

THE PURITY LOCUS OF MATRIX KLOOSTERMAN SUMS

MÁRTON ERDÉLYI, WILL SAWIN, AND ÁRPÁD TÓTH

ABSTRACT. We construct a perverse sheaf related to the the matrix exponential sums investigated in [9]. As this sheaf appears as a summand of certain tensor product of Kloosterman sheaves, we can establish the exact structure of the cohomology attached to the sums by relating it to the Springer correspondence and using the recursion formula of [9].

1. INTRODUCTION

Fix a finite field $\mathbb{F}_q$ of characteristic p and algebraic closure $\overline{\mathbb{F}}_q$, a complex valued additive character ψ of $\mathbb{F}_q$, and a natural number n . Let $\mathbb{F}_{q^m} \subset \overline{\mathbb{F}}_q$ be the degree m extension of $\mathbb{F}_q$ and $\psi_m = \psi \circ \text{Tr}_{\mathbb{F}_{q^m}|\mathbb{F}_q}$.

For a matrix $a \in M_n(\mathbb{F}_q) = \mathbb{F}_q^{n \times n}$ the exponential sums

$$K(a, \mathbb{F}_{q^m}) = K(a) = \sum_{x \in \text{GL}_n(\mathbb{F}_{q^m})} \psi_m(\text{tr}(ax + x^{-1}))$$

(here tr is the matrix trace) and the related compactly-supported cohomology complex (in the setting ℓ -adic étale cohomology for some $\ell \nmid q$)

$$H^*(a) = H_c^*(\text{GL}_{n, \overline{\mathbb{F}}_q}, \mathcal{L}_\psi(\text{tr}(ax^{-1} + x)))$$

are investigated in [9]. The relation between the exponential sum and the complex can be made explicit by the Grothendieck trace formula ([10]):

$$K(a, \mathbb{F}_{q^m}) = \sum_{i=0}^{2d} (-1)^i \sum_{j=1}^{d_i} (\lambda_j^i)^m,$$

where $d = \dim(\text{GL}_n) = n^2$, $d_i = \dim(H^i(a))$ and the λ_j^i -s are the eigenvalues of the Frobenius Frob_q on $H^i(a)$. Each λ_j^i has some integral weight $0 \leq w \leq i$, meaning that $|\iota(\lambda_j^i)| = q^{w/2}$, for all $\iota : \overline{\mathbb{Q}}_\ell \rightarrow \mathbb{C}$ field isomorphisms.

The original interest in these sums arose from work on making the equidistribution result on $\text{SL}_n(\mathbb{R})/\text{SL}_n(\mathbb{Z})$ in [7] effective. It was already demonstrated in dimension $n = 3$, where estimates for the classical Kloosterman sum $K(\alpha) = \sum_{x \in \mathbb{F}_q^*} \psi(\alpha x + x^{-1})$ suffice, that non-trivial power saving for the sums $K(a, \mathbb{F}_{q^m})$ yield such effective results. Such estimates were given in [9] and used successfully in [8] for exactly this purpose. However the paper [9] without clarifying the nature

2020 *Mathematics Subject Classification.* 11L05, 11L07, 14F10.

Key words and phrases. Kloosterman sums, perverse sheaves, Springer correspondence.

Erdélyi was supported by NKFIH Research Grants FK-127906 and K-135885 and by the Rényi Institute Lendület Analytic Number Theory and Representation Theory Research Group.

Sawin was supported by NSF grant DMS-2101491.

Tóth was supported by by the Rényi Institute Lendület Automorphic Research Group, and by NKFIH Research Grants K-135885.

of these connections already hinted that these sums exhibit some very interesting geometric phenomena.

An element $a \in M_n(\mathbb{F}_q)$ is regular if for all eigenvalues in $\overline{\mathbb{F}}_q$ the eigenspace is one-dimensional, a is semisimple if the algebra $\mathbb{F}_q[a]$ is semisimple, or equivalently a is diagonalizable in $\overline{\mathbb{F}}_q$. The regular semisimple locus is the subscheme $U \subset \mathbb{A}^{n^2}$ of regular semisimple elements.

An exponential sum or the corresponding cohomology complex is called pure of weight w if all λ_j^i are of weight $i + w$.

Although effective bounds were given for the sum or the weights of the Frob_q eigenvalues (not depending on the degree i), the exact structure of the cohomology was not determined (cf. [9], Remark 2.):

It was proven that $H^*(a)$ is pure of weight 0 and concentrated in the middle degree $i = n^2$, if a is invertible and regular semisimple in $\overline{\mathbb{F}}_q$.

In the invertible regular case it was pointed out that the sum admits square root cancellation, but the methods used there did not allow to decide whether in $H^*(a)$ is pure of weight 0 and concentrated in the middle degree or not (or equivalently the Frob_q eigenvalues in different degrees cancel or not).

The aim of this paper is to show that the cohomology groups are always pure, though of weight varying depending on a :

Theorem 1. *Let $a \in M_n(\mathbb{F}_q)$ be a matrix. Let k be the multiplicity of 0 as an eigenvalue of a . Then $H^*(a)$ is a pure complex of weight $-k$.*

Straightforward consequences of the purity are the following:

- (1) There can not be cancellation between the Frob_q eigenvalues of different degrees.
- (2) Knowing the value of $K(a, \mathbb{F}_{q^m})$ for all finite extension $\mathbb{F}_{q^m}|\mathbb{F}_q$ is equivalent to knowing the Frob_q eigenvalues of $H^*(a)$.

It can also be explicitly characterized when $H^*(a)$ is concentrated in the middle degree (and thus $K(a)$ admits square root cancellation):

Theorem 2. *The cohomology complex $H^*(a)$ is concentrated in the middle degree if and only if for each nonzero eigenvalue α of a , the eigenspace is 1-dimensional.*

In the case of regular matrices one can just refer to the computations in Section 22 of [9], but our methods put this in a more geometric context.

For the proof of the general case an explicit description of $K(a)$ and $H^*(a)$ is given for matrices a which have a unique eigenvalue $\alpha \in \mathbb{F}_q$ (compare with the recursion algorithm in Section 18 of [9]):

Theorem 3. *If a has a unique eigenvalue α and the classical Kloosterman sum is $K(\alpha) = -(\lambda + \bar{\lambda})$ with the Frob_q eigenvalues λ and $\bar{\lambda}$, then*

$$K(a) = (-1)^n q^{n(n-1)/2} \sum_{k=0}^n \# \{ W \leq \mathbb{F}_q^n \mid \dim(W) = k, aW = W \} \lambda^k \bar{\lambda}^{n-k}.$$

In order to prove these theorems, first a complex of sheaves is constructed on $\mathbb{A}^{n^2}$, whose fiber over the matrix a is the cohomology group $H^*(a)$.

This can be constructed as

$$K = Rpr_{1!} \mathcal{L}_\psi(\text{tr}(ax^{-1} + x))$$

where $pr_1 : \mathbb{A}^{n^2} \times GL_n \rightarrow \mathbb{A}^{n^2}$ is the projection.

Theorem 4. *The shifted complex $K[2n^2]$ is an irreducible pure perverse sheaf.*

Let $U \subset \mathbb{A}^{n^2}$ be the regular semisimple locus, and let $j : U \rightarrow \mathbb{A}^{n^2}$ be the open immersion. Let $\tilde{\mathbb{A}}^{n^2}$ be the space parameterizing pairs of a matrix $a \in \mathbb{A}^{n^2}$ and a complete flag F of linear subspaces F^i of $\mathbb{A}^n$ such that a preserves each of the subspaces F^i . Let $\pi : \tilde{\mathbb{A}}^{n^2} \rightarrow \mathbb{A}^{n^2}$ be the map forgetting the flag F . The map π is known as Grothendieck's simultaneous resolution because, for each degree n monic polynomial f , the inverse image under π of the variety of matrices with characteristic polynomial f is smooth, so it resolves the singularities of the variety of matrices with characteristic polynomial f . Indeed, it is not hard to check that the inverse image is a disjoint union over orderings of the roots of f of the space of pairs of a complete flag F and a matrix a such that a preserves each of the subspaces F^i and acts on F^i/F^{i-1} by scalar multiplication by the i th root of f , which is smooth. Grothendieck's simultaneous resolution may be defined for the Lie algebra of an arbitrary reducible group, but here we use only the case of $\mathfrak{gl}_n$.

Let $\tilde{U}$ be the inverse image of U under π , let $u : \tilde{U} \rightarrow \tilde{\mathbb{A}}^{n^2}$ be the open immersion, and let $\rho : \tilde{U} \rightarrow U$ be the projection. Note that $\tilde{U}$ parameterizes pairs (a, F) where a preserves F and a is regular semisimple. In this case, a has n one-dimensional eigenspaces for its n distinct eigenvalues, and F^i is invariant if and only if it is a sum of these eigenspaces, so choices of a complete flag are in one-to-one bijections with orderings of the eigenvalues.

Finally, let $\lambda_i : \tilde{\mathbb{A}}^{n^2} \rightarrow \mathbb{A}^1$ send a pair (a, F) to the unique eigenvalue of a acting on the one-dimensional space F^i/F^{i-1} . Thus, $\lambda_1(a, F), \dots, \lambda_n(a, F)$ are the eigenvalues of a with multiplicity.

By Grothendieck's simultaneous resolution the argument of [9] for regular semisimple elements can be translated to the fact that K is a summand of a complex of sheaves related to a tensor product of Kloosterman sheaves:

Theorem 5. *K is a summand of*

$$R\pi_* \bigotimes_{i=1}^n \lambda_i^* \mathcal{K}\ell_2[-n^2](n(n-1)/2),$$

where $\mathcal{K}\ell_2$ is the classical Kloosterman sheaf on $\mathbb{A}^1$ defined by Deligne ([3]).

Then Theorem 1 is a direct consequence of Springer's result on the cohomology of the Springer fiber [16]. Indeed, Theorem 1 concerns the fiber of the complex K at a point a , which by Theorem 5 is a summand of the fiber of the sheaf above at a . By proper base change, the fiber of the derived pushforward of a sheaf at a may be computed as the cohomology of $\pi^{-1}(a)$ with coefficients in the original sheaf. The variety $\pi^{-1}(a)$ is (the $\mathfrak{gl}_n$ special case of) a Springer fiber, and the purity of the cohomology of the constant sheaf on $\pi^{-1}(a)$ was established by Springer in [16]. The proof of Theorem 1 proceeds by checking that the sheaf $\bigotimes_{i=1}^n \lambda_i^* \mathcal{K}\ell_2$ is (geometrically) constant on each component.

When a is regular, then $\pi^{-1}(a)$ consists of finitely many points, which allows us to prove Theorem 2 in this case.

One can motivate Theorem 3 by relating the geometry of K to the Springer correspondence:

If the classical Kloosterman sum $K(\alpha)$ has Frob_q eigenvalues λ and $\bar{\lambda}$, then an n -fold tensor product of Kloosterman sheaves splits into one-dimensional summands with eigenvalues λ^k and $\bar{\lambda}^{n-k}$. The action of the Weyl group S_n permutes the terms and $S_k \times S_{n-k} \leq S_n$ shows up as the stabilizer of an eigenvector. This suggests that the invariants of $S_k \times S_{n-k}$ on the cohomology are given by the cohomology of the space of k -dimensional subspaces fixed by a , and that leads to the formula in Theorem 3.

This is proven using elementary arguments and a recursion formula for the sum $K(a)$, ([9], Theorem 18.1.), which also justifies the above geometric description, as it is the only one that matches the sum in Theorem 3.

The computation of some weights allows to conclude Theorem 2 for non-regular matrices.

One may be naturally inclined to speculate how the results generalize to reductive subgroups of GL_n . We will only comment on the question of purity, for which understanding the generic situation as in Theorem 5 is crucial. This result seems to be closely tied up with the underlying standard representation of the group GL_n . Other representations will definitely lead to more elaborate structures. To illustrate consider the subgroup $H = \text{Sym}^2(GL_2)$ of GL_3 and $K_H(a) = \sum_{x \in H} \psi(ax + x^{-1})$.

Let ρ be the underlying character $\text{tr } \text{Sym}^2 x$ on GL_2 and assume that $A = \text{Sym}^2(a)$ so that if summation in $K(A)$ is restricted to H it gives the exponential sum

$$K_{\text{Sym}^2(GL_2)}(a, \mathbb{F}_{q^m}) = \sum_{x \in GL_2(\mathbb{F}_{q^m})} \psi_m(\rho(ax) + \rho(x^{-1})).$$

Then for a regular semisimple with eigenvalues λ_1, λ_2 the Borel subgroup consisting upper diagonal matrices contributes (the small cell in the Bruhat decomposition) the exponential sum

$$\sum_{(t_1, t_2)} \psi(\lambda_1^2 t_1^2 + \lambda_1 \lambda_2 t_1 t_2 + \lambda_2^2 t_2^2 + t_1^{-2} + (t_1 t_2)^{-1} + t_2^{-2}),$$

on the torus consisting diagonal matrices, which is pure of weight 0 if $p \neq 3$ (for example by [1, 6]). However the rest (the open cell) also contributes

$$(q-1)G^2 K(\det a),$$

where G is the standard quadratic Gauss sum $\sum \psi(x^2)$ on $\mathbb{A}^1$ and so comparing the weights shows that the associated complex is not pure.

A brief description of the paper follows. Our argument naturally splits into two parts:

In Section 2 Theorems 4, 5, 1 and the regular case in 2 are proven by geometric arguments.

In Section 3 Theorem 3 is proven by combinatorial calculations and then the remaining part of Theorem 2.

2. GEOMETRIC ARGUMENTS

Proof of Theorem 4. Consider first the space $\mathbb{A}^{n^2} \times \mathbb{A}^{n^2}$ parameterizing pairs of $n \times n$ matrices (x, y) . Let $i: GL_n \rightarrow \mathbb{A}^{n^2} \times \mathbb{A}^{n^2}$, $x \mapsto (x, x^{-1})$. The map i is a closed immersion as its image is the closed set consisting of pairs of matrices whose product is the identity. Then since GL_n is smooth and irreducible of dimension n^2 ,

$\mathbb{Q}_\ell[n^2]$ is an irreducible perverse sheaf on GL_n . Then since i is a closed immersion, $i_*\mathbb{Q}_\ell[n^2]$ is an irreducible perverse sheaf on $\mathbb{A}^{n^2} \times \mathbb{A}^{n^2}$.

The Fourier transform is known to preserve irreducible perverse sheaves (perversity from [12, Corollary 2.1.5(ii)] and irreducibility from the fact that Fourier transform is an equivalence of categories [13, III, Theorem 8.1(3)]). Thus, the Fourier transform of $i_*\mathbb{Q}_\ell[n^2]$ is an irreducible perverse sheaf. Next, let us calculate this Fourier transform.

By definition [12, Definition 2.1.1], the Fourier transform of a complex L on $\mathbb{A}^N$ is given by $Rpr_{l!}(pr_r^*L \otimes \mathcal{L}_\psi(\mu))[N]$ where pr_l, pr_r are the two projections $\mathbb{A}^N \times \mathbb{A}^N \rightarrow \mathbb{A}^N$ and μ is a nondegenerate bilinear form on $\mathbb{A}^N \times \mathbb{A}^N$. We can specialize this formula for $N = 2n^2$, viewing each copy of $\mathbb{A}^N$ as $\mathbb{A}^{n^2} \times \mathbb{A}^{n^2}$. A suitable bilinear map takes two pairs of matrices $((a, b), (x, y))$ to $\text{tr}(ay + bx)$. Having done this, the Fourier transform of $i_*\mathbb{Q}_\ell[n^2]$ is

$$Rpr_{12!}(pr_{34}^*i_*\mathbb{Q}_\ell[n^2] \otimes \mathcal{L}_\psi(\text{tr}(ay + bx)))[2n^2].$$

Here pr_{12} and pr_{34} are the maps from $(\mathbb{A}^{n^2})^4$ to $(\mathbb{A}^{n^2})^2$ given by projection onto the first two and last two factors respectively.

Since i is a closed immersion, $i_* = i_!$. By proper base change [2, XVII, Proposition 5.2.8], $pr_{34}^*i_!\mathbb{Q}_\ell = ((id) \times i)_!\mathbb{Q}_\ell$ where $(id) \times i: \mathbb{A}^{n^2} \times \mathbb{A}^{n^2} \times GL_n \rightarrow (\mathbb{A}^{n^2})^4$ sends (a, b, x) to (a, b, x, x^{-1}) . By the projection formula [2, XVII, Proposition 5.2.9],

$$\begin{aligned} ((id) \times i)_!\mathbb{Q}_\ell \otimes \mathcal{L}_\psi(\text{tr}(ay + bx)) &= ((id) \times i)_!(id \times i)^*\mathcal{L}_\psi(\text{tr}(ay + bx)) = \\ &= ((id) \times i)_!\mathcal{L}_\psi(\text{tr}(ax^{-1} + bx)). \end{aligned}$$

Thus, the Fourier transform of $i_*\mathbb{Q}_\ell[n^2]$ is

$$Rpr_{12!}((id) \times i)_!\mathcal{L}_\psi(\text{tr}(ax^{-1} + bx))[3n^2]$$

which by functoriality of compactly supported pushforward [2, XVII, Theorem 5.1.8(a)] is

$$R(pr_{12} \circ (id) \times i)_!\mathcal{L}_\psi(\text{tr}(ax^{-1} + bx))[3n^2].$$

The composition $R(pr_{12} \circ (id) \times i)_!$ may also be called pr_{12} as it represents projection onto the first two factors a, b . So $Rpr_{12!}\mathcal{L}_\psi(\text{tr}(ax^{-1} + bx))[3n^2]$ is an irreducible perverse sheaf.

This irreducible sheaf remains irreducible when restricted to the open locus $\mathbb{A}^{n^2} \times GL_n$ where b is invertible. (By the structure theorem for irreducible perverse sheaves, the only other possibility is that it becomes zero on this subset, which is easy to rule out by the calculations already in [9], see for example Theorems 1.1 and 1.7.) On that locus, performing a change of variables substituting $b^{-1}x$ for x , we get

$$Rpr_{12!}\mathcal{L}_\psi(\text{tr}(ax^{-1}b + x))[3n^2] = Rpr_{12!}\mathcal{L}_\psi(\text{tr}(bax^{-1} + x))[3n^2].$$

By proper base change [2, XVII, Proposition 5.2.8] on the Cartesian square

$$\begin{array}{ccc} (\mathbb{A}^{n^2} \times GL_n) \times GL_n & \xrightarrow{m \times id} & \mathbb{A}^{n^2} \times GL_n \\ pr_{12} \downarrow & & \downarrow pr_1 \\ \mathbb{A}^{n^2} \times GL_n & \xrightarrow{m} & \mathbb{A}^{n^2}, \end{array}$$

where $m: (a, b) \mapsto ba$ from $\mathbb{A}^{n^2} \times GL_n$ to $\mathbb{A}^{n^2}$ is the multiplication map, the above sheaf is $m^*K[3n^2]$. Because m is smooth of relative dimension n^2 with nonempty, geometrically connected fibers, the functor $L \mapsto m^*L[n^2]$ preserves the perverse t -structure and moreover is a fully faithful functor on perverse sheaves [4, Proposition 4.2.5].

It follows that $K[2n^2]$ is an irreducible perverse sheaf: Assuming for contradiction that ${}^p\mathcal{H}^i(K[2n^2]) \neq 0$ for some $i \neq 0$ then since $L \mapsto m^*L[n^2]$ preserves the perverse t -structure and is fully faithful we have ${}^p\mathcal{H}^i(m^*K[3n^2]) = m^*{}^p\mathcal{H}^i(K)[2n^2] \neq 0$, a contradiction, so ${}^p\mathcal{H}^i(K[2n^2]) = 0$ for all $i \neq 0$, i.e. $K[2n^2]$ is perverse, and assuming for contradiction that $K[2n^2]$ is reducible, so there is a proper nontrivial subobject L of $K[2n^2]$, then $m^*L[n^2]$ would be a proper nontrivial subobject of $m^*K[3n^2]$, another contradiction, thus $K[2n^2]$ is irreducible.

Since $K[2n^2]$ is mixed by [5, Variant 6.2.3] and irreducible, it is pure by [4, Corollary 5.3.4]. $\square$

We state some general facts about perverse sheaves that are relatively standard but may not appear in the literature in exactly the form we need them.

Lemma 6. *Let K_1 and K_2 be two perverse sheaves on a variety X over a finite field $\mathbb{F}_q$. Assume that for each extension $\mathbb{F}_{q^e}$ of $\mathbb{F}_q$, for each $x \in X(\mathbb{F}_{q^e})$, we have the equality of trace functions*

$$(1) \quad \sum_i (-1)^i \operatorname{tr}(\operatorname{Frob}_{q^e}, \mathcal{H}^i(K_1)_x) = \sum_i (-1)^i \operatorname{tr}(\operatorname{Frob}_{q^e}, \mathcal{H}^i(K_2)_x)$$

and that K_2 is irreducible. Then $K_1 \cong K_2$.

For a similar statement see Proposition 1.1.2.1 in [14]

Proof. We first check that for bounded complexes K_1, K_2 of constructible ℓ -adic sheaves satisfying the equality of trace functions (1), the classes $\sum_i (-1)^i [\mathcal{H}^i(K_1)]$ and $\sum_i (-1)^i [\mathcal{H}^i(K_2)]$ in the Grothendieck group of constructible ℓ -adic sheaves of K_1 and K_2 are equal. Since both the assumption and conclusion involve an alternating sum, we reduce immediately to the case that K_1 and K_2 are constructible ℓ -adic sheaves. There exists a stratification of X such that both K_1 and K_2 are lisse on each stratum, and since we can write each as the sum in the Grothendieck group of its restriction to each stratum, it suffices to handle the case where K_1 and K_2 are lisse. In this case, we are working in the Grothendieck group of representations of the fundamental group. By the Chebotarev density theorem [15, Theorem 7], the equality of traces of Frobenius implies the equality of traces of every element of $\pi_1^{\text{ét}}(X)$. The result then follows from the fact, in character theory, that two finite dimensional continuous representations of a profinite group over $\mathbb{Q}_\ell$ with the same character are equal in the Grothendieck group of representations.

The class of a complex of perverse sheaves in the Grothendieck group is the alternating sum of the classes of its perverse homology sheaves. From this, one can see that K_1 and K_2 agree in the Grothendieck group of perverse sheaves. But, since the category of perverse sheaves is Artinian and Noetherian [4, Theorem 4.3.1(i)], the Grothendieck group of perverse sheaves is the free group on the isomorphism classes of irreducible perverse sheaves. It follows that, in the Jordan-Hölder decomposition of K_1 in the category of perverse sheaves, the only irreducible component is K_2 , and it occurs with multiplicity 1, i.e. K_1 is isomorphic to K_2 . $\square$

Lemma 7. *Let X be a variety over a finite field, U an open set of X , and $j: U \rightarrow X$ the corresponding open immersion. Let K_1 be an irreducible pure perverse sheaf on X and K_2 a pure complex on X .*

*Suppose that j^*K_1 is a nonzero summand of j^*K_2 . Then K_1 is a summand of K_2 .*

Proof. Since perversity, by definition, is preserved by restriction to an open set, we see that j^*K_1 is perverse and thus is a summand of ${}^p\mathcal{H}^0(j^*K_2) = j^*{}^p\mathcal{H}^0(K_2)$. Since ${}^p\mathcal{H}^0(K_2)$ is a summand of K_2 by [4, Theorem 5.4.5], it suffices to prove that K_1 is a summand of ${}^p\mathcal{H}^0(K_2)$, i.e. we can reduce to the case where K_2 is perverse.

By [4, Corollary 5.3.11], for i the closed immersion of the closed complement of U , we can write K_2 as a sum $j_{!*}K'_2 \oplus i_*K''_2$. We have

$$j^*K_2 = j^*(j_{!*}K'_2 \oplus i_*K''_2) = j^*j_{!*}K'_2 \oplus j^*i_*K''_2 = K'_2 \oplus 0$$

so j^*K_1 is a summand of K'_2 and thus $j_{!*}j^*K_1$ is a summand of $j_{!*}K'_2$ which is a summand of K_2 . So it suffices to show $j_{!*}j^*K_1 \cong K_1$.

But now applying [4, Corollary 5.3.11] to K_1 , we have $K_1 = j_{!*}K'_1 \oplus i_*K''_1$. Since K_1 is irreducible, only one of these terms can be nonzero, and since $j^*i_*K''_1 = 0$, if $j^*K_1 \neq 0$ then we must have $K''_1 = 0$, i.e. $K_1 = j_{!*}K'_1$. But then

$$j_{!*}j^*K_1 = j_{!*}(j^*j_{!*}K'_1) = j_{!*}K'_1 = K_1,$$

as desired. $\square$

Lemma 8. *Let X_1 and X_2 be two varieties over a field κ , and let K_1 and K_2 be geometrically irreducible perverse sheaves on X_1 and X_2 . Then the external tensor product $K_1 \boxtimes K_2$ is a geometrically irreducible perverse sheaf on $X_1 \times X_2$.*

Proof. By [4, Proposition 4.2.8], $K_1 \boxtimes K_2$ is perverse.

By passing to the algebraic closure, we may assume κ is algebraically closed, so geometrically irreducible is the same as irreducible.

By replacing X_1 and X_2 by the supports of K_1 and K_2 , we may assume K_1 is supported on X_1 and K_2 is supported on X_2 . Then K_1 and K_2 are middle extensions of irreducible local systems on smooth affine open subsets U_1 and U_2 of X_1 and X_2 [4, Theorem 4.3.2(ii)]. Restricted to the open set $U = U_1 \times U_2$, the perverse sheaf $K_1 \boxtimes K_2$ is the external tensor product of irreducible local systems and thus is irreducible.

Thus if $K_1 \boxtimes K_2$ is reducible, say it is the extension of a nontrivial quotient Q by a nontrivial subobject S , at least one of Q or S must vanish on restriction to U and thus have support contained in the complement Z of U . Since perverse sheaves and external tensor products are stable under duality, by dualizing everything we may assume without loss of generality that Q has support contained in Z , and thus can be written as i_*Q' for a perverse sheaf Q' on Z . We have a nontrivial morphism $K_1 \boxtimes K_2 \rightarrow Q = i_*Q'$, thus by adjunction a nontrivial morphism $i^*(K_1 \boxtimes K_2) \rightarrow Q'$.

However, we can check that $i^*(K_1 \boxtimes K_2)[-1]$ is semiperverse (i.e. concentrated in degree ≤ 0 in the perverse t -structure, or concretely having j th cohomology sheaf supported in dimension $\leq -j$ for all integers j). The complement of U is a union $(X_1 \setminus U_1) \times X_2 \cup X \times (X_2 \setminus U_2)$. The characterization of semiperversity in terms of support dimension implies that to check a complex on a union of two closed sets is perverse, it suffices to check its restriction to each closed set is perverse, since a union of two sets has dimension $\leq -j$ if and only if each set has dimension $\leq -j$. So it suffices to check that the restriction of $(K_1 \boxtimes K_2)[-1]$ to $(X_1 \setminus U_1) \times X_2$ and

$X \times (X_2 \setminus U_2)$ is each semiperverse. We handle $(X_1 \setminus U_1) \times X_2$, with the other case being symmetric. The restriction of $K_1[-1]$ to $(X_1 \setminus U_1)$ is perverse by [4, Corollary 4.1.12], so its external tensor product with K_2 is (semi)perverse by [4, Proposition 4.2.8].

Because $i^*(K_1 \boxtimes K_2)[-1]$ is semiperverse, the object $i^*(K_1 \boxtimes K_2)$ can have no nontrivial morphism to the perverse sheaf Q' by the definition of perverse t -structure. $\square$

Proof of Theorem [5]. We begin by proving the existence of an isomorphism

$$(2) \quad \rho^* j^* K[2n^2] = u^* \bigotimes_{i=1}^n \lambda_i^* \mathcal{K} \ell_2[n^2](-n(n-1)/2).$$

First we show how the formula of [9, Theorem 1.1(2)] implies the traces of Frob_q on the stalks of the two sides of (2) at any fixed point of $\tilde{U}(\mathbb{F}_q)$ are equal. A point of $\tilde{U}(\mathbb{F}_q)$ is a matrix $a \in M_n(\mathbb{F}_q)$ with distinct eigenvalues, together with an a -stable flag F on $\mathbb{F}_q^n$. The i th eigenvalue λ_i is the eigenvalue of the action of a on F^i/F^{i-1} , which, being the eigenvalue of a 1×1 matrix over $\mathbb{F}_q$, lies in $\mathbb{F}_q$. So a is a matrix with n distinct eigenvalues, all in $\mathbb{F}_q$.

The trace of Frob_q on the stalk of $\rho^* j^* K[2n^2]$ at (a, F) is the trace of Frobenius on the stalk of $K[2n^2]$ at $j(\rho((a, F))) = a$, which by the Lefschetz fixed point formula is the matrix Kloosterman sum $K(a)$.

On the other hand, the trace of Frob_q on the stalk of

$$u^* \bigotimes_{i=1}^n \lambda_i^* \mathcal{K} \ell_2[n^2](-n(n-1)/2)$$

at (a, F) is $(-1)^{n^2} q^{n(n-1)/2}$ times the trace of Frob_q on $u^* \bigotimes_{i=1}^n \lambda_i^* \mathcal{K} \ell_2$, which itself is the product for i from 1 to n of the trace of Frob_q on the stalk of $\mathcal{K} \ell_2$ at $\lambda_i((a, F))$. The stalk of $\mathcal{K} \ell_2$ at a point λ_i is, by definition and the Lefschetz fixed point formula, $(-1)^n$ times the Kloosterman sum $K(\lambda_i)$, so the equality of traces follows from [9, Theorem 1.1(2)]

$$K(a) = q^{n(n-1)/2} \prod_{j=1}^n K(\lambda_j)$$

once we realize that $(-1)^{n^2+n} = 1$. Note that this works for an arbitrary finite field $\mathbb{F}_q$.

It follows from Theorem [4] and the preservation of perversity under étale pull-backs [4, first line of p. 109, $d = 0$ case] that $\rho^* j^* K[2n^2]$ is perverse. If we let $(\prod_i \lambda_i)$ be the morphism $\tilde{\mathbb{A}}^{n^2} \rightarrow \mathbb{A}^n$ whose i th coordinate is λ_i , then

$$u^* \bigotimes_{i=1}^n \lambda_i^* \mathcal{K} \ell_2[n^2](-n(n-1)/2) = u^* \left(\prod_i \lambda_i \right)^* \boxtimes_{i=1}^n \mathcal{K} \ell_2[n^2](-n(n-1)/2)$$

and since $\mathcal{K} \ell_2[1]$ is perverse and geometrically irreducible since it arises by [11, Corollary 4.1.2(i)] from the construction of [4, 5.2.2(a)], $\boxtimes_{i=1}^n \mathcal{K} \ell_2[n]$ is perverse and geometrically irreducible by Lemma [8]. The map $(\prod_i \lambda_i)^*$ is smooth of relative dimension $n^2 - n$, and u is an open immersion, so $u^*(\prod_i \lambda_i)^* \boxtimes_{i=1}^n \mathcal{K} \ell_2[n^2]$ is perverse and geometrically irreducible by [4, Proposition 4.2.5]. Tate twisting does not affect perversity and geometrical irreducibility, since it does not change the complex over an algebraically closed field at all.

By Lemma 6, this implies 2. From 2 we obtain

$$\rho_* \rho^* j^* K[2n^2] = \rho_* u^* \bigotimes_{i=1}^n \lambda_i^* \mathcal{K} \ell_2[n^2](-n(n-1)/2).$$

Since ρ is a finite étale morphism, by [2, IX, (5.1.4)] there are morphisms $j^* K[2n^2] \rightarrow \rho_* \rho^* j^* K[2n^2] \rightarrow j^* K[2n^2]$ whose composition is multiplication by the degree of ρ and thus is invertible, meaning that $j^* K[2n^2]$ is a summand of $\rho_* \rho^* j^* K[2n^2]$. It follows that $j^* K[2n^2]$ is a summand of

$$\rho_* u^* \bigotimes_{i=1}^n \lambda_i^* \mathcal{K} \ell_2[n^2](n(n-1)/2),$$

which, since the map $\pi: \tilde{\mathbb{A}}^{n^2} \rightarrow \mathbb{A}^{n^2}$ forgetting the flag is proper, is isomorphic by proper base change [2, XVII, Proposition 5.2.8] to

$$(3) \quad j^* R\pi_* \bigotimes_{i=1}^n \lambda_i^* \mathcal{K} \ell_2[n^2](n(n-1)/2).$$

Again using the isomorphism $\bigotimes_{i=1}^n \lambda_i^* \mathcal{K} \ell_2 = (\prod_i \lambda_i)^* \boxtimes_{i=1}^n \mathcal{K} \ell_2$, since $\mathcal{K} \ell_2$ is pure of weight 1, $\boxtimes_{i=1}^n \mathcal{K} \ell_2$ is pure of weight n [4, (5.1.14.1) and (5.1.14.1*)], and because $(\prod_i \lambda_i)$ is smooth, $\bigotimes_{i=1}^n \lambda_i^* \mathcal{K} \ell_2$ is pure of weight n [4, Stabilities 5.1.14(i,i*)]. Because π is proper, $R\pi_*$ preserves weights, as does j^* [4, Stabilities 5.1.14(i,i*)]. The shift $[n^2]$ raises weights by n^2 , and the Tate twist $(-n(n-1)/2)$ raises them by $n(n-1)$, so altogether 3 is pure of weight $2n^2$.

We apply Lemma 7 with $K_1 = K[2n^2]$ and

$$K_2 = R\pi_* \bigotimes_{i=1}^n \lambda_i^* \mathcal{K} \ell_2[n^2](n(n-1)/2).$$

We have checked all the conditions except that $j^* K[2n^2]$ is nonzero, but $j^* K[2n^2] = 0$ is easy to rule out here since that would imply its trace function is identically zero, but the formula of [9, Theorem 1.1.2(2)] is manifestly not identically zero (noting that the standard Kloosterman sum is nonzero since it is congruent to $p-1 \equiv -1 \not\equiv 0$ modulo $e^{2\pi i/p} - 1$), so we conclude that $K[2n^2]$ is a summand of

$$R\pi_* \bigotimes_{i=1}^n \lambda_i^* \mathcal{K} \ell_2[n^2](n(n-1)/2)$$

and shifting both sides by $[-2n^2]$ we get the desired statement. $\square$

Corollary 9. *For any matrix a over $\mathbb{F}_q$, $H^j(a)$ is a summand of*

$$(4) \quad H^{j-n^2} \left(\pi^{-1}(a), \bigotimes_{i=1}^n \lambda_i^* \mathcal{K} \ell_2(n(n-1)/2) \right).$$

Proof. This follows immediately from Theorem 5 after taking stalks at a and applying proper base change [2, XVII, Proposition 5.2.8]. $\square$

This enables us to give geometric proofs of Theorems 1 and 2.

Proof of Theorem 2 for a regular. If a is regular then $\pi^{-1}(a)$ consists of finitely many points, as a has finitely many invariant subspaces of each dimension. Thus 4 vanishes unless $j = n^2$, and so $H^j(a)$ vanishes unless $j = n^2$ by Corollary 9. $\square$

Proof of Theorem [L]. We can write $\pi^{-1}(a)$ as a disjoint union of connected components, say $Z_1, \dots, Z_m$. Then we have

$$H^{j-n^2}\left(\pi^{-1}(a), \bigotimes_{i=1}^n \lambda_i^* \mathcal{K}\ell_2(n(n-1)/2)\right) = \bigoplus_{r=1}^m H^{j-n^2}\left(Z_r, \bigotimes_{i=1}^n \lambda_i^* \mathcal{K}\ell_2(n(n-1)/2)\right).$$

Now the maps λ_i can take only finitely many values on $\pi^{-1}(a)$, those being the eigenvalues of a . It follows that λ_i is constant on each connected component Z_r . (The image of a connected component under λ_i must be a connected component of the image of λ_i , but if the image is finite, then each connected component is a point.) Letting $\lambda_i(Z_r)$ be this constant value, then $\lambda_i^* \mathcal{K}\ell_2$ is the tensor product of the constant sheaf with the Galois representation $(\mathcal{K}\ell_2)_{\lambda_i(Z_r)}$. This gives

$$\begin{aligned} H^{j-n^2}\left(Z_r, \bigotimes_{i=1}^n \lambda_i^* \mathcal{K}\ell_2(n(n-1)/2)\right) &= H^{j-n^2}\left(Z_r, \overline{\mathbb{Q}}_\ell \otimes \bigotimes_{i=1}^n (\mathcal{K}\ell_2)_{\lambda_i(Z_r)}(n(n-1)/2)\right) \\ &= H^{j-n^2}\left(Z_r, \overline{\mathbb{Q}}_\ell\right) \otimes \bigotimes_{i=1}^n (\mathcal{K}\ell_2)_{\lambda_i(Z_r)}(n(n-1)/2). \end{aligned}$$

By [I6, Theorem 1], $H^{j-n^2}(Z_r, \overline{\mathbb{Q}}_\ell)$ is pure of weight $j - n^2$. Using the fact that the stalk of $\mathcal{K}\ell_2$ is pure of weight 1 except at 0 where it is pure of weight 0 [I1, Theorem 4.1.1(1) and Theorem 7.4.3], and the fact that the number of i with $\lambda_i = 0$ is the multiplicity of 0 as an eigenvalue and thus is equal to k , we see that $\bigotimes_{i=1}^n (\mathcal{K}\ell_2)_{\lambda_i(Z_r)}$ is pure of weight $n - k$. Finally $\overline{\mathbb{Q}}_\ell(n(n-1)/2)$ is pure of weight $n(n-1)$.

So their tensor product is pure of weight $(j - n^2) + (n - k) + n(n-1) = j - k$. Thus the same thing is true for $H^j(a)$, because $H^j(a)$ is a summand of a sum of these tensor products. $\square$

3. COMBINATORIAL ARGUMENTS

Now we turn to the proof of Theorem [3]. Therefore fix $a \in M_n(\mathbb{F}_q)$ with a unique nonzero eigenvalue α and let

$$r = \min\{j \mid (a - \alpha \text{id})^j = 0\}.$$

There is a filtration on $V = \mathbb{F}_q^n$ given by

$$(5) \quad V_j = \text{Ker}((a - \alpha \text{id})^j) \subset \mathbb{F}_q^n,$$

which we extend to any subspace $W \subset V = \mathbb{F}_q^n$, that is we let

$$(6) \quad W_j = \text{Ker}((a - \alpha \text{id})^j|_W) = W \cap V_j.$$

The associated numerical data $\mu_j = \dim(V_j/V_{j-1})$ defines a partition μ of $n = \dim V$,

$$(7) \quad \mu = [\mu_1 \geq \mu_2 \geq \dots \geq \mu_r] \vdash n.$$

The condition $\sum_{j=1}^r \mu_j = n$ is obvious, and for $0 \leq j < r-1$ the map $a - \alpha \text{id} : V/V_{j+1} \rightarrow V/V_j$ is injective and maps V_{j+2}/V_{j+1} into V_{j+1}/V_j which shows that $\mu_1 \geq \mu_2 \geq \dots \geq \mu_r$. Note that the dual partition μ' codifies the Jordan block structure of a , see e.g. [I7] Chapter 4.

The following counting result and its recursive version will be needed for the proof of Theorem 3. They involve a decomposition of the numbers

$$\# \{W \subset \mathbb{F}_q^n \mid \dim(W) = k, aW = W\}$$

in that theorem by their filtration type. Therefore assume that $\nu = [\nu_1 \geq \nu_2 \geq \dots \geq \nu_l]$ is a partition of k such that $l \leq r$ and $\nu_j \leq \mu_j$, for $j = 1, \dots, l$. We extend ν to length r by $\nu_j = 0$ for $j = l + 1, \dots, r$ and say that $\mu \succ \nu$. In this case for a of type μ as above we let

$$(8) \quad \mathcal{V}(\mu, \nu) := \{W \subset V \mid aW \subset W, \dim(W_{j+1}/W_j) = \nu_j \text{ for } 1 \leq j \leq r\}$$

and $V(\mu, \nu) = \#\mathcal{V}(\mu, \nu)$.

To state the main result recall that the number of k -dimensional subspaces of $\mathbb{F}_q^n$ is given by the q -binomial coefficient

$$\binom{n}{k}_q = \frac{(1 - q^n)(1 - q^{n-1}) \dots (1 - q^{n-k+1})}{(1 - q)(1 - q^2) \dots (1 - q^k)}.$$

Theorem 10. *Let $V = \mathbb{F}_q^n$, $a \in M_n(\mathbb{F}_q)$, and μ be as above. Let $\nu = [\nu_1 \geq \nu_2 \geq \dots \geq \nu_r] \vdash k$ be a partition, (with some ν_j -s possibly 0) such that $\mu \succ \nu$ and set $\nu_{r+1} = 0$. Then*

$$V(\mu, \nu) = \prod_{j=1}^r \binom{\mu_j - \nu_{j+1}}{\nu_j - \nu_{j+1}}_q \cdot q^{\nu_{j+1}(\mu_j - \nu_j)}.$$

Proof. We will proceed inductively. Note that by our convention $\mu_{r+1}, \nu_{r+1} = 0$, therefore we trivially have $\binom{\mu_r}{\nu_r}_q = \binom{\mu_r - \nu_{r+1}}{\nu_r - \nu_{r+1}}_q \cdot q^{\nu_{r+1}(\mu_r - \nu_r)}$. If $r = 1$ all ν_1 dimensional subspaces of V are a -invariant proving the claim in this case.

Let now assume that $r > 1$ and let $\bar{V} = V/V_1$, and denote the natural projection $V \rightarrow \bar{V}$ by π . Since $V_1 = \ker(a - \alpha \text{id})$, $\bar{V}$ has a natural induced action by a , which we denote by $\bar{a}$. Note that the $\bar{a}$ -module $\bar{V}$ is of type $(\mu_2, \dots, \mu_r)$ of length $r - 1$. If $W \subset V$ is of type $\nu = (\nu_1, \nu_2, \dots, \nu_r)$ then its image $\bar{W} = \pi(W)$ is of type $(\nu_2, \dots, \nu_r)$. By the inductive hypothesis the number of such $\bar{a}$ -invariant subspaces $\bar{W}$ is

$$\prod_{j=2}^r \binom{\mu_j - \nu_{j+1}}{\nu_j - \nu_{j+1}}_q \cdot q^{\nu_{j+1}(\mu_j - \nu_j)}.$$

Therefore it is enough to show that given $\bar{W}$ of type $(\nu_2, \dots, \nu_r)$ the number of a -invariant subspaces $W \subset V$ for which $\pi(W) = \bar{W}$ is

$$\binom{\mu_1 - \nu_2}{\nu_1 - \nu_2}_q \cdot q^{\nu_2(\mu_1 - \nu_1)}.$$

First note that the subspace aW is the same for all W for which $\pi(W) = \bar{W}$. To see this note that $\dim \bar{W}/a\bar{W} = \nu_2$ and choose elements $\bar{w}_1, \dots, \bar{w}_{\nu_2}$ whose images form a basis of $\bar{W}/a\bar{W}$. Choose $i(\bar{w}_j) \in \{w \in W \mid \pi(w) = \bar{w}_j\}$, then there is a unique a -invariant map i that extends this choice to give an isomorphism $i : \bar{W} \rightarrow W'$, to a subspace W' that satisfies $W' \oplus W_1 = W$. Clearly the $ai(\bar{w}_j)$ that generate aW are independent of W itself, proving the claim.

We will use this to count all possible W -s via counting the number of W_1 -s, and the possible decompositions $W = W' \oplus W_1$ arising from the maps i . First note that $W_1 \subset V_1$ is of dimension ν_1 and contains aW_2 of dimension ν_2 . The number of such subspaces is $\binom{\mu_1 - \nu_2}{\nu_1 - \nu_2}_q$.

Now fix such a choice of W_1 . If i_1, i_2 are maps as above, then

$$i_1(\overline{W}) \oplus W_1 = i_2(\overline{W}) \oplus W_1$$

if and only if $i_1(\overline{w}_j) - i_2(\overline{w}_j) \in W_1$ for all j . Therefore there are $q^{\nu_2(\mu_1 - \nu_1)}$ choices that lead to different W -s, proving the theorem. $\square$

We will now give a recursive formula for the numbers $V(\mu, \nu)$. Therefore for a partition μ we let

$$(9) \quad \begin{aligned} \mu' &= [\mu_1 \geq \mu_2 \geq \cdots \geq \mu_{r-1} \geq \mu_r - 1], \\ \mu'' &= \begin{cases} [\mu_1 \geq \mu_2 \geq \cdots \geq \mu_{r-1} - 1 \geq \mu_r - 1] & \text{if } r \geq 2 \\ \emptyset & \text{otherwise} \end{cases} \\ \mu''' &= \begin{cases} [\mu_1 \geq \mu_2 \geq \cdots \geq \mu_{r-1} \geq \mu_r - 2] & \text{if } \mu_r > 1 \\ \emptyset & \text{otherwise} \end{cases} \end{aligned}$$

Note that these operations may lead to partitions with some entries that are 0 or the non-partition $\emptyset$. For this reason for any partition λ we define

$$V(\lambda, \emptyset) = V(\emptyset, \lambda) = 0.$$

With this notation we have the following lemma.

Lemma 11.

$$V(\mu, \nu) = V(\mu', \nu) + V(\mu', \nu') - q^{\mu_r - 1} V(\mu'', \nu') + (q^{\mu_r - 1} - 1) V(\mu''', \nu').$$

Proof. The proof is a technical, but straightforward calculation. Set

$$P = \prod_{j=1}^{r-2} \binom{\mu_j - \nu_{j+1}}{\nu_j - \nu_{j+1}}_q \cdot q^{\nu_{j+1}(\mu_j - \nu_j)}.$$

This appears in the explicit formula for all the terms $V(\cdot, \cdot)$ above.

Then

$$\begin{aligned} V(\mu', \nu') + (q^{\mu_r - 1} - 1) V(\mu''', \nu') &= \\ \binom{\mu_{r-1} - \nu_r + 1}{\nu_{r-1} - \nu_r + 1}_q \cdot q^{(\nu_r - 1)(\mu_{r-1} - \nu_{r-1})} \left(\binom{\mu_r - 1}{\nu_r - 1}_q + (q^{\mu_r - 1} - 1) \binom{\mu_r - 2}{\nu_r - 1}_q \right) P &= \\ = q^{\mu_r - 1} \binom{\mu_r - 1}{\nu_r - 1}_q \binom{\mu_{r-1} - \nu_r + 1}{\nu_{r-1} - \nu_r + 1}_q \cdot q^{(\nu_r - 1)(\mu_{r-1} - \nu_{r-1} - 1)} P. \end{aligned}$$

And

$$\begin{aligned} V(\mu', \nu') + (q^{\mu_r - 1} - 1) V(\mu''', \nu') - q^{\mu_r - 1} V(\mu'', \nu') &= \\ = q^{\mu_r - 1} \binom{\mu_r - 1}{\nu_r - 1}_q \cdot q^{(\nu_r - 1)(\mu_{r-1} - \nu_{r-1} - 1)} \left(\binom{\mu_{r-1} - \nu_r + 1}{\nu_{r-1} - \nu_r + 1}_q - \binom{\mu_{r-1} - \nu_r}{\nu_{r-1} - \nu_r + 1}_q \right) P &= \\ = q^{\mu_r - \nu_r} \binom{\mu_r - 1}{\nu_r - 1}_q \binom{\mu_{r-1} - \nu_r}{\nu_{r-1} - \nu_r}_q \cdot q^{\nu_r(\mu_{r-1} - \nu_{r-1})} P. \end{aligned}$$

Thus

$$V(\mu', \nu) + V(\mu', \nu') + (q^{\mu_r-1} - 1)V(\mu''', \nu') - q^{\mu_r-1}V(\mu'', \nu') = \\ \left(\binom{\mu_r-1}{\nu_r}_q + q^{\mu_r-\nu_r} \binom{\mu_r-1}{\nu_r-1}_q \right) \binom{\mu_{r-1}-\nu_r}{\nu_{r-1}-\nu_r}_q \cdot q^{\nu_r(\mu_{r-1}-\nu_{r-1})} P = V(\mu, \nu).$$

□

Before we can turn to the proof of Theorem 3 we need a result from 9. There in Theorem 18.1. the Jordan block structure of a , i.e. the dual partition $\lambda = \mu'$ was used to describe the recursion algorithm. To state this theorem we denote $K(a)$ by $K_\lambda(\alpha)$, and will drop α from the notation since it is fixed for us. In particular $K_{[1]} = K_{[1]}(\alpha) = K(\alpha)$ is the classical Kloosterman. We also let $K_{[]} = 1$.

For example with this notation Theorem 1.4 of 9 says that for $\alpha \neq 0$

$$(10) \quad K_{[1^n]} = q^{n-1} K_{[1]} K_{[1^{n-1}]} + q^{2n-2} (q^{n-1} - 1) K_{[1^{n-2}]}.$$

In general the result in 9 gives for $\lambda = [n_l^{k_l}, n_{l-1}^{k_{l-1}}, \dots, n_1^{k_1}]$, with $n_l > n_{l-1} > \dots > n_1$ that

$$(11) \quad K_\lambda = q^{n-1} K_{[1]} K_{\lambda'} - q^{2n-2} K_{\lambda''} - (q^{k_l-1} - 1) q^{2n-2} (K_{\lambda''} - K_{\lambda'''}),$$

where $\lambda' = [n_l^{k_l-1}, n_l-1, n_{l-1}^{k_{l-1}}, \dots, n_1^{k_1}]$, $\lambda'' = [n_l^{k_l-1}, n_l-2, n_{l-1}^{k_{l-1}}, \dots, n_1^{k_1}]$ and $\lambda''' = [n_l^{k_l-2}, (n_l-1)^2, n_{l-1}^{k_{l-1}}, \dots, n_1^{k_1}]$ reordered into a monotonic sequence, if needed.

Clearly it is more convenient to use μ as in 7, and we reformulate this recursion formula for the matrix Kloosterman sums in terms of it. To denote this shift we will write

$$K^\mu = K_\lambda, \text{ whenever } \lambda = \mu'.$$

Then we have the following theorem.

Theorem 12. *Let $\mu = [\mu_1 \geq \mu_2 \geq \dots \geq \mu_r] \vdash n$ be a partition. Then*

$$K^\mu = q^{n-1} K^{[1]} K^{\mu'} - q^{2n+\mu_r-3} K^{\mu''} + (q^{\mu_r-1} - 1) q^{2n-2} K^{\mu'''}$$

Here $K^{[1]} = 1$ and $\mu' = [\mu_1 \geq \mu_2 \geq \dots \geq \mu_{r-1} \geq \mu_r - 1]$, $\mu'' = [\mu_1 \geq \mu_2 \geq \dots \geq \mu_{r-1} - 1 \geq \mu_r - 1]$ and $\mu''' = [\mu_1 \geq \mu_2 \geq \dots \geq \mu_{r-1} \geq \mu_r - 2]$ are as in 9.

Note that μ'' , μ''' may be the non-partition $\emptyset$, for which we define $K^\emptyset = 0$, i.e. those terms which correspond to $\emptyset$ are omitted from the sum.

Proof. Let λ be a partition and denote the dual partition $\mu = [\mu_1, \dots, \mu_r] = \lambda'$. Let $r_i = r_i(\lambda)$ be the number of parts of λ which are equal to $i \geq 1$. Then $\mu_i = \sum_{j \geq i} r_j$.

The proof of the reformulation is an elementary argument unfolding this relation, the details of which are omitted. □

Proof of Theorem 3. We want to show that

$$K^\mu = (-1)^n q^{n(n-1)/2} \sum_{k=0}^n \lambda^k \bar{\lambda}^{n-k} \sum_{\nu \vdash k} V(\mu, \nu).$$

For convenience we allow partitions ν such that $\nu_j > \mu_j$ for some j , by letting $V(\mu, \nu) = 0$ for them (this is consistent with the notation 8).

We have developed a recurrence relation for both sides, it is enough to check that they are compatible.

Since $K^{[\cdot]} = 1 = V([\cdot], [\cdot])$ and

$$K^{[1]} = -(\lambda + \bar{\lambda}) = -(V([1], [1])\lambda + V([1], [\cdot])\bar{\lambda}),$$

the initial conditions match.

Proceeding inductively we get

$$\begin{aligned} q^{n-1}K^{[1]}K^{\mu'} &= (-1)^n q^{n(n-1)/2} (\lambda + \bar{\lambda}) \sum_{k=0}^{n-1} \lambda^k \bar{\lambda}^{n-k-1} \sum_{\nu \vdash k} V(\mu', \nu) \\ q^{2n+\mu_r-3}K^{\mu''} &= (-1)^n q^{n(n-1)/2} \lambda \bar{\lambda} q^{\mu_r} \sum_{k=0}^{n-2} \lambda^k \bar{\lambda}^{n-k-2} \sum_{\nu \vdash k} V(\mu'', \nu) \\ q^{2n-2}(q^{\mu_r-1} - 1)K^{\mu'''} &= (-1)^n q^{n(n-1)/2} \lambda \bar{\lambda} (q^{\mu_r} - 1) \sum_{k=0}^{n-2} \lambda^k \bar{\lambda}^{n-k-2} \sum_{\nu \vdash k} V(\mu''', \nu). \end{aligned}$$

Now compare the coefficient of $\lambda^k \bar{\lambda}^{n-k}$. On the right-hand side we have

$$(12) \quad \sum_{\nu \vdash k} V(\mu', \nu) + \sum_{\nu' \vdash k-1} (V(\mu', \nu') - q^{\mu_r-1} V(\mu'', \nu') + (q^{\mu_r-1} - 1) V(\mu''', \nu')).$$

We need to show that the sum in (12) produces $\sum_{\nu \vdash k} V(\mu, \nu)$.

(As mentioned above we formally include many extra terms for which by definition $V(\mu, \nu) = 0$, $V(\mu', \nu) = 0$ or $V(\mu'', \nu') = V(\mu''', \nu') = 0$.)

Fix $\nu \vdash k$ and $\nu' = [\nu_1 \geq \nu_2 \geq \dots \geq \nu_{r-1} \geq \nu_r - 1] \vdash k-1$. Note that by Lemma 11 it is enough to prove that the contribution of the terms including ν' that arise from $[\nu_1 \geq \nu_2 \geq \dots \geq \nu_{r-1} \geq \nu_r - 1]$ for which $\nu_{r-1} = \nu_r$ is zero. We separate this vanishing into three cases.

Case 1. $\nu'_{r-1} = \nu'_r < \mu_r - 1$. Then $V(\mu', \nu') = V(\mu'', \nu') = V(\mu''', \nu')$, so

$$V(\mu', \nu') - q^{\mu_r-1} V(\mu'', \nu') + (q^{\mu_r-1} - 1) V(\mu''', \nu') = 0.$$

Case 2. $\nu'_{r-1} = \nu'_r > \mu_r - 1$. Then $V(\mu', \nu') = V(\mu'', \nu') = V(\mu''', \nu') = 0$.

Case 3. $\nu'_{r-1} = \nu'_r = \mu_r - 1$. Then $V(\mu, \nu''') = 0$. The product formula for $V(\mu', \nu')$ and $V(\mu'', \nu')$ differ only in the factor $j = r-1$ and so we let

$$P = \prod_{j \neq r-1} \binom{\mu_j - \nu_{j+1}}{\nu_j - \nu_{j+1}}_q \cdot q^{\nu_{j+1}(\mu_j - \nu_j)}$$

when we have

$$\begin{aligned} &V(\mu', \nu') - q^{\mu_r-1} V(\mu'', \nu') = \\ &P \left(\binom{\mu_{r-1} - \mu_r + 1}{0}_q \cdot q^{(\mu_r-1)(\mu_{r-1}-\mu_r+1)} - q^{\mu_r-1} \binom{\mu_{r-1} - \mu_r}{0}_q \cdot q^{(\mu_r-1)(\mu_{r-1}-\mu_r)} \right), \end{aligned}$$

which is also 0.

□

Corollary 13. $K^\mu = O(q^{n^2/2+d})$, where $d = \frac{1}{4} \sum_{j=1}^r (\mu_j^2 - \delta(\mu_j))$, where $\delta(m) = 1$ if m is odd and $\delta(m) = 0$ if m is even.

For this it is enough to note that

$$\binom{\mu_j - \nu_{j+1}}{\nu_j - \nu_{j+1}}_q \cdot q^{\nu_{j+1}(\mu_j - \nu_j)} = O(q^{\nu_j(\mu_j - \nu_j)}).$$

Thus to maximize the order of $V(\mu, \nu)$ the optimal choice for ν_j is the closest integer to $\mu_j/2$.

Corollary 14. *If $\mu_1 > 1$, then K^μ must admit Frob_q -eigenvalues of two different weights.*

This follows from the fact that K^μ is a polynomial of q , λ and $\bar{\lambda}$ and by the previous corollary we have that this polynomial has a monomial of weight more than n^2 and also one with weight n^2 (the one corresponding to the trivial irreducible subspaces).

Finally we treat the case of matrices a that are not regular:

Proof of Theorem 2. Assume a is not regular. Observe that $H^*(a)$ is the tensor product of the cohomology complexes corresponding to the matrix j_l containing the Jordan blocks for each eigenvalue λ_l of a (for $1 \leq l \leq r$) tensored with a power of the trivial sum (directly follows from [9], Proposition 3.18.):

$$H^*(a) = \left(\bigotimes_{l=1}^r H^*(j_l) \right) \otimes \left(\bigotimes_{m=1}^d H^*(\mathbb{A}^1) \right),$$

with $d = \sum_{1 \leq l < l' \leq r} ll'$.

Assume a has a nonzero eigenvalue α such that the eigenspace is not one-dimensional (that means $\mu_1 > 1$ in the above notation). As the term corresponding to α has two different weights (Corollary 14), so does $H^*(a)$, hence $K(a)$ is not concentrated in a single degree by Theorem 1. $\square$

REFERENCES

- [1] Adolphson, A., Sperber, S. (1989). Exponential sums and Newton polyhedra: cohomology and estimates. *Annals of Mathematics*, 130(2), 367-406.
- [2] Artin, M., Grothendieck, A., Verdier, J.-L., eds. (1972). *Séminaire de Géométrie Algébrique du Bois Marie - 1963-64 - Théorie des topos et cohomologie étale des schémas - (SGA 4) - vol. 3* (PDF). *Lecture Notes in Mathematics* (in French). Vol. 305. Berlin; New York: Springer-Verlag. pp. vi+640. [doi:10.1007/BFb0070714](https://doi.org/10.1007/BFb0070714). ISBN 978-3-540-06118-2. MR 0354654.
- [3] Deligne, P., Grothendieck, A., Verdier, J.-L., eds. (1977). *Séminaire de Géométrie Algébrique du Bois Marie - Cohomologie étale- (SGA 4½) - vol. 3* (PDF). *Lecture Notes in Mathematics* (in French). Vol. 569. Berlin; New York: Springer-Verlag. pp. iv+312. [doi:10.1007/bfb0091526](https://doi.org/10.1007/bfb0091526). ISBN 978-3-540-08066-X. MR 0463174.
- [4] Beilinson, A.A., Bernstein, J., Deligne, P. 1982. *Faisceaux Pervers*. *Astérisque* 100, pp. 1-174.
- [5] Deligne, P. 1980. La conjecture de Weil: II. *Publications mathématiques de l'I.H.É.S.* 52, pp. 137-252
- [6] Denef, J., Loeser, F. (1991). Weights of exponential sums, intersection cohomology, and Newton polyhedra. *Inventiones mathematicae*, 106(1), 275-294.
- [7] Einsiedler, M., Mozes, S., Shah, N., Shapira, U. 2016. Equidistribution of primitive rational points on expanding horospheres. *Compositio Mathematica*, 152(4), 667-692.
- [8] El-Baz, D., Lee, M., Strömbergsson, A. 2022. Effective equidistribution of primitive rational points on expanding horospheres. *arXiv preprint arXiv:2212.07408*.
- [9] Erdélyi, M., Tóth, Á, 2021. Matrix Kloosterman sums. [arXiv:2109.00762](https://arxiv.org/abs/2109.00762)
- [10] Grothendieck, A., 1964. Formule de Lefschetz et rationalité des fonctions L. *Séminaire Bourbaki*, 9, pp.41-55.

- [11] Katz, N.M., 1988. Gauss sums, Kloosterman sums, and monodromy groups (No. 116). Princeton university press.
- [12] Katz, N.M. and Laumon, G., 1985. Transformation de Fourier et majoration de sommes exponentielles. Publications Mathématiques de l'Institut des Hautes Études Scientifiques, 62(1), pp. 145-202.
- [13] Kiehl, R., Weissauer, R., 2001. Weil Conjectures, Perverse Sheaves and ℓ -adic Fourier Transform. Volume 42 of *Ergebnisse der Mathematik und ihrer Grenzgebiete*. Springer-Verlag, Berlin, 2001.
- [14] Laumon, G. 1987. Transformation de Fourier, constantes d'équations fonctionnelles et conjecture de Weil, *Publ. Math. IHES*, 65, pp 131-210.
- [15] Serre, J.-P. 1965. Zeta and L -functions. in *Arithmetical Algebraic Geometry, Proceedings of a Conference held at Purdue University, December 5-7, 1963*, edited by O.F.G. Schilling, pp. 82-92.
- [16] Springer, T. A., 1984. A purity result for fixed point varieties in flag manifolds. *Journal of the Faculty of Science, the University of Tokyo*, 31, pp. 271-282.
- [17] Springer, T. A. and Steinberg, R. Conjugacy classes. In *Seminar on Algebraic Groups and Related Finite Groups (The Institute for Advanced Study, Princeton, N.J., 1968/69)*, pages 167–266. Springer, Berlin, 1970. *Lecture Notes in Mathematics*, Vol. 131.

Email address: merdelyi@math.bme.hu

Email address: sawin@math.columbia.edu

Email address: arpad.toth@ttk.elte.hu