forms on $D(K) \setminus D(A_K)/D(K_\infty)(1+\mathcal{P}_1^2)\prod_{v\neq 1,\infty} D(\mathcal{O}_v)$ spanned by the functions f_{ψ_x}, $x \in S_a$, as defined above. Then V_a is an eigenspace of all Hecke operators T_v, $v \neq 1, \infty$, with eigenvalue $-Kl(\mathbb{F}_v; a/(t-1))$. Moreover, the L-function attached to the automorphic representation associated to the space V_a is*

$$L_a(s) = \prod_{v \neq 1, \infty} \frac{1}{1 + Kl(\mathbb{F}_v; a/(t-1))Nv^{-s} + Nv^{1-2s}} \cdot \frac{1}{1 - N\infty^{-s}}.$$

(4.3.1) Remark Implicitly contained in the theorem above are many identities on character sums. For example, take a place v of degree 2 with $\pi_v = c_2(t-1)^2 + c_1(t-1) + 1$, we may choose the representatives in T_v to consist of $1 + \alpha(t-1) + \beta j$, where $\alpha, \beta \in \mathbb{F}(i)$ satisfy the relations $\alpha\bar{\alpha} = c_2$ and $\alpha + \bar{\alpha} - \beta\bar{\beta} = c_1$. Of these, two of α's, namely, the roots of $c_2 - c_1 x + x^2$, have trace equal to c_1 so that the corresponding $\beta = 0$, and the remaining α's yield nonzero β. Note that replacing β in a representative by $u\beta$ for any $u \in S_1$ yields another representative. In view of the action of the Hecke operators as analyzed above, we see that the eigenvalue of T_v on V_a is $2 + \sum_b \sum_{x \in S_{ab}} \psi(x)$, where b runs through all elements in $\mathbb{F}^\times$ so that $b = \mathrm{Tr}_{\mathbb{F}(i)/\mathbb{F}}(\alpha) - c_1$ for some $\alpha \in S_{c_2}$. There are $(q-1)/2$ such b. This character sum is equal to $-Kl(\mathbb{F}_v, a/(t-1))$ by Theorem 4.3. In other words, $kl(\mathbb{F}_v, a/(t-1)) = 2 + \sum_b kl(\mathbb{F}, ab)$ with b as above.

References

[1] A. Booker. A test for identifying Fourier coefficients of automorphic forms and application to Kloosterman sums. *Jour. of Experimental Math.*, 9:571–581, 2000.

[2] J. W. Cogdell and I. I. Piatetski-Shapiro. Converse theorems for GL_n. *Publ. Math. IHES*, 79:157–214, 1994.

[3] J. W. Cogdell and I. I. Piatetski-Shapiro. A converse theorem for GL_4. *Math. Research Letters*, 3:67–76, 1996.

[4] J. W. Cogdell and I. I. Piatetski-Shapiro. Converse theorems for GL_n. II. *Jour. reine angwante Math.*, 507:165–188, 1999.

[5] P. Deligne. Les constantes des équations fonctionelles des fonctions L. In *Modular Functions of One Variable.* II, volume 349 of *Lecture Notes in Math.*, pages 501–595. Springer-Verlag, 1973.

[6] P. Deligne. *Cohomologie Étale (SGA $4\frac{1}{2}$)*, volume 569 of *Lecture Notes in Math.* Springer-Verlag, 1977.

[7] P. Deligne. Le conjecture de Weil, II. *Publ. Math. IHES*, 52:138–252, 1980.

[8] P. Deligne and N. Katz. *Groupes de Monodromie en Géométrie Algébrique (SGA 7) II*, volume 340 of *Lecture Notes in Math.* Springer-Verlag, 1973.

[9] V. G. Drinfel'd. Elliptic modules II. *Math. USSR Sb.*, 31:159–170, 1977.

[10] V. G. Drinfel'd. The proof of Petersson's conjecture for $GL(2)$ over a global field of characteristic p. *Functional Anal. Appl.*, 22:28–43, 1988.

[11] S. Gelbart and H. Jacquet. Forms of GL(2) from the analytic point of view. In *Automorphic Forms, Representations, and L-functions, Part 1*, volume 33 of *Proc. Symp. Pure Math.*, pages 213–251. Amer. Math. Soc., 1979.

[12] E. Hecke. Über eine neue Art von Zetafunktionen und ihre Beziehungen zur Verteilung der Primzahlen I. *Math. Z.*, 1:357–376, 1918.

[13] E. Hecke. Über eine neue Art von Zetafunktionen und ihre Beziehungen zur Verteilung der Primzahlen II. *Math. Z.*, 4:11–21, 1920.

[14] H. Jacquet and R. P. Langlands. *Automorphic Forms on $GL(2)$*, volume 114 of *Lecture Notes in Math.* Springer-Verlag, 1970.

[15] H. Jacquet, I. I Piatetski-Shapiro, and J. Shalika. Automorphic forms on GL_3, I & II. *Ann. of Math.*, 109:169–258, 1979.

[16] N. Katz. *Sommes Exponentiells, rédigé par G. Laumon*, volume 79 of *Astérisque.* 1980.

[17] N. Katz. *Gauss Sums, Kloosterman Sums, and Monodromy Groups*, volume 116 of *Annals of Math. Studies.* Princeton Univ. Press, 1988.

[18] L. Lafforgue. La correspondance de langlands sur les corps de fonctions. 1999. manuscript.

[19] G. Lauman. La correspondance de langlands sur les corps de fonctions. *Séminaire Bourbaki*, (873), 1999–2000.

[20] G. Laumon. Semi-continuité du conducteur de Swan. *Astérisque*, 82–83:173–219, 1981.

[21] G. Laumon. Transformation de Fourier, constantes d'équations fonctionelles et conjecture de weil. *Publ. Math. I.H.E.S*, 65:173–219, 1987.

[22] W.-C. W. Li. Character sums and abelian Ramanujan graphs. *J. Number Theory*, 41:199–217, 1992.

[23] W.-C. W. Li. *Number Theory with Applications*. World Scientific, 1996.

[24] W.-C. W. Li. A survey of Ramanujan graphs. In *Arithmetic, Geometry and Coding Theory, Proceedings of International Conference held at Luminy, France, June 28 - July 2, 1993*, pages 127–143. de Gruyter, 1996.

[25] W.-C. W. Li. Elliptic curves, Kloosterman sums and Ramanujan graphs. In D. A. Buell and J. T. Teitelbaum, editors, *Computational Perspectives on Number Theory: Proceedings of a Conference in honor of A. O. L. Atkin*, volume 7 of *AMS/IP Studies in Advanced Math.* Amer. Math. Soc., 1998.

[26] W.-C. W. Li. Character sums over p-adic fields. *J. Number Theory*, 74:181–229, 1999.

[27] W.-C. W. Li. Eigenvalues of Ramanujan graphs. In *Emerging Applications of Number Theory* (Minneapolis, MN, 1996), pages 387-403, IMA Vol. Math. Appl. 109, Springer, New York, 1999.

[28] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8:261–277, 1988.

[29] G. Margulis. Explicit group theoretic constructions of combinatorial schemes and their application to the design of expanders and concentrators. *J. Prob. of Info. Trans.*, pages 39–46, 1988.

[30] C. Moreno. The Kloosterman conjecture in characteristic two. preprint.

[31] M. Morgenstern. Existence and explicit constructions of $q + 1$ regular Ramanujan graphs for every prime power q. *J. Comb. Theory, series B*, 62:44–62, 1994.

[32] L. Szpiro. Propriétés numériques du faisceau dualisant relatif. In *Pinceaux de Courbes de Genre au Moins Deux*, volume 86 of *Astérisque*, pages 44–78, 1981.

[33] J. Tate. Fourier analysis in number fields and Hecke's zeta-functions. In J. W. S. Cassel and A. Fröhlich, editors, *Algebraic Number Theory*, pages 305–347. Academic Press, 1967.

[34] J. Tate. Number theoretic background. In *Automorphic Forms, Representations, and L-functions, Part 2,*, volume 33 of *Proc. Symp. Pure Math.*, pages 3–26. Amer. Math. Soc., 1979.

[35] A. Weil. *Courbes algébriques et variétés abéliennes.* Hermann, 1948, 1971.

[36] H. Yoshida. On an analogue of the sato-tate conjecture. *Invent. Math.*, 19:261–277, 1973.

Ching-Li Chai
Department of Mathematics
University of Pennsylvania
Philadelphia, PA 19104
U.S.A.
Email: `chai@math.upenn.edu`

Wen-Ching Winnie Li
Department of Mathematics
Pennsylvania State University
University Park, PA 16802
U.S.A.
Email: `wli@math.psu.edu`