THE HASSE NORM PRINCIPLE FOR A_4-QUARTIC EXTENSIONS OF GLOBAL FUNCTION FIELDS

ANAND DEOPURKAR, RACHEL NEWTON, VAIDEHEE THATTE, ROSA WINTER

Abstract. For a finite extension of global fields K/k, the norm map $N_{K/k} : K^\times \to k^\times$ extends to a map on idèle groups. The Hasse norm principle holds if every element of $k^\times$ that is a norm everywhere locally is also a norm globally. In this paper, we study the statistics of the Hasse norm principle in a setting that is out of reach in the number field context, namely that of A_4-quartic extensions. We show that failures of the Hasse norm principle are generally rare for A_4-quartic extensions of global function fields $\mathbb{F}_q(t)$. We achieve this by introducing a decorated Hurwitz space parametrising the failures of the Hasse norm principle and then using the Chebotarev density theorem to estimate their frequency.

1. Introduction

Let K/k be a finite extension of global fields. Denote by $N_{K/k}\colon K^\times \to k^\times$ the norm map and recall that it extends to a map on idèle groups $N_{K/k}\colon \mathbb{A}_K^\times \to \mathbb{A}_k^\times$. The *Hasse norm principle* is said to hold for K/k if, for an element of $k^\times$, being a global norm from K is equivalent to being a norm everywhere locally. In other words, the Hasse norm principle holds if the so-called *knot group*, defined as

$$\mathfrak{K}(K/k) = (k^\times \cap N_{K/k}\mathbb{A}_K^\times)/N_{K/k}K^\times,$$

is trivial. Historically, most of the work on the Hasse norm principle has been presented in the context of number fields, although many of the results also extend to the setting of global fields of arbitrary characteristic. For example, Hasse's original norm theorem states that the Hasse norm principle holds for cyclic extensions of number fields, but it is also valid for cyclic extensions of global function fields.

Beyond the cyclic case, the Hasse norm principle has been proven to hold, for example, when $[K : k] = p$ for p prime [3]; when $[K : k] = 2p$ for p a prime with $2p-1$ not a prime power [39]; and when $[K : k] = n$ with Galois group either D_n [2], or S_n [48] (see also [46]), or A_n with $n \geq 5$ [30]. Again, while they may be stated for number fields, these results hold for extensions of global fields of arbitrary characteristic — see the references in the proof of Lemma 2.1.

On the other hand, Hasse already showed that not all extensions satisfy the Hasse norm principle by demonstrating that in the abelian extension $\mathbb{Q}(\sqrt{-3}, \sqrt{13})/\mathbb{Q}$ the number 3 gives a counterexample [21]. In [8], Browning and the second author calculated the frequency of counterexamples to the Hasse norm principle for a finite extension $K/\mathbb{Q}$ in terms of the cardinality of the knot group, which they computed for bicyclic extensions. In a similar vein, the authors of [35] computed the proportion of polynomials of degree d in $\mathbb{F}_q[t]$ that are everywhere locally norms but fail to be global norms from a finite extension $K/\mathbb{F}_q(t)$.

The Hasse norm principle has also been studied in families of number fields. Together with Frei and Loughran, the second author showed in [16] that for every non-cyclic finite abelian group G and number field k, there exists a G-extension of k for which the Hasse norm principle fails. In the same paper, the authors showed that, when ordered by discriminant, the Hasse norm principle fails for a positive proportion of finite abelian G-extensions if and only if $G/G[p]$ is not cyclic, where p is the smallest prime dividing $|G|$. These results were strengthened by Koymans and Rome in [23].

2020 *Mathematics Subject Classification.* 11R58, 11N45 (primary), 14G12, 14H30, 14H10 (secondary) .

Asymptotics for the number of biquadratic extensions of $\mathbb{Q}$ for which the Hasse norm principle fails were given by Rome [40]. Similar asymptotics were given for a larger class of extensions by Koymans and Rome in [24]. Finally, Frei, Loughran and the second author proved that, for a fixed number field k and a fixed abelian group G, the Hasse norm principle holds for 100% of extensions of k with Galois group G, when ordered by conductor [17].

Moving beyond the abelian setting is challenging because results on counting number fields, even without imposing the local conditions needed to guarantee validity (or failure) of the Hasse norm principle, are few and far between. Moreover, in some non-abelian families where counting results are available, such as cubic [13, 12, 42] or sextic [6] extensions with Galois group S_3, the Hasse norm principle is known to hold for all extensions in the family, so there is no statistical analysis to be done. The only studies of the statistics of the Hasse norm principle in families of non-abelian extensions of number fields of which we are aware are those carried out by Macedo for D_4-octics [31] and by the second author and Varma for extensions with Galois group S_4 or S_5 [38].

In this paper, we demonstrate the advantages of the global function field setting by studying the statistics of the Hasse norm principle in a setting that is currently out of reach in the number field context, namely that of A_4-quartic extensions. By *an A_4-quartic extension*, we mean a degree 4 extension with Galois group A_4. Our main result is the following.

Theorem 1.1. *Let q be a power of a prime $p \geq 5$. Given $\epsilon > 0$, if n is sufficiently large, then*

$$\limsup_{m\to\infty} \frac{\left|\left\{\begin{array}{l}\text{Isomorphism classes of } A_4\text{-quartic extensions of } \mathbb{F}_{q^m}(t) \text{ with}\\ \text{discriminant of degree } 2n \text{ that fail the Hasse norm principle}\end{array}\right\}\right|}{\left|\left\{\begin{array}{l}\text{Isomorphism classes of } A_4\text{-quartic extensions of } \mathbb{F}_{q^m}(t) \text{ with}\\ \text{discriminant of degree } 2n\end{array}\right\}\right|} \leq (4+\epsilon)n^{-1/3}.$$

Remark 1.2.
(1) For any field k, a finite extension of $k(t)$ is equivalent to a finite cover $f\colon C \to \mathbb{P}^1_k$, where C is a smooth projective curve over k. In our case, the degree of f is 4 and the characteristic of k is at least 5, so f is separable. By the *discriminant* of the extension, we mean the branch divisor of f. For A_4-quartics, the branch divisor is even. In [15], "discriminant" is used slightly differently: there, it means q raised to the degree of the ramification divisor. (Note that the degree of the ramification divisor is the same as the degree of the branch divisor.)
(2) In Theorem 1.1 we require n to be sufficiently large. This comes from Proposition 5.8 and Lemma 6.1; see Remark 6.3.

We obtain the following immediate corollary.

Corollary 1.3. *Let q be a power of a prime $p \geq 5$. Then we have*

$$\lim_{n\to\infty}\limsup_{m\to\infty} \frac{\left|\left\{\begin{array}{l}\text{Isomorphism classes of } A_4\text{-quartic extensions of } \mathbb{F}_{q^m}(t) \text{ with}\\ \text{discriminant of degree } 2n \text{ that fail the Hasse norm principle}\end{array}\right\}\right|}{\left|\left\{\begin{array}{l}\text{Isomorphism classes of } A_4\text{-quartic extensions of } \mathbb{F}_{q^m}(t) \text{ with}\\ \text{discriminant of degree } 2n\end{array}\right\}\right|} = 0.$$

Note that if one replaces the base field $\mathbb{F}_{q^m}(t)$ in Theorem 1.1 and Corollary 1.3 with a number field k, then even the total number of A_4-quartic extensions of k appearing in the denominator is not known: it is a famous open case of Malle's conjecture on the asymptotic distribution of G-extensions of bounded discriminant [33, 34]. For A_4-quartic extensions of $\mathbb{Q}$, the lower bound predicted by Malle has recently been proven by Loughran and Paterson [29] but the upper bound remains out of reach — see [5] for the best result to date.

An A_4-quartic extension of $\mathbb{F}_q(t)$ with discriminant of degree $2n$ corresponds to a degree 4 cover of curves $f\colon C \to \mathbb{P}^1_{\mathbb{F}_q}$ with Galois group A_4 and branch divisor of degree $2n$. Such covers are parametrised by a Hurwitz space which we denote by $H(n)$ (see Section 3 and 6.1 for its definition). Ellenberg and Venkatesh [15] were the first to use the topology of Hurwitz spaces over $\mathbb{C}$ to control their point counts over finite fields and thus obtain heuristics for the number of finite (geometrically connected) G-extensions of $\mathbb{F}_q(t)$, initiating a great deal of activity and progress in this area. Türkelli [44] extended the work of Ellenberg and Venkatesh to tackle covers of $\mathbb{P}^1_{\mathbb{F}_q}$ that are not necessarily geometrically connected, and showed that the resulting heuristic suggested a modification of Malle's conjecture that avoided the counterexamples due to Klüners [22] that were known at the time. Interestingly, despite counterexamples to Türkelli's modification of Malle's conjecture being found by Wang in the number field setting [49], it has now been proven for finite G-extensions of $\mathbb{F}_q(t)$, with q sufficiently large and coprime to $|G|$. Landesman and Levy [26] proved that the order of magnitude of the counting function for such G-extensions is as predicted by Türkelli, improving on an upper bound given by Ellenberg, Tran and Westerland [14]. An asymptotic formula with an explicit description of the leading constant was given by Santens [41].

In order to study the statistics of the Hasse norm principle in families of G-extensions of $\mathbb{F}_q(t)$, one needs to count extensions with local conditions imposed. Ramification conditions are readily captured by the Hurwitz space machinery. Imposing conditions on residue degrees is more challenging as this information is lost in the geometric setting when one base changes to $\overline{\mathbb{F}}_q$. The first example of counting extensions of global function fields with a residue degree condition was in [28] where the authors required that their extensions be completely split at all places over infinity.

In this paper, we translate the conditions for failure of the Hasse norm principle into prescriptions for the ramification indices and residue degrees of our A_4-quartics. We define a decorated Hurwitz space $H^{\mathrm{fail}}(n)$ admitting a finite map $\phi : H^{\mathrm{fail}}(n) \to H(n)$, and show that the Hasse norm principle fails for an A_4-quartic cover if and only if it lies in the image of ϕ. Thus, we are led to consider the quotient

$$\frac{\left|\phi(H^{\mathrm{fail}}(n)(\mathbb{F}_q))\right|}{|H(n)(\mathbb{F}_q)|}.$$

We have an obvious upper bound

$$\frac{\left|\phi(H^{\mathrm{fail}}(n)(\mathbb{F}_q))\right|}{|H(n)(\mathbb{F}_q)|} \le \frac{\left|H^{\mathrm{fail}}(n)(\mathbb{F}_q)\right|}{|H(n)(\mathbb{F}_q)|}$$

and one can understand the right-hand side using existing technology for point counts on Hurwitz spaces. But this upper bound turns out to be too crude for our purposes, giving a positive constant in place of the asymptotically zero result of Corollary 1.3. The key point is that the size of $\phi(H^{\mathrm{fail}}(n)(\mathbb{F}_q))$ is vastly smaller than the size of $H^{\mathrm{fail}}(n)(\mathbb{F}_q)$; the $\mathbb{F}_q$-points of $H^{\mathrm{fail}}(n)$ tend to cluster together in fibres over the $\mathbb{F}_q$-points of $H(n)$. The Chebotarev density theorem allows us to get a handle on this clustering and thus prove our main result, Theorem 1.1.

We expect the techniques introduced in this paper to work in any setting where one can translate the failure of the Hasse norm principle into conditions on ramification indices and residue degrees. In these cases, one can parametrise the failures by a Hurwitz space analogous to our fail space $H^{\mathrm{fail}}(n)$. For example, this is the case for biquadratic extensions, since analogues of Lemmas 2.1 and 2.2 hold by a result of Tate, see [9, p.198]. However, biquadratic extensions can also be tackled by more elementary methods. In [32, Proposition 1.8], local conditions for the validity of the Hasse norm principle are given for extensions with Galois group S_4, S_5, A_4 or A_5. Although [32, Proposition 1.8] is stated for number fields, it also holds for extensions of global function fields, see the proof of Lemma 2.1. For example, for Galois group A_4, the only extensions for which the Hasse norm principle can fail are those of degrees $4, 6$ and 12. We expect the methods showcased here for A_4-quartic extensions to work for A_4-extensions of degrees 6 and 12 as well.

Outline of the paper. The paper is organised as follows. Section 2 contains the necessary background on ramification theory and the Chebotarev density theorem for function fields. We determine conditions for the validity of the Hasse norm principle for an A_4-quartic extension of

$\mathbb{F}_q(t)$ in terms of ramification indices and residue degrees (Lemma 2.2). We use this in Section 3, where we introduce a fail space $H^{\text{fail}}(n_1, n_2)$ as a modification of the Hurwitz space $H(n_1, n_2)$. These spaces depend on parameters n_1 and n_2 corresponding to different types of ramification. We prove that the A_4-quartics for which the Hasse norm principle fails can be parametrised by the image of a forgetful map $H^{\text{fail}}(n_1, n_2)(\mathbb{F}_q) \to H(n_1, n_2)(\mathbb{F}_q)$ (Theorem 3.7). In Section 4, we show how to estimate the ratio of the size of this image to the size of the codomain $H(n_1, n_2)(\mathbb{F}_q)$ using the Chebotarev density theorem. We analyse connected components of our various Hurwitz spaces in Section 5, allowing us to compute this ratio in the large power of q limit (Theorem 5.10). Finally, in Section 6 we analyse the limit as $n = n_1 + n_2$ goes to infinity, proving Theorem 1.1.

Notation and conventions. For a field extension K/k and a place v' of K above a place v of k, we denote by $e(v'/v)$ and $f(v'/v)$ the ramification index and inertia degree, respectively. When K and k are function fields, we write $e(p'/p)$ and $f(p'/p)$ for the ramification index and residue degree, respectively, of the closed point p' corresponding to v' lying over the closed point p corresponding to v.

By Schemes we mean the category of locally Noetherian schemes over $\mathbb{Z}[1/6]$. A *scheme* means an object in this category. If S is a scheme, then an S-scheme is a scheme together with a morphism to S. If X is an S-scheme and $T \to S$ is a morphism, we denote by X_T the T-scheme $X \times_S T \to T$.

A *degree d cover* means a finite flat morphism of degree d.

Acknowledgements. We are grateful to Jordan Ellenberg, Aaron Landesman, Scott Mullane and Melanie Matchett Wood for helpful discussions and to Harmeet Singh for pointing out some useful references. This research was supported by UKRI Future Leaders Fellowship MR/T041609/1, MR/T041609/2 and UKRI1060. Additionally, the first author was supported by the Australian Research Council grant DE180101360 and the last author by ERC Horizon 2020 Consolidator Grant ID 101001051 and MSCA Postdoctoral Fellowship 101148712.

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

2. Preliminaries

This section gathers some necessary background and preliminary results that will eventually allow us to encode failures of the Hasse norm principle via Hurwitz spaces, and to measure the frequency of such failures.

2.1. Local conditions.

In this section we classify the A_4-quartic extensions of $\mathbb{F}_q(t)$ for which the Hasse norm principle holds, in terms of the splitting of primes. We first set up notation and background results.

Let p be a prime, q a power of p, and K a finite extension of $k = \mathbb{F}_q(t)$ of degree coprime to p, with Galois closure L. Let G be the Galois group $\text{Gal}(L/k)$. We assume $(p, |G|) = 1$, and that L/k does not contain any non-trivial constant sub-extensions. Let $H = \text{Gal}(L/K)$.

For a place v of k, and a place w of L above v, we denote by k_v and L_w the completions of k and L with respect to v and w, respectively, and by $\mathbb{F}_v$ and $\mathbb{F}_w$ the corresponding residue fields. We write $D_{w/v} \subset G$ for the decomposition group of w in G, that is, $D_{w/v} = \{g \in G \colon g(w) = w\}$. This is the Galois group of L_w over k_v [37, Prop. II.9.6]. The following diagram summarises what we have defined so far.

$$
\begin{tikzcd}[column sep=large]
\text{field extensions} & \text{places} & \text{completions} & \text{residue fields} \\
L \arrow[d, dash] \arrow[d, dash, bend left=60, "H"] \arrow[dd, dash, bend right=60, "G"'] & w \arrow[d, dash] & L_w \arrow[d, dash] \arrow[d, dash, bend left=60, "D_{w/v}\cap H"] \arrow[dd, dash, bend right=60, "D_{w/v}"'] & \mathbb{F}_w \arrow[d, dash] \\
K \arrow[d, dash] & v' \arrow[d, dash] & K_{v'} \arrow[d, dash] & \mathbb{F}_{v'} \arrow[d, dash] \\
k = \mathbb{F}_q(t) & v & k_v & \mathbb{F}_v \arrow[d, dash] \\
 & & & \mathbb{F}_q
\end{tikzcd}
$$

We denote by $I_{w/v}$ the inertia group of w, that is, the subgroup of $D_{w/v}$ that acts trivially on the residue field $\mathbb{F}_w$. We have a short exact sequence [37, Prop. II.9.9]:

$$0 \longrightarrow I_{w/v} \longrightarrow D_{w/v} \longrightarrow \mathrm{Gal}(\mathbb{F}_w/\mathbb{F}_v) \longrightarrow 0. \tag{2.1}$$

We write D_v to denote a choice of $D_{w/v}$ for some place w of L above a place v of k, and we use the notation I_v similarly. The groups D_v and I_v are well defined up to conjugacy in G.

Lemma 2.1. *Let K/k be an A_4-quartic, i.e. a quartic extension whose normal closure L/k has Galois group A_4. Then the Hasse norm principle holds for K/k if and only if there exists a place v of k whose decomposition group $D_v \subset A_4$ contains a copy of V_4.*

Proof. In the number field setting, this is a result of Kunyavskiĭ, see [25], or a special case of [32, Proposition 1.8], which is a consequence of [32, Theorems 1.1 and 1.3]. In fact, these results also hold true in the setting of extensions of global function fields. Many of the results used in their proofs are purely group theoretic or class field theoretic. A few of them assume characteristic zero in order to make use of resolution of singularities, for example the work of Voskresenskiĭ in [45]. However, the required results have in fact been proven in the setting of global fields of arbitrary characteristic, see e.g. [11, Corollaire 1]. See [43] for a useful collection of extensions to the setting of arbitrary global fields of many results in this area that had previously appeared in the context of number fields. More precisely, for the extension to this more general setting of [32, Theorem 2.1], see [47, Section 11.6]; for [32, Theorem 2.2], see [10, Theorem 5.4]; and for the result on Picard groups used in the proof of [32, Theorem 3.3], see [10, Proposition 3.1].

For the reader's convenience, we also give a direct proof of the result we need here. Tate ([9, p. 198]) has shown that $\mathfrak{K}(L/k)$ is dual to $\ker(\mathrm{H}^3(G,\mathbb{Z}) \to \prod_v \mathrm{H}^3(D_v,\mathbb{Z}))$ where $G = \mathrm{Gal}(L/k) \cong A_4$ and v runs over all places of k. One computes $\mathrm{H}^3(A_4,\mathbb{Z}) = \mathbb{Z}/2\mathbb{Z}$ and uses the fact that V_4 is the Sylow 2-subgroup of A_4 along with [7, Theorem III.10.3] to deduce that $\mathfrak{K}(L/k) \hookrightarrow \mathbb{Z}/2\mathbb{Z}$ is trivial if and only if there exists a place v of k with $V_4 \hookrightarrow D_v$. Now observe that $\mathfrak{K}(K/k)$ is killed by $[K:k] = 4$ and apply [32, Proposition 3.9] to obtain an isomorphism $\mathfrak{K}(K/k) \cong \mathfrak{K}(L/k)$. □

Lemma 2.2. *Let K/k be an A_4-quartic. Then there is a place v of k with $V_4 \hookrightarrow D_v$ if and only if there is a place v' of K above v with $e(v'/v) = f(v'/v) = 2$. Moreover, if this is the case then $D_v \cong V_4$.*

Proof. Let v be a place of k and w a place of L lying above v. Since D_v is the Galois group of the extension of local fields L_w/k_v, we have a filtration of D_v formed by the ramification groups

$$D_v \rhd G_0 \rhd G_1 \rhd G_2 \rhd \ldots,$$

where $G_0 = I_v$, the quotient $I_v/G_1 \hookrightarrow \mathbb{F}_w^\times$ is cyclic of order coprime to p, and $G_i/G_{i+1} \hookrightarrow (\mathbb{F}_w, +)$ is a p-group for $i \geq 1$ [37, Prop. II.10.2]. Therefore, the group G_1 is a p-group, and since $(p, |G|) = 1$ by assumption, it follows that G_1 is trivial, hence I_v is cyclic. Since D_v/I_v is the Galois group of the field extension $\mathbb{F}_w/\mathbb{F}_v$ by (2.1), it is also cyclic. As before, we denote by H the Galois group $\mathrm{Gal}(L/K)$; it is a subgroup of A_4 of order 3.

Now assume that $V_4 \hookrightarrow D_v$. Since D_v is a subgroup of $G \cong A_4$, we deduce that D_v is isomorphic to V_4 or A_4. As both I_v and D_v/I_v are cyclic, we find $D_v \cong V_4$, and $I_v \cong C_2$. Therefore, $e(w/v) = f(w/v) = 2$ and $D_v \cap H = \{1\}$. Now let v' be a place of K below w, and note that $K_{v'} \subset L_w$ is the fixed field of the intersection $D_v \cap H \subset A_4$. Since $D_v = V_4$, it follows that $K_{v'}$ is the fixed field of the trivial group, hence $K_{v'} = L_w$. We conclude that $e(v'/v) = f(v'/v) = 2$, as required.

Conversely, let v' be a place of K above v and assume that $e(v'/v) = f(v'/v) = 2$. Then 4 divides $|D_v|$ and by considering the subgroups of A_4 we conclude that $V_4 \hookrightarrow D_v$. $\square$

Remark 2.3. It follows from the proof of Lemma 2.2 that the inertia group at a ramified place is a cyclic subgroup of A_4, hence it is either generated by a 3-cycle or by a product of two disjoint 2-cycles. We refer to the former as type $(3,1)$ and the latter as type $(2,2)$. In the notation of [4, § 2], the former has splitting symbol $(1^3 1^1)$ and the latter has splitting symbol $(1^2 1^2)$ if $f(v'/v) = 1$ and (2^2) if $f(v'/v) = 2$.

2.2. **The Chebotarev density theorem.** We recall the Chebotarev density theorem for function fields. Let G be a finite group. Let X and Y be $\mathbb{F}_q$-schemes and $\phi\colon X \to Y$ a G-torsor defined over $\mathbb{F}_q$. Fix an algebraic closure $\mathbb{F}_q \subset \overline{\mathbb{F}}_q$ and let $\mathrm{Frob}_q \in \mathrm{Gal}(\overline{\mathbb{F}}_q/\mathbb{F}_q)$ be the Frobenius. Given $y \in Y(\mathbb{F}_q)$, consider the $\mathbb{F}_q$-scheme X_y. The map $\phi\colon X_y \to y = \operatorname{Spec} \mathbb{F}_q$ is a G-torsor. Therefore, given $\overline{x} \in X_y(\overline{\mathbb{F}}_q)$, there is a unique $g \in G$ such that

$$\mathrm{Frob}_q(\overline{x}) = g \cdot \overline{x}.$$

A different choice of $\overline{x} \in X_y(\overline{\mathbb{F}}_q)$ changes g by a conjugate in G. So the conjugacy class of g is a well-defined invariant associated to ϕ and y, which we denote by $\varphi(y)$.

Theorem 2.4 (Chebotarev density theorem). *Let G be a finite group, $\phi\colon X \to Y$ a G-torsor defined over $\mathbb{F}_q$, and $\Phi \subset G$ a conjugacy class.*

(1) Suppose that X is geometrically connected. Then

$$\lim_{m\to\infty} \frac{|\{y \in Y(\mathbb{F}_{q^m}) \mid \varphi(y) = \Phi\}|}{|Y(\mathbb{F}_{q^m})|} = \frac{|\Phi|}{|G|}.$$

(2) More generally, suppose that Y is geometrically connected and X has r geometric components, all defined over $\mathbb{F}_q$. Then

$$\lim_{m\to\infty} \frac{|\{y \in Y(\mathbb{F}_{q^m}) \mid \varphi(y) = \Phi\}|}{|Y(\mathbb{F}_{q^m})|} \leq r\frac{|\Phi|}{|G|},$$

with equality if and only if Φ is contained in the stabiliser of a geometric component of X.

Proof. The first assertion is well known; see, for example, [27, Théorème 1]. It has a short proof using the Lang–Weil bounds; see [36].

For the second assertion, let $X_0 \subset X$ be a geometrically connected component. Let $H \subset G$ be the stabiliser of X_0. Then X_0 is geometrically connected, $X_0 \to Y$ is an H-torsor, and the index of H in G is r.

Associated to a $y \in Y(\mathbb{F}_{q^m})$, we now have two conjugacy classes. The first is a conjugacy class of G from the G-torsor $X \to Y$. The second is a conjugacy class of H from the H-torsor $X_0 \to Y$. Denote the first by $\varphi^G(y)$ and the second by $\varphi^H(y)$. We claim that $\varphi^H(y) \subset \varphi^G(y)$. To see this, recall that the definition of $\varphi^G(y)$ involves a choice of an arbitrary $\overline{x} \in X(\overline{\mathbb{F}}_q)$ over $y \in Y(\mathbb{F}_{q^m})$.

If we take $\overline{x}$ to be in X_0, then the unique $g \in G$ such that $\mathrm{Frob}_{q^m}(\overline{x}) = g \cdot \overline{x}$ lies in $H \subset G$. By definition, $\varphi^G(y)$ is the conjugacy class of G represented by g and $\varphi^H(y)$ is the conjugacy class of H represented by g. So $\varphi^H(y) \subset \varphi^G(y)$.

The subset $\Phi \cap H \subset H$ is preserved by conjugation by elements of H. Write it as a (possibly empty) union of distinct conjugacy classes of H, say

$$\Phi \cap H = \Phi_1 \sqcup \cdots \sqcup \Phi_n.$$

Since $\varphi^H(y) \subset \varphi^G(y)$, we see that $\varphi^G(y) = \Phi$ if and only if $\varphi^H(y) = \Phi_i$ for some i. By the Chebotarev density theorem applied to the H-torsor $X_0 \to Y$, we have

$$\lim_{m\to\infty} \frac{|\{y \in Y(\mathbb{F}_{q^m}) \mid \varphi^H(y) = \Phi_i\}|}{|Y(\mathbb{F}_{q^m})|} = \frac{|\Phi_i|}{|H|}.$$

Since $\varphi^G(y) = \Phi$ is equivalent to $\varphi^H(y) = \Phi_i$ for some i, we get

$$\lim_{m\to\infty} \frac{|\{y \in Y(\mathbb{F}_{q^m}) \mid \varphi^G(y) = \Phi\}|}{|Y(\mathbb{F}_{q^m})|} = \frac{|\Phi_1| + \cdots + |\Phi_n|}{|H|} = \frac{|\Phi \cap H|}{|H|}.$$

Recalling that the index of H in G is r, we have

$$\frac{|\Phi \cap H|}{|H|} \leq \frac{|\Phi|}{|H|} = r\frac{|\Phi|}{|G|},$$

with equality if and only if $\Phi \subset H$. The proof is now complete. $\square$

Let X and Y be $\mathbb{F}_q$-schemes, $\phi\colon X \to Y$ a G-torsor defined over $\mathbb{F}_q$, and $\Lambda \subset G$ a subgroup. Let Z be the quotient of X by Λ. This quotient exists as a scheme by [1, Expose V, Proposition 3.1]. Since $X \to Y$ is finite and étale, so is the induced map $\phi\colon Z \to Y$.

Proposition 2.5. *In the set-up above, let $\Phi \subset G$ be the union $\Phi = \bigcup_{h\in G} h^{-1}\Lambda h$. Then*

$$\phi(Z(\mathbb{F}_q)) = \{y \in Y(\mathbb{F}_q) \mid \varphi_y \subset \Phi\}.$$

Furthermore, if Y is geometrically connected and X has r geometric components, all defined over $\mathbb{F}_q$, then

$$\lim_{m\to\infty} \frac{|\phi(Z(\mathbb{F}_{q^m}))|}{|Y(\mathbb{F}_{q^m})|} \leq r\frac{|\Phi|}{|G|},$$

with equality if and only if Φ is contained in the stabiliser of a geometric component of X.

Proof. Let $y \in Y(\mathbb{F}_q)$. Let $X_y(\overline{\mathbb{F}}_q)$ be the set of geometric points of X over y. Choose $\overline{x} \in X_y(\overline{\mathbb{F}}_q)$. Since $X \to Y$ is a G-torsor, we have

$$X_y(\overline{\mathbb{F}}_q) = \{g \cdot \overline{x} \mid g \in G\}.$$

Since $Z = X/\Lambda$, the elements of $Z_y(\overline{\mathbb{F}}_q)$ correspond to right Λ-cosets

$$Z_y(\overline{\mathbb{F}}_q) = \{\Lambda h \cdot \overline{x} \mid h \in G\}.$$

Let $g \in G$ be such that $\mathrm{Frob}_q(\overline{x}) = g \cdot \overline{x}$. By definition, $\varphi(y)$ is the conjugacy class of g. The action of Frob_q on $Z_y(\overline{\mathbb{F}}_q)$ sends the point corresponding to the coset $\Lambda h \cdot \overline{x}$ to the point corresponding to $\Lambda h \cdot g\overline{x}$. This point is fixed by Frob_q if and only if $\Lambda h = \Lambda hg$, which in turn is equivalent to $g \in h^{-1}\Lambda h$. We conclude that $y \in Y(\mathbb{F}_q)$ lies in the image of $Z(\mathbb{F}_q)$ if and only if there exists an $h \in G$ and $g \in h^{-1}\Lambda h$ such that φ_y is the conjugacy class of g. This statement is clearly equivalent to $\varphi_y \subset \Phi$. We have thus proved the first assertion. Given the first assertion, the second follows from Theorem 2.4. $\square$

3. Hurwitz spaces and the Hasse norm principle

In this section, we define a Hurwitz space of covers of the projective line, and a particular covering of this space that accounts for the failure of the Hasse norm principle.

Fix $n_1 \in \mathbb{Z}_{>0}$ and $n_2 \in \mathbb{Z}_{\geq 0}$. For $d \in \mathbb{Z}_{\geq 0}$, we interpret $\mathbb{P}^d$ as the moduli space of effective divisors of degree d on $\mathbb{P}^1$. Let $U(n_1, n_2) \subset \mathbb{P}^{n_1} \times \mathbb{P}^{n_2}$ be the open subscheme whose S-points for a scheme S are (B_1, B_2), where

(1) $B_i \subset \mathbb{P}^1_S$ is finite and étale over S of degree n_i, and
(2) B_1 and B_2 are disjoint.

Let $\mathcal{U}(n_1, n_2)$ be the functor represented by $U(n_1, n_2)$.

An A_4-*quartic cover of* $\mathbb{P}^1$ over a scheme S consists of

$$(C \xrightarrow{\pi} S, B_1, B_2, C \xrightarrow{f} \mathbb{P}^1_S),$$

where

(1) π is a proper smooth morphism whose geometric fibres are connected and of dimension 1,
(2) $B_1, B_2 \subset \mathbb{P}^1_S$ are disjoint subschemes, finite and étale over S of degrees n_1 and n_2, respectively,
(3) f is a finite flat morphism of degree 4 that is étale over the complement of $B_1 \cup B_2$;
(4) for every geometric point $s \to S$, the degree 4 map $f_s \colon C_s \to \mathbb{P}^1_s$ has ramification type $(3, 1)$ over every point of $B_{1,s}$ and type $(2, 2)$ over every point of $B_{2,s}$. In the notation of [4, § 2], the splitting symbol is $(1^3 1^1)$ over every (geometric) point of $B_{1,s}$ and $(1^2 1^2)$ over every (geometric) point of $B_{2,s}$.

We remind the reader that our schemes are over $\mathbb{Z}[1/6]$, and therefore a cover of degree 4 is automatically tame.

Remark 3.1. Connected degree 4 covers with ramification of type $(3, 1)$ over at least one point and of type $(2, 2)$ over zero or more points are precisely the covers with Galois group A_4. Indeed, let k be an algebraically closed field of characteristic greater than 3, and let

$$(C \to \operatorname{Spec} k, B_1, B_2, C \to \mathbb{P}^1_k)$$

be an A_4-quartic cover of $\mathbb{P}^1_k$. Then the Galois group of $k(C)/k(\mathbb{P}^1)$ is a transitive subgroup of the symmetric group S_4 generated by at least one 3-cycle (since $n_1 = \deg B_1 > 0$) and zero or more $(2, 2)$-cycles. The only such subgroup is the alternating group A_4. Conversely, any degree 4 cover $f \colon C \to \mathbb{P}^1_k$ with Galois group A_4 must have ramification types $(3, 1)$ or $(2, 2)$, and $(3, 1)$ must appear at least once.

Let $(C \to S, B_1, B_2, C \xrightarrow{f} \mathbb{P}^1_S)$ and $(C' \to S, B_1', B_2', C' \xrightarrow{f'} \mathbb{P}^1_S)$ be two A_4-quartic covers of $\mathbb{P}^1_S$. We say that they are *isomorphic* if for $i = 1, 2$ we have $B_i = B_i'$ and there exists an S-isomorphism $\psi \colon C \to C'$ such that the following diagram commutes:

$$\begin{tikzcd} C \arrow[r, "\psi"] \arrow[d, "f"] & C' \arrow[d, "f'"] \\ \mathbb{P}^1_S \arrow[r, equal] & \mathbb{P}^1_S. \end{tikzcd}$$

Let $\mathcal{H}(n_1, n_2)$ be the contravariant functor from <u>Schemes</u> to <u>Sets</u> that sends S to the set of isomorphism classes of A_4-quartic covers of $\mathbb{P}^1_S$.

Remark 3.2. A more "correct" formulation of the functor of A_4-quartic covers would have been as a functor valued in groupoids. However, it is easy to check that A_4-quartic covers of $\mathbb{P}^1$ have trivial automorphism groups. So the groupoid-valued functor is equivalent to a set-valued functor.

THEOREM 3.3. *The functor $\mathcal{H}(n_1, n_2)$ is represented by a smooth quasi-projective $\mathbb{Z}[1/6]$-scheme $H(n_1, n_2)$ of relative dimension $n_1 + n_2$. The forgetful natural transformation $\mathcal{H}(n_1, n_2) \to \mathcal{U}(n_1, n_2)$ is represented by a finite étale morphism*

$$H(n_1, n_2) \to U(n_1, n_2).$$

Proof. Follows from arguments analogous to the construction of the Hurwitz scheme of simple coverings due to Fulton [19, § 6] (see also, [50, § 4]). We note that $U(n_1, n_2) \to \operatorname{Spec} \mathbb{Z}[1/6]$ is smooth of relative dimension $n_1 + n_2$. Therefore, $H(n_1, n_2) \to \operatorname{Spec} \mathbb{Z}[1/6]$ is also smooth of the same relative dimension. □

Let

$$(\mathcal{C} \xrightarrow{\pi} H(n_1, n_2), \mathcal{B}_1, \mathcal{B}_2, \mathcal{C} \xrightarrow{f} \mathbb{P}^1 \times H(n_1, n_2))$$

be the universal A_4-quartic cover of $\mathbb{P}^1$ over $H(n_1, n_2)$. Let $\mathcal{C}_2 \subset \mathcal{C}$ be the reduced subscheme underlying $f^{-1}(\mathcal{B}_2)$.

PROPOSITION 3.4. *As divisors on $\mathcal{C}$, we have*

$$2 \cdot \mathcal{C}_2 = f^* \mathcal{B}_2.$$

Furthermore, the map $f \colon \mathcal{C}_2 \to \mathcal{B}_2$ is finite étale of degree 2.

Proof. This is a simple local calculation using the structure theorem for finite tame covers [19, § 4]. □

Remark 3.5. Let S be a reduced scheme and let $(C \xrightarrow{\pi} S, B_1, B_2, C \xrightarrow{f} \mathbb{P}^1_S)$ be an A_4-quartic cover of $\mathbb{P}^1$ over S. Let $S \to H(n_1, n_2)$ be the corresponding morphism. Then $\mathcal{C}_2 \times_{H(n_1,n_2)} S$ is simply the reduced scheme underlying $f^{-1}(B_2)$. Indeed, both are reduced schemes with the same support.

Let S be a scheme and let X_1, X_2, and Y be S-schemes with morphisms $X_1 \to Y$ and $X_2 \to Y$. By $\operatorname{Isom}_S(X_1 \to Y, X_2 \to Y)$ we denote the functor that sends an S-scheme T to the set of isomorphisms

$$i \colon X_{1,T} \to X_{2,T}$$

that make the following diagram commute

$$\begin{array}{ccc} X_{1,T} & \xrightarrow{i} & X_{2,T} \\ \downarrow & & \downarrow \\ Y_T & = & Y_T. \end{array}$$

If $Y = S$, we write $\operatorname{Isom}_S(X_1, X_2)$ for $\operatorname{Isom}_S(X_1 \to S, X_2 \to S)$.

We let $\mathcal{H}^{\text{fail}}(n_1, n_2)$ be the functor

$$\mathcal{H}^{\text{fail}}(n_1, n_2) = \operatorname{Isom}_{H(n_1,n_2)}(\{\pm 1\} \times \mathcal{B}_2 \to \mathcal{B}_2, \mathcal{C}_2 \to \mathcal{B}_2).$$

Let us pause to understand the functor over $S = \operatorname{Spec} k$ for an algebraically closed field k. In this case, an element of $\mathcal{H}^{\text{fail}}(n_1, n_2)(S)$ is an A_4-quartic cover of $\mathbb{P}^1_k$, say $(C \to \operatorname{Spec} k, B_1, B_2, C \xrightarrow{f} \mathbb{P}^1_k)$ together with an isomorphism

$$\{\pm 1\} \times B_2 \to \mathcal{C}_{2,k}$$

over B_2. From Remark 3.5, we know that $C_2 = \mathcal{C}_{2,k}$ is the reduced pre-image of B_2 in C. An isomorphism over B_2 from $\{\pm 1\} \times B_2$ to C_2 is simply a labelling by $\{\pm 1\}$ of the two points of C_2 over every point of B_2. Thus, $\mathcal{H}^{\text{fail}}(n_1, n_2)(S)$ is the set of isomorphism classes of A_4-quartic covers $(C \to \operatorname{Spec} k, B_1, B_2, C \xrightarrow{f} \mathbb{P}^1_k)$ of $\mathbb{P}^1_k$ together with a labelling of the two points of C over each point of B_2.

Proposition 3.6. *The functor $\mathcal{H}^{\mathrm{fail}}(n_1, n_2)$ is represented by an $H(n_1, n_2)$-scheme $H^{\mathrm{fail}}(n_1, n_2)$. The natural map $H^{\mathrm{fail}}(n_1, n_2) \to H(n_1, n_2)$ is finite and étale of degree 2^{n_2}.*

Proof. The representability of Isom follows from standard arguments using the Hilbert scheme (for example, [20, § 4 Variantes]). In fact, it is even simpler in our case since the morphisms involved are finite.

To see that the natural map is finite étale of the given degree, we make an étale base change $S \to H(n_1, n_2)$ such that we have an isomorphism $\{1, \dots, n_2\} \times S \to \mathcal{B}_{2,S}$ over S and an isomorphism $\{\pm 1\} \times \mathcal{B}_{2,S} \to \mathcal{C}_{2,S}$ over $\mathcal{B}_{2,S}$. Let F be the subset of the permutations of $\{\pm 1\} \times \{1, \dots, n_2\}$ that commute with the projection to $\{1, \dots, n_2\}$. Then, it is easy to see that $F = (\mathbb{Z}/2\mathbb{Z})^{n_2}$ and $H^{\mathrm{fail}}(n_1, n_2)_S = F \times S$. □

We now relate the map $H^{\mathrm{fail}}(n_1, n_2) \to H(n_1, n_2)$ to the failure of the Hasse norm principle.

Theorem 3.7. *Let k be a finite field of characteristic $\neq 2, 3$. Let $(C, B_1, B_2, C \xrightarrow{f} \mathbb{P}^1_k)$ be an A_4-quartic cover of $\mathbb{P}^1$ over k. The Hasse norm principle fails for f if and only if the k-point of $H(n_1, n_2)$ represented by f lies in the image of a k-point of $H^{\mathrm{fail}}(n_1, n_2)$.*

Proof. Let $K = k(C)$ be the function field of C and $k(\mathbb{P}^1) = k(t)$ the function field of $\mathbb{P}^1_k$. Let $C_2 \subset C$ be the reduced pre-image of B_2.

Lemmas 2.1 and 2.2 show that the Hasse norm principle fails for $K/k(t)$ if and only if there is no place v' of K lying above a place v of $k(t)$ with $e(v'/v) = f(v'/v) = 2$. A place v of $k(t)$ corresponds to a closed point p of $\mathbb{P}^1_k$; a place of K over v corresponds to a closed point p' of C over p. If p lies in the complement of $B_1 \cup B_2$, then $e(p'/p) = 1$ since $C \to \mathbb{P}^1_k$ is unramified there. If p lies in B_1, then $e(p'/p) = 1$ or 3 since the ramification type over points of B_2 is $(3, 1)$.

For a closed point $p \in B_2$, we have $\sum e(p'/p) f(p'/p) = 4$, where the sum runs over all closed points p' of C_2 lying over p. Since the ramification type over points of B_2 is $(2, 2)$, we have $e(p'/p) = 2$ for all p' over p, and hence $\sum f(p'/p) = 2$. This means we are in one of the following two cases:

(1) there is a unique p' lying over p with $f(p'/p) = 2$; or
(2) there are two closed points p'_1 and p'_2 of C lying over p with $f(p'_1/p) = f(p'_2/p) = 1$.

If Case 1 occurs for some closed point p of B_2, then the Hasse norm principle holds for f by Lemma 2.2. In this case, the residue field of C at p' is a degree 2 extension of the residue field of $\mathbb{P}^1$ at p. In particular, there is no isomorphism $\{\pm 1\} \times p \to p'$ and, therefore, no isomorphism $\{\pm 1\} \times B_2 \to C_2$. Therefore, the k-point of $H(n_1, n_2)$ represented by f is not the image of a k-point of $H^{\mathrm{fail}}(n_1, n_2)$.

If Case 2 occurs for every closed point p of B_2, then the Hasse norm principle fails for f by Lemma 2.2. In this case, the labelling $+1 \mapsto p'_1, -1 \mapsto p'_2$ of the two points of C over p gives an isomorphism $\{\pm 1\} \times p \to C_{2,p}$. Letting p vary over the closed points of B_2 gives an isomorphism $\{\pm 1\} \times B_2 \to C_2$. This isomorphism gives a k-point of $H^{\mathrm{fail}}(n_1, n_2)$ that maps to the k-point represented by f. □

4. Hurwitz spaces for the Chebotarev density theorem

Let $\phi \colon H^{\mathrm{fail}}(n_1, n_2) \to H(n_1, n_2)$ be the natural forgetful map of Hurwitz spaces defined in Section 3. Our goal is to use the Chebotarev density theorem to estimate the ratio

$$\frac{|\phi(H^{\mathrm{fail}}(n_1, n_2)(\mathbb{F}_q))|}{|H(n_1, n_2)(\mathbb{F}_q)|}.$$

By Theorem 3.7, this ratio represents the proportion of A_4-quartic covers of $\mathbb{P}^1_{\mathbb{F}_q}$ for which the Hasse norm principle fails. To estimate this ratio, we express $H^{\mathrm{fail}}(n_1, n_2)$ as the quotient of a

G-torsor over $H(n_1, n_2)$. This G-torsor is another Hurwitz space, with some more decoration. We now define it.

Recall that

$$(\mathcal{C} \xrightarrow{\pi} H(n_1, n_2), \mathcal{B}_1, \mathcal{B}_2, \mathcal{C} \xrightarrow{f} \mathbb{P}^1 \times H(n_1, n_2))$$

is the universal object over $H(n_1, n_2)$. The map $\mathcal{B}_2 \to H(n_1, n_2)$ is étale of degree n_2. Set

$$\widetilde{H}(n_1, n_2) = \mathrm{Isom}_{H(n_1,n_2)}(\{1, \ldots, n_2\} \times H(n_1, n_2), \mathcal{B}_2), \tag{4.1}$$

and

$$\widetilde{H}^{\mathrm{fail}}(n_1, n_2) = H^{\mathrm{fail}}(n_1, n_2) \times_{H(n_1,n_2)} \widetilde{H}(n_1, n_2). \tag{4.2}$$

The following pull-back diagram summarises the relationships between the spaces defined so far. All the arrows are finite and étale of the indicated degree.

$$\begin{tikzcd} \widetilde{H}^{\mathrm{fail}}(n_1, n_2) \arrow[r, "n_2!"] \arrow[d, "2^{n_2}"'] & H^{\mathrm{fail}}(n_1, n_2) \arrow[d, "2^{n_2}"] \\ \widetilde{H}(n_1, n_2) \arrow[r, "n_2!"'] & H(n_1, n_2). \end{tikzcd} \tag{4.3}$$

For a scheme S, let us also recall the S-points of each of the spaces above.

$H(n_1, n_2)$: An isomorphism class of an A_4-quartic cover of $\mathbb{P}^1_S$, say

$$(C \to S, B_1, B_2, C \xrightarrow{f} \mathbb{P}^1_S).$$

Let $C_2 \subset C$ be the pull-back of $\mathcal{C}_2$.

$H^{\mathrm{fail}}(n_1, n_2)$: The data above together with an S-isomorphism $\{\pm 1\} \times B_2 \to C_2$ compatible with the projection to B_2.

$\widetilde{H}(n_1, n_2)$: A point of $H(n_1, n_2)$ as above together with an S-isomorphism

$$\{1, \ldots, n_2\} \times S \to B_2.$$

$\widetilde{H}^{\mathrm{fail}}(n_1, n_2)$: A point of $H(n_1, n_2)$ as above together with S-isomorphisms

$$\nu\colon \{\pm 1\} \times \{1, \ldots, n_2\} \times S \to C_2 \text{ and } \overline{\nu}\colon \{1, \ldots, n_2\} \times S \to B_2$$

such that the following diagram commutes

$$\begin{tikzcd} \{\pm 1\} \times \{1, \ldots, n_2\} \times S \arrow[r, "\nu"] \arrow[d] & C_2 \arrow[d, "f"] \\ \{1, \ldots, n_2\} \times S \arrow[r, "\overline{\nu}"] & B_2. \end{tikzcd}$$

Having described the S-points, we now define natural group actions on the spaces above under which three of the arrows in (4.3) become torsors. First, the natural action of the symmetric group S_{n_2} on $\{1, \ldots, n_2\}$ induces an action on $\widetilde{H}(n_1, n_2)$. Under this action, $\widetilde{H}(n_1, n_2) \to H(n_1, n_2)$ is an S_{n_2}-torsor. By pull-back, $\widetilde{H}^{\mathrm{fail}}(n_1, n_2) \to H^{\mathrm{fail}}(n_1, n_2)$ is also an S_{n_2}-torsor.

Let $G \subset \mathrm{Perm}(\{\pm 1\} \times \{1, \ldots, n_2\})$ be the subgroup consisting of permutations α such that for all $i \in \{1, \ldots, n_2\}$, the second coordinates of $\alpha(+1, i)$ and $\alpha(-1, i)$ agree. We have a natural map $G \to S_{n_2}$ that sends α to the permutation $\overline{\alpha}$ that takes i to the second coordinate of $\alpha(\pm 1, i)$. This map is clearly surjective. In fact, it has a left inverse that sends $\sigma \in S_{n_2}$ to the α that sends $(\pm 1, i)$ to $(\pm 1, \sigma(i))$. It is easy to see that the kernel of $G \to S_{n_2}$ is $(\mathbb{Z}/2\mathbb{Z})^{n_2}$, and G is the semidirect product

$$G = (\mathbb{Z}/2\mathbb{Z})^{n_2} \rtimes S_{n_2}$$

for the permutation action of S_{n_2} on $(\mathbb{Z}/2\mathbb{Z})^{n_2}$.

We have an action of G on $\widetilde{H}^{\text{fail}}(n_1, n_2)$. The element $\alpha \in G$ acts by $\nu \mapsto \nu \circ \alpha^{-1}$ and $\overline{\nu} \mapsto \overline{\nu} \circ \overline{\alpha}^{-1}$. The map $\widetilde{H}^{\text{fail}}(n_1, n_2) \to H(n_1, n_2)$ is a G-torsor under this action. Observe that the action of the subgroup $S_{n_2} \subset G$ on $\widetilde{H}^{\text{fail}}(n_1, n_2)$ is precisely the action induced from the action on $\widetilde{H}(n_1, n_2)$.

We now redraw the diagram (4.3) and decorate each arrow with the group under which it is a torsor. (The right vertical arrow is not a torsor.)

$$\begin{tikzcd} \widetilde{H}^{\text{fail}}(n_1, n_2) \arrow[r, "S_{n_2}"] \arrow[d, "(\mathbb{Z}/2\mathbb{Z})^{n_2}"'] \arrow[dr, "G"] & H^{\text{fail}}(n_1, n_2) \arrow[d] \\ \widetilde{H}(n_1, n_2) \arrow[r, "S_{n_2}"'] & H(n_1, n_2). \end{tikzcd} \tag{4.4}$$

We have expressed $H^{\text{fail}}(n_1, n_2)$ as the quotient of a $G = (\mathbb{Z}/2\mathbb{Z})^{n_2} \rtimes S_{n_2}$-torsor by the action of S_{n_2}. To apply the Chebotarev density theorem, we must now understand two quantities:

(1) the number of elements of G that are conjugate to an element of $S_{n_2} \subset G$, and
(2) the number of geometric components of $\widetilde{H}^{\text{fail}}(n_1, n_2)$ over each geometric component of $H(n_1, n_2)$.

We take up the first issue here and the second in the next section. To lighten notation, set $\ell = n_2$ so that $G = (\mathbb{Z}/2\mathbb{Z})^{\ell} \rtimes S_{\ell}$. Let $\Phi \subset G$ be the union $\Phi = \bigcup_{h \in G} h S_{\ell} h^{-1}$. Let $s(\ell, i)$ be the unsigned Sterling number of the first kind. This is the number of elements in S_{ℓ} whose disjoint cycle decomposition has exactly i cycles.

Proposition 4.1. *An element $(u, \sigma) \in G$ lies in Φ if and only if there exists $w \in (\mathbb{Z}/2\mathbb{Z})^{\ell}$ such that $w + \sigma(w) = u$. Consequently, the size of Φ is*

$$|\Phi| = \sum_{i=1}^{\ell} 2^{\ell - i} s(\ell, i).$$

Proof. Consider an arbitrary $(w, \tau) \in G$. We have

$$(w, \tau)(u, \sigma)(w, \tau)^{-1} = (w + \tau(u) + \tau\sigma\tau^{-1}(w), \tau\sigma\tau^{-1}).$$

This element lies in S_{ℓ} if and only if

$$w + \tau(u) + \tau\sigma\tau^{-1}(w) = 0.$$

Apply τ^{-1} and replace $\tau^{-1}(w)$ by w to get

$$w + \sigma(w) = u.$$

We now find the size of Φ. Let us first fix $\sigma \in S_{\ell}$ and find the number of u such that $(u, \sigma) \in \Phi$. We have proved that $(u, \sigma) \in \Phi$ if and only if $u \in (\mathbb{Z}/2\mathbb{Z})^{\ell}$ is in the image of the linear map $w \mapsto w + \sigma(w)$. Let $e_1, \dots, e_{\ell} \in (\mathbb{Z}/2\mathbb{Z})^{\ell}$ be the standard basis vectors. Given $T \subset \{1, \dots, \ell\}$, set $e_T = \sum_{t \in T} e_t$. It is easy to check that the kernel of the map $w \mapsto w + \sigma(w)$ is spanned by the vectors e_T as T ranges over the cycles in the disjoint cycle decomposition of σ. Thus, if σ has i cycles, then the kernel of $w \mapsto w + \sigma(w)$ has dimension i, and hence its image has dimension $\ell - i$. We conclude that the number of u such that $(u, \sigma) \in \Phi$ is $2^{\ell - i}$. We now take the sum as σ varies. ☐

Let us estimate the quantity in Proposition 4.1. For functions $f, g \colon \mathbb{Z}_{\geq 0} \to \mathbb{R}$, we write $f \sim g$ if $\lim_{n \to \infty}(f(n)/g(n))$ is a positive (finite) number.

Proposition 4.2. *Let $b_{\ell} = \sum_{i=1}^{\ell} 2^{-i} s(\ell, i)/\ell!$. Then $b_{\ell} < \ell^{-1/2}$ and $b_{\ell} \sim \ell^{-1/2}$.*

Proof. The unsigned Sterling numbers satisfy the identity
$$(1-t)^{-x} = \sum_{\ell=0}^{\infty}\sum_{i=1}^{\ell} \frac{s(\ell,i)}{\ell!} t^\ell x^i.$$
Substituting $x = 1/2$ gives
$$(1-t)^{-1/2} = \sum_{\ell=0}^{\infty}\sum_{i=1}^{\ell} \frac{s(\ell,i)}{\ell!} t^\ell 2^{-i}.$$
Taking the ℓ-th derivative of both sides and setting $t = 0$ gives
$$\frac{1}{2}\cdot\frac{3}{2}\cdots\frac{2\ell-1}{2} = \sum_{i=1}^{\ell} s(\ell,i)2^{-i}.$$
Dividing both sides by $\ell!$ yields
$$\left(1-\frac{1}{2}\right)\left(1-\frac{1}{4}\right)\cdots\left(1-\frac{1}{2\ell}\right) = b_\ell.$$
Since $1 - x < \exp(-x)$ for $x > 0$, we get
$$b_\ell < \exp\left(-\sum_{i=1}^{\ell}\frac{1}{2i}\right). \tag{4.5}$$
Furthermore, since
$$\log \ell < \sum_{i=1}^{\ell} 1/i, \tag{4.6}$$
the right-hand side of (4.5) is bounded above by
$$\exp\left(-\frac{1}{2}\log\ell\right) = \ell^{-1/2}.$$
It is standard that the ratio of the two sides of (4.5) and the difference of the two sides of (4.6) both approach constants as $\ell \to \infty$. It follows that $b_\ell \sim \ell^{-1/2}$. □

5. Connected components of the Hurwitz spaces

We now turn to the number of geometric components of the Hurwitz spaces and their fields of definition. We have the following comparison theorem (see also [26, Lemma 2.3.5] and [28, Lemma 10.3]).

Proposition 5.1. *Let H be any finite étale cover of $U(n_1, n_2)$, for example, any of the four spaces in* (4.4)*. Reduction to characteristic p gives a bijection between the geometric components of $H \otimes \mathbb{Q}$ and the geometric components of $H \otimes \mathbb{F}_p$.*

Proof. We have a morphism $\mathbb{P}^{n_1} \times \mathbb{P}^{n_2} \to \mathbb{P}^{n_1+n_2}$ that sends a pair of effective divisors (B_1, B_2) of degrees n_1 and n_2 to the divisor $B_1 + B_2$. Let $\Delta \subset \mathbb{P}^{n_1+n_2}$ be the discriminant divisor. Then $U(n_1, n_2) \to \mathbb{P}^{n_1+n_2} - \Delta$ is a finite étale cover. Thus, H is a finite étale cover of $\mathbb{P}^{n_1+n_2} - \Delta$. The result now follows from Fulton's reduction theorem [19, Theorem 3.3, Lemma A.3], following an argument similar to [19, Corollary 7.5]. □

We analyse the connected components of the four spaces in (4.4) over $\mathbb{C}$. We take a two-step approach. In the first step, we understand the components of $\widetilde{H}(n_1, n_2)$, $\widetilde{H}^{\text{fail}}(n_1, n_2)$, and $H^{\text{fail}}(n_1, n_2)$ lying over a given connected component of $H(n_1, n_2)$. In the second step, we understand the connected components of $H(n_1, n_2)$.

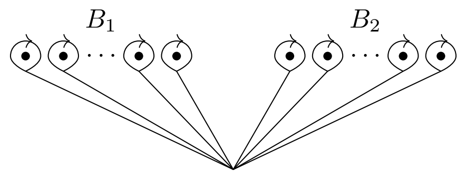


FIGURE 5.1. We describe a point of $H(n_1,n_2)_{\mathbb{C}}$ over a point of $U(n_1,n_2)_{\mathbb{C}}$ in terms of monodromy data consisting of elements of S_4 associated to each loop as shown. We describe a point of $H^{\text{fail}}(n_1,n_2)_{\mathbb{C}}$ similarly, using decorated monodromy data.

Choose the following base point $B=(B_1,B_2)\in U(n_1,n_2)_{\mathbb{C}}$: identify the $\mathbb{C}$-points of $\mathbb{P}^1-\{\infty\}$ with the points of the complex plane $\mathbb{C}$; our B_1 consists of the points $(-n_1+\mathbf{i}),\dots,(-1+\mathbf{i})$ and B_2 consists of the points $(1+\mathbf{i}),\dots,(n_2+\mathbf{i})$ (see Figure 5.1 for a rough sketch). An A_4-quartic cover of $\mathbb{P}^1_{\mathbb{C}}$ branched over (B_1,B_2) is defined by the monodromy map

$$\mu\colon \pi_1(\mathbb{P}^1_{\mathbb{C}}-(B_1\cup B_2),0)\to S_4.$$

We choose standard generators $(\alpha_1,\dots,\alpha_{n_1},\beta_1,\dots,\beta_{n_2})$ for $\pi_1(\mathbb{P}^1_{\mathbb{C}}-(B_1\cup B_2),0)$ as shown in Figure 5.1. The only relation they satisfy is

$$\alpha_1\cdots\alpha_{n_1}\cdot\beta_1\cdots\beta_{n_2}=1.$$

(Our convention is that the product above is represented by the loop that follows α_1, then α_2, and so on. Then the product is homotopic to a giant clockwise loop based at 0.) The monodromy map μ is equivalent to an (n_1+n_2)-tuple of permutations in S_4, say $(\sigma_1,\dots,\sigma_{n_1},\tau_1,\dots,\tau_{n_2})$ such that

$$\sigma_1\cdots\sigma_{n_1}\cdot\tau_1\cdots\tau_{n_2}=\mathrm{id}\,.$$

Since the cover has ramification of type $(3,1)$ over points of B_1, each σ_i is a 3-cycle, and similarly, each τ_j is a $(2,2)$-cycle (a product of two disjoint 2-cycles). Two such tuples $(\sigma_1,\dots,\sigma_{n_1},\tau_1,\dots,\tau_{n_2})$ and $(\sigma'_1,\dots,\sigma'_{n_1},\tau'_1,\dots,\tau'_{n_2})$ define isomorphic A_4-quartic covers of $\mathbb{P}^1_{\mathbb{C}}$ branched over B if and only if there exists $g\in S_4$ such that $g\sigma_ig^{-1}=\sigma'_i$ and $g\tau_jg^{-1}=\tau'_j$ for all $i\in\{1,\dots,n_1\}$ and all $j\in\{1,\dots,n_2\}$.

A point of $H^{\text{fail}}(n_1,n_2)$ over $B\in U(n_1,n_2)_{\mathbb{C}}$ consists of an A_4-quartic cover of $\mathbb{P}^1_{\mathbb{C}}$ branched over B described by the monodromy μ as above, together with a labelling by $\{\pm1\}$ of the two points of the quartic cover over every point of B_2. Consider the j-th point of B_2, around which the monodromy τ_j is the product of two disjoint 2-cycles $(pq)(rs)$. The two points of the A_4-quartic cover over this point correspond naturally to the 2-cycles (pq) and (rs). Labelling these two points is equivalent to labelling these 2-cycles. We indicate the labelling of the cycles by a decoration—we underline the 2-cycle labelled by $+1$. We use $\hat{\tau}_j$ to refer to a $(2,2)$-cycle τ_j together with the underlining of one of its 2-cycles.

The conjugation action of S_4 on the set of $(2,2)$-cycles lifts to an action on the set of decorated $(2,2)$-cycles as follows. We set

$$g\cdot\underline{(pq)}(rs)\cdot g^{-1}=\underline{g(pq)g^{-1}}\cdot g(rs)g^{-1}.$$

Then the points of $H^{\text{fail}}(n_1,n_2)$ over B are given by decorated monodromy data

$$\hat{\mu}=(\sigma_1,\dots,\sigma_{n_1},\hat{\tau}_1,\dots,\hat{\tau}_{n_2}),$$

modulo simultaneous conjugation by S_4. The map $H^{\text{fail}}(n_1,n_2)\to H(n_1,n_2)$ simply forgets the decoration.

The fundamental group $\pi_1(U(n_1,n_2)_{\mathbb{C}},B)$ acts on the $\mathbb{C}$-points of the fibres over B of the maps $H(n_1,n_2)_{\mathbb{C}}\to U(n_1,n_2)_{\mathbb{C}}$ and $H^{\text{fail}}(n_1,n_2)_{\mathbb{C}}\to U(n_1,n_2)_{\mathbb{C}}$. These actions are given by well-known Hurwitz moves; see for example [19, §1.5]. We recall some important Hurwitz moves for the convenience of the reader. An anti-clockwise half-twist that exchanges the ith and $(i+1)$th point

Space	Fibre over $B = (B_1, B_2)$
$H(n_1, n_2)$	Tuples $(\sigma_1, \dots, \sigma_{n_1}, \tau_1, \dots, \tau_{n_2})$ where $\sigma_i \in S_4$ is a 3-cycle and $\tau_j \in S_4$ is a $(2,2)$-cycle with $\sigma_1 \dots \sigma_{n_1} \cdot \tau_1 \dots \tau_{n_2} = 1$, up to conjugation by S_4.
$H^{\mathrm{fail}}(n_1, n_2)$	Tuples as above together with the underlining of a 2-cycle in each τ_j, denoted by $(\sigma_1, \dots, \sigma_{n_1}, \hat{\tau}_1, \dots, \hat{\tau}_{n_2})$, up to conjugation by S_4.
$\widetilde{H}(n_1, n_2)$	Tuples $(\sigma_1, \dots, \sigma_{n_1}, \tau_1, \dots, \tau_{n_2})$ as above together with a numbering of the points of B_2, up to conjugation by S_4.
$\widetilde{H}^{\mathrm{fail}}(n_1, n_2)$	Tuples $(\sigma_1, \dots, \sigma_{n_1}, \hat{\tau}_1, \dots, \hat{\tau}_{n_2})$ as above together with a numbering of the points of B_2, up to conjugation by S_4.

TABLE 1. A description of the fibres of the Hurwitz spaces over the fixed base point $B \in U(n_1, n_2)_{\mathbb{C}}$.

of B_1 defines a closed loop in $U(n_1, n_2)_{\mathbb{C}}$ based at B. The action of this loop on the points of $H^{\mathrm{fail}}(n_1, n_2)_{\mathbb{C}}$ over B is given in terms of decorated monodromy by

$$(\sigma_1, \dots, \sigma_{n_1}, \hat{\tau}_1, \dots, \hat{\tau}_{n_2}) \mapsto (\sigma_1, \dots, \sigma_{i-1}, \sigma_i \sigma_{i+1} \sigma_i^{-1}, \sigma_i, \dots, \sigma_{n_1}, \hat{\tau}_1, \dots, \hat{\tau}_{n_2}). \tag{5.1}$$

Similarly, an anti-clockwise half-twist that exchanges the jth and $(j+1)$th points of B_2 acts by

$$(\sigma_1, \dots, \sigma_{n_1}, \hat{\tau}_1, \dots, \hat{\tau}_{n_2}) \mapsto (\sigma_1, \dots, \sigma_{n_1}, \hat{\tau}_1, \dots, \hat{\tau}_{j-1}, \tau_j \hat{\tau}_{j+1} \tau_j^{-1}, \hat{\tau}_j, \dots, \hat{\tau}_{n_2}). \tag{5.2}$$

An anti-clockwise half-twist exchanging the last point of B_1 and the first point of B_2 is not a closed loop in $U(n_1, n_2)_{\mathbb{C}}$. However, the anti-clockwise full twist is a closed loop; this full twist acts by

$$(\sigma_1, \dots, \sigma_{n_1}, \hat{\tau}_1, \dots, \hat{\tau}_{n_2}) \mapsto (\sigma_1, \dots, \sigma_{n_1 - 1}, (\sigma_{n_1} \tau_1)^2 \sigma_{n_1}^{-1}, \sigma_{n_1} \hat{\tau}_1 \sigma_{n_1}^{-1}, \hat{\tau}_2, \dots, \hat{\tau}_{n_2}). \tag{5.3}$$

The action on the fibres of $H(n_1, n_2)_{\mathbb{C}} \to U(n_1, n_2)_{\mathbb{C}}$ is given by the same formulas, but without the decoration.

A point of $\widetilde{H}(n_1, n_2)_{\mathbb{C}}$ over $B \in U(n_1, n_2)_{\mathbb{C}}$ consists of a point of $H(n_1, n_2)_{\mathbb{C}}$ together with a numbering of the points of B_2. Likewise, a point of $\widetilde{H}^{\mathrm{fail}}(n_1, n_2)_{\mathbb{C}}$ over $B \in U(n_1, n_2)_{\mathbb{C}}$ consists of a point of $H^{\mathrm{fail}}(n_1, n_2)_{\mathbb{C}}$ together with a numbering of the points of B_2. The action of $\pi_1(U(n_1, n_2)_{\mathbb{C}}, B)$ on the numbering is via the natural homomorphism $\pi_1(U(n_1, n_2)_{\mathbb{C}}, B) \to S_{n_2}$. We summarise the discussion above in Table 1.

We begin with a useful lemma.

LEMMA 5.2. *Fix the basepoint $B \in U(n_1, n_2)_{\mathbb{C}}$ as above. Let $H \subset H(n_1, n_2)_{\mathbb{C}}$ be a connected component. There exists a point of H over B whose monodromy $(\sigma_1, \dots, \sigma_{n_1}, \tau_1, \dots, \tau_{n_2})$ has $\tau_i = \tau_j$ for all $i, j \in \{1, \dots, n_2\}$.*

Proof. Start with any point of H over B given by a monodromy $\mu = (\sigma_1, \dots, \sigma_{n_1}, \tau_1, \dots, \tau_{n_2})$. We modify it so that all τ_i's become equal.

Consider the anti-clockwise full twist in $U(n_1, n_2)_{\mathbb{C}}$ in the last point of B_1 and the first point of B_2. Acting by this loop on μ yields a μ' that differs from μ only at the positions of σ_{n_1} and τ_1. Using (5.3), we compute the new σ_{n_1} and τ_1 after zero, one, and two full twists, starting with $\sigma_{n_1} = (123)$ and $\tau_1 = (12)(34)$ (our convention for multiplying permutations is that $p \cdot q = p \circ q$).

	σ_{n_1}	τ_1
Initial	(123)	(12)(34)
After one twist	(243)	(14)(23)
After two twists	(142)	(13)(24)

We see that by zero, one, or two full twists, we can change the $(2,2)$-cycle τ_1 to any $(2,2)$-cycle.

Since the $(2,2)$-cycles commute with each other, a half-twist exchanging two adjacent points of B_2 simply exchanges the monodromy elements.

So, if we have $\tau_1 = \tau_2 = \cdots = \tau_j \neq \tau_{j+1}$, we do a sequence of half-twists to bring the $(j+1)$-th point to the first position; do one or more full twists around the last point of B_1 to make the monodromy around this point equal to τ_j, and do the reverse sequence of half-twists to bring it back to the $(j+1)$th position. By repeating this procedure, we make $\tau_1 = \cdots = \tau_{n_2}$. $\square$

Proposition 5.3. *Let $H \subset H(n_1,n_2)_{\mathbb{C}}$ be a connected component. The pre-image of H in $\widetilde{H}(n_1,n_2)_{\mathbb{C}}$ is connected.*

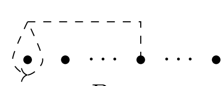
Proof. Recall that the points of $\widetilde{H}(n_1,n_2)_{\mathbb{C}}$ are the points of $H(n_1,n_2)_{\mathbb{C}}$ together with a numbering of the points of B_2. Choose a point $f \in H$ over the fixed basepoint B as above, whose monodromy is $(\sigma_1,\dots,\sigma_{n_1},\tau_1,\dots,\tau_{n_2})$ with $\tau_1 = \cdots = \tau_{n_2}$. This is possible by Lemma 5.2. The pre-image of f in $\widetilde{H}(n_1,n_2)_{\mathbb{C}}$ consists of copies of f equipped with different numberings of the n_2 points of B_2. We will show that the action of $\pi_1(H,f)$ on the points of the fibre of $\widetilde{H}(n_1,n_2)_{\mathbb{C}}$ over f is transitive. By the path lifting property for covering spaces, this will imply path-connectedness, and hence connectedness, of the pre-image of H.

Consider the half-twist that interchanges two adjacent points of B_2. Since the monodromy around all points of B_2 is the same, this half-twist does not change the monodromy, and thus lifts to a loop in H. The action of this loop on the pre-image of f simply exchanges the numbers associated with the two interchanged points of B_2. By a sequence of pairwise interchanges, we may change a given numbering of the points of B_2 to any other numbering. $\square$

Proposition 5.4. *Let $H \subset H(n_1,n_2)_{\mathbb{C}}$ be a connected component. The pre-image of H in $\widetilde{H}^{\mathrm{fail}}(n_1,n_2)_{\mathbb{C}}$ has at most two connected components. Furthermore, the action of S_{n_2} on this preimage preserves the connected components.*

Proof. Let $\widetilde{H}$ be the pre-image of H in $\widetilde{H}(n_1,n_2)_{\mathbb{C}}$. By Proposition 5.3, $\widetilde{H}$ is connected. Choose a point $\widetilde{f} \in \widetilde{H}$ given by the monodromy $(\sigma_1,\dots,\sigma_{n_1},\tau_1,\dots,\tau_{n_2})$ with the left-to-right numbering of the points of B_2. The fibre of $\widetilde{f}$ in $\widetilde{H}^{\mathrm{fail}}(n_1,n_2)_{\mathbb{C}}$ consists of the 2^{n_2} possible decorations of $\tau_1,\dots,\tau_{n_2}$. We show that the action of $\pi_1(\widetilde{H},\widetilde{f})$ on this set has at most two orbits.

By an argument similar to the proof of Lemma 5.2, we may assume that $\tau_2 = \cdots = \tau_{n_2}$ and $\tau_1 \neq \tau_2$. Without loss of generality, take $\tau_1 = (12)(34)$ and $\tau_2 = \cdots = \tau_{n_2} = (13)(24)$. Consider a decoration $(\hat{\tau}_1,\dots,\hat{\tau}_{n_2})$ of $(\tau_1,\dots,\tau_{n_2})$. Given $j \in \{2,\dots,n_2\}$, we consider the loop in $U(n_1,n_2)_{\mathbb{C}}$ that moves point j up, then left towards point 1 of B_2, takes it in a full twist around point 1, and then takes it back to its original position along the same path (see below):



B_2

This loop has no effect on the monodromy or on the numbering of B_2. So it lifts to a closed loop in $\widetilde{H}$ based at $\widetilde{f}$. Let us analyse its effect on the decorated monodromy. It is easy to see that the only possible change is in $\hat{\tau}_1$ and $\hat{\tau}_j$. In particular, we see that $\hat{\tau}_j$ changes to the conjugate $\tau_1\hat{\tau}_j\tau_1^{-1}$. We have

$$(12)(34)\cdot(13)\underline{(24)}\cdot(12)(34) = \underline{(13)}(24);$$

that is, the conjugate $\tau_1\hat{\tau}_j\tau_1^{-1}$ is the same $(2,2)$-cycle as $\hat{\tau}_j$ but with the other underlining. By a sequence of such loops, we can ensure that the decorated monodromy around the points $2,\dots,n_2$ is $\underline{(13)}(24)$. We conclude that under the action of $\pi_1(\widetilde{H},\widetilde{f})$, any decoration $(\hat{\tau}_1,\dots,\hat{\tau}_{n_2})$ lies in the orbit of

$$(\underline{(12)}(34), \underline{(13)}(24), \dots, \underline{(13)}(24)) \text{ or } ((12)\underline{(34)}, \underline{(13)}(24), \dots, \underline{(13)}(24)).$$

It follows that the pre-image of $\widetilde{H}$ has at most two components.

We now prove that S_{n_2} preserves the connected components. Let $f \in H(n_1, n_2)_{\mathbb{C}}$. It suffices to prove that for all $\hat{f}$ in $\widetilde{H}^{\text{fail}}(n_1, n_2)_{\mathbb{C}}$ over f and $p \in S_{n_2}$, the points $\hat{f}$ and $p \cdot \hat{f}$ are in the same connected component of $\widetilde{H}^{\text{fail}}(n_1, n_2)_{\mathbb{C}}$. Since $\widetilde{H}^{\text{fail}}(n_1, n_2)_{\mathbb{C}}$ has at most two connected components over the connected component of $H(n_1, n_2)_{\mathbb{C}}$ containing f, it suffices to prove this for any one $\hat{f}$. We choose f branched over $B_1 \cup B_2$ with the monodromy $(\sigma_1, \dots, \sigma_{n_1}, \tau_1, \dots, \tau_{n_2})$ satisfying $\tau_i = \tau_j$ for all i, j (see Lemma 5.2). We then choose $\hat{f}$ to have the left-to-right numbering of the points of B_2 and the same 2-cycle underlined in all the τ_i. Let $p \in S_{n_2}$ be the transposition $(j, j+1)$. A half-twist in the j-th and $(j+1)$-th points of B_2 lifts to a path in $\widetilde{H}^{\text{fail}}(n_1, n_2)_{\mathbb{C}}$ connecting $\hat{f}$ and $p \cdot \hat{f}$. We conclude that $\hat{f}$ and $p \cdot \hat{f}$ lie in the same connected component. Since the transpositions $(j, j+1)$ generate S_{n_2}, the same follows for any $p \in S_{n_2}$. □

Proposition 5.5. *Let $H \subset H(n_1, n_2)_{\mathbb{C}}$ be a connected component. The map $\widetilde{H}^{\text{fail}}(n_1, n_2)_{\mathbb{C}} \to H^{\text{fail}}(n_1, n_2)_{\mathbb{C}}$ is a bijection on the connected components over H. In particular, $H^{\text{fail}}(n_1, n_2)_{\mathbb{C}}$ also has at most two connected components over H.*

Proof. Recall that $\widetilde{H}^{\text{fail}}(n_1, n_2)_{\mathbb{C}} \to H^{\text{fail}}(n_1, n_2)_{\mathbb{C}}$ is the quotient by S_{n_2}. The statement follows from Proposition 5.4. □

Having discussed the connected components of $\widetilde{H}(n_1, n_2)$, $\widetilde{H}^{\text{fail}}(n_1, n_2)$ and $H^{\text{fail}}(n_1, n_2)$ over a connected component of $H(n_1, n_2)$, we now address the connected components of $H(n_1, n_2)$. Let $v(n_1, n_2)$ be the number of connected components of $H(n_1, n_2)_{\mathbb{C}}$. Set $n = n_1 + n_2$ and let $U \subset \mathbb{P}^n$ be the complement of the discriminant divisor. Recall that we have a finite étale morphism $H(n_1, n_2) \to U$. Consider the point $B \in U(n_1, n_2)_{\mathbb{C}}$ also as a point of $U_{\mathbb{C}}$. Let $V(n_1, n_2)$ be the set

$$V(n_1, n_2) = \left\{ (g_1, \dots, g_n) \;\middle|\; \begin{array}{l} g_i \in S_4, \quad g_1 \cdots g_n = 1, \\ n_1 \text{ of the } g_i \text{ have cycle type } (3,1), \\ n_2 \text{ of the } g_i \text{ have cycle type } (2,2). \end{array} \right\}. \tag{5.4}$$

The group S_4 acts on $V(n_1, n_2)$ by simultaneous conjugation

$$h\colon (g_1, \dots, g_n) \mapsto (hg_1h^{-1}, \dots, hg_nh^{-1}).$$

The points of $H(n_1, n_2)_{\mathbb{C}}$ over $B \in U_{\mathbb{C}}$ are in natural bijection with $V(n_1, n_2)$ modulo the conjugation action of S_4.

The fundamental group $\pi_1(U_{\mathbb{C}}, B)$ is the braid group $\mathbb{B}_n$. The ith generator s_i represents the anti-clockwise half-twist that interchanges the ith and $(i+1)$th points of B. The group $\mathbb{B}_n$ acts on $V(n_1, n_2)$ by

$$s_i\colon (g_1, \dots, g_n) \mapsto (g_1, \dots, g_{i-1}, g_i g_{i+1} g_i^{-1}, g_i, g_{i+2}, \dots, g_n).$$

The connected components of $H(n_1, n_2)_{\mathbb{C}}$ are in natural bijection with the orbits of the $\mathbb{B}_n \times S_4$ action on $V(n_1, n_2)$ [18, Section 1.3]. In particular, $v(n_1, n_2)$ is the number of orbits. This is bounded above by the number of $\mathbb{B}_n$-orbits. We analyse the latter using the lifting invariant and the Conway–Parker theorem, which we first recall from [51].

Let G be a finite group and let $c \subset G$ be a subset that generates G, does not contain the identity, and is closed under conjugation by G. Let $U(G, c)$ be the group with generators $[g]$ for every $g \in c$ and relations

$$[x][y][x]^{-1} = [xyx^{-1}],$$

for every $x, y \in c$. We have a homomorphism $U(G, c) \to G$ that sends the generator $[g]$ to g. Let D be the set of conjugacy classes in c. We have a homomorphism $U(G, c) \to \mathbb{Z}^D$ that sends the generator $[g]$ to the conjugacy class of g. We also have a homomorphism $\mathbb{Z}^D \to G^{\mathrm{ab}}$ that sends the conjugacy class of g to the image of g in G^{ab}. The homomorphisms above combine to give a homomorphism

$$U(G,c) \to G \times_{G^{\mathrm{ab}}} \mathbb{Z}^D.$$

Given a positive integer n, set

$$V_n^G = \{(g_1,\dots,g_n) \mid g_i \in c,\ g_1 \dots g_n = 1 \text{ and } g_1,\dots,g_n \text{ generate } G\}.$$

We have a map

$$\Pi\colon V_n^G \to \ker(U(G,c) \to G)$$

that sends $(g_1,\dots,g_n)$ to $[g_1]\cdots[g_n]$. This map is constant on $\mathbb{B}_n$ orbits and hence descends to a map $V_n^G/\mathbb{B}_n \to \ker(U(G,c) \to G)$. By composition, we have a map $V_n^G/\mathbb{B}_n \to \mathbb{Z}^D$. Let $V_{\underline{m}} \subset \bigcup_{n\geq 0} V_n^G/\mathbb{B}_n$ and $U(G,c)_{\underline{m}} \subset U(G,c)$ be the pre-images of $\underline{m} \in \mathbb{Z}^D$.

Theorem 5.6 (Conway–Parker). *There exists a constant $M \in \mathbb{Z}_{>0}$ such that if all entries of $\underline{m} \in \mathbb{Z}^D$ are at least M, the map Π induces a bijection*

$$V_{\underline{m}} \to U(G,c)_{\underline{m}} \cap \ker(U(G,c) \to G).$$

Proof. See [28, Theorem 12.5], [51, Theorem 3.1] and its proof. □

We take $G = A_4$ and $c = A_4 - \{\mathrm{id}\}$. Then c is the union of 3 conjugacy classes, represented by (123), (132), and (12)(34). The commutator subgroup of G is the Klein 4-group $\{\mathrm{id}, (12)(34), (13)(24), (14)(23)\}$ and $G^{\mathrm{ab}} \cong C_3$. Choosing the image of (123) as a generator fixes an isomorphism $G^{\mathrm{ab}} \to C_3 = \mathbb{Z}/3\mathbb{Z}$. Then the map $\mathbb{Z}^D = \mathbb{Z}^3 \to C_3$ sends the basis vector e_1 to 1, e_2 to -1, and e_3 to 0. Write $U(A_4) = U(A_4, c)$.

Proposition 5.7. *The natural homomorphism $U(A_4) \to A_4 \times_{C_3} \mathbb{Z}^3$ is an isomorphism.*

Proof. In [51, Theorem 2.5], Wood shows that $U(A_4)$ is isomorphic to $S_c \times_{C_3} \mathbb{Z}^3$, where S_c is the reduced Schur cover defined before [51, Lemma 2.2]. As in the proof of [14, Proposition 8.11], viewing A_4 as $\mathrm{PSL}_2(\mathbb{F}_3)$, one sees that a Schur cover for A_4 is $\mathrm{SL}_2(\mathbb{F}_3)$ and deduces that $S_c = A_4$. The proof of [51, Theorem 2.5] shows that the natural homomorphism $U(A_4) \to A_4 \times_{C_3} \mathbb{Z}^3$ is an isomorphism. □

Observe that a 3-cycle and a $(2,2)$-cycle are enough to generate A_4. So, as long as $n_1, n_2 > 0$, for every $(g_1,\dots,g_n) \in V(n_1,n_2)$, the g_i generate A_4. That is, we have the inclusion $V(n_1,n_2) \subset V_n^{A_4}$. Given non-negative integers a, b with $a+b = n_1$, let $V(a,b,n_2) \subset V(n_1,n_2)$ be the pre-image of $(a,b,n_2) \in \mathbb{Z}^3$ under the composition $V_n^{A_4} \xrightarrow{\Pi} U(A_4) \twoheadrightarrow \mathbb{Z}^3$. Then $V(n_1,n_2)$ is the disjoint union

$$V(n_1,n_2) = \bigsqcup_{a+b=n_1} V(a,b,n_2).$$

The $\mathbb{B}_n$-action on $V(n_1,n_2)$ preserves $V(a,b,n_2)$. The S_4-action preserves the union

$$V(\{a,b\},n_2) = V(a,b,n_2) \cup V(b,a,n_2). \tag{5.5}$$

The even permutations preserve the two sets individually, but the odd permutations interchange them.

Proposition 5.8. *(1) There exists $\lambda \in \mathbb{Z}_{>0}$ such that for all $a,b,n_2,n \in \mathbb{Z}_{\geq 0}$ with $a+b+n_2 = n$,*

$$|V(a,b,n_2)/\mathbb{B}_n| \leq \lambda.$$

(2) There exists $M \in \mathbb{Z}_{>0}$ such that for all $a, b, n_2, n \in \mathbb{Z}_{\geq 0}$ with $a, b, n_2 \geq M$ and $a+b+n_2 = n$,

$$|V(a, b, n_2)/\mathbb{B}_n| = \begin{cases} 1 & \text{if } a \equiv b \pmod 3 \\ 0 & \text{otherwise.} \end{cases}$$

Proof. The first assertion follows from [15, Lemma 3.3]. We prove the second assertion. Let $n_1 = a + b$. Recall that, as long as $n_1, n_2 > 0$, we have $V(a, b, n_2) \subset V(n_1, n_2) \subset V_n^{A_4}$. For $a, b, n_2 \geq M$, the Conway–Parker theorem (Theorem 5.6) gives a bijection

$$V(a, b, n_2)/B_n \to U(A_4)_{(a,b,n_2)} \cap \ker(U(A_4) \to A_4).$$

The isomorphism $U(A_4) \cong A_4 \times_{C_3} \mathbb{Z}^3$ from Proposition 5.7 gives

$$\ker(U(A_4) \to A_4) = \{(a, b, c) \in \mathbb{Z}^3 \mid a \equiv b \pmod 3\}.$$

The statement follows. □

Proposition 5.8 implies that certain geometric components of $H(n_1, n_2) \otimes \mathbb{F}_p$ are defined over $\mathbb{F}_p$, as we now explain. Recall that $\mathbb{B}_n \times S_4$ orbits of $V(n_1, n_2)$ correspond to connected components of $H(n_1, n_2)_{\mathbb{C}}$. Given non-negative integers $a \leq b$, the subset $V(\{a, b\}, n_2) \subset V(n_1, n_2)$ is a union of $\mathbb{B}_n \times S_4$ orbits and therefore corresponds to a union of connected components of $H(n_1, n_2)_{\mathbb{C}}$.

Proposition 5.9. *The union of the connected components of $H(n_1, n_2)_{\mathbb{C}}$ corresponding to $V(\{a, b\}, n_2)$ is defined over $\mathbb{Q}$. In particular, if $V(\{a, b\}, n_2)$ forms one orbit under $\mathbb{B}_n \times S_4$, then the unique connected component of $H(n_1, n_2)_{\mathbb{C}}$ corresponding to this orbit is defined over $\mathbb{Q}$.*

Proof. We may replace $\mathbb{C}$ by $\overline{\mathbb{Q}}$. Let $H \subset H(n_1, n_2)_{\overline{\mathbb{Q}}}$ be the union of connected components corresponding to $V(\{a, b\}, n_2)$. Let us prove that H is defined over $\mathbb{Q}$. Consider a $\overline{\mathbb{Q}}$-point of H corresponding to the isomorphism class of an A_4-quartic cover $(C, B_1, B_2, C \xrightarrow{f} \mathbb{P}^1_{\overline{\mathbb{Q}}})$. Choose a point $0 \in \mathbb{P}^1_{\overline{\mathbb{Q}}}$ away from $B_1 \cup B_2$; fix a numbering of $f^{-1}(0)$; and let

$$\mu\colon \pi_1(\mathbb{P}^1_{\mathbb{C}} - (B_1 \cup B_2), 0) \to S_4$$

be the monodromy of f. Recall that the image of μ is $A_4 \subset S_4$.

For each $x \in B_1$, let $D_x \subset \mathbb{P}^1_{\mathbb{C}}$ be a small disc centred at x, and $0_x \in D_x$ a point other than x. Let $\gamma_x \in \pi_1(\mathbb{P}^1_{\mathbb{C}} - (B_1 \cup B_2), 0_x)$ be the element represented by an anti-clockwise loop in D_x around x based at 0_x. Recall that a path from 0 to 0_x in $\mathbb{P}^1_{\mathbb{C}} - (B_1 \cup B_2)$ gives an isomorphism

$$\pi_1(\mathbb{P}^1_{\mathbb{C}} - (B_1 \cup B_2), 0_x) \xrightarrow{\sim} \pi_1(\mathbb{P}^1_{\mathbb{C}} - (B_1 \cup B_2), 0);$$

two isomorphisms given by different paths are conjugate to each other. Thus, the conjugacy class of the image of γ_x in $\pi_1(\mathbb{P}^1_{\mathbb{C}} - (B_1 \cup B_2), 0)$ is well defined independently of the path. Denote it by $[\gamma_x] \subset \pi_1(\mathbb{P}^1_{\mathbb{C}} - (B_1 \cup B_2), 0)$. Then $[\mu(\gamma_x)] \subset A_4$ is a well-defined conjugacy class of A_4. Let $M(f)$ be the multiset

$$M(f) = \{[\mu(\gamma_x)] \mid x \in B_1\}.$$

Changing the numbering of $f^{-1}(0)$ changes the multiset by simultaneous conjugation by an element of S_4. The $f \in H(n_1, n_2)_{\mathbb{C}}$ lying in H are characterised by the condition that, up to simultaneous conjugation by S_4, we have

$$M(f) = \{a \cdot [(123)], b \cdot [(132)]\}.$$

Take $\varphi \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ and let $\chi \in \varprojlim(\mathbb{Z}/m\mathbb{Z})^\times$ be its cyclotomic character. Then, by [50, § 4.3.1], we have

$$M(f^\varphi) = \{[\mu(\gamma_x)]^\chi \mid x \in B_1\}.$$

We see that, up to conjugation by S_4, the multiset $M(f^\varphi)$ is also $\{a \cdot [(123)], b \cdot [(132)]\}$. So f^φ also lies in H. Thus, H is invariant under $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ and hence is defined over $\mathbb{Q}$. □

Proposition 5.9 enables us to define $H(\{a,b\},n_2) \subset H(n_1,n_2)$ as the open and closed subspace such that $H(\{a,b\},n_2)_{\mathbb{C}}$ is the union of the connected components of $H(n_1,n_2)_{\mathbb{C}}$ corresponding to $V(\{a,b\},n_2)$.

We now have all the tools to apply the Chebotarev density theorem. Recall from Proposition 4.2 that $b_\ell = \sum_{i=1}^{\ell} 2^{-i} s(\ell,i)/\ell!$.

Theorem 5.10. *Let $n_1 > 0$ and $n_2 \geq 0$ be integers and $p \geq 5$ a prime. Let q be a power of p such that all geometric components of $\widetilde{H}^{\mathrm{fail}}(n_1,n_2) \otimes \mathbb{F}_p$ are defined over $\mathbb{F}_q$. Let $\phi\colon H^{\mathrm{fail}}(n_1,n_2) \to H(n_1,n_2)$ be the natural forgetful map. Let $H \subset H(n_1,n_2) \otimes \mathbb{F}_q$ be a (geometric) component and let H^{fail} be the pre-image of H in $H^{\mathrm{fail}}(n_1,n_2) \otimes \mathbb{F}_q$. Then,*

$$\lim_{m\to\infty} \frac{|\phi(H^{\mathrm{fail}}(\mathbb{F}_{q^m}))|}{|H(\mathbb{F}_{q^m})|} = r b_{n_2},$$

where $r \leq 2$ is the number of geometric components of $\widetilde{H}^{\mathrm{fail}}(n_1,n_2) \otimes \mathbb{F}_q$ lying above H. Consequently,

$$\lim_{m\to\infty} \frac{|\phi(H^{\mathrm{fail}}(n_1,n_2)(\mathbb{F}_{q^m}))|}{|H(n_1,n_2)(\mathbb{F}_{q^m})|} \leq 2b_{n_2}.$$

Proof. Let $\widetilde{H}^{\mathrm{fail}}$ be the pre-image of H in $\widetilde{H}^{\mathrm{fail}}(n_1,n_2) \otimes \mathbb{F}_q$. Let $G = (\mathbb{Z}/2\mathbb{Z})^{n_2} \rtimes S_{n_2}$. We know from (4.4) that the map $\widetilde{H}^{\mathrm{fail}} \to H$ is a G-torsor and $\widetilde{H}^{\mathrm{fail}} \to H^{\mathrm{fail}}$ is the quotient by $S_{n_2} \subset G$. By Proposition 4.1, the union of the conjugates of $S_{n_2} \subset G$ has size $b_{n_2} \cdot |G|$.

We will apply the Chebotarev density theorem (Proposition 2.5) with $G = (\mathbb{Z}/2\mathbb{Z})^{n_2} \rtimes S_{n_2}$ and $\Phi = \bigcup_{h\in G} hS_{n_2}h^{-1}$. First, we show that Φ is contained in the stabiliser G_0 of a geometric component of $\widetilde{H}^{\mathrm{fail}}$. By Propositions 5.1 and 5.4, $[G : G_0] \leq 2$ and $S_{n_2} \subset G_0$. Therefore, G_0 is a normal subgroup of G containing S_{n_2}, which implies that $\Phi \subset G_0$, as claimed.

Now, by Proposition 2.5, we get

$$\lim_{m\to\infty} \frac{|\phi(H^{\mathrm{fail}}(\mathbb{F}_{q^m}))|}{|H(\mathbb{F}_{q^m})|} = r b_{n_2},$$

with $r \leq 2$, thanks to Propositions 5.1 and 5.4. Summing over all geometric components of $H(n_1,n_2) \otimes \mathbb{F}_q$ yields the second statement. □

Remark 5.11. Let k be a finite field of characteristic $\neq 2,3$. By Theorem 3.7, the image of the k-points of $H^{\mathrm{fail}}(n_1,n_2)$ are precisely the isomorphism classes of A_4-quartic covers of $\mathbb{P}^1_k$ that fail the Hasse norm principle. Theorem 5.10 says that, asymptotically in m, the proportion of isomorphism classes of A_4-quartic covers of $\mathbb{P}^1_{\mathbb{F}_{q^m}}$ that fail the Hasse norm principle is bounded above by $2b_{n_2}$. By Proposition 4.2, b_{n_2} is within a constant factor of $n_2^{-1/2}$.

Corollary 5.12. *In the set-up of Theorem 5.10, we have*

$$\lim_{n_2\to\infty} \lim_{m\to\infty} \frac{|\phi(H^{\mathrm{fail}}(n_1,n_2)(\mathbb{F}_{q^m}))|}{|H(n_1,n_2)(\mathbb{F}_{q^m})|} = 0.$$

Proof. Use $b_{n_2} \sim n_2^{-1/2}$ from Proposition 4.2. □

Recall that $H(\{a,b\},n_2) \subset H(n_1,n_2)$ is the open and closed subspace corresponding to the set $V(\{a,b\},n_2)$ (see Proposition 5.9). Let $H^{\mathrm{fail}}(\{a,b\},n_2) \subset H^{\mathrm{fail}}(n_1,n_2)$ be the pre-image of $H(\{a,b\},n_2)$. Under the one-orbit hypothesis of Proposition 5.13 below, we can take $q = p$ in Theorem 5.10. By Proposition 5.8, the one-orbit hypothesis holds if a, b and n_2 are sufficiently large.

Proposition 5.13. *Suppose a, b, n_2 are such that $V(\{a,b\}, n_2)$ forms one orbit under $\mathbb{B}_n \times S_4$. Then $H(\{a,b\}, n_2) \otimes \mathbb{F}_p$ is geometrically connected and*

$$\limsup_{m\to\infty} \frac{|\phi(H^{\mathrm{fail}}(\{a,b\},n_2)(\mathbb{F}_{p^m}))|}{|H(\{a,b\},n_2)(\mathbb{F}_{p^m})|} = \lim_{m\to\infty} \frac{|\phi(H^{\mathrm{fail}}(\{a,b\},n_2)(\mathbb{F}_{p^{2m}}))|}{|H(\{a,b\},n_2)(\mathbb{F}_{p^{2m}})|} = rb_{n_2},$$

where $r \leq 2$ is the number of geometric components of $\widetilde{H}^{\mathrm{fail}}(n_1, n_2) \otimes \mathbb{F}_p$ lying above $H(\{a,b\}, n_2) \otimes \mathbb{F}_p$.

Proof. The first statement follows from Propositions 5.1 and 5.9. By Proposition 5.4, $\widetilde{H}^{\mathrm{fail}}(n_1, n_2) \otimes \mathbb{F}_p$ has at most two geometric components over $H(\{a,b\}, n_2) \otimes \mathbb{F}_p$. If these geometric components are defined over $\mathbb{F}_p$ then we apply the Chebotarev density theorem as in the proof of Theorem 5.10. Now suppose they are not defined over $\mathbb{F}_p$. That is, there are two geometric components and they are interchanged by Frob_p. By Proposition 5.5, $H^{\mathrm{fail}}(\{a,b\}, n_2)$ also has two geometric components and they are interchanged by Frob_p. Then for any odd m, we have $H^{\mathrm{fail}}(\{a,b\}, n_2)(\mathbb{F}_{p^m}) = 0$, so the quantity in the limit is zero. For even m, we apply the Chebotarev density theorem as in the proof of Theorem 5.10. □

6. Proof of Theorem 1.1

When combined with Theorem 3.7 and Proposition 4.2, Theorem 5.10 and Corollary 5.12 yield analogues of Theorem 1.1 and Corollary 1.3 but with the bound in Theorem 1.1 replaced by $2n_2^{-\frac{1}{2}}$ and the limit $n \to \infty$ in Corollary 1.3 replaced by $n_2 \to \infty$. Our final task is to go from n_2 to $n = n_1 + n_2$ and thus prove our main results: Theorem 1.1 and Corollary 1.3.

So, without further ado, let n, n_1, and n_2 be non-negative integers such that $n = n_1 + n_2$ and $n_1 > 0$. Let $H(n)$ be the disjoint union

$$H(n) = \bigsqcup_{\substack{n_1+n_2=n \\ n_1>0, n_2\geq 0}} H(n_1, n_2). \tag{6.1}$$

Define $H^{\mathrm{fail}}(n)$ analogously. Our goal in this section is to combine the bounds in Section 5 to obtain a bound on

$$\lim_{m\to\infty} \frac{|\phi(H^{\mathrm{fail}}(n)(\mathbb{F}_{p^m}))|}{|H(n)(\mathbb{F}_{p^m})|}.$$

Recall that $v(n_1, n_2)$ is the number of connected components of $H(n_1, n_2)_{\mathbb{C}}$. Equivalently, it is the number of $\mathbb{B}_n \times S_4$-orbits of the set $V(n_1, n_2)$ defined in (5.4). Given non-negative integers a, b with $a + b = n_1$, let $v(\{a,b\}, n_2)$ be the number of $\mathbb{B}_n \times S_4$-orbits of the set $V(\{a,b\}, n_2)$ defined in (5.5). We estimate $\sum v(\{a,b\}, n_2)$. For two functions $f, g\colon \mathbb{Z}_{\geq 0} \to \mathbb{R}$, we write $f \approx g$ if

$$\lim_{n\to\infty} f(n)/g(n) = 1.$$

Lemma 6.1. *Fix an integer m. We have*

$$\sum_{\substack{a+b+n_2=n \\ a,b,n_2\geq m,\ a\leq b}} v(\{a,b\}, n_2) \approx n^2/12.$$

Furthermore, for $0 < \alpha < 1$, we have

$$\sum_{\substack{a+b+n_2=n \\ a,b,n_2\geq m,\ a\leq b \\ n_2<n^{\alpha}}} v(\{a,b\}, n_2) \approx n^{\alpha+1}/6.$$

Proof. From Proposition 5.8 it follows that there is a constant λ such that

$$v(\{a,b\},n_2) \leq \lambda, \tag{6.2}$$

and there is a constant M such that if $a,b,n_2 \geq M$, then

$$v(\{a,b\},n_2) = \begin{cases} 1 & \text{if } a \equiv b \pmod 3 \\ 0 & \text{otherwise.} \end{cases} \tag{6.3}$$

We can thus break up the sum into two parts: the sum of the $v(\{a,b\},n_2)$ where at least one of a,b,n_2 is less than M and the sum of the $v(\{a,b\},n_2)$ where all of a,b,n_2 are at least M. Since $a+b+n_2=n$, we see that the number of terms in the first part is a linear function in n, and by (6.2), their sum is bounded by a linear function in n. Therefore, it suffices to show that the second part has the stated asymptotics. The second part is the sum

$$\sum_{\substack{a+b+n_2=n \\ a,b,n_2 \geq M,\ a \leq b}} v(\{a,b\},n_2). \tag{6.4}$$

By (6.3) this sum is the cardinality of the set

$$\{(a,b,n_2) \in \mathbb{Z}^3 \mid a \equiv b \pmod 3,\ a+b+n_2=n,\ a \leq b, \text{ and } a,b,n_2 \geq M\}.$$

It is easy to check that this cardinality is $\approx n^2/12$.

For the second assertion, we replace (6.4) by

$$\sum_{\substack{a+b+n_2=n \\ a,b,n_2 \geq M,\ a \leq b \\ n_2 < n^\alpha}} v(\{a,b\},n_2). \tag{6.5}$$

By (6.3) this sum is the cardinality of the set

$$\{(a,b,n_2) \in \mathbb{Z}^3 \mid a \equiv b \pmod 3,\ a+b+n_2=n,\ a \leq b,\ n_2 < n^\alpha \text{ and } a,b,n_2 \geq M\}.$$

Again, it is easy to check that this cardinality is $\approx n^{\alpha+1}/6$. $\square$

We now have the tools to prove the main result.

Theorem 6.2. *Let $p \geq 5$ be a prime. Given $\epsilon > 0$, if n is sufficiently large, then*

$$\limsup_{m\to\infty} \frac{|\phi(H^{\text{fail}}(n)(\mathbb{F}_{p^m}))|}{|H(n)(\mathbb{F}_{p^m})|} \leq (4+\epsilon)n^{-1/3}.$$

Proof. All the summations that follow are over a,b,n_2 satisfying $a+b+n_2=n$, $n_2 \geq 0$ and $a+b>0$ and $a \leq b$. Under the summation signs, we drop these conditions and only write additional conditions. Fix α to be any real number in $(0,1)$. Let M be an integer such that $v(\{a,b\},n_2)=1$ for $a,b,n_2 \geq M$ (see Proposition 5.8). Using (6.1) and

$$H^{\text{fail}}(n) = \bigsqcup_{\substack{a+b+n_2=n \\ n_2 \geq 0 \\ a+b>0, a \leq b}} H^{\text{fail}}(\{a,b\},n_2),$$

when $n^\alpha \geq M$, we have

$$
\begin{aligned}
|\phi(H^{\mathrm{fail}}(n)(\mathbb{F}_{p^m}))| = &\sum_{\substack{a,b\geq M\\ n_2\geq n^\alpha}} |\phi(H^{\mathrm{fail}}(\{a,b\},n_2)(\mathbb{F}_{p^m}))| \\
&+ \sum_{\substack{a,b,n_2\geq M\\ n_2< n^\alpha}} |\phi(H^{\mathrm{fail}}(\{a,b\},n_2)(\mathbb{F}_{p^m}))| \\
&+ \sum_{a,b,\text{ or } n_2<M} |\phi(H^{\mathrm{fail}}(\{a,b\},n_2)(\mathbb{F}_{p^m}))|.
\end{aligned} \tag{6.6}
$$

We consider the three parts of the sum separately, in sequence.

Let us consider the first part. If $n_2 \geq n^\alpha$ and $a, b \geq M$, Proposition 5.13 and Proposition 4.2 imply

$$\limsup_{m\to\infty} \frac{|\phi(H^{\mathrm{fail}}(\{a,b\},n_2)(\mathbb{F}_{p^m}))|}{|H(\{a,b\},n_2)(\mathbb{F}_{p^m})|} < 2n^{-\alpha/2}.$$

Therefore,

$$\limsup_{m\to\infty} \frac{\sum_{\substack{a,b\geq M\\ n_2\geq n^\alpha}} |\phi(H^{\mathrm{fail}}(\{a,b\},n_2)(\mathbb{F}_{p^m}))|}{|H(n)(\mathbb{F}_{p^m})|} < 2n^{-\alpha/2}. \tag{6.7}$$

Now let us consider the second and third parts. We have the bound

$$|\phi(H^{\mathrm{fail}}(\{a,b\},n_2)(\mathbb{F}_{p^m}))| \leq |H^{\mathrm{fail}}(\{a,b\},n_2)(\mathbb{F}_{p^m})|.$$

By the Lang–Weil bounds, we have

$$\limsup_{m\to\infty} \frac{|H^{\mathrm{fail}}(\{a,b\},n_2)(\mathbb{F}_{p^m})|}{p^{mn}} \leq v(\{a,b\},n_2).$$

Therefore, for n sufficiently large, we have

$$
\begin{aligned}
&\limsup_{m\to\infty} \frac{1}{p^{mn}} \left(\sum_{\substack{a,b,n_2\geq M\\ n_2< n^\alpha}} |\phi(H^{\mathrm{fail}}(\{a,b\},n_2)(\mathbb{F}_{p^m}))| + \sum_{a,b,\text{ or } n_2<M} |\phi(H^{\mathrm{fail}}(\{a,b\},n_2)(\mathbb{F}_{p^m}))| \right) \\
&\leq \sum_{\substack{a,b,n_2\geq M\\ n_2< n^\alpha}} v(\{a,b\},n_2) + \sum_{a,b,\text{ or } n_2<M} v(\{a,b\},n_2) \approx n^{\alpha+1}/6.
\end{aligned} \tag{6.8}
$$

The last approximation follows from Lemma 6.1 applied to the first sum. By Proposition 5.8, the second sum is much smaller—bounded by a multiple of n.

By Propositions 5.8 and 5.13, the components $H(\{a,b\},n_2)\otimes\mathbb{F}_p$ of $H(n)\otimes\mathbb{F}_p$ for $a,b,n_2\geq M$ are geometrically connected, and again by the Lang–Weil bounds, we have

$$\liminf_{m\to\infty} \frac{1}{p^{mn}} |H(n)(\mathbb{F}_{p^m})| \geq \sum_{a,b,n_2\geq M} v(\{a,b\},n_2) \approx n^2/12. \tag{6.9}$$

The last approximation is again thanks to Lemma 6.1.

By combining the inequalities (6.7), (6.8), and (6.9), we get

(6.10)

$$
\begin{aligned}
\limsup_{m\to\infty} \frac{|\phi(H^{\mathrm{fail}}(n)(\mathbb{F}_{p^m}))|}{|H(n)(\mathbb{F}_{p^m})|} &< 2n^{-\alpha/2} + \frac{\sum_{\substack{a,b,n_2\geq M\\ n_2<n^\alpha}} v(\{a,b\},n_2) + \sum_{a,b,\text{ or } n_2<M} v(\{a,b\},n_2)}{\sum_{a,b,n_2\geq M} v(\{a,b\},n_2)} \\
&\approx 2n^{-\alpha/2} + 2n^{\alpha-1}.
\end{aligned}
$$

Taking $\alpha = 2/3$ yields the result. □

Remark 6.3. An inspection of the proof of Theorem 6.2 shows that determining an explicit bound beyond which n is considered "sufficiently large" would involve determining the constants λ and M from Proposition 5.8 and then obtaining more precise estimates for the sums in Lemma 6.1. The proof of [15, Lemma 3.3] shows that one can take $\lambda = |A_4|^{|A_4|^2} = 12^{12^2}$. It is less clear how to make M explicit.

Proof of Theorem 1.1. Combine Theorem 3.7, (6.1) and Theorem 6.2. □

References

[1] *Revêtements étales et groupe fondamental (SGA 1)*, volume 3 of *Documents Mathématiques (Paris)*. Société Mathématique de France, Paris, 2003. Séminaire de géométrie algébrique du Bois Marie 1960–61, Directed by Alexandre Grothendieck, With two papers by Michèle Raynaud, Updated and annotated reprint of the 1971 original [Lecture Notes in Math., 224, Springer, Berlin; MR0354651 (50 #7129)].

[2] Hans-Jochen Bartels. Zur Arithmetik von Diedergruppenerweiterungen. *Mathematische Annalen*, 256:465–474, 1981.

[3] Hans-Jochen Bartels. Zur Arithmetik von Konjugationsklassen in algebraischen Gruppen. *Journal of Algebra*, 70(1):179–199, 1981.

[4] Manjul Bhargava. Mass formulae for extensions of local fields, and conjectures on the density of number field discriminants. *Int. Math. Res. Not.*, 2007(17):20, 2007. Id/No rnm052.

[5] Manjul Bhargava, Arul Shankar, Takashi Taniguchi, Frank Thorne, Jacob Tsimerman, and Yongqiang Zhao. Bounds on 2-torsion in class groups of number fields and integral points on elliptic curves. *J. Amer. Math. Soc.*, 33(4):1087–1099, 2020.

[6] Manjul Bhargava and Melanie Matchett Wood. The density of discriminants of S_3-sextic number fields. *Proc. Amer. Math. Soc.*, 136(5):1581–1587, 2008.

[7] Kenneth S. Brown. *Cohomology of groups*, volume 87 of *Graduate Texts in Mathematics*. Springer-Verlag, New York-Berlin, 1982.

[8] Tim D. Browning and Rachel Newton. The proportion of failures of the Hasse norm principle. *Mathematika*, 62(2):337–347, 2016.

[9] John W. S. Cassels and Albrecht Fröhlich, editors. *Algebraic number theory*. Academic Press, London; Thompson Book Co., Inc., Washington, DC, 1967.

[10] Jean-Louis Colliot-Thélène. Lectures on linear algebraic groups, beijing lectures, morning side centre, april 2007. Available from `https://www.math.u-psud.fr/colliot/BeijingLectures2Juin07.pdf`.

[11] Jean-Louis Colliot-Thélène, David Harari, and Alexei N. Skorobogatov. Compactification équivariante d'un tore (d'après Brylinski et Künnemann). *Expo. Math.*, 23(2):161–170, 2005.

[12] Boris Datskovsky and David J. Wright. Density of discriminants of cubic extensions. *J. Reine Angew. Math.*, 386:116–138, 1988.

[13] Harold Davenport and Hans Heilbronn. On the density of discriminants of cubic fields. II. *Proc. Roy. Soc. London Ser. A*, 322(1551):405–420, 1971.

[14] Jordan S. Ellenberg, TriThang Tran, and Craig Westerland. Fox–Neuwirth–Fuks cells, quantum shuffle algebras, and Malle's conjecture for function fields. Preprint, arXiv:1701.04541v2.

[15] Jordan S. Ellenberg and Akshay Venkatesh. Counting extensions of function fields with bounded discriminant and specified Galois group. In *Geometric methods in algebra and number theory*, pages 151–168. Basel: Birkhäuser, 2005.

[16] Christopher Frei, Daniel Loughran, and Rachel Newton. The Hasse norm principle for abelian extensions. *Amer. J. Math.*, 140(6):1639–1685, 2018. Corrigendum in: *Amer. J. Math.*, 148(4):1199-1227, 2026.

[17] Christopher Frei, Daniel Loughran, and Rachel Newton. Number fields with prescribed norms. *Comment. Math. Helv.*, 97(1):133–181, 2022. With an appendix by Yonatan Harpaz and Olivier Wittenberg.

[18] Michael D. Fried and Helmut Völklein. The inverse Galois problem and rational points on moduli spaces. *Math. Ann.*, 290(4):771–800, 1991.

[19] William Fulton. Hurwitz schemes and irreducibility of moduli of algebraic curves. *Ann. Math. (2)*, 90:542–575, 1969.

[20] Alexandre Grothendieck. Techniques de construction et théoremes d'existence en géométrie algébrique. IV: Les schemas de Hilbert. Sem. Bourbaki 13(1960/61), No. 221, 28 p. (1961)., 1961.

[21] Helmut Hasse. Beweis eines Satzes und Wiederlegung einer Vermutung über das allgemeine Normenrestsymbol. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1931:64–69, 1931.

[22] Jürgen Klüners. A counterexample to Malle's conjecture on the asymptotics of discriminants. *C. R. Math. Acad. Sci. Paris*, 340(6):411–414, 2005.

[23] Peter Koymans and Nick Rome. A note on the Hasse norm principle. *Bull. Lond. Math. Soc.*, 56(3):1004–1013, 2024.

[24] Peter Koymans and Nick Rome. Weak approximation on the norm one torus. *Compos. Math.*, 160(6):1304–1348, 2024.

[25] Boris È. Kunyavskiĭ. Arithmetic properties of three-dimensional algebraic tori. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)*, 116:102–107, 163, 1982. Integral lattices and finite linear groups.

[26] Aaron Landesman and Ishan Levy. Homological stability for Hurwitz spaces and applications. Preprint, arXiv:2503.03861.

[27] Serge Lang. Sur les séries L d'une variété algébrique. *Bull. Soc. Math. Fr.*, 84:385–407, 1956.

[28] Yuan Liu, Melanie Matchett Wood, and David Zureick-Brown. A predicted distribution for Galois groups of maximal unramified extensions. *Invent. Math.*, 237(1):49–116, 2024.

[29] Daniel Loughran and Ross Paterson. Lower bounds for counting A_4-quartic fields. Preprint, arXiv:2510.05248.

[30] André Macedo. The Hasse norm principle for A_n-extensions. *Journal of Number Theory*, 211:500–512, 2020.

[31] André Macedo. *Local-global principles for norms*. PhD thesis, University of Reading, 2021. Available from `https://drive.google.com/file/d/1gcXSbm5GXFZsYT2hU6tKPSsFUuopv_4m/view`.

[32] André Macedo and Rachel Newton. Explicit methods for the Hasse norm principle and applications to A_n and S_n extensions. *Mathematical Proceedings of the Cambridge Philosophical Society*, 172(3):489–529, 2022.

[33] Gunter Malle. On the distribution of Galois groups. *J. Number Theory*, 92(2):315–329, 2002.

[34] Gunter Malle. On the distribution of Galois groups. II. *Experiment. Math.*, 13(2):129–135, 2004.

[35] Adelina Mânzăţeanu, Rachel Newton, Ekin Ozman, Nicole Sutherland, and Rabia Gülşah Uysal. The Hasse Norm Principle in Global Function Fields. In *Women in Numbers Europe III: Research Directions in Number Theory*, pages 275–290. Springer International Publishing, Cham, 2021.

[36] Steve Meagher. A simple proof of Chebotarev's density theorem over finite fields. *Bull. Aust. Math. Soc.*, 98(2):196–202, 2018.

[37] Jürgen Neukirch. *Algebraic number theory*, volume 322 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, Berlin, 1999. Translated from the 1992 German original.

[38] Rachel Newton and Ila Varma. Counting S_4 and S_5 extensions satisfying the Hasse norm principle. Preprint, forthcoming.

[39] Yasuhiro Oki, Harmeet Singh, and Gareth Tracey. The Hasse norm principle for extensions of degree pq. Preprint, forthcoming.

[40] Nick Rome. The Hasse norm principle for biquadratic extensions. *J. Théor. Nombres Bordeaux*, 30(3):947–964, 2018.

[41] Tim Santens. The leading constant in Malle's conjecture over function fields. Preprint, arXiv:2512.12838v2.

[42] Arul Shankar and Frank Thorne. On the asymptotics of cubic fields ordered by general invariants. *Comment. Math. Helv.*, 99(4):769–797, 2024.

[43] Harmeet Singh. *The Hasse norm principle: proportions of successes and degrees of failure*. PhD thesis, King's College London, 2026.

[44] Seyfi Türkelli. Connected components of Hurwitz schemes and Malle's conjecture. *Journal of Number Theory*, 155:163–201, 2015.

[45] Valentin E. Voskresenskiĭ. Birational properties of linear algebraic groups. *Math. USSR-Izv*, 4:1–17, 1970.

[46] Valentin E. Voskresenskiĭ. Maximal tori without effect in semisimple algebraic groups. *Mathematical notes of the Academy of Sciences of the USSR*, 44:651–655, 1988.

[47] Valentin E. Voskresenskiĭ. *Algebraic groups and their birational invariants*, volume 179 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 1998. Translated from the Russian manuscript by Boris Kunyavski [Boris È. Kunyavskiĭ].

[48] Valentin E. Voskresenskiĭ and Boris E. Kunyavskii. Maximal tori in semisimple algebraic groups. Deposited in VINITI March 5, 1984, No. 1269-84 Dep. (Ref. Zh. Mat. (1984), 7A405 Dep.).

[49] Jiuya Wang. Counterexamples for Türkelli's modification on Malle's conjecture. Preprint, arXiv:2502.04261.

[50] Stefan Wewers. *Construction of Hurwitz spaces*. PhD thesis, Essen: Univ.-GHS Essen, Fachbereich 6 Mathematik und Informatik, 1998.

[51] Melanie Matchett Wood. An algebraic lifting invariant of Ellenberg, Venkatesh, and Westerland. *Research in the Mathematical Sciences*, 8(21), 2021.