Algebra & Number Theory

Volume 4
2010
No. 6

mathematical sciences publishers

Algebra & Number Theory

www.jant.org

EDITORS

MANAGING EDITOR

Bjorn Poonen
Massachusetts Institute of Technology
Cambridge, USA

EDITORIAL BOARD CHAIR

David Eisenbud
University of California
Berkeley, USA

BOARD OF EDITORS

Georgia Benkart	University of Wisconsin, Madison, USA
Dave Benson	University of Aberdeen, Scotland
Richard E. Borcherds	University of California, Berkeley, USA
John H. Coates	University of Cambridge, UK
J-L. Colliot-Thélène	CNRS, Université Paris-Sud, France
Brian D. Conrad	University of Michigan, USA
Hélène Esnault	Universität Duisburg-Essen, Germany
Hubert Flenner	Ruhr-Universität, Germany
Edward Frenkel	University of California, Berkeley, USA
Andrew Granville	Université de Montréal, Canada
Joseph Gubeladze	San Francisco State University, USA
Ehud Hrushovski	Hebrew University, Israel
Craig Huneke	University of Kansas, USA
Mikhail Kapranov	Yale University, USA
Yujiro Kawamata	University of Tokyo, Japan
János Kollár	Princeton University, USA
Hendrik W. Lenstra	Universiteit Leiden, The Netherlands
Yuri Manin	Northwestern University, USA
Barry Mazur	Harvard University, USA
Susan Montgomery	University of Southern California, USA
Shigefumi Mori	RIMS, Kyoto University, Japan
Andrei Okounkov	Princeton University, USA
Raman Parimala	Emory University, USA
Victor Reiner	University of Minnesota, USA
Karl Rubin	University of California, Irvine, USA
Peter Sarnak	Princeton University, USA
Michael Singer	North Carolina State University, USA
Ronald Solomon	Ohio State University, USA
Vasudevan Srinivas	Tata Inst. of Fund. Research, India
J. Toby Stafford	University of Michigan, USA
Bernd Sturmfels	University of California, Berkeley, USA
Richard Taylor	Harvard University, USA
Ravi Vakil	Stanford University, USA
Michel van den Bergh	Hasselt University, Belgium
Marie-France Vignéras	Université Paris VII, France
Kei-Ichi Watanabe	Nihon University, Japan
Andrei Zelevinsky	Northeastern University, USA
Efim Zelmanov	University of California, San Diego, USA

PRODUCTION

ant@mathscipub.org

Silvio Levy, Scientific Editor

Andrew Levy, Production Editor

See inside back cover or www.jant.org for submission instructions.

The subscription price for 2010 is US $140/year for the electronic version, and $200/year (+$30 shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to Mathematical Sciences Publishers, Department of Mathematics, University of California, Berkeley, CA 94720-3840, USA.

Algebra & Number Theory (ISSN 1937-0652) at Mathematical Sciences Publishers, Department of Mathematics, University of California, Berkeley, CA 94720-3840 is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

ANT peer review and production are managed by EditFLOW™ from Mathematical Sciences Publishers.

PUBLISHED BY

mathematical sciences publishers

http://www.mathscipub.org

A NON-PROFIT CORPORATION

Typeset in LaTeX

Copyright ©2010 by Mathematical Sciences Publishers

ALGEBRA AND NUMBER THEORY 4:6(2010)

Generalized moonshine I: Genus-zero functions

Scott Carnahan

We introduce a notion of Hecke-monicity for functions on certain moduli spaces associated to torsors of finite groups over elliptic curves, and show that it implies strong invariance properties under linear fractional transformations. Specifically, if a weakly Hecke-monic function has algebraic integer coefficients and a pole at infinity, then it is either a holomorphic genus-zero function invariant under a congruence group or of a certain degenerate type. As a special case, we prove the same conclusion for replicable functions of finite order, which were introduced by Conway and Norton in the context of monstrous moonshine. As an application, we introduce a class of Lie algebras with group actions, and show that the characters derived from them are weakly Hecke-monic. When the Lie algebras come from chiral conformal field theory in a certain sense, then the characters form holomorphic genus-zero functions invariant under a congruence group.

Introduction 649
1. Equivariant Hecke operators 653
2. Hecke-monicity 659
3. Modular equations 661
4. Finite level 665
5. Replicability 669
6. twisted denominator formulas 673
Acknowledgments 677
References 677

Introduction

We define a holomorphic genus-zero function to be a holomorphic function $f : \mathfrak{H} \to \mathbb{C}$ on the complex upper half-plane, with finite-order poles at cusps, such that there exists a discrete group $\Gamma_f \subset \mathrm{SL}_2(\mathbb{R})$ for which f is invariant under the action of Γ_f by Möbius transformations, inducing a dominant injection $\mathfrak{H}/\Gamma_f \to \mathbb{C}$.

Keywords: moonshine, replicable function, Hecke operator, generalized moonshine.
This material is partly based upon work supported by the National Science Foundation under grant DMS-0354321.

A holomorphic genus-zero function f therefore generates the field of meromorphic functions on the quotient of $\mathfrak{H}$ by its invariance group. In this paper, we are interested primarily in holomorphic congruence genus-zero functions, especially those f for which $\Gamma(N) \subset \Gamma_f$ for some $N > 0$. These functions are often called Hauptmoduln.

The theory of holomorphic genus-zero modular functions began with Jacobi's work on elliptic and modular functions in the early 1800's, but did not receive much attention until the 1970's, when Conway and Norton found numerical relationships between the Fourier coefficients of a distinguished class of these functions and the representation theory of the largest sporadic finite simple group $\mathbb{M}$, called the monster. Using their own computations together with work of Thompson and McKay, they formulated the monstrous moonshine conjecture, which asserts the existence of a graded representation $V^\natural = \bigoplus_{n\geq -1} V_n$ of $\mathbb{M}$ such that for each $g \in \mathbb{M}$, the graded character $T_g(\tau) := \sum_{n\geq -1} \mathrm{Tr}(g|V_n)q^n$ is a normalized holomorphic genus-zero function invariant under some congruence group $\Gamma_0(N)$, where the normalization indicates a q-expansion of the form $q^{-1} + O(q)$. More precisely, they gave a list of holomorphic genus-zero functions f_g as candidates for T_g, whose first several coefficients arise from characters of the monster, and whose invariance groups Γ_{f_g} contain some $\Gamma_0(N)$ [Conway and Norton 1979]. By unpublished work of Koike, the power series expansions of f_g satisfy a condition known as complete replicability, given by a family of recurrence relations, and the relations determine the full expansion of f_g from only the first seven coefficients of f_{g^n} for n ranging over powers of two.

Borcherds [1992] proved this conjecture using a combination of techniques from the theory of vertex algebras and infinite-dimensional Lie algebras: $V^\natural$ was constructed by Frenkel, Lepowsky, and Meurman [1988] as a vertex operator algebra, and Borcherds used it to construct the monster Lie algebra, which inherits an action of the monster. Since the monster Lie algebra is a generalized Kac–Moody algebra with a homogeneous action of $\mathbb{M}$, it admits twisted denominator formulas, which relate the coefficients of T_g to characters of powers of g acting on the root spaces. In particular, each T_g is completely replicable, and Borcherds completed the proof by checking that the first seven coefficients matched the expected values.

Knowing this theorem and some additional data, one can ask at least two natural questions:

(1) The explicit checking of coefficients at the end of the proof has been called a "conceptual gap" in [Cummins and Gannon 1997], and this problem has been rectified in some sense by replacing that step with noncomputational theorems:

- Borcherds [1992] pointed out that the twisted denominator formulas imply that the functions T_g are completely replicable.

- Kozlov [1994] showed that completely replicable functions satisfy lots of modular equations.
- Cummins and Gannon [1997] showed that power series satisfying enough modular equations are either holomorphic genus-zero and invariant under $\Gamma_0(N)$, or of a particular degenerate type resembling trigonometric functions.
- One can eliminate the degenerate types, either by appealing to a result of Martin [1996] asserting that completely replicable series that are "J-final" (a condition that holds for all T_g, since $T_1 = J$) are invariant under $\Gamma_0(N)$ for some N, or by using a result of Dong, Li, and Mason [2000] that restricts the form of the q-expansions at other cusps.

Since modular functions live on moduli spaces of structured elliptic curves, one might ask how these recursion relations and replicability relate to group actions and moduli of elliptic curves.

(2) One might wonder if similar behavior applies to groups other than the monster. Conway and Norton [1979] suggested that other sporadic groups may exhibit properties resembling moonshine, and [Queen 1981] produced strong computational evidence for this. Norton [Mason 1987, Appendix] organized this data into the generalized moonshine conjecture, which asserts the existence of a generalized character Z that associates a holomorphic function on $\mathfrak{H}$ to each commuting pair of elements of the monster, satisfying the following conditions:

- $Z(g, h, \tau)$ is invariant under simultaneous conjugation of g and h.
- For any $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$, there exists a nonzero constant γ (said to be a twenty-fourth root of unity in [Norton 2001]) such that
$$Z(g^a h^c, g^b h^d, \tau) = \gamma Z\Big(g, h, \frac{a\tau + b}{c\tau + d}\Big).$$
- The coefficients of the q-expansion of $Z(g, h, \tau)$ for fixed g form characters of a graded representation of a central extension of $C_{\mathbb{M}}(g)$.
- $Z(g, h, \tau)$ is either constant or holomorphic congruence genus-zero.
- $Z(g, h, \tau) = j(\tau) - 744 = q^{-1} + 196884q + 21493760q^2 + \cdots$ if and only if $g = h = 1$.

This conjecture is still open, but if we fix $g = 1$, it reduces to the original moonshine conjecture. One might hope that techniques similar to those used in [Borcherds 1992] can be applied to attack this conjecture in other cases, and the answer seems to be affirmative. For example, Höhn [2003] has proved it for the case when g is an involution in conjugacy class 2A, using a construction of a vertex algebra with baby monster symmetry, and roughly following the outline of Borcherds' proof. However, there are obstructions to making this technique work in general,

since there are many elements of the monster for which we do not know character tables of centralizers or their central extensions. One might ask whether there is a reasonably uniform way of generating holomorphic congruence genus-zero functions from actions of groups on certain Lie algebras.

This paper is an attempt to unify the two questions, and set the stage for a more detailed study of the infinite-dimensional algebraic structures involved. The main result is that modular functions (and more generally, singular q-expansions with algebraic integer coefficients) that are holomorphic on $\mathfrak{H}$ and satisfy a certain Hecke-theoretic property are holomorphic congruence genus-zero or degenerate in a specified way. As a special case, we find that finite-order replicable functions with algebraic integer coefficients, as defined in Section 4, satisfy the same property. The algebraic integer condition is sufficient for our purposes, since we intend to use this theorem in the context of representations of finite groups. Since modular functions with algebraic coefficients that are holomorphic on $\mathfrak{H}$ and invariant under a congruence group have bounded denominators (see [Shimura 1971, Theorem 3.52] and divide by a suitable power of Δ), it is reasonable to conjecture that all holomorphic congruence genus-zero functions whose poles have integral residue and constant term have algebraic integer coefficients.

We apply the theory to show that when a group acts on an infinite-dimensional Lie algebra with a special form, the character functions are holomorphic congruence genus-zero. We call these algebras *Fricke compatible* because they have the form we expect from elements $g \in \mathbb{M}$ for which the function T_g is invariant under a Fricke involution $\tau \mapsto -1/N\tau$. Later papers in this series will focus on constructing these and other (*non-Fricke compatible*) Lie algebras, first by generators-and-relations, and then by applying a version of the no-ghost theorem to abelian intertwiner algebras. At the time of writing, this strategy does not seem to yield a complete proof of generalized moonshine, because of some subtleties in computing eigenvalue multiplicities for certain cyclic groups of composite order acting on certain irreducible twisted modules of $V^\natural$. It is possible that some straightforward method of controlling these multiplicities has escaped our attention, but for the near future we plan to rest the full result on some precisely stated assumptions.

Most of the general ideas in the proof are not new, but our specific implementation bears meaningful differences from the existing literature. In fact, Hecke operators have been related to genus-zero questions since the beginning of moonshine, under the guise of replicability, and the question of relating replication to holomorphic genus-zero modular functions was proposed in the original paper [Conway and Norton 1979]. However, the idea of using an interpretation via moduli of elliptic curves with torsors is relatively recent, and arrives from algebraic topology. Equivariant Hecke operators, or more generally, isogenies of (formal) groups, can be used to describe operations on complex-oriented cohomology theories like elliptic

cohomology, and they were introduced in various forms in [Ando 1995; Baker 1998]. More precise connections to generalized moonshine were established in [Ganter 2009].

Summary. In Section 1, we introduce Hecke operators, first as operators on modular functions, and then on general power series. In Section 2, we define Hecke-monicity and prove elementary properties of Hecke-monic functions. In Section 3, we relate Hecke-monicity to equivariant modular equations. Most of this step is a minor modification of part of Kozlov's master's thesis [1994]. In Section 4, we prove a holomorphic congruence genus-zero theorem, and our proof borrows heavily from [Cummins and Gannon 1997]. Most of the arguments require minimal alteration from the form given in that paper, so in those cases we simply indicate which changes need to be made. In Section 5, we focus on the special case of replicable functions, and we show that those with finite order and algebraic integer coefficients are holomorphic congruence genus-zero or of a specific degenerate type. In Section 6, we conclude with an application to groups acting on Lie algebras, and show that under certain conditions arising from conformal field theory, the characters from the action on homology yield holomorphic congruence genus-zero functions.

1. Equivariant Hecke operators

The aims of this section are to introduce a combinatorial formula for equivariant Hecke operators for functions that are not necessarily modular, and to prove some elementary properties. The geometric language of stacks and torsors is only used in this section, and only to justify the claim that these Hecke operators occur naturally. It is not strictly necessary for understanding the formula, and the reader may skip everything in this section except for the statements of the lemmata without missing substantial constituents of the main theorem.

Let G be a finite group, and let $\mathcal{M}_{\mathrm{Ell}}^{G}$ denote the analytic stack of elliptic curves equipped with G-torsors (also known as the Hom stack $\underline{\mathrm{Hom}}(\mathcal{M}_{\mathrm{Ell}}, BG)$). Objects in the fibered category are diagrams

$$P \to E \overset{e}{\leftrightarrows} S$$

of complex analytic spaces satisfying:

- $P \to E$ is a G-torsor (that is, an analytically locally trivial principal G-bundle).
- $E \to S$ is a smooth proper morphism, whose fibers are genus-one curves.
- e is a section of $E \to S$.

Morphisms are fibered diagrams satisfying the condition that the torsor maps are G-equivariant. This is a smooth Deligne–Mumford stack (in the sense of [Behrend

and Noohi 2006]). For each positive integer n, we consider the degree-n G-Hecke correspondence, given by the diagram

$$\mathcal{I}_n^G \underset{t}{\overset{s}{\rightrightarrows}} \mathcal{M}_{\mathrm{Ell}}^G$$

defined as follows: $\mathcal{I}_n^G$ is the stack of n-isogenies of elliptic curves with G-torsors. Its objects are diagrams

$$\begin{tikzcd}
 & & P_2 \arrow[d, "\pi"] \\
E_1 \arrow[rr, "f"] \arrow[dr] & & E_2 \arrow[dl] \\
 & S \arrow[ul, bend left, "e_1"] \arrow[ur, bend right, "e_2"'] &
\end{tikzcd}$$

of complex analytic spaces, where:

- π is a G-torsor.
- $E_1 \to S$ and $E_2 \to S$ are smooth proper morphisms, whose geometric fibers are genus-one curves.
- e_1 and e_2 are sections of the corresponding maps.
- f is an n-isogeny, that is, a homomorphism whose kernel is a finite flat T-group scheme of length n (in particular, f makes the evident triangle diagrams commute).

As before, morphisms are fibered diagrams satisfying the condition that the torsor maps are G-equivariant. The two canonical maps $s, t : \mathcal{I}_n^G \to \mathcal{M}_{\mathrm{Ell}}^G$ are defined by $s(E_1, E_2, P_2, S) = (E_1 \times_{E_2} P_2 \to E_1 \leftrightarrows S)$ and $t(E_1, E_2, P_2, S) = (P_2 \to E_2 \leftrightarrows S)$ for objects, and the evident diagrams are given for morphisms. One can show that s and t are finite étale morphisms of degree $\psi(n) = \prod_{p \mid n}(1 + 1/p)$, essentially by transferring the arguments of [Katz and Mazur 1985, Proposition 6.5.1] to the analytic setting.

The Hecke operator nT_n is defined as the canonical trace map $s_* t^*$ on the structure sheaf of $\mathcal{M}_{\mathrm{Ell}}^G$. Over each point, it satisfies the formula

$$nT_n(f)(P \xrightarrow{G} E) = \sum_{\substack{0 \to H \to E' \xrightarrow{\pi} E \to 0 \\ |H| = n}} f(\pi^* P \xrightarrow{G} E'),$$

where the sum is over all degree-n isogenies to E. When G is trivial, this is the usual weight-zero Hecke operator.

We wish to describe these operators in terms of functions on the complex upper half-plane, and this requires an analytic uniformization of the moduli problem.

Following the unpublished book [Conrad $\geq$ 2010] and [Deligne 1971], the upper half-plane classifies pairs (π, ψ), where $\pi : E \to S$ is an elliptic curve, and

$$\psi : R^1\pi_*\underline{\mathbb{Z}} \to \underline{\mathbb{Z}}^2$$

is an isomorphism whose exterior square induces the (negative of the) canonical isomorphism $R^2\pi_*\underline{\mathbb{Z}(1)} \to \underline{\mathbb{Z}}$. By dualizing, one has a universal diagram $\mathbb{Z}^2 \times \mathfrak{H} \to \mathbb{C} \times \mathfrak{H} \to \mathfrak{H}$ defining a family of elliptic curves $E \to \mathfrak{H}$ equipped with oriented bases of fiberwise H_1. There is an $\mathrm{SL}_2(\mathbb{Z})$-action from the left via Möbius transformations (equivalently, changing the oriented homology basis of a curve), that induces a surjection onto $\mathcal{M}_{\mathrm{Ell}}$. When we consider G-torsors over elliptic curves equipped with homology bases, we find that they are classified up to isomorphism by their monodromy along the basis, given by a conjugacy class of a pair of commuting elements in G. Since a pair of commuting elements is a homomorphism from $\mathbb{Z} \times \mathbb{Z} \to G$, there is also an action of $\mathrm{SL}_2(\mathbb{Z})$ on commuting pairs of elements from the right via $(g, h)\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = (g^a h^c, g^b h^d)$. We obtain an identification on the level of points:

$$\mathcal{M}_{\mathrm{Ell}}^G \cong \mathrm{Hom}(\mathbb{Z} \times \mathbb{Z}, G)/G \underset{\mathrm{SL}_2(\mathbb{Z})}{\times} \mathfrak{H},$$

where the quotient by G arises from the action by conjugation on the target. The identification can be promoted to an equivalence of analytic stacks by choosing a uniformizing moduli problem of triples $(P \to E \to S, \psi, \tilde{e} : S \to P)$, where ψ is as above, and $\tilde{e}$ is a lift of $e : S \to E$ to the G-torsor. It is represented by a disjoint union of upper half-planes in bijection with $\mathrm{Hom}(\mathbb{Z} \times \mathbb{Z}, G)$, and one obtains the quotient via commuting actions of G (on the set of lifts $\tilde{e}$) and $\mathrm{SL}_2(\mathbb{Z})$ (on the set of ψ).

With this presentation, we can recast the Hecke operators in terms of holomorphic functions on the complex upper half-plane $\mathfrak{H}$. We can write any $f : \mathcal{M}_{\mathrm{Ell}}^G \to \mathbb{C}$ as $f(g, h, \tau)$, for g and h commuting elements of G, and $\tau \in \mathfrak{H}$. Also, f is invariant under simultaneous conjugation on g and h, and satisfies $f(g^a h^c, g^b h^d, \tau) = f(g, h, (a\tau+b)/(c\tau+d))$. In particular, for fixed g and h, $f(g, h, \tau)$ is a holomorphic modular function, invariant under $\Gamma(\mathrm{lcm}(|g|, |h|))$. Following [Ganter 2009], we map the homology basis to $(-1, \tau)$, so (g, h, τ) describes an elliptic curve $\mathbb{C}/\langle -1, \tau\rangle$ equipped with a G-torsor with monodromy (g, h). (Many texts use the basis $(1, \tau)$ when studying modular functions, mostly because τ then becomes the ratio of periods, but our convention is what we need for the left $\mathrm{SL}_2(\mathbb{Z})$ action to work correctly.) Any degree-n isogeny from an elliptic curve E' to $\mathbb{C}/\langle -1, \tau\rangle$ can be described as the identity map on $\mathbb{C}$, where E' is the quotient by a unique index-n sublattice of $\langle -1, \tau\rangle$. Since we are assuming $\mathrm{SL}_2(\mathbb{Z})$-equivariance of f, we can choose any basis, and get the same value from f. We preferentially choose bases

$(-d, a\tau + b)$, where d is the index in $\mathbb{Z}$ of the intersection of the sublattice with $\mathbb{Z}$, and we get

$$T_n(f)(g, h, \tau) = \frac{1}{n} \sum_{\substack{ad=n \\ 0 \le b < d}} f\Big(g^d, g^{-b}h^a, \frac{a\tau + b}{d}\Big).$$

Suppose we wanted to extend the notion of Hecke operator to a larger class of functions, particularly ones that are not a priori completely independent of the choice of homology basis of our elliptic curve. One might hope that we could have a good notion for all functions on $\mathrm{Hom}(\mathbb{Z}\times\mathbb{Z}, G)/G\times\mathfrak{H}$. Unfortunately, there is no canonical choice of homology basis for E' (that is, a basis for the index-n sublattice of $\langle -1, \tau\rangle$), and it is difficult make choices in a systematic way that makes the sum a canonical quantity, so we do not know of any definition of Hecke operator for arbitrary functions on $\mathrm{Hom}(\mathbb{Z}\times\mathbb{Z}, G)/G\times\mathfrak{H}$ that is particularly natural. However, there is an intermediate form of equivariance for which we *can* make a canonical definition, using the subgroup $\pm\mathbb{Z} := \{\pm\left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right) : n \in \mathbb{Z}\} \cong \mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$. Invariance under this group implies that for a function f on

$$\mathrm{Hom}(\mathbb{Z} \times \mathbb{Z}, G)/G \underset{\pm\mathbb{Z}}{\times} \mathfrak{H}, \tag{1}$$

we have $f(g, gh, \tau) = f(g, h, \tau + 1)$, so the Fourier expansion of $f(g, h, \tau)$ is a power series in $q^{1/|g|}$ that converges on the punctured open unit disc parametrized by $q^{1/|g|}$, $|q| < 1$. We will assume the existence of a lower bound on exponents, that is, that all of our power series are Laurent series.

We can interpret this geometrically. The quotient (1) is a disjoint union of punctured unit discs, and parametrizes G-torsors over elliptic curves that are equipped with a distinguished primitive element of H_1 (up to sign — the -1 automorphism inverts the monodromy and fixes the curve, so as long as we remember that any function is invariant under this transformation, we can safely ignore it). This element functions as the first element in the homology basis, since the action of $\pm\mathbb{Z}$ renders all choices of second oriented basis element equivalent. It also uniquely determines a multiplicative uniformization $\mathbb{C}^\times \xrightarrow{\pi} E$ with kernel $\langle q\rangle$, $|q| < 1$. One can classify the G-torsors over an elliptic curve with multiplicative uniformization by studying its monodromy. Monodromy along the primitive homology element gives a distinguished element g, up to conjugacy. Monodromy along a path from 1 to q in $\mathbb{C}^\times$ yields a commuting element h that is unique up to conjugation that is simultaneous with g. However, the set of homotopy classes of paths to q is a $\mathbb{Z}$-torsor given by winding number around zero, and the action changes this monodromy by powers of g, so the equivalence classes of G-torsors are determined by assigning a commuting element h not to q, but to a choice of $q^{1/|g|}$.

Definition. Given an elliptic curve E equipped with a multiplicative uniformization, a restricted degree-n isogeny is a pullback diagram

$$\begin{tikzcd} \mathbb{C}^\times \arrow[r] \arrow[d] & \mathbb{C}^\times \arrow[d] \\ E' \arrow[r] & E \end{tikzcd}$$

where the bottom row is a degree-n isogeny of elliptic curves.

The map on top is then given by a d-th power map, for some $d \mid n$. If we examine kernels of the uniformization, we find that this induces an inclusion $\mathbb{Z} \to \mathbb{Z}$ by multiplication by $a := n/d$. If we let q generate the kernel of the uniformization on the target, the isogenies that pull back to the d-th power map on $\mathbb{C}^\times$ are then classified by d-th roots of q^a in the source, and there are exactly d of them. In particular, there is a bijection between degree-n isogenies in the classical sense and degree-n restricted isogenies. Each restricted isogeny then has the form

$$\mathbb{C}^\times/q^{(a/d)\mathbb{Z}} \xrightarrow{z \mapsto z^d} \mathbb{C}^\times/q^{\mathbb{Z}} .$$

We can rephrase this using lattices: A uniformized elliptic curve is given by an equivalence class of lattices $\langle -1, \tau \rangle$, where we consider two lattices equivalent if the second elements differ by an integer. The isogeny condition is equivalent to demanding that the distinguished homology basis element is $-d$ for some $d \mid n$, as we chose before.

We can now define our Hecke operators by summing over pullbacks along our restricted isogenies.

Lemma 1.1. *Given a function f on (1), define the function $n\hat{T}_n f$ on the same space by assigning to each elliptic curve equipped with a G-torsor and multiplicative uniformization the sum of f evaluated on the sources of restricted isogenies of degree n. Then*

$$n\hat{T}_n f(g, h, \tau) = \sum_{ad=n} \sum_{0 \le b < d} f\Big(g^d, g^{-b}h^a, \frac{a\tau + b}{d}\Big),$$

that is, we get the same formula for Hecke operators as we would over $\mathcal{M}_{\mathrm{Ell}}^G$.

Proof. Fix an elliptic curve E with a multiplicative uniformization and a G-torsor. We may assume that $E \cong \mathbb{C}/\langle -1, \tau \rangle$ for some $\tau \in \mathfrak{H}$, where the path from 0 to -1 along $\mathbb{R}$ maps to the distinguished homology element. Let g be the monodromy of the G-torsor along the image of this path, and let h be the monodromy along the image of a path from 0 to τ. Fix a restricted isogeny, that is, an index-n sublattice of $\langle -1, \tau \rangle$ together with a (uniquely defined) fixed negative integer $-d$, $d \mid n$. The path from 0 to $-d$ is the chosen primitive homology element of the source elliptic

curve. The monodromy of the G-torsor along the image of this path is g^d. The sublattice is characterized by a second homology generator $a\tau+b$ for $a=n/d$, and b is uniquely determined modulo d. The generator then has monodromy $h^a g^{-b}$, and by applying the $(1/d)$-dilation homothety, the elliptic curve is given by the point $(a\tau+b)/d$. To show that the formula above holds, it suffices to show that f, evaluated on these generators, does not depend on which coset representative modulo d we choose. This independence arises from the $\pm\mathbb{Z}$-equivariance, that is, if we choose b and b' such that $b-b'=kd$, then

$$\begin{aligned} f\Big(g^d, g^{-b'}h^a, \frac{a\tau+b'}{d}\Big) &= f\Big(g^d, g^{-b}h^a g^{kd}, \frac{a\tau+b-kd}{d}\Big) \\ &= f\Big(g^d, g^{-b}h^a, \frac{a\tau+b}{d}-k+k\Big) \\ &= f\Big(g^d, g^{-b}h^a, \frac{a\tau+b}{d}\Big). \end{aligned} \qquad \square$$

From now on, we will use the notation nT_n for this Hecke operator, instead of $n\hat{T}_n$.

Lemma 1.2. *If f is a function on* $\mathrm{Hom}(\mathbb{Z}\times\mathbb{Z}, G)/G \underset{\pm\mathbb{Z}}{\times} \mathfrak{H}$, *then*

$$T_k T_m f(g,h,\tau) = \sum_{t\mid(k,m)} \frac{1}{t} T_{km/t^2} f(g^t, h^t, \tau).$$

Proof.

$$\begin{aligned} T_k T_m f(g,h,\tau) &= T_k \frac{1}{m} \sum_{\substack{ad=m\\ 0\le b<d\\ (a,b,d)=1}} f\Big(g^d, g^{-b}h^a, \frac{a\tau+b}{d}\Big) \\ &= \frac{1}{km} \sum_{\substack{a'd'=k\\ 0\le b'<d'}} \sum_{\substack{ad=m\\ 0\le b<d}} f\Big(g^{dd'}, g^{-bd'-ab'}h^{aa'}, \frac{aa'\tau+ab'+bd'}{dd'}\Big) \\ &= \frac{1}{km} \sum_{\substack{a'd'=k\\ ad=m\\ t=(a,d')}} \sum_{\substack{0\le b'<d'\\ 0\le b<d}} f\Big(g^{t(dd'/t)}, g^{t((-bd'-b'a)/t)}h^{t(aa'/t)}, \frac{a'\frac{a}{t}\tau+b'\frac{a}{t}+b\frac{d'}{t}}{d\frac{d'}{t}}\Big) \\ &= \frac{1}{km} \sum_{t\mid(k,m)} \sum_{\substack{a'd'=k/t\\ ad=m/t\\ (a,d')=1}} \sum_{\substack{0\le b'<td'\\ 0\le b<d}} f\Big(g^{tdd'}, g^{t(-bd'-ab')}h^{taa'}, \frac{aa'\tau+ab'+bd'}{dd'}\Big) \\ &= \frac{1}{km} \sum_{t\mid(k,m)} \sum_{a''d''=km/t^2} t \sum_{0\le b''<d''} f\Big(g^{td''}, g^{-tb''}h^{ta''}, \frac{a''\tau+b''}{d''}\Big) \\ &= \sum_{t\mid(k,m)} \frac{1}{t} T_{km/t^2} f(g^t, h^t, \tau). \end{aligned}$$

We explain the second-to-last equality using Kozlov's argument [1994]. In this step, we substitute $a'' = aa'$, $d'' = dd'$, and b'' for any solution to the congruence $ab' + bd' \equiv b'' \pmod{dd'}$. By $\pm\mathbb{Z}$-invariance, it remains to show that for any $0 \le b'' < dd'$, this congruence has exactly t solutions. There are exactly tdd' possible values of b and b' satisfying $0 \le b < d$, $0 \le b' < td'$, and dd' values of b'' satisfying $0 \le b'' < d''$. The first value is t times the second value, so it suffices to show that for any fixed admissible pair (b, b') there are exactly t solutions (c, c') satisfying $0 \le c < d$ and $0 \le c' < td'$ to the congruence $ab' + bd' \equiv ac' + cd' \pmod{dd'}$. Any such solution yields the identity $dd' \mid a(b' - c') + d'(b - c)$, so $d' \mid a(b' - c')$. Since $(a, d') = 1$, $d' \mid b' - c'$, so we write $c' = b' + sd'$, and there are t choices of s that satisfy $0 \le c' < td'$. Canceling d' in the identity yields $d \mid as + b - c$, so each choice of s gives a uniquely defined value of c satisfying $0 \le c < d$. □

It is also possible to prove this by working one prime at a time, or by invoking the moduli interpretation and enumerating restricted isogenies.

2. Hecke-monicity

Definition. Let f be a holomorphic function on (1). We say that f is Hecke-monic if on each connected component, the restriction of $nT_n(f)$ is a monic polynomial of degree n in the restriction of f, for all positive integers n.

Remark. Since we only require our functions to admit translation-equivariance, and the Hecke operators only involve transformations of the form $\tau \mapsto (a\tau + b)/d$, Hecke-monicity only depends on the values of f when the monodromy around the first homology basis element lies in a subset of G that is closed under taking power maps. We will find it useful to weaken the condition that f be defined on all components. For example, if we choose $g \in G$, we only need to consider the functions $\{f(1, g^i, \tau)\}_{i>0}$ to define Hecke operators on $f(1, g, \tau)$.

Definition. Let $g, h \in G$ be commuting elements, and let f be a function on the connected components of (1) corresponding to pairs $(g^d, g^{-b}h^a)$ for $a, b, d > 0$. We say that f is weakly Hecke-monic for (g, h) if for all $n > 0$, $nT_n f(g, h, \tau)$ is a monic polynomial of degree n in $f(g, h, \tau)$. We say that f is semiweakly Hecke-monic for (g, h) if for all $n > 0$, $nT_n f(g^d, g^{-b}h^a, \tau)$ is a monic polynomial of degree n in $f(g^d, g^{-b}h^a, \tau)$ for all $a, b, d > 0$.

We use the notation $e(x)$ to denote $e^{2\pi i x}$ for the rest of this paper.

Lemma 2.1. *Let f be a weakly Hecke-monic function for (g, h), and let $N > 0$ satisfy $g^N = h^N = 1$. If $f(g, h, \tau)$ has a singularity at infinity, then its q-expansion has the form $\zeta q^{C/|g|} + O(1)$ for C a negative integer and ζ a root of unity satisfying $\zeta^N = 1$ if N is even and $\zeta^{2N} = 1$ if N is odd.*

Proof. Let $f(g,h,\tau)=\sum_{n\in(1/|g|)\mathbb{Z}} a_n q^n = a_{n_0}q^{n_0}+a_{n_1}q^{n_1}+\cdots$ for a_{n_0} nonzero, $n_0<0$, and let p be a prime congruent to 1 mod N. Then

$$\begin{aligned} pT_p f(g,h,\tau) &= f(g,h^p,p\tau)+\sum_{b=0}^{p-1} f\Big(g^p, g^{-b}h, \frac{\tau+b}{p}\Big) \\ &= f(g,h,p\tau)+\sum_{b=0}^{p-1} f\Big(g,h,\frac{\tau+b}{p}-b\Big) \\ &= \sum_n a_n q^{pn} + \sum_n a_n \sum_b e\Big(n\Big(\frac{\tau+b}{p}-b\Big)\Big) \\ &= \sum_n a_n q^{pn} + \sum_n a_n q^{n/p} \sum_b e\Big(nb\frac{1-p}{p}\Big). \end{aligned}$$

Hecke-monicity implies that $a_{n_0}q^{pn_0}=(a_{n_0}q^{n_0})^p$ for all p congruent to 1 mod N, so $\zeta=a_{n_0}$ is an M-th root of unity, where M is the greatest common divisor of $\{p-1 : p \text{ prime}, p\equiv 1(N)\}$. $M=kN$ for some integer k, and $(k,N)=1$, since otherwise $(k/(k,N))N+1$ would be a residue class mod M that is coprime to M but not congruent to 1. By the Chinese Remainder Theorem, k must have a unique residue class mod k that is coprime to k. Therefore, the only possible values of k are 1 or 2, and $k=2$ is only possible when N is odd.

If $f(g,h,\tau)$ is a singular monomial $\zeta q^{C/|g|}$ with $C<0$, then we are done. Otherwise, we assume $a_{n_1}\neq 0$, and from the calculation above, we have

$$pT_p f(g,h,\tau) = \begin{cases} a_{n_0}q^{pn_0}+a_{n_0}q^{n_0/p}\sum_b e\Big(n_0 b\frac{p-1}{p}\Big)+\cdots & n_1>n_0/p^2 \\ a_{n_0}q^{pn_0}+a_{n_1}q^{pn_1}+\cdots & n_1<n_0/p^2. \end{cases}$$

We will not bother with the case of equality, because we will let p become large. If $n_1<0$, then the second case will hold for almost all p congruent to 1 mod N, and if $n_1\geq 0$, then the first case will hold for all such p. If $n_1<0$ and p is sufficiently large, then

$$\begin{aligned} a_{n_0}q^{pn_0}+a_{n_1}q^{pn_1}+\cdots &= (a_{n_0}q^{n_0}+a_{n_1}q^{n_1}+\cdots)^p + c(a_{n_0}q^{n_0}+\cdots)^{p-1}+\cdots \\ &= a_{n_0}q^{pn_0}+pa_{n_0}^{p-1}a_{n_1}q^{(p-1)n_0+n_1}+\cdots \\ &= a_{n_0}q^{pn_0}+pa_{n_1}q^{(p-1)n_0+n_1}+\cdots. \end{aligned}$$

This yields an equality $a_{n_1}q^{pn_1}=pa_{n_1}q^{(p-1)n_0+n_1}$, which under our assumptions is a contradiction. Therefore, $n_1\geq 0$, and we are done. □

Lemma 2.2. *Let f be a weakly Hecke-monic function for (g,h) such that*

$$f(g,h,\tau)=\zeta q^{C/|g|}+O(1)$$

for some $C < 0$ and some root of unity ζ. Then there exists some N such that $pT_p f(g,h,\tau) = \zeta^p q^{Cp/|g|} + O(1)$ for all primes $p > N$.

Proof. Since $\langle g,h\rangle$ has finite order, we can choose N such that $N/|g|$ is greater than the order of any pole of $f(g^k, g^l h^m, \tau)$ at infinity, as k and l range over $\mathbb{Z}/|g|\mathbb{Z}$ and m ranges over $\mathbb{Z}/|h|\mathbb{Z}$. Suppose $p > N$, and write any singular functions $f(g^p, g^{-b}h, \tau)$ as $\zeta_b q^{C_b/|g|} + O(1)$. Then

$$\begin{aligned} pT_p f(g,h,\tau) &= f(g,h^p,p\tau) + \sum_{0\le b<d} f\Big(g^p, g^{-b}h, \frac{\tau+b}{p}\Big) \\ &= f(g,h^p,p\tau) + \sum_b \zeta_b e\Big(\frac{bC_b}{p|g|}\Big) q^{C_b/p|g|} + O(1). \end{aligned}$$

Since $p > |C_b|$, $C_b/p|g| > -1/|g|$ for all b such that $f(g^p, g^{-b}h, \tau)$ has a pole at infinity. However, the above sum is a polynomial in $f(g,h,\tau)$, and therefore a power series in $q^{1/|g|}$, so the only contribution with a negative power of q comes from $f(g,h^p,p\tau)$. We have $f(g,h^p,p\tau) = \zeta' q^{C'p/|g|} + O(1)$ for some ζ' and C', and since this is a monic polynomial of degree p in $f(g,h,\tau)$, we have $\zeta' = \zeta^p$ and $C' = C$. □

Proposition 2.3. *Let f be a weakly Hecke-monic function for (g,h), such that $f(g,h,\tau) = \zeta q^{C/|g|} + O(1)$ for some $C < 0$ and some root of unity ζ. Then $f(g,h,\tau)$ is invariant under translation by $|g|/C$, that is, the only nonzero terms in the q-expansion are those with integer powers of $q^{C/|g|}$.*

Proof. Suppose $f(g,h,\tau)$ is not a power series in $q^{C/|g|}$, and let n_0 be the smallest integer such that n_0 is not a multiple of C, and the coefficient a_{n_0} of $q^{n_0/|g|}$ in the q-expansion of $f(g,h,\tau)$ is nonzero. Choose N as in Lemma 2.2, and let p be a prime satisfying $p > N$, $p \equiv 1 \pmod{|g|}$ and $(p-1)C + n_0 < 0$ (that is, p is large).

By the lemma, $pT_p f(g,h,\tau)$ has q-expansion $\zeta^p q^{Cp/|g|} + O(1)$. However, $pT_p f(g,h,\tau)$ is a monic polynomial of degree p in $f(g,h,\tau)$, so we can write its q-expansion as a sum of a series in $q^{C/|g|}$ and a series with initial term

$$p\zeta^{p-1} a_{n_0} q^{(p-1)C+n_0}.$$

Since the coefficient is nonzero and the exponent is negative, we have a contradiction. □

3. Modular equations

Cummins and Gannon found a characterization of holomorphic genus-zero functions invariant under $\Gamma_0(N)$ as power series satisfying many modular equations. We show that weakly Hecke-monic functions satisfy a similar condition, and we

modify the first half of their proof to get global symmetries. In particular, any Hecke-monic function on $\mathcal{M}^G_{\mathrm{Ell}}$ is holomorphic congruence genus-zero on nonconstant components.

Lemma 3.1. *Fix a positive integer n, and let f be weakly Hecke-monic for (g^t, h^t) for all $t|n$. Then the power sum symmetric polynomials in*

$$\left\{ f\left(g^d, g^{-b}h^a, \frac{a\tau+b}{d}\right) : ad = n,\ 0 \le b < d \right\}$$

are polynomials in $f(g^t, h^t, \tau)$ for t ranging over positive integers dividing n. Furthermore, the term with highest degree in $f(g, h, \tau)$ has coefficient equal to one. In particular, if n is a prime satisfying $g^n = g$ and $h^n = h$, then the power sums are polynomials in $f(g, h, \tau)$.

Proof. This is essentially the same as in [Kozlov 1994]. We apply nT_n to the equation $f^m = mT_m(f) - a_{m-1}f^{m-1} - \cdots - a_1 f - a_0$ to find that the power sum

$$\sum_{\substack{ad=n \\ 0\le b<d}} f\left(g^d, g^{-b}h^a, \frac{a\tau+b}{d}\right)^m = nT_n(f(g, h, \tau)^m)$$

can be written as a sum of $mnT_nT_m(f)(g, h, \tau)$ and a linear combination of T_n applied to lower degree polynomials in $f(g, h, \tau)$. By induction on m, these are polynomials in $f(g^t, h^t, \tau)$ for $t \mid n$. □

Lemma 3.2. *Fix $n \ge 2$ square-free, and let f be a weakly Hecke-monic function for (g^t, h^t) for all $t \mid n$. Then there exists a monic polynomial $F_n(x)$ of degree $n\prod_{p\mid n}(p+1)/p$, whose coefficients are polynomials in $f(g^t, h^t, \tau)$ for $t \mid n$, and with roots $\{f(g^d, g^{-b}h^a, (a\tau+b)/d) : ad = n, 0 \le b < d, (a, b, d) = 1\}$ for any τ.*

Proof. Since n is square-free, the condition $(a, b, d) = 1$ is a consequence of $ad = n$. The power sums generate the ring of symmetric polynomials in

$$\left\{ f\left(g^d, g^{-b}h^a, \frac{a\tau+b}{d}\right) \right\},$$

from which we draw the coefficients of F_n. □

A holomorphic function f on $\mathfrak{H}$ is said to satisfy a modular equation of order n if there exists a monic polynomial $F_n(x)$ of degree $n\prod_{p\mid n}(p+1)/p$ whose coefficients are polynomials in f, and with roots $f((a\tau+b)/d)$ for a, b, d satisfying $ad = n, 0 \le b < d, (a, b, d) = 1$. We will use a slightly altered notion to account for invariance under congruence groups other than $\Gamma_0(N)$.

Definition. Let $g, h \in G$ be a commuting pair, and let f be a function on

$$(\{(g, g^n h)\}_{n\in\mathbb{Z}}) \underset{\pm\mathbb{Z}}{\times} \mathfrak{H}.$$

If p is a prime satisfying $g^p = g$ and $h^p = h$, we say $f(g, h, \tau)$ satisfies an equivariant modular equation of order p if there exists a monic polynomial $F_n(x)$ of degree $p+1$, whose coefficients are polynomials in $f(g, h, \tau)$ and whose roots are $f(g, g^{-b}h, (a\tau + b)/d)$ for a, b, d satisfying $ad = p, 0 \le b < d$.

When $g = 1$ and $f(1, h, \tau)$ has q-expansion $q^{-1} + O(q)$, this agrees with the nonequivariant notion.

Proposition 3.3. *Suppose $g, h \in G$ commute. If a function f is weakly Hecke-monic for (g, h), then $f(g, h, \tau)$ satisfies equivariant modular equations of order p for all primes p congruent to* 1 *modulo* $\operatorname{lcm}(|g|, |h|)$.

Proof. Since $g^p = g$ and $h^p = h$, this is a special case of Lemma 3.2. □

If $f(g, h, \tau)$ satisfies an equivariant modular equation of order p, we can write the polynomial $F_p(x)$ as a two-variable polynomial $F_p(y, x) \in \mathbb{C}[x, y]$, where we set $y = f(g, h, \tau)$, and expand the coefficients of $F_p(x)$ as polynomials in $f(g, h, \tau)$. If $f(g, h, \tau)$ is a nonconstant holomorphic function, then $F_p(y, x)$ is uniquely defined by the properties that it is monic of degree $p+1$ in x and that it vanishes under the substitutions $f(g, h, \tau)$ for y and $f(g, g^{-b}h, (a\tau + b)/d)$ for x for any $\tau \in \mathfrak{H}$. This is because a polynomial in one variable is uniquely determined by its values on a nonempty open subset of $\mathbb{C}$, and the coefficients of $F_p(x)$ are polynomials in $f(g, h, \tau)$, which includes such an open subset in its range.

Lemma 3.4. *Let p be a prime satisfying $g^p = g$ and $h^p = h$. Suppose $f(g, h, \tau)$ is a nonconstant holomorphic function satisfying an equivariant modular equation of order p. Then $F_p(y, x) = F_p(x, y)$.*

Proof. This is a modification of [Kozlov 1994, Proposition 3.2].

If $d = 1$, then $F_p(f(g, h, \tau), f(g, h, p\tau)) = 0$. We make the substitution $\tau := (\tau' + b)/p - b$ for $0 \le b < p$, and we get

$$\begin{aligned} 0 &= F_p\Big(f\Big(g, h, \frac{\tau'+b}{p} - b\Big), f(g, h, \tau' + b - pb)\Big) \\ &= F_p\Big(f\Big(g, g^{-b}h, \frac{\tau'+b}{p}\Big), f(g, h, \tau')\Big). \end{aligned}$$

Then $f(g, g^{-b}h, (\tau' + b)/p)$ is a root of $F_p(y, f(g, h, \tau'))$.

If $d = p$, then $F_p(f(g, h, \tau), f(g, h, (\tau + b)/p - b)) = 0$. We make the substitution $\tau = p\tau' + pb - b$ for $0 \le b < p$, and we get

$$0 = F_p\big(f(g, h, p\tau' + pb - b), f(g, h, \tau')\big) = F_p\big(f(g, h, p\tau'), f(g, h, \tau')\big).$$

Then $f(g, h, p\tau')$ is a root of $F_p(y, f(g, h, \tau'))$.

This proves that $F_n(y, f(g, h, \tau))$ has roots $f(g, g^{-b}h, (a\tau + b)/d)$, which means that for any fixed $\tau \in \mathfrak{H}$, $F_p(f(g, h, \tau), x) = F_p(x, f(g, h, \tau)) \in \mathbb{C}[x]$.

The coefficients of $F_p(x) = F_p(x, f(g,h,\tau))$ are polynomials in f, so they are uniquely determined by finitely many values. If f is nonconstant and holomorphic on some nonempty open set, then the coefficients of $F_p(x,y)$ match those of $F_p(y,x)$, so we get a polynomial equality. □

Proposition 3.5. *If f is weakly Hecke-monic for (g,h) and $f(g,h,\tau)$ has a pole at infinity, $f(g,h,\tau)$ admits global symmetries; in other words, if $f(g,h,\tau_1) = f(g,h,\tau_2)$ for given $\tau_1, \tau_2 \in \mathfrak{H}$, there exists $\gamma \in \mathrm{SL}_2(\mathbb{R})$ such that $\tau_1 = \gamma\tau_2$, and $f(g,h,\tau) = f(g,h,\gamma\tau)$ for all $\tau \in \mathfrak{H}$.*

Proof. By Proposition 3.3, $f(g,h,\tau)$ satisfies equivariant modular equations of degree p for infinitely many primes p congruent to 1 modulo $\mathrm{lcm}(|g|,|h|)$.

We give a list of modifications of the first half of [Cummins and Gannon 1997] (up to Proposition 4.6) to allow equivariance. Note that the summands for $d = p$ are $f(g, g^{-b}h, (\tau+b)/p) = f(g,h,(\tau+b)/p - b)$, $0 \le b < p$, so we make the global modification that

$$A(p) = \left\{ \begin{pmatrix} 1 & (1-p)b \\ 0 & p \end{pmatrix} : 0 \le b < p \right\} \cup \left\{ \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} \right\}.$$

The proof of existence of global symmetries in [Cummins and Gannon 1997] needs the following cosmetic changes:

- The statement of Lemma 2.2, Condition 1 should be changed from $z_1 - z_2 \in \mathbb{Z}$ to $z_1 - z_2 \in (|g|/C)\mathbb{Z}$.
- In Lemma 2.5, β should be changed to $\begin{pmatrix} n/d & r-dr \\ 0 & d \end{pmatrix}$; the proof uses the symmetry of $F_p(x,y)$ (Lemma 3.4).
- The statement of Lemma 3.2 requires the form of $\beta \in A(p)$ to be changed as above.
- All instances of $\mathbb{Z}$ in the proof of Lemma 3.3 should be replaced by $(|g|/C)\mathbb{Z}$.
- The phrase "translating by integers if necessary" in the proof of Proposition 4.3 should become "translating by integer multiples of $(|g|/C)$ if necessary."
- In the proof of Proposition 4.6, the form of $\beta \in A(p)$ needs to be suitably adjusted. □

Corollary 3.6. *Let f be a Hecke-monic function on $\mathcal{M}^G_{\mathrm{Ell}}$. If $f(g,h,\tau)$ is nonconstant, then it is a holomorphic congruence genus-zero function.*

Proof. From our hypotheses, we know that $f(g,h,\tau)$ is invariant under some $\Gamma(N)$, and we are assuming that f has no essential singularities at cusps. Therefore, if $f(g,h,\tau)$ is nonconstant, then there is some $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ such that $f(g,h,(a\tau+b)/(c\tau+d))$ has a pole at infinity. Then $f(g^a h^c, g^b h^d, \tau)$ has a pole at infinity, and satisfies the hypotheses of the proposition. This implies $f(g^a h^c, g^b h^d, \tau)$ is holomorphic congruence genus-zero, so $f(g,h,\tau)$ is also. □

4. Finite level

Theorem 1.3 of [Cummins and Gannon 1997] asserts that any series $q^{-1} + O(q)$ with algebraic integer coefficients satisfying modular equations of all orders coprime to some n is either holomorphic genus-zero and invariant under some $\Gamma_0(N)$, or a function of the form $q^{-1} + \zeta q$ for ζ either zero or a twenty-fourth root of unity. The hypotheses we use to prove Theorem 4.6 are weaker, since the functions satisfy equivariant modular equations only for primes congruent to 1 (mod n), and the functions have the form $q^{C/|g|} + O(1)$. However, our conclusions are weaker, since even if we normalize to an integral-powered q-series, we only have invariance under $\Gamma_1(N)$, much like the situation in [Cummins 2002].

Definition. Let G be a subgroup of $\mathrm{SL}_2(\mathbb{R})$, and let M, N, and C be nonzero integers such that $M \mid N$. We say that the quadruple (G, M, N, C) satisfies properties (1)–(3) if:

(1) G is a discrete group.

(2) The stabilizer of infinity $G_\infty \subset G$ is $\langle -\mathrm{Id}, \left(\begin{smallmatrix} 1 & M \\ 0 & 1 \end{smallmatrix}\right)\rangle$.

(3) For all primes p congruent to 1 mod N, and all $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in G$, there exist integers l and k such that $l \mid p$, $0 \le -k < p/l$, and such that

$$\begin{pmatrix} \dfrac{ap}{l} & \dfrac{k(1-p)a}{C} + lb \\ \dfrac{c}{l} & \dfrac{1}{p}\left(\dfrac{k(1-p)c}{C} + ld\right) \end{pmatrix} \in G.$$

Lemma 4.1. *If (G, M, N, C) satisfies properties (1)–(3), and $\gamma \in G$, then there exists $\lambda \in \mathbb{R}$ such that $\lambda\gamma \in GL_2^+(\mathbb{Q})$.*

Proof. This is a minor variation of [Cummins and Gannon 1997, Lemma 5.4]. Our G_∞ is a subgroup of theirs, so double coset invariants surject. We define $r_m \equiv a/c$ (mod M) instead of mod 1, but it is still a G_∞ double coset invariant for our G_∞. The proof there uses a slightly different property (3) for G, but the two left entries of the matrices match, and that is what is needed. □

Following [Cummins and Gannon 1997], we say that $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ is primitive if a, b, c, d are integers with no common factors. By the previous lemma, there exists for any $\gamma \in G$ some $\lambda \in \mathbb{R}$ (unique up to sign) such that $\lambda\gamma$ is primitive, and we define $|\gamma|$ to be the determinant of $\lambda\gamma$. This is an invariant of the double coset $G_\infty \gamma G_\infty$.

Lemma 4.2. *Let (G, M, N, C) satisfy properties (1)–(3), let $\gamma_1 \in G$, let*

$$\lambda\gamma_1 = \begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix}$$

be primitive, and assume $c_1 \neq 0$. *Choose a prime* $p \equiv 1$ (*mod* NC), *and choose a sequence of elements* $\{\gamma_n\}_{n\geq 1} \subset G$ *by iteratively applying property* (3). *Define*

$$\lambda\gamma_i = \begin{pmatrix} a_i & b_i \\ c_i & d_i \end{pmatrix} \in M_2(\mathbb{Q}),$$

and let $l_i, k_i \in \mathbb{Z}$ *be the corresponding integers arising in each application of property* (3).

(a) *The sequence* $\{c_i\}_{i\geq 1}$ *eventually stabilizes to some* $c_\infty = c_1/\prod_{i\geq 1} l_i \in \mathbb{Q}$, *that is, all but finitely many* l_i *are equal to* 1.

(b) *If* $c_\infty \in \mathbb{Z}$, *then* $d_i \in \mathbb{Z}$ *for all* $i \geq 1$.

(c) *If* c_∞ *is a nonzero integer multiple of* p, *then* p *divides* d_i *for all* $i \geq 1$.

(d) *There exists* $W > 0$, *depending only on* c_1 *and* λ, *such that if* $p > W$, *then* $l_i = 1$ *and* $d_i \in \mathbb{Z}$ *for all* $i \geq 1$.

Proof. This is a minor alteration of [Cummins and Gannon 1997, Lemma 5.7], and we point out the necessary changes.

Statement (a) follows from [Shimura 1971, Lemma 1.25], which asserts that the lower left entries of elements of a discrete subgroup of $\mathrm{SL}_2(\mathbb{R})$ that don't fix infinity are bounded away from zero.

Statements (b) and (c) can be proved by following the proofs of [Cummins and Gannon 1997, Lemma 5.7a and 5.7b], and changing n to p, k_i to $k_i(1-p)/C$ (which is an integer by our assumption on p), and $p^{2((j-i)\eta+s)+s'}$ to $p^{2(i-i_0+s)+s'}$. The last alteration is mostly to rectify a typographical error.

The assertion about $l_i = 1$ in (d) follows from [Shimura 1971, Lemma 1.25], and the assertion about d_i follows from (b). □

Lemma 4.3. *Suppose* (G, M, N, C) *satisfies properties* (1)–(3), *and* G *does not stabilize infinity. Then* G *contains an element of the form* $\left(\begin{smallmatrix} 1 & 0 \\ n & 1 \end{smallmatrix}\right)$ *for* n *a nonzero multiple of* NC.

Proof. Since G does not stabilize infinity, then G contains γ such that the primitive $\lambda\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ has $c \neq 0$, and hence

$$\gamma' = \gamma\begin{pmatrix} 1 & -NC|\gamma| \\ 0 & 1 \end{pmatrix}\gamma^{-1} = \begin{pmatrix} 1+NCac & -NCa^2 \\ NCc^2 & 1-NCac \end{pmatrix} \in G \cap \Gamma(N).$$

Let $W(\gamma')$ be the constant given by the fourth part of Lemma 4.2. After translating on the right by multiples of $\left(\begin{smallmatrix} 1 & NC \\ 0 & 1 \end{smallmatrix}\right)$, we find that G has an element

$$g = \begin{pmatrix} 1+NCac & b' \\ NCc^2 & p \end{pmatrix},$$

with $p \equiv 1 \pmod{NC}$ a prime larger than $W(\gamma')$. Since both matrices have primitive multipliers $\lambda = 1$ and the same bottom left entries, $W(\gamma') = W(g)$. We apply Lemma 4.2(d) and use property (3) to find that G contains

$$\begin{pmatrix} (1+NCac)p & k(1-p)(1-NCac)/C + b' \\ NCc^2 & k(1-p)Nc^2 + 1 \end{pmatrix} \in \Gamma(NC)$$

for some $0 \le k < p$. Since $M \mid N$, we can multiply on the right by $\left(\begin{smallmatrix} 1 & kmN \\ 0 & 1 \end{smallmatrix}\right) \in G$, where $m = (p-1)/(NC) \in \mathbb{Z}$, and this yields

$$\begin{pmatrix} * & * \\ NCc^2 & 1 \end{pmatrix} \in \Gamma(N).$$

We multiply on the left by a suitable multiple of $\left(\begin{smallmatrix} 1 & N \\ 0 & 1 \end{smallmatrix}\right)$ to get $\left(\begin{smallmatrix} 1 & 0 \\ NCc^2 & 1 \end{smallmatrix}\right) \in G$. □

Lemma 4.4. *Let X be a set of matrices $\begin{pmatrix} 1+an & bn \\ cn & 1+dn \end{pmatrix} \in \Gamma(n)$, satisfying:*

- *For every integer c_0, there exists an element of X as above with $c = c_0$.*
- *For all nonzero c, and all a_0 and d_0 satisfying $(1+a_0n, cn) = (1+d_0n, cn) = 1$, there exists an element of X as above such that $a \equiv a_0$ (mod $|c|n$) and $d \equiv d_0$ (mod $|c|n$).*

Then X is a complete set of double coset representatives for $\Gamma(n)$ with respect to the subgroup $\langle \left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right) \rangle$.

Proof.
$$\begin{pmatrix} 1 & en \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1+an & bn \\ cn & 1+dn \end{pmatrix}\begin{pmatrix} 1 & fn \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1+an+cen^2 & (b+e+f)n + (af+de)n^2 + cefn^3 \\ cn & 1+dn+cfn^2 \end{pmatrix}.$$

To find all double coset representatives, it suffices to cover the possible lower triangular entries, since for $c \neq 0$, the top right entry is uniquely determined by the fact that the determinant is one. Any element of X satisfying $c = 0$ lies in the group $\langle \left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right) \rangle$, so the corresponding double coset is equal to this group. □

Lemma 4.5. *Suppose (G, M, N, C) satisfies properties (1)–(3), and suppose G contains $\left(\begin{smallmatrix} 1 & 0 \\ n & 1 \end{smallmatrix}\right)$ for some nonzero integer n for $NC \mid n$. Then G contains $\Gamma(n)$.*

Proof. It suffices to produce double coset representatives with respect to translations, so suppose we are given a, c, d satisfying the conditions in the above lemma, with $c \neq 0$. Let $r > 0$ be a lower bound on absolute value of nonzero lower left entries of elements of G, guaranteed by [Shimura 1971, Lemma 1.25]. By Dirichlet, there exist primes p and q such that $p \equiv 1 + an \pmod{|c|n^2}$, $q \equiv 1 + dn$

(mod $|c|n^2$), and $p > \max(|c|n/r, |c|n)$. Since $(1+an)(1+dn)-bcn^2 = 1$, there exists an integer m such that $pq = mcn^2+1$. Then

$$\begin{pmatrix} 1 & 0 \\ cn & 1 \end{pmatrix}\begin{pmatrix} 1 & mn \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & mn \\ cn & pq \end{pmatrix} \in G.$$

By Lemma 4.2(d), any application of property (3) for our choice of p to this matrix requires $l = 1$, so G contains

$$\begin{pmatrix} p & k(1-p)p/C + mn \\ cn & k(1-p)cn/pC + q \end{pmatrix},$$

with $(k(1-p)cn)/pC \in \mathbb{Z}$. By our assumptions on p, $(1-p)cn$ is coprime to p, so $k = 0$. Therefore, G contains $\left(\begin{smallmatrix} p & mn \\ cn & q \end{smallmatrix}\right)$, and this is the desired double coset representative. □

We say that a function on $\mathfrak{H}$ is of trigonometric type if after some transformation $\tau \mapsto a\tau + b$, it has the form $q^{-1} + a_0 + \zeta q$, for ζ a root of unity or zero.

Theorem 4.6. *Let f be weakly Hecke-monic for (g, h), and suppose $f(g, h, \tau)$ has a pole at infinity, and q-expansion coefficients that are algebraic integers. Then $f(g, h, \tau)$ is either of trigonometric type or holomorphic congruence genus-zero.*

Proof. By Proposition 2.3, the q-expansion of $f(g, h, \tau)$ has the form $\zeta q^{C/|g|} + O(1) \in \overline{\mathbb{Q}}((q^{-C/|g|}))$ for some root of unity ζ and some negative integer C. By Proposition 3.3, $f(g, h, \tau)$ satisfies equivariant modular equations for all primes p satisfying $g^p = g$, $h^p = h$. Following the proof of [Cummins and Gannon 1997, Lemma 7.1] (changing $(az+b)/d$ to $(az+b(1-d))/d$ and $\mathbb{Z}$ to $(|g|/C)\mathbb{Z}$), we find that $f(g, h, \tau)$ is invariant under a discrete subgroup of $\mathrm{SL}_2(\mathbb{R})$. By Proposition 3.5, $f(g, h, \tau)$ admits global symmetries, and in particular, an altered version of [Cummins and Gannon 1997, Lemma 3.2] holds, where $A(p)$ is replaced by the equivariant version. In summary, the group G of global symmetries of $f(g, h, \tau)$ satisfies the following three conditions:

(1) G is a discrete group.

(2) The stabilizer of infinity $G_\infty \subset G$ is $\left\langle -\mathrm{Id}, \left(\begin{smallmatrix} 1 & |g|/C \\ 0 & 1 \end{smallmatrix}\right)\right\rangle$.

(3) For all primes p congruent to 1 mod $\mathrm{lcm}(|g|, |h|)$, and all $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in G$, there exist integers l and k such that $l \mid p$, $0 \le -k < p/l$, and such that

$$\begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} l & k(p-1) \\ 0 & p/l \end{pmatrix}^{-1} = \begin{pmatrix} ap/l & k(1-p)a + lb \\ c/l & (k(1-p)c + ld)/p \end{pmatrix} \in G.$$

We now consider the function $f(g, h, -\tau/C)$, which is a power series in $q^{1/|g|}$. Let G' denote the subgroup of $\mathrm{SL}_2(\mathbb{R})$ that fixes $f(g, h, -\tau/C)$, so

$$G' = \left(\begin{smallmatrix} -1/C & 0 \\ 0 & 1 \end{smallmatrix}\right) G \left(\begin{smallmatrix} -C & 0 \\ 0 & 1 \end{smallmatrix}\right).$$

The quadruple $(G', |g|, \mathrm{lcm}(|g|, |h|), C)$ then satisfies conditions (1)–(3).

If G' does not fix infinity, then by the previous lemmata, G' contains $\Gamma(n)$ for some n, and G contains some congruence group. Therefore $f(g, h, \tau)$ is a holomorphic congruence genus-zero function.

If G' fixes infinity, then $G = G_\infty = \left\langle -\mathrm{Id}, \left(\begin{smallmatrix} 1 & |g|/C \\ 0 & 1 \end{smallmatrix}\right) \right\rangle$. The proof that $f(g, h, \tau)$ is of trigonometric type is given by following the first half of the proof of [Cummins and Gannon 1997, Lemma 7.2], and replacing modular equations with equivariant modular equations, and q with $q^{-C/|g|}$. □

Remark. The proof of [Cummins and Gannon 1997, Lemma 7.2] allows us to make a slightly stronger statement: If we don't necessarily have algebraic integer coefficients, but we know that $f(g, h, \tau) = \zeta q^{C/|g|} + \sum_{n \geq 0} a_n q^{-nC/|g|}$ satisfies $\sum n\,|a_n|^2 > 1$, then $f(g, h, \tau)$ is a holomorphic congruence genus-zero function.

5. Replicability

In this paragraph, we summarize some results and assertions in [Norton 1984]. One can start with a formal power series $f(q) = q^{-1} + \sum_{n>0} a_n q^n$ and define numbers $H_{m,n}$ for $m, n \in \mathbb{Z}_{>0}$ by the bivarial transform:

$$\log \frac{f(p) - f(q)}{p^{-1} - q^{-1}} = -\sum_{m,n=1}^{\infty} H_{m,n} p^m q^n.$$

For each $n > 0$, there is a unique normalized Faber polynomial $\Phi_n(x)$ (depending on f), defined by the property that $\Phi_n(f(q)) = q^{-n} + O(q)$. The polynomial in $f(q)$ that is n times the coefficient of p^n in the formal series $-\log p(f(p) - f(q))$ also has this form, so we have $\Phi_n(f(q)) = q^{-n} + n \sum_m H_{m,n} q^m$. We say that $f(q)$ is replicable if and only if for any $t > 0$ there exists a series $f^{(t)}(q) = q^{-1} + O(q)$ such that $\Phi_n(f(q)) = \sum_{ad=n, 0 \leq b < d} f^{(a)}((a\tau + b)/d)$. The series $f^{(t)} = \sum_{n>0} a_n^{(t)} q^n$ is called the t-th replicate of f, and by suitable use of induction, one can show that it is unique if it exists, and its coefficients satisfy the relation

$$H_{m,n} = \sum_{t \mid (m,n)} \frac{1}{t} a^{(t)}_{mn/t^2}$$

In particular, if f is replicable, $H_{m,n}$ only depends on (m, n) and mn. Another induction argument implies the converse of this, that is, that one can define replicability by this independence.

Note. Replicability was originally defined only for power series with rational integer coefficients, and for more general series, there is some disagreement in the literature regarding the correct definition. Norton has proposed a definition of replicability for series that have irrational cyclotomic integer coefficients, and it

seems to involve a Galois action. This is potentially useful when considering functions invariant under some $\Gamma_0(N)$. One can instead extend one of the integral definitions above to allow arbitrary complex coefficients, without alteration of the formulas, and this was done in [Kozlov 1994]. We use the latter generalization of the condition here, because we can prove something about series that satisfy it.

Definition. A replicable function is a replicable power series that converges on the open unit q-disc, that is, one that expands to a holomorphic function on $\mathfrak{H}$. A replicable function has order n if $f^{(m)} = f^{(m+n)}$ for all $m > 0$.

We note that a replicable function of finite order has a unique minimal order, but not a unique order. If f is a replicable function, then all $\Phi_m(f)$ are holomorphic on $\mathfrak{H}$, and hence all $f^{(m)}$ are also holomorphic on $\mathfrak{H}$.

We would like to relate replicability to Hecke-monicity.

Lemma 5.1. *If f is a weakly Hecke-monic function for $(1, g)$ such that*

$$f(1, g, \tau) = q^{-1} + O(q),$$

then $f(1, g^m, \tau)$ has the form $q^{-1} + O(1)$, and is uniquely defined by the Hecke-monic property up to a constant.

Proof. For the purposes of induction, we assume $f(1, g^k, \tau) = q^{-1} + O(1)$ for all $k < m$. Then

$$\begin{aligned} mT_m f(1, g, \tau) &= \sum_{\substack{ad=m \\ 0\le b<d}} f\left(1, g^a, \frac{a\tau+b}{d}\right) \\ &= f(1, g^m, m\tau) + \sum_{\substack{d\mid m \\ d<m}} \sum_{0\le b<d} e(b/d) q^{m/d^2} + O(1). \end{aligned}$$

Since $mT_m f(1, g, \tau)$ is monic of degree m in $f(1, g, \tau)$, the leading term is q^{-m}, and all of the other summands have poles of lower order. By subtracting those summands, we find that the leading term of $f(1, g^m, m\tau)$ is q^{-m}, so $f(1, g^m, \tau)$ has leading term q^{-1}. Since $f(1, g^m, \tau)$ is a power series in q, it has the form we want.

To show uniqueness, suppose there were some $f'(1, g^m, \tau) = q^{-1} + O(1)$ such that $f'(1, g^m, m\tau) + \sum_{ad=m,\, d<m,\, 0\le b<d} f(1, g^d, (a\tau + b)/d)$ is monic of degree m in $f(1, g, \tau)$. Since this sum and $mT_m f(1, g, \tau)$ have the same coefficients in negative degree, $f'(1, g^m, m\tau) - f(1, g^m, m\tau) = O(1)$. However, this difference must be a polynomial in $f(1, g, \tau)$, so it is constant. □

Lemma 5.2. *If f is a weakly Hecke-monic function for $(1, g)$ such that*

$$f(1, g^m, \tau) = q^{-1} + O(q)$$

for all $m > 0$, then $nT_n f(1, g, \tau)$ is the unique polynomial in $f(1, g, \tau)$ whose expansion is $q^{-n} + O(q)$.

Proof.
$$\begin{aligned} nT_n f(1, g, \tau) &= \sum_{\substack{ad=n \\ 0\le b<d}} f\Big(1, g^a, \frac{a\tau+b}{d}\Big) \\ &= \sum_{d\mid n} \sum_{0\le b<d} e\Big(\frac{-(n/d)\tau-b}{d}\Big) + O(q^{1/n}) \\ &= \sum_{d\mid n} e(-n\tau/d^2) \sum_{0\le b<d} e(-b/d) + O(q^{1/n}) \\ &= \sum_{d\mid n} e(-n\tau/d^2)\delta_{d,1} + O(q^{1/n}) = q^{-n} + O(q^{1/n}). \end{aligned}$$

Because $f(1, g, \tau)$ is a power series in q and $nT_n f(1, g, \tau)$ is a polynomial in $f(1, g, \tau)$, we can refine the $O(q^{1/n})$ to $O(q)$. If we add any other polynomial in $f(1, g, \tau)$, the leading term will yield a nontrivial contribution to the nonpositive powers in the expansion, so the polynomial is unique. □

Proposition 5.3. *The map $f^{(m)}(\tau) \mapsto f(1, g^m, \tau)$ induces a bijection between replicable functions of order N and weakly Hecke-monic functions for $(1, g)$ on*

$$\mathrm{Hom}(\mathbb{Z}\times\mathbb{Z}, \mathbb{Z}/N\mathbb{Z}) \underset{\pm\mathbb{Z}}{\times} \mathfrak{H}$$

whose expansions at infinity have the form $q^{-1} + O(q)$, where g is a generator of $\mathbb{Z}/N\mathbb{Z}$.

Proof. We first assume that $f^{(1)}$ is replicable, so $\Phi_n(f^{(1)}) = \sum_{\substack{ad=n \\ 0\le b<d}} f^{(a)}\Big(\frac{a\tau+b}{d}\Big)$.

Then

$$\Phi_n(f^{(1)}(\tau)) = \sum_{\substack{ad=n \\ 0\le b<d}} f^{(a)}\Big(\frac{a\tau+b}{d}\Big) = \sum_{\substack{ad=n \\ 0\le b<d}} f\Big(1, g^a, \frac{a\tau+b}{d}\Big) = nT_n f(1, g, \tau).$$

Therefore, $nTnf(1, g, \tau)$ is a monic polynomial in $f(1, g, \tau)$ for all n, and f is weakly Hecke-monic for $(1, g)$.

Now let f be a weakly Hecke-monic function for $(1, g)$ satisfying $f(1, g^i, \tau) = q^{-1} + O(q)$. By Lemma 5.2, $mT_m f(1, g, \tau) = q^{-m} + O(q)$, and is a monic polynomial in $f(1, g, \tau)$, so it is equal to $q^{-m} + m\sum_k H_{k,m} q^k = \Phi_n(f)$. If we assume for the purposes of induction that $f(1, g^k, \tau) = f^{(k)}(\tau)$ for all $k \mid m$, $k \ne m$, then

$$\sum_{\substack{ad=m \\ 0\le b<d}} f\Big(1, g^a, \frac{a\tau+b}{d}\Big) = mT_m f(1, g, \tau) = \Phi_m(f) = \sum_{\substack{ad=m \\ 0\le b<d}} f^{(a)}\Big(\frac{a\tau+b}{d}\Big)$$

implies $f^{(m)} = f(1, g^m, \tau)$. □

Corollary 5.4. *If f is a replicable function of finite order with algebraic integer coefficients, then f is either of trigonometric type or holomorphic congruence genus-zero for a group containing $\Gamma_1(N)$ for some N.*

Proof. By the above proposition, f together with its replicates forms a weakly Hecke-monic function for $(1, g)$, where g generates a cyclic group whose order is that of f. By Theorem 4.6, $f(1, g, \tau)$ is either of trigonometric type or holomorphic congruence genus-zero and invariant under a group containing $\Gamma(N)$ for some N. Since f is invariant under translation by 1, it is invariant under $\Gamma_1(N)$. □

Norton [1984] also defined a stronger notion: f is completely replicable if all $f^{(t)}$ are replicable, or equivalently, if the s-th replication power of $f^{(t)}$ is $f^{(st)}$ for all s and t [Kozlov 1994, Proposition 2.5]. He also pointed out that

$$-J(z+\tfrac{1}{2}) = q^{-1} + 196884q - 21493760q^2 + \cdots$$

is a function that is replicable but not completely replicable.

Corollary 5.5. *The above bijection specializes to a bijection between completely replicable functions of order N and semiweakly Hecke-monic functions for $(1, g)$ on*

$$\mathrm{Hom}(\mathbb{Z}\times\mathbb{Z}, \mathbb{Z}/N\mathbb{Z}) \underset{\pm\mathbb{Z}}{\times} \mathfrak{H}$$

whose expansions at infinity have the form $q^{-1} + O(q)$.

Proof. Using the proposition, we get a chain of equivalent statements:

- $f^{(1)}$ is completely replicable.
- $f^{(m)}$ is replicable for all m.
- f is weakly Hecke-monic for all $(1, g^m)$.
- f is semiweakly Hecke-monic for $(1, g)$. □

Corollary 5.6. *The above bijection specializes to a bijection between completely replicable functions $f^{(1)}$ with rational integer coefficients invariant under $\Gamma_0(N)$ and Hecke-monic functions f on $\mathcal{M}_{\mathrm{Ell}}^{\mathbb{Z}/N\mathbb{Z}}$ satisfying the property that the q-expansions of $f(1, g^i, \tau)$ have the form $q^{-1} + O(q)$, with rational integer coefficients.*

Proof. It suffices to show that if $f^{(1)}$ is invariant under $\Gamma_0(N)$, then $f^{(m)}$ is invariant under $\Gamma_0(N/(m, N))$. The completely replicable functions with integer coefficients were exhaustively enumerated in [Alexander et al. 1992], and their fixing groups were found to obey this condition in [Ferenbaugh 1993]. □

Replicable functions without a specified order also have an interpretation in terms of Hecke-monicity, if we allow our group G to be infinite. If we let g generate

a copy of $\mathbb{Z}$, we can think of replicable functions together with their replicable powers as weakly Hecke-monic functions for $(1, g)$ on

$$\mathrm{Hom}(\mathbb{Z} \times \mathbb{Z}, \mathbb{Z}) \underset{\pm\mathbb{Z}}{\times} \mathfrak{H}.$$

Unfortunately, the finite order condition is essential for our techniques to produce a genus-zero statement.

6. twisted denominator formulas

Given a Lie algebra of a rather specialized form described below, we can make strong statements about certain characters of automorphisms acting on homology. When this Lie algebra arises from conformal field theory in a certain way, we show that in fact the characters are holomorphic congruence genus-zero functions. The particular constraints on the Lie algebra force it to be "mostly free" in the sense that its higher homology is very small. This is somewhat related to work of Jurisich [1998] on free Lie subalgebras of generalized Kac–Moody algebras like the monster Lie algebra. Some connections to elliptic cohomology appear in unpublished work of Lurie [2005] concerning exponential operations on elliptic λ-rings.

Let G be a finite group, and let g be an element of order N in the center of G. Suppose we have a collection

$$\mathcal{V} = \{V_k^{i,j/N} : i, j \in \mathbb{Z}/N\mathbb{Z}, k \in (1/N)\mathbb{Z}\}$$

of G-modules, such that the action of g on $V_k^{i,j/N}$ is given by constant multiplication by the root of unity $e(j/N)$, and such that $\dim V_k^{i,j/N}$ grows subexponentially with k, that is, for any $\epsilon > 0$, there is some $C > 0$ such that $\dim V_k^{i,j/N} < Ce^{\epsilon k}$ for all i, j, k.

Note. We occasionally write $V_k^{i,j/N}$ where i and j are given as integers, tacitly reducing modulo N, so $V_k^{i,j/N}$ is the same G-module as $V_k^{i+aN,j/N+b}$ for all integers a and b.

Definition. A complex Lie algebra E is Fricke compatible with $\mathcal{V}$ if the following conditions are satisfied:

- E is graded by $\mathbb{Z}_{>0} \times \frac{1}{N}\mathbb{Z}$, with finite- dimensional homogeneous components $E_{i,j}$. We introduce degree indicator symbols p and q, which denote grading shifts by $(1, 0)$ and $(0, \frac{1}{N})$, respectively, and write the graded vector space decomposition as $E = \bigoplus_{i>0,\, j\in(1/N)\mathbb{Z}} E_{i,j}\, p^i q^j$. We can view this as a character decomposition of E under an action of a two-dimensional torus H.
- E admits a homogeneous action of G by Lie algebra automorphisms, such that we have G-module isomorphisms $E_{i,j} \cong V_{1+ij}^{i,j}$.

- The homology of E is given by $H_0(E) = \mathbb{C}$,

$$H_1(E) = \bigoplus_{n\in(1/N)\mathbb{Z}} V^{1,n}_{1+n} pq^n, \quad H_2(E) = p\bigoplus_{m=0}^{\infty} V^{1,-1/N}_{1-1/N} \otimes V^{m,1/N}_{1+m/N} p^m,$$

  and $H_i(E) = 0$ for $i > 2$.
- $E_{1,-1/N} \cong V^{1,-1/N}_{1-1/N}$ is one-dimensional.

Remark. Our use of the term "Fricke compatible" is motivated by considerations from conformal field theory. If g is a Fricke element of the monster, that is, if the McKay–Thompson series $T_g(\tau) = \mathrm{Tr}(gq^{L_0-1}|V^\natural)$ is invariant under the transformation $\tau \mapsto -1/N\tau$ for some N, and if G is a central extension of the centralizer of g in the monster, then we expect the Lie algebra of physical states of the g-orbifold intertwiner algebra to be a generalized Kac–Moody algebra whose positive subalgebra is isomorphic to E as a Lie algebra with a homogeneous action of G by automorphisms, and we expect the unique irreducible g-twisted module of $V^\natural$ to be isomorphic to $H_1(E)$ as a graded G-module. If g is a non-Fricke element, then we expect the compatible Lie algebra to have a large abelian subalgebra and higher homology described by its exterior powers. We explore this further in [Carnahan 2009; ≥ 2010].

Proposition 6.1 (Twisted denominator formula). *Suppose E is Fricke compatible with $\mathcal{V}$. Then for any $h \in G$,*

$$p^{-1} + \sum_{m>0} \mathrm{Tr}\left(h|V^{1,-1/N}_{1-1/N}\right) \mathrm{Tr}\left(h|V^{m,1/N}_{1+m/N}\right) p^m - \sum_{n\in(1/N)\mathbb{Z}} \mathrm{Tr}\left(h|V^{1,n}_{n+1}\right) q^n$$
$$= p^{-1} \exp\left(-\sum_{i>0} \sum_{\substack{m>0\\ n\in(1/N)\mathbb{Z}}} \mathrm{Tr}\left(h^i|V^{m,n}_{1+mn}\right) p^{im} q^{in}/i\right).$$

Proof. This is essentially identical to [Borcherds 1992, Section 8]. The Chevalley–Eilenberg resolution yields the equation $H(E) = \bigwedge(E)$ of virtual $H \times G$-representations, and the left side is given by taking traces on the homology groups given above. By Adams' exponential formula from K-theory, we have

$$\bigwedge(U) = \exp\left(-\sum_{i>0} \psi^i(U)/i\right)$$

for any finite-dimensional $H \times G$-module U (which we take to be the homogeneous components $E_{i,j}$ or finite sums thereof). The ψ^i are the i-th Adams operations, which satisfy the identity $\mathrm{Tr}(g|\psi^i(U)) = \mathrm{Tr}(g^i|U)$. The right side of the equation is then given by extending this to a formal sum on the infinite-dimensional direct sum of homogeneous components, and this is allowed because their degrees are supported in a strict half-space. □

For any $h \in G$, we define *formal orbifold partition functions*:

$$Z(g^k, g^l h^m, \tau) := \sum_{n \in (1/N)\mathbb{Z}} \sum_{\substack{r \in (1/N)\mathbb{Z}/\mathbb{Z} \\ n \in kr + \mathbb{Z}}} \operatorname{Tr}\left(g^l h^m | V_{1+n}^{k,r}\right) e(n\tau).$$

We refer to the collection of these functions as Z, and they converge on $\mathfrak{H}$, by the subexponential growth condition. We can then define equivariant Hecke operators:

$$T_n Z(g, h, \tau) = \frac{1}{n} \sum_{\substack{ad=n \\ 0 \le b < d}} Z\left(g^d, g^{-b} h^a, \frac{a\tau + b}{d}\right).$$

Proposition 6.2. *Suppose E is Fricke compatible with $\mathcal{V}$. Then Z is weakly Hecke-monic for (g, h).*

Proof. We multiply both sides of the twisted denominator formula by p, and viewing the equality as an identification of formal expansions, we take logarithms.

$$\begin{aligned}
&\log\left(1 - p \sum_{n \in (1/N)\mathbb{Z}} \operatorname{Tr}\left(h | V_{1+n}^{1,n}\right) q^n + \sum_{m>0} \operatorname{Tr}\left(h | V_{1-1/N}^{1,-1/N}\right) \operatorname{Tr}\left(h | V_{1+m/N}^{m,1/N}\right) p^{m+1}\right) \\
&\quad = -\sum_{i>0} \sum_{\substack{m>0 \\ n \in (1/N)\mathbb{Z}}} \operatorname{Tr}\left(h^i | V_{1+mn}^{m,n}\right) p^{im} q^{in}/i \\
&\quad = -\sum_{m>0} \sum_{a \mid m} \frac{1}{a} \sum_{n \in (1/N)\mathbb{Z}} \operatorname{Tr}\left(h^a | V_{1+mn/a}^{m/a,n}\right) p^m q^{an} \\
&\quad = -\sum_{m>0} \sum_{ad=m} \frac{1}{a} \sum_{0 \le b < d} \frac{1}{d} \sum_{n \in (1/N)\mathbb{Z}} \operatorname{Tr}\left(h^a | V_{1+dn}^{d,n}\right) p^m q^{an} \\
&\quad = -\sum_{m>0} \sum_{ad=m} \frac{1}{a} \sum_{0 \le b < d} \frac{1}{d} \sum_{n \in (1/N)\mathbb{Z}} \sum_{\substack{r \in (1/N)\mathbb{Z}/\mathbb{Z} \\ n \in dr + \mathbb{Z}}} e(-br) \operatorname{Tr}\left(h^a | V_{1+n}^{d,r}\right) e(br) q^{an/d} p^m \\
&\quad = -\sum_{m>0} \frac{1}{m} \sum_{\substack{ad=m \\ 0 \le b < d}} \sum_{n \in (1/N)\mathbb{Z}} \sum_{\substack{r \in (1/N)\mathbb{Z}/\mathbb{Z} \\ n \in dr + \mathbb{Z}}} \operatorname{Tr}\left(g^{-b} h^a | V_{1+n}^{d,r}\right) e\left(n \frac{a\tau + b}{d}\right) p^m \\
&\quad = -\sum_{m>0} \frac{1}{m} \sum_{\substack{ad=m \\ 0 \le b < d}} Z\left(g^d, g^{-b} h^a, \frac{a\tau + b}{d}\right) p^m = -\sum_{m>0} T_m Z(g, h, \tau) p^m.
\end{aligned}$$

Isolating the terms that are degree k in p on the first line yields a polynomial of degree k in $Z(g, h, \tau)$, with leading coefficient $-1/k$. This implies that for all k, $kT_k Z(g, h, \tau)$ as a formal q-series is a monic polynomial of degree k in the q-expansion of $Z(g, h, \tau)$. All of the formal orbifold partition functions uniquely define holomorphic functions on $\mathfrak{H}$, so for all k, $kT_k Z(g, h, \tau)$ is a monic polynomial of degree k in $Z(g, h, \tau)$, where they are viewed as functions on $\mathfrak{H}$. $\square$

We describe a connection to generalized moonshine. Recall that one of the key hypotheses in the conjecture was the existence of certain representations of central extensions of centralizers of elements. An interpretation of these representations was given in [Dixon et al. 1988], where they were said to be twisted Hilbert spaces of an orbifold conformal field theory. In our language, these are twisted modules of the vertex operator algebra $V^\natural$. The theoretical details of vertex operator algebras and twisted modules are outside the scope of this paper, but we can think of these objects as graded vector spaces, where the grading is given by eigenvalues of a semisimple operator L_0. When a vertex operator algebra has a unique irreducible g-twisted module for some automorphism g, Schur's lemma produces a natural action of some central extension of the centralizer of g on the twisted module. The two facts we need concerning twisted modules are from [Dong et al. 2000]:

(1) (Theorem 10.3) If V is a holomorphic C_2-cofinite vertex operator algebra with central charge 24, and g is a conformal automorphism of finite order, then there exists a unique irreducible g-twisted module $V(g)$ up to isomorphism.

(2) (Theorems 5.4, 6.4, and 8.1) Let $M > 0$ satisfy $g^M = h^M = 1$, and suppose we have a G-module isomorphism

$$V(g^i) \cong \bigoplus_{k\in(1/N)\mathbb{Z}} \bigoplus_{j\in\mathbb{Z}/N\mathbb{Z}} V_k^{i,j/N},$$

where $V(g^i)$ is the irreducible g^i-twisted module, and the outer sum gives the L_0-eigenvalue decomposition. For any $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$,

$$Z\left(g^i, h, \frac{a\tau+b}{c\tau+d}\right)$$

lies in a certain space of holomorphic functions on $\mathfrak{H}$, each element of which is annihilated by a differential operator of the form

$$\left(\frac{d}{d\tau}\right)^m + \sum_{j=0}^{m-1} r_j(q)\left(\frac{d}{d\tau}\right)^j,$$

where $m > 0$ and $r_j(q) \in \mathbb{C}[\![q^{1/M}]\!]$ converges on $\mathfrak{H}$.

Proposition 6.3. *Suppose that E is a Lie algebra Fricke compatible with $\mathcal{V}$, and suppose that G acts conformally on a holomorphic C_2-cofinite vertex operator algebra V of central charge 24, such that for all $i \in \mathbb{Z}/N\mathbb{Z}$, we have G-module isomorphisms $V(g^i) \cong \bigoplus_{k\in(1/N)\mathbb{Z}} \bigoplus_{j\in\mathbb{Z}/N\mathbb{Z}} V_k^{i,j/N}$ as in Fact (2). Then $Z(g,h,\tau)$ is a holomorphic congruence genus-zero function.*

Proof. By Proposition 6.2, Z is weakly Hecke-monic for (g, h). Since $E_{1,-1/N} = V_{1-1/N}^{1,-1/N}$ is one-dimensional, the trace of h on this space is nonzero, so $Z(g, h, \tau)$ has a pole at infinity. By Theorem 4.6, $Z(g, h, \tau)$ is then either a holomorphic congruence genus-zero function, or of trigonometric type. However, the expansion of any function of trigonometric type at a cusp other than infinity is not annihilated by any differential operator of the form given in Fact (2). □

The hypotheses for this proposition are quite strong, but it is not a vacuous statement. When $G = \mathbb{M}$ and $g = 1$, this implies the McKay–Thompson series are holomorphic congruence genus-zero modular functions, assuming the positive subalgebra of the monster Lie algebra is Fricke compatible with $V^\natural$. This compatibility was proved in [Borcherds 1992, Section 8]. When $G = 2.B$, the nontrivial central extension of the baby monster simple group, and g is the central element of order two, this yields holomorphic congruence genus-zero characters for the conjugacy class 2A case of generalized moonshine, assuming there exists a Lie algebra Fricke compatible with the suitable twisted modules. The holomorphic congruence genus-zero result was proved in [Höhn 2003] using a construction of a Fricke compatible Lie algebra, and Proposition 6.3 allows one to eliminate the explicit computations in the final step of the proof, which involved matching the first 25 coefficients of the character for every conjugacy class of G with Norton's list of known replicable functions.

Acknowledgments

The author thanks Richard Borcherds, Gerald Höhn, Jacob Lurie, and Arne Meurman. Borcherds suggested generalized moonshine as a dissertation project, and gave much useful advice and perspective. Höhn offered many helpful comments on an earlier draft, from which this paper was drawn. Lurie provided inspiring conversations, and suggested that twisted denominator formulas appear to be constructed from equivariant Hecke operators. Meurman kindly mailed a copy of Kozlov's thesis [1994] across the Atlantic Ocean. The author would also like to thank the referees for many helpful suggestions and corrections, and an anonymous member of the editorial board for suggestive hints concerning complex analytic moduli.

References

[Alexander et al. 1992] D. Alexander, C. Cummins, J. McKay, and C. Simons, "Completely replicable functions", pp. 87–98 in *Groups, combinatorics and geometry* (Durham, 1990), edited by M. Liebeck and J. Saxl, London Math. Soc. Lecture Note Ser. **165**, Cambridge University Press, 1992. MR 94g:11029 Zbl 0831.11032

[Ando 1995] M. Ando, "Isogenies of formal group laws and power operations in the cohomology theories E_n", *Duke Math. J.* **79**:2 (1995), 423–485. MR 97a:55006 Zbl 0862.55004

[Baker 1998] A. Baker, "Hecke algebras acting on elliptic cohomology", pp. 17–26 in *Homotopy theory via algebraic geometry and group representations* (Evanston, IL, 1997), edited by M. Mahowald and S. Priddy, Contemp. Math. **220**, Amer. Math. Soc., Providence, RI, 1998. MR 99h:55004 Zbl 0911.55002

[Behrend and Noohi 2006] K. Behrend and B. Noohi, "Uniformization of Deligne-Mumford curves", *J. Reine Angew. Math.* **599** (2006), 111–153. MR 2007k:14017 Zbl 1124.14004

[Borcherds 1992] R. E. Borcherds, "Monstrous moonshine and monstrous Lie superalgebras", *Invent. Math.* **109**:2 (1992), 405–444. MR 94f:11030 Zbl 0799.17014

[Carnahan 2009] S. Carnahan, "Generalized moonshine II: Borcherds products", preprint, 2009. arXiv 0908.4223

[Carnahan ≥ 2010] S. Carnahan, "Generalized moonshine, V: Comparisons", in preparation.

[Conrad ≥ 2010] B. Conrad, *Modular forms and the Ramanujan conjecture*, Cambridge University Press, New York. To appear.

[Conway and Norton 1979] J. H. Conway and S. P. Norton, "Monstrous moonshine", *Bull. London Math. Soc.* **11**:3 (1979), 308–339. MR 81j:20028 Zbl 0424.20010

[Cummins 2002] C. J. Cummins, "Modular equations and discrete, genus-zero subgroups of $\mathrm{SL}(2,\mathbb{R})$ containing $\Gamma(N)$", *Canad. Math. Bull.* **45**:1 (2002), 36–45. MR 2002k:11057 Zbl 1044.11028

[Cummins and Gannon 1997] C. J. Cummins and T. Gannon, "Modular equations and the genus zero property of moonshine functions", *Invent. Math.* **129**:3 (1997), 413–443. MR 98k:11046 Zbl 0985.11017

[Deligne 1971] P. Deligne, "Formes modulaires et représentations ℓ-adiques", in *Séminaire Bourbaki* 1968/1969 (Exposé 355), Springer, Berlin, 1971. Zbl 0206.49901

[Dixon et al. 1988] L. Dixon, P. Ginsparg, and J. Harvey, "Beauty and the beast: superconformal symmetry in a Monster module", *Comm. Math. Phys.* **119**:2 (1988), 221–241. MR 90b:81119 Zbl 0657.17011

[Dong et al. 2000] C. Dong, H. Li, and G. Mason, "Modular-invariance of trace functions in orbifold theory and generalized moonshine", *Comm. Math. Phys.* **214**:1 (2000), 1–56. MR 2001k:17043 Zbl 1061.17025

[Ferenbaugh 1993] C. R. Ferenbaugh, "The genus-zero problem for $n|h$-type groups", *Duke Math. J.* **72**:1 (1993), 31–63. MR 94i:20030 Zbl 0828.20016

[Frenkel et al. 1988] I. Frenkel, J. Lepowsky, and A. Meurman, *Vertex operator algebras and the Monster*, Pure and Applied Mathematics **134**, Academic Press, Boston, MA, 1988. MR 90h:17026 Zbl 0674.17001

[Ganter 2009] N. Ganter, "Hecke operators in equivariant elliptic cohomology and generalized moonshine", pp. 173–209 in *Groups and symmetries*, edited by J. Harnad and P. Winternitz, CRM Proc. Lecture Notes **47**, Amer. Math. Soc., Providence, RI, 2009. MR 2500561 Zbl 05657063

[Höhn 2003] G. Höhn, "Generalized moonshine for the baby monster", preprint, 2003, available at http://www.math.ksu.edu/~gerald/papers/baby8.ps.

[Jurisich 1998] E. Jurisich, "Generalized Kac–Moody Lie algebras, free Lie algebras and the structure of the Monster Lie algebra", *J. Pure Appl. Algebra* **126**:1-3 (1998), 233–266. MR 99b:17032 Zbl 0898.17011

[Katz and Mazur 1985] N. M. Katz and B. Mazur, *Arithmetic moduli of elliptic curves*, Annals of Mathematics Studies **108**, Princeton University Press, 1985. MR 86i:11024 Zbl 0576.14026

[Kozlov 1994] D. Kozlov, *On completely replicable functions*, Master's thesis, Lund University, 1994.

[Lurie 2005] J. Lurie, personal communication, 2005.

[Martin 1996] Y. Martin, "On modular invariance of completely replicable functions", pp. 263–286 in *Moonshine, the Monster, and related topics* (South Hadley, MA, 1994), edited by C. Dong and G. Mason, Contemp. Math. **193**, Amer. Math. Soc., Providence, RI, 1996. MR 96m:11032 Zbl 0860.11023

[Mason 1987] G. Mason, "Finite groups and modular functions", pp. 181–210 in *The Arcata Conference on Representations of Finite Groups* (Arcata, CA, 1986), edited by P. Fong, Proc. Sympos. Pure Math. **47**, Amer. Math. Soc., Providence, RI, 1987. With an appendix by S. P. Norton. MR 89c:11066 Zbl 0642.10025

[Norton 1984] S. P. Norton, "More on moonshine", pp. 185–193 in *Computational group theory* (Durham, 1982), edited by M. D. Atkinson, Academic Press, London, 1984. MR 86h:20020 Zbl 0541.20009

[Norton 2001] S. Norton, "From Moonshine to the Monster", pp. 163–171 in *Proceedings on moonshine and related topics* (Montréal, 1999), edited by J. McKay and A. Sebbar, CRM Proc. Lecture Notes **30**, Amer. Math. Soc., Providence, RI, 2001. MR 2002k:20025 Zbl 1009.20023

[Queen 1981] L. Queen, "Modular functions arising from some finite groups", *Math. Comp.* **37**:156 (1981), 547–580. MR 83d:20008 Zbl 0514.20009

[Shimura 1971] G. Shimura, *Introduction to the arithmetic theory of automorphic functions*, Kanô Memorial Lectures **1**, Iwanami Shoten, Tokyo, 1971. MR 47 #3318 Zbl 0221.10029

Communicated by Edward Frenkel
Received 2008-12-28 Revised 2009-10-02 Accepted 2010-01-05

carnahan@math.mit.edu *Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, United States*
http://math.mit.edu/~carnahan/

ALGEBRA AND NUMBER THEORY 4:6(2010)

Integral trace forms associated to cubic extensions

Guillermo Mantilla-Soler

Given a nonzero integer d, we know by Hermite's Theorem that there exist only finitely many cubic number fields of discriminant d. However, it can happen that two nonisomorphic cubic fields have the same discriminant. It is thus natural to ask whether there are natural refinements of the discriminant which completely determine the isomorphism class of the cubic field. Here we consider the trace form $q_K : \mathrm{tr}_{K/\mathbb{Q}}(x^2)|_{O_K^0}$ as such a refinement. For a cubic field of fundamental discriminant d we show the existence of an element T_K in Bhargava's class group $\mathrm{Cl}(\mathbb{Z}^2 \otimes \mathbb{Z}^2 \otimes \mathbb{Z}^2; -3d)$ such that q_K is completely determined by T_K. By using one of Bhargava's composition laws, we show that q_K is a complete invariant whenever K is totally real and of fundamental discriminant.

1.	Introduction	681
2.	Basic facts	684
3.	Galois fields and rational 3-torsion	685
4.	Cubic fields with fundamental discriminant	688
5.	Trace form and class groups	690
6.	From cubic fields to cubes and trace forms	694
	Acknowledgements	698
	References	698

1. Introduction

Generalities. A difference between quadratic and nonquadratic number fields is that the former are totally characterized by their discriminant. One natural choice for a *refined discriminant* is given by the isometry class with respect to the trace form of the lattice defined by the maximal order. The purpose of this paper is to give a detailed analysis of this refinement for cubic extensions, and to show under which conditions this refinement characterizes the field. Given a number field K with maximal order O_K, we consider the trace form $\mathrm{tr}_{K/\mathbb{Q}}(x^2)|_{O_K}$.

MSC2000: primary 11E12; secondary 11R29, 11R16, 11E76.

Keywords: integral trace forms, cubic fields, Bhargava's class group, discriminants of number fields.

Question 1.1. *Do there exist two nonisomorphic number fields K and L such that their corresponding trace forms are isomorphic*?

In this paper we analyze this question in the case of cubic extensions.

Definition 1.2. Let K be a number field and let O_K be its maximal order. The *trace zero module* O_K^0 is the set $\{x \in O_K : \mathrm{tr}_{K/\mathbb{Q}}(x) = 0\}$.

Our main result is this:

Theorem 6.5. *Let K be a cubic number field of positive, fundamental discriminant. Let L be a number field such that there exists an isomorphism of quadratic modules*

$$\langle O_K^0, \mathrm{tr}_{K/\mathbb{Q}}(x^2)|_{O_K^0}\rangle \cong \langle O_L^0, \mathrm{tr}_{L/\mathbb{Q}}(x^2)|_{O_L^0}\rangle,$$

and assume $9 \nmid d_L$. Then $K \cong L$.

Outline of the paper. We start by analyzing Question 1.1 for general cubic fields. For this purpose we consider first the case in which the common discriminant of K and L is not fundamental.[1]

Nonfundamental discriminants. In this case, we find that our proposed refinement does not characterize the field. In other words, for nonfundamental discriminants we have an affirmative answer to Question 1.1. We divide the class of nonfundamental discriminants into two groups according to sign. We further divide the positive discriminants into two groups: those that are perfect squares, and those that are not. For each one of these cases we show that there are some nonfundamental discriminants such that Question 1.1 has an affirmative answer.

(i) *Negative nonfundamental discriminants*. We define a sequence of positive integers Σ and a family of triples $\{K_m, L_m, E_m\}_{m\in\Sigma}$ with the following properties (see Proposition 3.4):

- K_m and L_m are two nonisomorphic cubic fields with discriminant $-3n^2$, where n is a positive integer depending only on m.
- An elliptic curve E_m defined over $\mathbb{Q}$ such that $E_m[3](\mathbb{Q})$ determines completely a ternary quadratic form equivalent to both $\mathrm{tr}_{K/\mathbb{Q}}(x^2)|_{O_{K_m}}$ and $\mathrm{tr}_{K/\mathbb{Q}}(x^2)|_{O_{L_m}}$.

(ii) *Square discriminants*. In this case we generalize in Theorem 3.1 a result of Conner and Perlis [1984, Theorem IV.1.1 with $p = 3$]. Let K and L be two Galois cubic number fields of the same discriminant and let M be either O_K or O_K^0. Then $\mathrm{tr}_{K/\mathbb{Q}}(x^2)|_M$ and $\mathrm{tr}_{L/\mathbb{Q}}(x^2)|_M$ are equivalent. Since there are examples of nonisomorphic Galois cubic fields of the same discriminant, Question 1.1 has a positive answer for such cases.

[1] Recall that d is a fundamental discriminant if it is the discriminant of a quadratic field.

(iii) *Positive, nonfundamental, nonsquare discriminants*. See Example 3.6 for two fields with positive, non-square-free, non-perfect-square discriminant and isometric integral trace forms.

Main results. For fields of fundamental discriminant we see, thanks to Lemma 2.5, that the binary quadratic form $\mathrm{tr}_{K/\mathbb{Q}}(x^2)|_{O_K^0}$ is a refinement of the discriminant. Hence, we reformulate Question 1.1.

Question 1.3. *Do there exist two nonisomorphic cubic fields K and L such that the forms $\mathrm{tr}_{K/\mathbb{Q}}(x^2)|_{O_K^0}$ and $\mathrm{tr}_{L/\mathbb{Q}}(x^2)|_{O_L^0}$ are isomorphic?*

Although this question has relevance for us only for fundamental discriminants, we note that the examples (i), (ii) and (iii) described above also answer 1.3 in an affirmative way. On the other hand, for fundamental discriminants (see diagram 4-1), class field theory provides examples of nonisomorphic cubic fields of the same discriminant. Among the fields with negative discriminants we found examples giving an affirmative answer to Question 1.3.

It is clear, thanks to the results developed so far, that one should consider working over cubic fields of fundamental discriminant. We show for such discriminants that the trace form is equal, as an element of a narrow class group, to the Hessian multiplied by an element that only depends on the discriminant.

Theorem 5.5. *Let K be a cubic field with discriminant d_K. Assume that d_K is fundamental and that $3 \nmid d_K$. Let $F_K = (a, b, c, d)$ be a cubic in the $\mathrm{GL}_2(\mathbb{Z})$-equivalence class defined by K. Then $\frac{1}{2}q_K * C_{d_K} = H_K^{\pm 1}$ as elements of $\mathrm{Cl}^+_{\mathbb{Q}(\sqrt{-3d_K})}$, where $C_{d_K} = (3, 0, d_K/4)$ or $C_{d_K} = (3, 3, (d_K+3)/4)$ in accordance with whether $d_k \equiv 0 \pmod 4$ or $d_k \equiv 1 \pmod 4$.*

By reformulating all of this in the language of Bhargava's composition of cubes [2004], we show that the trace form arises naturally as a projection of a cube determined by the field.

Theorem 6.2. *Let K be a cubic field with discriminant d_K and associated cubic form $F_K = (a, b, c, d)$. Assume that d_K is fundamental and that 3 does not ramify. Then there exists $T_{F_K} \in \mathrm{Cl}(\mathbb{Z}^2 \otimes \mathbb{Z}^2 \otimes \mathbb{Z}^2; -3d_K)$ such that $(\pi_1 \circ \phi)(T_{F_K})^{\pm 1} = \frac{1}{2}q_K$ as elements of $\mathrm{Cl}^+_{\mathbb{Q}(\sqrt{-3d_K})}$.*

In this setting, Theorem 6.5 follows from Theorem 5.11, which is the modern version of a theorem of Eisenstein [1844]. By reformulating Theorem 6.5 (see Theorem 6.8 and its corollary), we obtain one inequality of the classical Scholz reflection principle [1932].

Theorem 6.5 can be obtained with the tools developed by Eisenstein [1844]. However, we have decided to use Bhargava's theory of $2 \times 2 \times 2$ orbits of cubes, to suggest that it might be possible to use some other prehomogeneous spaces to "generalize" Theorem 6.5 to higher dimensions.

2. Basic facts

Definition 2.1. Let G be a free abelian group. We say that a map $q : G \to \mathbb{Z}$ is a *quadratic form* if

- $q(nx) = n^2 q(x)$ for all integer n, and
- the map $B_q : G \times G \to \frac{1}{2}\mathbb{Z}$ defined as $B_q(x, y) = \frac{1}{2}(q(x+y) - q(x) - q(y))$ is $\mathbb{Z}$-bilinear.

Remark 2.2. Let $\langle G, q\rangle$ be a quadratic $\mathbb{Z}$-module of rank $n = \operatorname{rank}(G)$. After choosing a basis, we can think of q as a homogeneous polynomial in n variables of degree two, that is, $q \in (\operatorname{Sym}^2\mathbb{Z}^n)^*$. There is a natural action of $\mathrm{GL}_2(\mathbb{Z})$ on $(\operatorname{Sym}^2\mathbb{Z}^n)^*$. Under this action, q and q_1 belong to the same orbit if and only if $\langle G, q\rangle$ is isometric to $\langle G_1, q_1\rangle$. Abusing notation, we denote this by $q \sim_{\mathrm{GL}_2(\mathbb{Z})} q_1$.

Let K be a number field and let O_K be its maximal order. The map

$$\tilde{q}_K : O_K \to \mathbb{Z}, \quad x \mapsto \operatorname{tr}_{K/\mathbb{Q}}(x^2),$$

defines a quadratic form with corresponding bilinear form

$$B_K(x, y) = \operatorname{tr}_{K/\mathbb{Q}}(xy)|_{O_K}.$$

Thus, $\langle O_K, \tilde{q}_K\rangle$ is a quadratic $\mathbb{Z}$-module and its discriminant is precisely the discriminant of K. Hence, if K and L are two number fields such that $\langle O_K, \tilde{q}_K\rangle$ and $\langle O_L, \tilde{q}_L\rangle$ are isomorphic quadratic $\mathbb{Z}$-modules, we have

$$[K : \mathbb{Q}] = [L : \mathbb{Q}] \quad \text{and} \quad \operatorname{Disc} K = \operatorname{Disc} L.$$

Therefore the isomorphism class of $\langle O_K, \tilde{q}_K\rangle$ is to us a natural refinement of the discriminant.

Lemma 2.3. *Let K be a number field of degree n and let $G_K = \mathbb{Z} + O_K^0$. We have*

$$|O_K/G_K| = |\operatorname{tr}_{K/\mathbb{Q}}(O_K)/nZ|.$$

Corollary 2.4. *Let K and L be number fields. If*

$$f : \langle O_K, B_K\rangle \to \langle O_L, B_L\rangle$$

is an isomorphism, then $\operatorname{Disc} G_K = \operatorname{Disc} G_L$.

Proof. Since $\operatorname{tr}_{L/\mathbb{Q}}(f(x)f(y)) = \operatorname{tr}_{K/\mathbb{Q}}(xy)$ for all $x, y \in O_K$ we have that $\operatorname{tr}_{K/\mathbb{Q}} : O_K \twoheadrightarrow \mathbb{Z}$ implies $\operatorname{tr}_{L/\mathbb{Q}} : O_L \twoheadrightarrow \mathbb{Z}$. Since f is an isometry, the argument is symmetric in K and L. By Lemma 2.3 we have $|O_K/G_K| = |O_L/G_L|$. Hence

$$\operatorname{Disc} G_K = |O_K/G_K|^2 \operatorname{Disc} O_K = |O_L/G_L|^2 \operatorname{Disc} O_L = \operatorname{Disc} G_L. \qquad \square$$

For a number field K, we set $q_k = \tilde{q}_K|_{O_K^0}$.

Lemma 2.5. *Let K and L be two number fields of degree n. Assume they both have discriminants that are square-free at all primes dividing n. Further, suppose that $\langle O_K^0, q_K\rangle$ and $\langle O_L^0, q_L\rangle$ are isomorphic. Then K and L have the same discriminant.*

Proof. Since $\operatorname{Disc} G_K = \operatorname{Disc} G_L$, we have that

$$|O_K/G_K|^2 \operatorname{Disc} O_K = |O_L/G_L|^2 \operatorname{Disc} O_L.$$

The result now follows from Lemma 2.3. □

Proposition 2.6. *Let K be a Galois number field of prime degree p. Then p ramifies in K if and only if $\operatorname{tr}_{K/\mathbb{Q}}(O_K) = p\mathbb{Z}$.*

Proof. It is clear that $\operatorname{tr}_{K/\mathbb{Q}}(O_K) = p\mathbb{Z}$ implies that p ramifies in K. Next, assuming that p ramifies, let P be the unique prime of O_K lying above p. By hypothesis, we have that $|O_K/P| = p$. In particular, P is a maximal $\mathbb{Z}$-submodule of O_K. Since $1 \notin P$, we must have $O_K = \mathbb{Z} + P$. Since P is Galois invariant, $\operatorname{tr}_{K/\mathbb{Q}}(P) \subseteq P \cap \mathbb{Z} = p\mathbb{Z}$. Thus $\operatorname{tr}_{K/\mathbb{Q}}(O_K) = \operatorname{tr}_{K/\mathbb{Q}}(\mathbb{Z} + P) \subseteq p\mathbb{Z}$. □

3. Galois fields and rational 3-torsion

In this section we explain some situations in which Questions 1.1 and 1.3 have positive answers. The examples in this section are characterized by having discriminants with a nontrivial square factor.

The following result is a generalization of the case $p = 3$ of [Conner and Perlis 1984, Theorem IV.1.1].

Theorem 3.1. *Let K and L be two Galois, cubic number fields of discriminant $D = d^2$. We have*

$$\langle O_K^0, q_K\rangle \cong \langle O_L^0, q_L\rangle \cong \begin{cases} 2d(x^2+xy+y^2) & \text{if } 3 \nmid d, \\ \frac{2}{3}d(x^2+xy+y^2) & \text{otherwise.} \end{cases}$$

The isometry can be chosen so it extends to one between $\langle O_K, \tilde{q}_K\rangle$ and $\langle O_L, \tilde{q}_L\rangle$.

Proof. Assume first that 3 does not divide D. By Theorem 132 of Hilbert [1900], write $O_K = e_1\mathbb{Z} \oplus e_2\mathbb{Z} \oplus e_3\mathbb{Z}$, where $\sigma(e_1) = e_2$, $\sigma(e_2) = e_3$, and σ is a generator of $\operatorname{Gal}(K/\mathbb{Q})$. Because 3 does not ramify, Proposition 2.6 implies that $\operatorname{tr}_{F/\mathbb{Q}}(e_1) = 1$, and furthermore that $O_K^0 = (e_1 - e_2)\mathbb{Z} \oplus (e_1 - e_3)\mathbb{Z}$. Let $a = \operatorname{tr}_{F/\mathbb{Q}}(e_1^2)$ and $b = \operatorname{tr}_{F/\mathbb{Q}}(e_1e_2)$. Then

$$M = \begin{pmatrix} (1+2a-2b)/3 & a-b & a-b \\ a-b & 2a-2b & a-b \\ a-b & a-b & 2a-2b \end{pmatrix} \quad \text{and} \quad M_0 = (a-b)\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

represent respecitvely the trace form over O_K in the basis $\{e_1, e_1-e_2, e_1-e_3\}$, and the trace form over O_K^0 in the basis $\{e_1-e_2, e_2-e_3\}$. Note that $a+2b =$

$(\mathrm{tr}_{F/\mathbb{Q}}(e_1))^2 = 1$; thus $D = \det M = (a-b)^2(a+2b) = (a-b)^2$. By the Cauchy–Schwartz inequality, $a-b > 0$, and hence $d = a-b$, which implies that $a = (1+2d)/3$ and $b = (1-d)/3$. Thus, every cubic field of discriminant d^2, with $3 \nmid d$, has an integral basis for which the trace form over O_K has representative matrix M (respectively trace form over O_K^0 has representative matrix M_0).

On the other hand, if 3 divides d, then Proposition 2.6 and Lemma 2.3 imply that $O_K = \mathbb{Z} \oplus O_K^0$. Hence, $\tilde{q}_K$ is totally determined by $q_K = \tilde{q}_K|_{O_K^0}$. Since every integral quadratic form of discriminant -3 is $\mathrm{SL}_2(\mathbb{Z})$-equivalent to (x^2+xy+y^2), the result follows from the following claim. □

Claim. $\dfrac{3}{2d} q_K$ *is an integral, primitive, binary quadratic form of discriminant* -3.

Proof of claim. Let $\{\alpha, \beta\}$ an integral basis for O_K^0. Let $O_\alpha \subseteq O_K^0$ be the $\mathbb{Z}$-module generated by $\{\alpha, \sigma(\alpha)\}$, where σ is a generator for $\mathrm{Gal}(K/\mathbb{Q})$. Since $\alpha \notin \mathbb{Z}$, we know that α and $\sigma(\alpha)$ are distinct elements of O_K with the same norm. In particular, $\sigma(\alpha)$ cannot be a rational multiple of α, so $\mathrm{rank}_{\mathbb{Z}}(O) = 2$. Thus, $[O_K^0 : O_\alpha]$ is finite, and moreover $\sigma(\alpha) = m\alpha + [O_K^0 : O_\alpha]\beta$ for some integer m. Note that $\big(\mathrm{tr}_{K/\mathbb{Q}}(\alpha^2), 2\,\mathrm{tr}_{K/\mathbb{Q}}(\alpha\beta), \mathrm{tr}_{K/\mathbb{Q}}(\beta^2)\big)$, $\big(\mathrm{tr}_{K/\mathbb{Q}}(\alpha^2), 2\,\mathrm{tr}_{K/\mathbb{Q}}(\alpha\sigma(\alpha)), \mathrm{tr}_{K/\mathbb{Q}}(\sigma(\alpha)^2)\big)$ represent q_K in the bases $\{\alpha, \beta\}$ and $\{\alpha, \sigma(\alpha)\}$ respectively. Hence

$$\begin{aligned}\mathrm{tr}_{K/\mathbb{Q}}(\alpha^2)\,\mathrm{tr}_{K/\mathbb{Q}}(\sigma(\alpha)^2) - \mathrm{tr}^2_{K/\mathbb{Q}}(\alpha\sigma(\alpha))& \\ = [O_K^0 : O_\alpha]^2\big(\mathrm{tr}_{K/\mathbb{Q}}(\alpha^2)\,\mathrm{tr}_{K/\mathbb{Q}}(\beta^2) - \mathrm{tr}_{K/\mathbb{Q}}(\alpha\beta)\big).& \quad (3\text{-}1)\end{aligned}$$

Since $\mathrm{Disc}\, K = d^2$ and $O_K = \mathbb{Z} + O_K^0$, $\frac{1}{3}d^2 = \mathrm{tr}_{K/\mathbb{Q}}(\alpha^2)\,\mathrm{tr}_{K/\mathbb{Q}}(\beta^2) - \mathrm{tr}_{K/\mathbb{Q}}(\alpha\beta)$. On the other hand, since $\alpha \in O_K^0$, $\mathrm{tr}_{K/\mathbb{Q}}(\alpha^2) = -2\,\mathrm{tr}_{K/\mathbb{Q}}(\alpha\sigma(\alpha))$, and the left side of (3-1) is $3\,\mathrm{tr}^2_{K/\mathbb{Q}}(\alpha\sigma(\alpha))$. Thus,

$$\mathrm{tr}_{K/\mathbb{Q}}(\alpha\sigma(\alpha)) = \pm[O_K^0 : O_\alpha]\frac{d}{3}. \quad (3\text{-}2)$$

In particular we see that $\frac{d}{3}$ divides $\frac{1}{2}\,\mathrm{tr}_{K/\mathbb{Q}}(\alpha^2)$. Exchanging the roles of α and β we see that $\frac{d}{3}$ also divides $\frac{1}{2}\,\mathrm{tr}_{K/\mathbb{Q}}(\beta^2)$. Now consider $\sigma(\alpha) = m\alpha + [O_K^0 : O_\alpha]\beta$. Multiplying both sides by α and then taking traces, we see that $\frac{d}{3}$ divides $\mathrm{tr}_{K/\mathbb{Q}}(\alpha\beta)$. We conclude that $(\mathrm{tr}_{K/\mathbb{Q}}(\alpha^2), 2\,\mathrm{tr}_{K/\mathbb{Q}}(\alpha\beta), \mathrm{tr}_{K/\mathbb{Q}}(\beta^2))$ can be written as $\frac{2d}{3} f$, with f an integral quadratic form of discriminant -3. □

Example 3.2. Let K and L be cubic fields defined by $x^3 + 6x^2 - 9x + 1$ and $2x^3 + 3x^2 - 9x + 2$ respectively. One sees by direct computation that K and L are nonisomorphic fields of discriminant 3969; for instance, they have different regulators.

We conclude that the trace form does not characterize the field in the case where the discriminant is a square. Proposition 3.4 below is an indication that the case of the square discriminant is not the only case that should be reconsidered, but also the

non-square-free case. Cubic fields of a fixed discriminant Δ can be parametrized by a subset of rational points on a certain elliptic curve. Assume that $L = \mathbb{Q}(\beta)$ is a cubic field defined by the equation $x^3 + px + q \in \mathbb{Z}[x]$. If $O_L = \mathbb{Z}[\beta]$, then Disc $L = -27q^2 - 4p^3$. Hence if K is a cubic field of discriminant Δ, one could try to find a cubic field L of the same discriminant by finding rational points $(-\frac{1}{3}p, \pm\frac{1}{2}q)$ of $y^2 = x^3 - \frac{1}{108}\Delta$. Using this idea, we construct a family of nonisomorphic cubic fields with prescribed discriminant. We need the following result from algebraic number theory.

Proposition 3.3 [Marcus 1977]. *Let m be a nonperfect cube integer and α a root of $x^3 - m$. Write $m = m_f m_s^2$, with m_f square-free and* $\gcd(m_f, m_s) = 1$. *Suppose that $m \not\equiv \pm 1 \pmod 9$. Then $\{1, \alpha, \alpha^2/m_s\}$ is an integral basis for $K_m = \mathbb{Q}(\alpha)$; in particular*, Disc $K_m = -27(m_s m_f)^2$.

Let $\Sigma = \{m \in \mathbb{N} \setminus \mathbb{N}^3 \mid m_s \neq 1, m_f m_s \not\equiv \pm 1 \pmod 9, m \not\equiv \pm 1 \pmod 9\}$.

Proposition 3.4. *Let $m \in \Sigma$ and K_m, L_m be the cubic fields defined by $x^3 - m$ and $x^3 - m_f m_s$ respectively, with m_f, m_s as in Proposition 3.3. Then K_m, L_m are cubic fields with equivalent trace forms, and have discriminant $-3(3m_f m_s)^2$.*

Proof. By the discussion above and Proposition 3.3, we have that K_m defines the rational elliptic curve $E_m : y^2 = x^3 + \frac{1}{4}m_f^2 m_s^2$. A simple calculation shows that $E_m[3](\mathbb{Q}) = \{\infty, (0, \frac{1}{2}m_f m_s), (0, -\frac{1}{2}m_f m_s)\}$, and these points define the field L_m. Let P be a generator of $E_m[3](\mathbb{Q})$ and let

$$M_m = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 0 & 6y(p) \\ 0 & 6y(p) & 0 \end{pmatrix}.$$

Then M_m represents simultaneously the trace form in O_{K_m} and O_{L_m} with respect to the bases given by Proposition 3.3. □

The pair of number fields given by Proposition 3.4 need not be isomorphic, as the following example demonstrates.

Example 3.5. Let $m = 12$ so that K_{12} and L_{12} are the cubic fields defined by $x^3 - 12$ and $x^3 - 6$ respectively. Then $\langle O_K, \tilde{q}_{K_{12}}\rangle$ and $\langle O_L, \tilde{q}_{L_{12}}\rangle$ are isomorphic to $\langle \mathbb{Z}^3, 3x^2 + 36yz\rangle$. We see that K_{12} and L_{12} are nonisomorphic fields of discriminant $-2^2 3^5$ by direct computation; for instance, 7 splits in L_{12} but is inert in K_{12}.

Recall that for Galois cubic fields of fixed discriminant, there is only one possibility for the trace form (see Theorem 3.1), since after a suitable scaling we are left with a binary quadratic form of discriminant -3. Inspired by this, we began looking for discriminants D of totally real cubic fields satisfying four conditions: (i) D is a nonperfect square; (ii) D is nonfundamental; (iii) up to square factors

and factors of 3, $-D$ defines an imaginary quadratic field of class number 1. and (iv) there are at least two cubic fields of discriminant D.

It turns out that the first D satisfying these conditions (see tables at the end of [Ennola and Turunen 1985]) is $D = 66825 = 3^5 5^2 11$. For this value of D we have:

Example 3.6. Let K and L be the cubic fields defined by $2x^3 + 3x^2 - 21x + 4$ and $x^3 + 9x^2 - 18x - 3$ respectively. Then $\langle O_K, \tilde{q}_K \rangle$ and $\langle O_L, \tilde{q}_L \rangle$ are isomorphic to $\langle \mathbb{Z}^3, 3x^2 + 90(y^2 + yz + 3z^2) \rangle$. One sees by direct computation that K and L are nonisomorphic fields of discriminant $3^5 5^2 11$ (they have different regulators).

None of our results so far yield positive answers to Questions 1.1 or 1.3 with fundamental discriminant. It is thus natural to ask whether those questions have negative answers in the special case where the discriminant of the cubic field is fundamental. Under these circumstances we exhibit a more convenient refinement. To describe this, let K be a cubic number field and recall our notation $q_k = \tilde{q}_K|_{O_K^0}$. Then q_K is an integral, binary quadratic form. Moreover, under the fundamental discriminant hypothesis, the isometry class of $\langle O_K^0, q_K \rangle$ is a refinement of the discriminant, as shown in Lemma 2.5.

4. Cubic fields with fundamental discriminant

In this section, all cubic fields are assumed to have fundamental discriminant. The first question that comes to mind is this: for which fundamental discriminants d does there exist a cubic field with discriminant d? Moreover, we would like to know for which values of d there is more than one isomorphism class of cubic fields of discriminant d. It turns out that class field theory gives nice answers to these questions. Let K be a cubic field of fundamental discriminant d and Galois closure $\widetilde{K}$. Clearly, $\mathbb{Q}(\sqrt{d}) \subseteq \widetilde{K}$, and this extension is unramified. Since d is a fundamental discriminant, $\mathrm{Gal}(\widetilde{K}/\mathbb{Q}) \cong S_3$. Hence $[\widetilde{K} : \mathbb{Q}(\sqrt{d})] = 3$, and $\widetilde{K}/\mathbb{Q}(\sqrt{d})$ is abelian. Therefore, if H_d denotes the Hilbert class field of $\mathbb{Q}(\sqrt{d})$, and $\mathrm{Cl}_{\mathbb{Q}(\sqrt{d})}$ denotes the ideal class group of $\mathbb{Q}(\sqrt{d})$, we have this diagram:

$$\begin{tikzcd}
 & H_d \arrow[d, dash, "H_K"] \arrow[ddr, dash, bend left, "\mathrm{Cl}_{\mathbb{Q}(\sqrt{d})}"] & \\
 & \widetilde{K} \arrow[dl, dash, "\mathbb{Z}/2\mathbb{Z}"'] \arrow[dr, dash, "\mathbb{Z}/3\mathbb{Z}"] \arrow[dd, dash, "S_3"] & \\
K \arrow[dr, dash] & & \mathbb{Q}(\sqrt{d}) \arrow[dl, dash, "\mathbb{Z}/2\mathbb{Z}"] \\
 & \mathbb{Q} &
\end{tikzcd} \tag{4-1}$$

Thus, if we start with K as above, we obtain H_K, an index-3 subgroup of $\mathrm{Cl}_{\mathbb{Q}(\sqrt{d})}$. Conversely, it can be shown [Hasse 1930] that the fixed field of an index-3 subgroup of $\mathrm{Cl}_{\mathbb{Q}(\sqrt{d})}$ corresponds to the Galois closure of a cubic field of discriminant d. Hence:

Proposition 4.1 [Hasse 1930]. *The number of isomorphism classes of cubic fields of discriminant d is $(3^{\mathrm{r}_3(d)}-1)/2$, where $\mathrm{r}_3(d) = \dim_{\mathbb{F}_3}(\mathrm{Cl}_{\mathbb{Q}(\sqrt{d})} \otimes_{\mathbb{Z}} \mathbb{F}_3)$.*

Corollary 4.2 [Hasse 1930]. *There exists a cubic field K of discriminant d if and only if $\mathrm{Cl}_{\mathbb{Q}(\sqrt{d})}[3] \neq 0$.*

Section 3 gave affirmative answers to Questions 1.1 and 1.3 for nonfundamental discriminants. Example 4.3 shows that among fundamental discriminants, one still finds positive answers to Questions 1.1 and 1.3.

Example 4.3. The fundamental discriminant of least absolute value with $r_3(d) > 1$ is $d = -3299$. For this value of d, $\mathrm{Cl}_{\mathbb{Q}(\sqrt{d})} \cong \mathbb{Z}/3\mathbb{Z} \oplus \mathbb{Z}/9\mathbb{Z}$; hence there exist four nonisomorphic cubic fields of discriminant -3299. Among these four fields, the ones defined by $x^3+2x+11$ and $x^3-16x+27$ have isometric trace-0 parts.

Cubic fields with square-free discriminants lead us to 3-torsion of class groups of quadratic fields. Another very well-known source of class groups of quadratic fields is binary quadratic forms. Let us recall briefly how these two are connected. Let Δ be a non-perfect-square integer and let Γ_Δ (respectively Γ^1_Δ) be the set of $\mathrm{GL}_2(\mathbb{Z})$-equivalence classes (respectively $\mathrm{SL}_2(\mathbb{Z})$-equivalence classes) of primitive, binary quadratic forms of discriminant Δ. Gauss composition gives a group structure to Γ^1_Δ, and furthermore this group is isomorphic to the *narrow class group* $\mathrm{Cl}^+_{\mathbb{Q}(\sqrt{\Delta})}$. In particular, $|\Gamma_\Delta| \le |\mathrm{Cl}^+_{\mathbb{Q}(\sqrt{\Delta})}|$. Now, let K be a cubic field of discriminant d not divisible by 3. According to the next lemma, the $\mathrm{GL}_2(\mathbb{Z})$-equivalence class of $\left[\frac{1}{2}q_K\right]$ defines an element of Γ_{-3d}. Thus, if we denote by $\mathscr{C}_d$ the set of isomorphism classes of cubic fields of discriminant d, we have the map

$$\Phi_d : \mathscr{C}_d \to \Gamma_{-3d}, \quad K \mapsto \left[\tfrac{1}{2}q_K\right].$$

Since $\mathrm{Cl}^+_{\mathbb{Q}(\sqrt{9897})} \cong \mathbb{Z}/3\mathbb{Z}$ and $|\mathscr{C}_{-3299}| = 4$, the previous example can be restated as the noninjectivity of Φ_{-3299}.

Lemma 4.4. *Let K be a cubic field with fundamental discriminant d. Then $\frac{1}{2}q_K$ is an integral, binary quadratic form of discriminant $-3d$.*

Proof. Note that $\operatorname{Disc} q_K = -4 \operatorname{Disc} O_K^0 = -\frac{4}{3}|O_K/G_K|^2 d$. Since d is fundamental, $9 \nmid d$. In particular, $\mathrm{tr}_{K/\mathbb{Q}}$ is a surjection from O_K to $\mathbb{Z}$, and thanks to Lemma 2.3, we have $\operatorname{Disc} q_K = -12d$. Note that if $x \in O_K^0$, then $\mathrm{tr}_{K/\mathbb{Q}}(x^2) = \mathrm{tr}_{K/\mathbb{Q}}(x^2) - \mathrm{tr}^2_{K/\mathbb{Q}}(x) \in 2\mathbb{Z}$, and hence $\frac{1}{2}q_K$ is integral. □

Remark 4.5. In fact $\frac{1}{2}q_K$ is primitive if $3 \nmid d$, as seen in Corollary 5.4.

Often it is more convenient to work with primitive forms than general ones. Since $q_K \sim_{\mathrm{GL}_2(\mathbb{Z})} q_L$ if and only if $aq_K \sim_{\mathrm{GL}_2(\mathbb{Z})} aq_L$ for any nonzero rational number a, Remark 4.5 will allow us to restrict ourselves to primitive forms.

5. Trace form and class groups

In this section we calculate q_K explicitly, and then show that for positive fundamental discriminants, q_K characterizes the field. We start by recalling [Delone and Faddeev 1964; Gan et al. 2002; Belabas and Cohen 1998] on the parametrization of cubic rings. Every conjugacy class of a cubic ring R has associated to it a unique integral binary cubic form $(a, b, c, d) := F(x, y) = ax^3 + bx^2y + cxy^2 + dy^3$ up to $\mathrm{GL}_2(\mathbb{Z})$-equivalence. Let K be a cubic number field and F the form associated to its maximal order. Among the properties of F, we have:

- $K = \mathbb{Q}(\theta)$, where $\theta \in K$ is a root of $F_K(x, 1)$.
- $d_K := \operatorname{Disc} K = \operatorname{Disc}(a, b, c, d) = b^2c^2 - 27a^2d^2 + 18abcd - 4ac^3 - 4b^3d$.
- The Hessian form of F, $H_F = (P, Q, R) := Px^2 + Qxy + Ry^2$, has discriminant $-3d_K$, where
$$P = b^2 - 3ac, \quad Q = bc - 9ad, \quad R = c^2 - 3bd.$$
- H_F is covariant with respect to the $\mathrm{GL}_2(\mathbb{Z})$-action on binary cubic forms and on binary quadratic forms.
- $\mathcal{B} = \{1, -a\theta, d/\theta\}$ is a $\mathbb{Z}$-basis of O_K.
- If d_K is fundamental, then H_F is a primitive, binary quadratic form.

Lemma 5.1. *Let $\alpha = -a\theta$ and $\beta = d/\theta$. Then H_F is realized as the integral quadratic form $\frac{3}{2}\operatorname{tr}_{K/\mathbb{Q}}(X^2)$ over the $\mathbb{Z}$-module*
$$O_K^{\mathcal{B}} = \operatorname{Span}_{\mathbb{Z}}\left\{\alpha - \frac{\operatorname{tr}_{K/\mathbb{Q}}(\alpha)}{3}, \beta - \frac{\operatorname{tr}_{K/\mathbb{Q}}(\beta)}{3}\right\}.$$

Proof. Note that $a^2F(x/a, 1)$ and $d^2F(1, x/d)$ are the minimal polynomials over $\mathbb{Q}$ of α and β respectively. Hence, $\operatorname{tr}_{K/\mathbb{Q}}(\alpha) = b$, $\operatorname{tr}_{K/\mathbb{Q}}(\beta) = -c$, $\operatorname{tr}_{K/\mathbb{Q}}(\alpha\beta) = -3ad$, $\operatorname{tr}_{K/\mathbb{Q}}(\alpha^2) = b^2 - 2ac$, and $\operatorname{tr}_{K/\mathbb{Q}}(\beta^2) = c^2 - 2bd$. From this and a simple calculation the result follows. □

Proposition 5.2. *Let $\alpha_0 = \alpha - \frac{1}{3}\operatorname{tr}_{K/\mathbb{Q}}(\alpha)$ and $\beta_0 = \beta - \frac{1}{3}\operatorname{tr}_{K/\mathbb{Q}}(\beta)$. Then*
$$O_k^0 = \begin{cases} O_1 = \operatorname{Span}_{\mathbb{Z}}\{\alpha_0, 3\beta_0\} & \text{if } b \equiv 0 \pmod 3, \\ O_2 = \operatorname{Span}_{\mathbb{Z}}\{3\alpha_0, \beta_0\} & \text{if } c \equiv 0 \pmod 3, \\ O_3 = \operatorname{Span}_{\mathbb{Z}}\{\alpha_0 - \beta_0, 3\beta_0\} & \text{if } b \equiv -c \pmod 3, \\ O_4 = \operatorname{Span}_{\mathbb{Z}}\{\alpha_0 + \beta_0, 3\beta_0\} & \text{if } b \equiv c \pmod 3. \end{cases}$$

Proof. By Lemma 5.1, $(\frac{3}{2}\operatorname{tr}_{K/\mathbb{Q}}(X^2)|O_K^{\mathcal{B}}) = -3d_K$ or, equivalently,

$$\left(\tfrac{1}{2}\operatorname{tr}_{K/\mathbb{Q}}(X^2)|O_K^{\mathcal{B}}\right) = -\tfrac{1}{3}d_K.$$

On the other hand,

$$-3d_K = \left(\tfrac{1}{2}\operatorname{tr}_{K/\mathbb{Q}}(X^2)|O_K^0\right) = [O_K^{\mathcal{B}} : O_K^0]^2\left(\tfrac{1}{2}\operatorname{tr}_{K/\mathbb{Q}}(X^2)|O_K^{\mathcal{B}}\right). \tag{5-1}$$

It follows that $[O_K^{\mathcal{B}} : O_K^0] = 3$. Notice that for each i, the given congruence conditions on b and c imply that $O_i \subseteq O_K^0$. Since $[O_K^{\mathcal{B}} : O_i] = 3$ for $i \in \{1, 2, 3, 4\}$, the result follows. □

Corollary 5.3. *Let K be a cubic field and let $F_K = (a, b, c, d)$ be a cubic form associated to K. Let $H_K = (P, Q, R)$ be the Hessian of F_K. Then the binary quadratic form $\frac{1}{2}\operatorname{tr}_{K/\mathbb{Q}}(X^2)$ on the lattice O_K^0 can be explicitly described as follows*:

$$\begin{cases} (P/3, Q, 3R) & \text{if } b \equiv 0 \pmod 3, \\ (3P, Q, R/3) & \text{if } c \equiv 0 \pmod 3, \\ (3P, 2P - Q, \frac{1}{3}(P + R - Q)) & \text{if } b \equiv -c \pmod 3, \\ (3P, 2P + Q, \frac{1}{3}(P + Q + R)) & \text{if } b \equiv c \pmod 3. \end{cases}$$

Proof. By Lemma 5.1, the matrix of $\frac{3}{2}\operatorname{tr}_{K/\mathbb{Q}}(X^2)$ over $O_K^{\mathcal{B}}$ in the basis $\{\alpha_0, \beta_0\}$ is given by

$$M = \begin{pmatrix} P & Q/2 \\ Q/2 & R \end{pmatrix}.$$

Let $N_1 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 3 \end{smallmatrix}\right)$, $N_2 = \left(\begin{smallmatrix} 3 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, $N_3 = \left(\begin{smallmatrix} 1 & -1 \\ 0 & 3 \end{smallmatrix}\right)$, and $N_4 = \left(\begin{smallmatrix} 1 & 0 \\ -1 & 3 \end{smallmatrix}\right)$. Then the coordinates of the vector $N_i(\alpha_0, \beta_0)^t$ form a basis of O_i, for $i \in \{1, 2, 3, 4\}$. Hence, $\frac{1}{3}N_i M N_i^t$ is the matrix that represents $\frac{1}{2}\operatorname{tr}_{K/\mathbb{Q}}(X^2)$ over O_i in such a basis. After applying Proposition 5.2, the result follows. □

From now on, whenever we choose a cubic form F_K in the $\mathrm{GL}_2(\mathbb{Z})$-class given by the field K, what we mean by $\frac{1}{2}q_K$ is the quadratic form in the coordinates given by Corollary 5.3.

Corollary 5.4. *If K is a cubic field with fundamental discriminant d not divisible by 3, then $\frac{1}{2}q_K$ is a primitive, integral, binary quadratic form of discriminant $-3d$.*

Proof. By Lemma 4.4, it remains only to prove that $\frac{1}{2}q_K$ is primitive. Since H_k is primitive and $9 \nmid -3d$, the result follows from Corollary 5.3. □

For a fixed F_K in the $\mathrm{GL}_2(\mathbb{Z})$-class given by the field K, we have found explicit relations between the binary quadratic forms $\frac{1}{2}q_K$ and H_K. Since they have the same discriminant, namely $-3d_K$, one might ask what their relation is as elements of the group $\mathrm{Cl}^+_{\mathbb{Q}(\sqrt{-3d_K})}$. A small objection to this question is that even though H_K represents a valid element of this group, $\frac{1}{2}q_K$ need not, since it may not be

primitive. Yet, as Corollary 5.4 shows, $\frac{1}{2}q_K$ is primitive whenever 3 does not ramify in K. In this setting we are able to find the following connection between forms.

Theorem 5.5. *Let K be a cubic field with discriminant d_K. Assume that d_K is fundamental and that 3 does not divide d_K. Let $F_K = (a, b, c, d)$ be a cubic in the $\mathrm{GL}_2(\mathbb{Z})$-equivalence class defined by K. Then $\frac{1}{2}q_K * C_{d_K} = H_K^{\pm 1}$ as elements of $\mathrm{Cl}^+_{\mathbb{Q}(\sqrt{-3d_K})}$, where $C_{d_K} = \left(3, 0, \frac{1}{4}d_K\right)$ or $C_{d_K} = \left(3, 3, \frac{1}{4}(d_K + 3)\right)$ in accordance with whether $d_k \equiv 0 \pmod 4$ or $d_k \equiv 1 \pmod 4$.*

Proof. We work out the case when $d_k \equiv 1 \pmod 4$, the other case being completely analogous. By Arndt's composition algorithm [Buell 1989, Theorem 4.10],

$$\begin{cases} C_K * (P, Q, R) = (P/3, Q, 3R) & \text{if } b \equiv 0 \pmod 3, \\ C_K * (3P, Q, R/3) = (P, Q, R) & \text{if } c \equiv 0 \pmod 3, \\ C_K * (3P, 2P-Q, (P+R-Q)/3) = (P, 2P-Q, P+R-Q) & \text{if } b \equiv -c \pmod 3, \\ C_K * (3P, 2P+Q, (P+Q+R)/3) = (P, 2P+Q, P+R+Q) & \text{if } b \equiv c \pmod 3. \end{cases}$$

Using the matrix $\left(\begin{smallmatrix} 1 & -1 \\ 0 & 1 \end{smallmatrix}\right)$, we see that we have the identities in $\mathrm{Cl}^+_{\mathbb{Q}(\sqrt{-3d_K})}$

$$(P, 2P - Q, P + R - Q) = H_K^{-1} \quad \text{and} \quad (P, 2P + Q, P + R + Q) = H_K.$$

Since C_K is its own inverse, the result follows from the explicit description of $\frac{1}{2}q_K$ given in Corollary 5.3. □

Remark 5.6. Note that given K, we have freedom in choosing F_K in such a way that $b \not\equiv -c \pmod 3$. Hence Theorem 5.5 can be actually interpreted as saying that $\frac{1}{2}q_K * C_{d_K} = H_K$.

Remark 5.7. We denote the form C_K by C_{d_K} in order to stress the fact that this form only depends on the discriminant of K.

Bhargava's composition laws on cubes and their relation to the trace form. We have related the trace form, in the cubic case, to class groups of quadratic fields. There is a well-known generalization of Gauss's composition of quadratic forms to cubic forms. Inspired by this generalization, we expected some connection between the cubic forms attached to cubic number fields, and the quadratic forms given by the traces of these fields. We briefly recall some of the basics of Bhargava's laws on cubes and then we explain how to get such a connection (see Theorem 6.2).

In his Ph.D. thesis [2004], Bhargava generalizes the composition laws on binary quadratic forms of a fixed discriminant Δ discovered by Gauss. Bhargava defines a $\mathrm{SL}_2(\mathbb{Z}) \times \mathrm{SL}_2(\mathbb{Z}) \times \mathrm{SL}_2(\mathbb{Z})$-action on the set of $2 \times 2 \times 2$ integral cubes of discriminant Δ. Let $\mathrm{Cl}(\mathbb{Z}^2 \otimes \mathbb{Z}^2 \otimes \mathbb{Z}^2; \Delta)$ be the space of orbits given of action. Using

the generalization of Gauss's composition mentioned above, Bhargava discovered a composition law on $\mathrm{Cl}(\mathbb{Z}^2 \otimes \mathbb{Z}^2 \otimes \mathbb{Z}^2; \Delta)$.

In explicit terms, one can think of a $2\times 2\times 2$ integral cube $\mathscr{C}$ as a pair of 2×2 integral matrices (A, B), where A is the front face and B is the back face. Let $Q_1(\mathscr{C}) = -\det(Ax + By)$, $Q_2(\mathscr{C}) = -\det([\begin{smallmatrix} x \\ y \end{smallmatrix}]|B[\begin{smallmatrix} x \\ y \end{smallmatrix}])$ and $Q_3(\mathscr{C}) = -\det(A^t[\begin{smallmatrix} x \\ y \end{smallmatrix}]|B^t[\begin{smallmatrix} x \\ y \end{smallmatrix}])$.

It can be verified that $\mathrm{Disc}\, Q_1 = \mathrm{Disc}\, Q_2 = \mathrm{Disc}\, Q_3$. This common discriminant Δ is precisely the definition of the discriminant of $\mathscr{C}$. If

$$g := (g_1, g_3, g_3) \in \Gamma := \mathrm{SL}_2(\mathbb{Z}) \times \mathrm{SL}_2(\mathbb{Z}) \times \mathrm{SL}_2(\mathbb{Z})$$

and (A, B) is a cube, then

$$g \cdot (A, B) := g_1 \begin{pmatrix} g_3 A g_2^t \\ g_3 B g_2^t \end{pmatrix}.$$

This action preserves the discriminant. Moreover, if Q_1, Q_2, Q_3 are primitive forms, one has that $Q_1 * Q_2 * Q_3 = 0$ as elements of $\mathrm{Cl}^+_{\mathbb{Q}(\sqrt{\Delta})}$. Conversely, let (Q_1, Q_2, Q_3) be a triple of primitive, binary quadratic forms of discriminant Δ such that $Q_1 * Q_2 * Q_3 = 0$. Then there is a unique class on $\mathrm{Cl}(\mathbb{Z}^2 \otimes \mathbb{Z}^2 \otimes \mathbb{Z}^2; \Delta)$ giving rise to (Q_1, Q_2, Q_3) as above. With this in hand, it is simple to define a composition law on cubes: $(A, B) + (A', B')$ is the cube that corresponds to the triple $(Q_1 * Q_1', Q_2 * Q_2', Q_3 * Q_3')$. Furthermore:

Theorem 5.8 [Bhargava 2004]. *There is an isomorphism*

$$\phi : \mathrm{Cl}(\mathbb{Z}^2 \otimes \mathbb{Z}^2 \otimes \mathbb{Z}^2; \Delta) \to \mathrm{Cl}^+_{\mathbb{Q}(\sqrt{\Delta})} \times \mathrm{Cl}^+_{\mathbb{Q}(\sqrt{\Delta})}$$

defined by $(A, B)_\Gamma \mapsto ([Q_1]_{\mathrm{SL}_2(\mathbb{Z})}, [Q_2]_{\mathrm{SL}_2(\mathbb{Z})})$.

Definition 5.9. A binary cubic form $f(x, y) \in \mathbb{Z}[x, y]$ is called a *Gaussian cubic form* if it is of the form $(a_0, 3a_1, 3a_2, a_3)$. The set of Gaussian cubic forms is denoted by $\mathrm{Sym}^3\mathbb{Z}^2$.

One may naturally associate to a Gaussian cubic form $f = (a_0, 3a_1, 3a_2, a_3)$ a triple symmetric cube:

$$f \mapsto \begin{tikzcd}[column sep=small] & a_1 \arrow[rr, dash] \arrow[dd, dash] & & a_2 \arrow[dd, dash] \\ a_0 \arrow[ur, dash] \arrow[rr, dash] \arrow[dd, dash] & & a_1 \arrow[ur, dash] \arrow[dd, dash] & \\ & a_2 \arrow[rr, dash] & & a_3 \\ a_1 \arrow[ur, dash] \arrow[rr, dash] & & a_2 \arrow[ur, dash] & \end{tikzcd}$$

The correspondence between cubic forms and cubes is identified with a map

$$\iota : \mathrm{Sym}^3\mathbb{Z}^2 \to \mathbb{Z}^2 \otimes \mathbb{Z}^2 \otimes \mathbb{Z}^2.$$

If we replace f by a Gaussian form in the same $\mathrm{SL}_2(\mathbb{Z})$ equivalence class as f, one obtains a well defined element under the Γ-action on cubes.

Let

$$\mathrm{Cl}(\mathrm{Sym}^3\mathbb{Z}^2;\, \Delta)$$

be the set of Gaussian forms, up to $\mathrm{SL}_2(\mathbb{Z})$-action, such that the corresponding cubes have fundamental discriminant Δ.

Remark 5.10. One must distinguish between the notions of the discriminant of cubic forms and the discriminant of cubes. For example, let f be a Gaussian form of discriminant D. Then the cube corresponding to f has discriminant $\Delta = -\frac{1}{27}D$.

It turns out that $\mathrm{Cl}(\mathrm{Sym}^3\mathbb{Z}^2;\, \Delta)$ is an abelian group. Furthermore,

$$[\iota] : [f]_{\mathrm{SL}_2(\mathbb{Z})} \mapsto [\iota(f)]_\Gamma$$

is a group homomorphism. By composing the homomorphisms

$$\mathrm{Cl}(\mathrm{Sym}^3\mathbb{Z}^2;\, \Delta) \xrightarrow{[\iota]} \mathrm{Cl}(\mathbb{Z}^2 \otimes \mathbb{Z}^2 \otimes \mathbb{Z}^2;\, \Delta) \xrightarrow{\phi} \mathrm{Cl}^+_{\mathbb{Q}(\sqrt{\Delta})} \times \mathrm{Cl}^+_{\mathbb{Q}(\sqrt{\Delta})} \xrightarrow{\pi_1} \mathrm{Cl}^+_{\mathbb{Q}(\sqrt{\Delta})},$$

Bhargava obtains:

Theorem 5.11 [Bhargava 2004; Hoffman and Morales 2000]. *There is a surjective homomorphism*

$$\phi_1 : \mathrm{Cl}(\,\mathrm{Sym}^3\mathbb{Z}^2;\, \Delta) \twoheadrightarrow \mathrm{Cl}^+_{\mathbb{Q}(\sqrt{\Delta})}[3],$$

where ϕ_1 is the first projection of ϕ composed with $[\iota]$. The cardinality of the kernel is equal to $|U/U^3|$, where U denotes the group of units in $\mathbb{Q}(\sqrt{\Delta})$. In other words, the kernel has order 1 if $\Delta < -3$, or 3 otherwise.

This theorem was in essence first obtained by Eisenstein [1844], but he incorrectly asserted that the kernel of the map was always trivial. Later Arnt and Cayley pointed out that it is not a bijection if $\Delta \geq -3$.

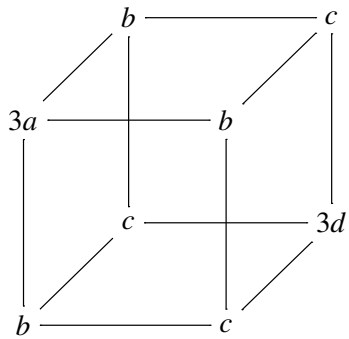
Remark 5.12. Explicitly,

$$\phi_1(a_0, 3a_1, 3a_2, a_3) = (a_1^2 - a_0a_2,\, a_1a_2 - a_0a_3,\, a_2^2 - a_1a_3).$$

6. From cubic fields to cubes and trace forms

Given K, a cubic field of discriminant d_K, and representative form $F_K(x, y) = (a, b, c, d)$, we naturally associate a cube as follows:

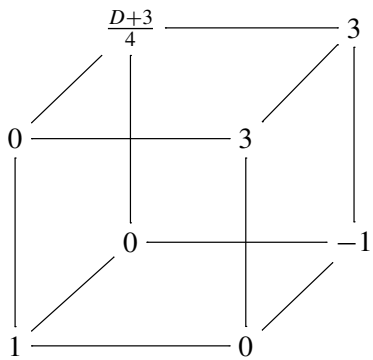
$$K : ax^3 + bx^2y + cxy^2 + dy^3 \longmapsto$$



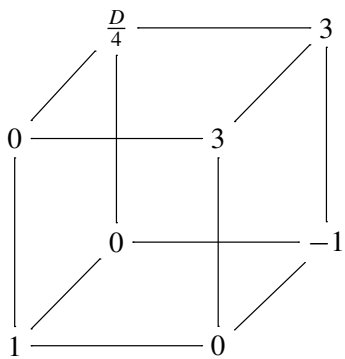
We obtain in this way an element

$$\mathcal{K}_F \in [\iota](\mathrm{Cl}(\mathrm{Sym}^3\mathbb{Z}^2; -3d_K)) \subseteq \mathrm{Cl}(\mathbb{Z}^2 \otimes \mathbb{Z}^2 \otimes \mathbb{Z}^2; -3d_K).$$

Let D be a fundamental discriminant. Let $\mathscr{C}_D \in \mathrm{Cl}(\mathbb{Z}^2 \otimes \mathbb{Z}^2 \otimes \mathbb{Z}^2; -3D)$ be given by



or



in accordance with whether $D \equiv 0 \pmod 4$ or $D \equiv 1 \pmod 4$.

Lemma 6.1. *Let K be a cubic field with a fixed cubic form $F = (a, b, c, d)$. Then $Q_1(\mathcal{K}_F) = H_F$ and $Q_1(\mathscr{C}_{d_K}) = C_{d_K}$.*

Proof. The result follows easily using the definition $Q_1(A, B) = -\det(Ax + By)$ for a cube (A, B). □

Theorem 6.2. *Let K be a cubic field with discriminant d_K and associated cubic form $F_K = (a, b, c, d)$. Assume that d_K is fundamental and that 3 does not ramify. Let $T_{F_K} = \mathcal{K}_F + \mathscr{C}_{d_K}$. Then $(\pi_1 \circ \phi)(T_{F_K})^{\pm 1} = \frac{1}{2} q_K$ as elements of $\mathrm{Cl}^+_{\mathbb{Q}(\sqrt{-3d_K})}$.*

Proof. Since ϕ is a group homomorphism, we have

$$\phi(T_{F_K}) = \phi((K)_F) * \phi(\mathscr{C}_{d_K}).$$

Projecting to the first component by π_1, we get that $(\pi_1 \circ \phi)(T_{F_K}) = H_K * C_{d_K}$. Since all of the functions involved are group homomorphisms, the result follows

from Theorem 5.5. In other "words",

$$\begin{tikzcd}[column sep=small]
 & b \arrow[rr, dash] \arrow[dl, dash] \arrow[dd, dash] & & c \arrow[dl, dash] \arrow[dd, dash] \\
3a \arrow[rr, dash] \arrow[dd, dash] & & b \arrow[dd, dash] & \\
 & c \arrow[rr, dash] \arrow[dl, dash] & & 3d \arrow[dl, dash] \\
b \arrow[rr, dash] & & c &
\end{tikzcd}
+
\begin{tikzcd}[column sep=small]
 & \frac{D+3}{4} \arrow[rr, dash] \arrow[dl, dash] \arrow[dd, dash] & & 3 \arrow[dl, dash] \arrow[dd, dash] \\
0 \arrow[rr, dash] \arrow[dd, dash] & & 3 \arrow[dd, dash] & \\
 & 0 \arrow[rr, dash] \arrow[dl, dash] & & -1 \arrow[dl, dash] \\
1 \arrow[rr, dash] & & 0 &
\end{tikzcd}
\overset{\phi_1}{\mapsto} \tfrac{1}{2}\operatorname{tr}_K(x^2). \qquad \square$$

Remark 6.3. We could choose F_K (see Remark 5.6) so that the conclusion of Theorem 6.2 is $(\pi_1 \circ \phi)(T_{F_K}) = \frac{1}{2} q_K$.

Theorem 6.4. *Let K be a cubic field with discriminant d_K, and let $F_K(x, y) = (a, b, c, d)$ be a cubic form associated to K. Assume that d_K is fundamental and that 3 ramifies in $K/\mathbb{Q}$. Then we have*

$$\begin{aligned} \phi_1 : \mathrm{Cl}(\operatorname{Sym}^3 \mathbb{Z}^2; -d_K/3) &\to \mathrm{Cl}^+_{\mathbb{Q}(\sqrt{-d_K/3})}[3] \\ (f_K)_{\mathrm{SL}_2(\mathbb{Z})} &\mapsto (\tfrac{1}{6} q_K)_{\mathrm{SL}_2(\mathbb{Z})}, \end{aligned}$$

where

$$f_K(x, y) := \begin{cases} \frac{1}{3} F(x, 3y) & \text{if } b \equiv 0 \pmod 3, \\ \frac{1}{3} F(3x, y) & \text{if } c \equiv 0 \pmod 3, \\ \frac{1}{3} F(x, 3(y-x)) & \text{if } b \equiv -c \pmod 3, \\ \frac{1}{3} F(x, 3(y+x)) & \text{if } b \equiv c \pmod 3. \end{cases}$$

Proof. Replacing $F(x, y)$ with either $F(y, x)$, $F(x, y-x)$ or $F(x, y+x)$, we may assume that $b \equiv 0 \pmod 3$. With this in hand, we have that $d_K \equiv -ac^3 \pmod 3$, and since 3 ramifies, $ac \equiv 0 \pmod 3$. On the other hand, since d_K is fundamental, we see that $3|a$. By Corollary 5.3, $\frac{1}{2} q_K = ((b^2 - 3ac)/3, bc - 9ad, 3(c^2 - 3bd))$, and thus $\frac{1}{6} q_K = ((\frac{1}{3}b)^2 - \frac{1}{3}ac, \frac{1}{3}bc - \frac{a}{3}9d, (c^2 - \frac{b}{3}9d))$, which is $\phi_1(\frac{1}{3}F(x, 3y))$. $\square$

Theorem 6.5. *Let K be a cubic number field of positive, fundamental discriminant, and let L be a number field such there exists an isomorphism of quadratic modules*

$$\langle O_K^0, q_K \rangle \cong \langle O_L^0, q_L \rangle.$$

Further assume $9 \nmid d_L$. Then $K \cong L$.

Proof. By Lemma 2.5, we have $d_K = d_L$. As usual, fix cubic forms $F_K(x, y)$ and $F_L(x, y)$ in the classes given by K and L respectively. Suppose first that $3 \nmid d_K$.

Since the isometry between the forms need not be proper, we can only ensure that, as elements of $\mathrm{Cl}^+_{\mathbb{Q}(\sqrt{-3d_K})}$,

$$\tfrac{1}{2}q_K = (\tfrac{1}{2}q_L)^{\pm 1}.$$

By Theorem 6.2, we have $(\pi_1 \circ \phi)(T_{F_K})^{\pm 1} = (\pi_1 \circ \phi)(T_{F_L})$. Replacing $F_K(x, y)$ by $F_K(x, -y)$ has the effect of replacing $H_{F_K}(x, y)$ by $H_{F_K}(x, -y)$. On the other hand, $H_{F_K}(x, -y)$ is inverse to H_{F_K} in the narrow class group. Since C_{d_K} has order 2, Theorem 5.5 says that we may replace $F_K(x, y)$ by $F_K(x, -y)$, if necessary, so we may assume that

$$(\pi_1 \circ \phi)(T_{F_K}) = (\pi_1 \circ \phi)(T_{F_L}).$$

Equivalently,

$$(\pi_1 \circ \phi)(\mathcal{K}_{F_K}) = (\pi_1 \circ \phi)(\mathcal{K}_{F_L}).$$

Notice that $\mathcal{K}_F = \iota(3F)$, so $\phi_1(3F_K) = \phi_1(3F_L)$. Since $d_K > 1$, Theorem 5.11 implies that $3F_K$ and $3F_L$ are $\mathrm{SL}_2(\mathbb{Z})$-equivalent. Since we could have replaced $F_K(x, y)$ by $F_K(x, -y)$, the equivalence between $3F_K$ and $3F_L$ is up to $\mathrm{GL}_2(\mathbb{Z})$. In any case this implies that $K \cong L$. If $3 \mid d_K$, we apply Theorem 6.4 and the argument follows the same lines as in the case without 3-ramification. □

Observations. Given $\Delta \in \mathbb{Z}$, let X_Δ be the set of integral, primitive, binary quadratic forms of discriminant Δ. Recall our notation $\Gamma_\Delta = \mathrm{GL}_2(\mathbb{Z}) \setminus X_\Delta$ and $\Gamma^1_\Delta = \mathrm{SL}_2(\mathbb{Z}) \setminus X_\Delta$.

Let d be a positive fundamental discriminant, $n_d := \gcd(3, d)$, and $\mathscr{C}_d$ the set of isomorphism classes of cubic fields of discriminant d.

Remark 6.6. Theorem 6.5 is equivalent to the injectivity of

$$\Phi_d : \mathscr{C}_d \to \Gamma_{-3d/n_d^2}, \quad K \mapsto \left[\frac{1}{2n_d} q_K\right].$$

Since Gauss's composition induces a group isomorphism between $\mathrm{Cl}^+_{\mathbb{Q}(\sqrt{-3d/n_d^2})}$ and Γ^1_{-3d/n_d^2}, we have a double cover

$$\pi : \mathrm{Cl}^+_{\mathbb{Q}(\sqrt{-3d/n_d^2})} \to \Gamma_{-3d/n_d^2}$$

with the property that the fiber of every point consists of an element and its inverse. Therefore, even though $q_K/2n_d$ does not define a point in $\mathrm{Cl}^+_{\mathbb{Q}(\sqrt{-3d/n_d^2})}$, it defines a cyclic subgroup, the one generated by $\pi^{-1}(\Phi_d(K))$. Corollary 5.3 and Theorem 6.4 provide us with a generator of this group. Let g_k be such a generator. Using Arndt's composition algorithm [Buell 1989], one sees that $g_K^3 = C_K$ when $3 \nmid d$, and that g_K has order 3 otherwise. Since C_{d_K} has order 2, it follows that $\langle \pi^{-1}(\Phi_d(K)) \rangle$ has order $2n_d$.

Proposition 6.7. *Let $d > 0$ be a fundamental discriminant. The map $K \mapsto \langle g_K \rangle$ is injective.*

Proof. Since $\langle g_K \rangle$ has order 3 or 6, its set of generators is $\{g_K^{\pm 1}\}$. Thus, if $\langle g_K \rangle = \langle g_L \rangle$, then $g_K^{\pm 1} = g_L$. Projecting under π, we obtain $\Phi_d(K) = \Phi_d(L)$, and the result follows from Remark 6.6. □

The unique subgroup of order 3 of $\langle g_K \rangle$ is given by $\langle g_K^2 \rangle$. From Proposition 6.7 we thus have:

Theorem 6.8. *Let $d > 0$ be a fundamental discriminant such that $\mathscr{C}_d \neq \varnothing$. Let $\mathscr{P}_3(\mathrm{Cl}^+_{\mathbb{Q}(\sqrt{-3d})})$ be the set of subgroups of size 3 of $\mathrm{Cl}^+_{\mathbb{Q}(\sqrt{-3d})}$. Then*

$$\Theta_d : \mathscr{C}_d \to \mathscr{P}_3(\mathrm{Cl}_{\mathbb{Q}(\sqrt{-3d})}), \quad K \mapsto \langle g_K^2 \rangle$$

is injective.

The injection Θ_d provides an alternative proof for one inequality of the Scholz reflection principle [1932].

Corollary 6.9. *Let d be a positive fundamental discriminant, and let $r = \mathrm{r}_3(-3d)$ and $s = \mathrm{r}_3(d)$ (recall our notation $\mathrm{r}_3(d) = \dim_{\mathbb{F}_3}(\mathrm{Cl}_{\mathbb{Q}(\sqrt{d})} \otimes_{\mathbb{Z}} \mathbb{F}_3)$). Then $s \leq r$.*

Proof. $(3^s - 1)/2 = |\mathscr{C}_d|$ and $(3^r - 1)/2 = |\mathscr{P}_3(\mathrm{Cl}_{\mathbb{Q}(\sqrt{-3d})})|$. □

Acknowledgements

I thank Jordan Ellenberg for introducing me to this subject, and for many helpful discussions and suggestions during the writing of this paper. I also thank Manjul Bhargava, Amanda Folsom, and Yongqiang Zhao for thorough and helpful comments on an earlier version of this paper.

References

[Belabas and Cohen 1998] K. Belabas and H. Cohen, "Binary cubic forms and cubic number fields", pp. 191–219 in *Computational perspectives on number theory* (Chicago, 1995), edited by D. A. Buell and J. T. Teitelbaum, AMS/IP Stud. Adv. Math. **7**, Amer. Math. Soc., Providence, 1998. MR 98m:11027 Zbl 0915.11024

[Bhargava 2004] M. Bhargava, "Higher composition laws, I: A new view on Gauss composition, and quadratic generalizations", *Ann. of Math.* (2) **159**:1 (2004), 217–250. MR 2005f:11062a

[Buell 1989] D. A. Buell, *Binary quadratic forms*, Springer, New York, 1989. MR 92b:11021 Zbl 0698.10013

[Conner and Perlis 1984] P. E. Conner and R. Perlis, *A survey of trace forms of algebraic number fields*, Series in Pure Mathematics **2**, World Scientific, Singapore, 1984. MR 86g:11021 Zbl 0551.10017

[Delone and Faddeev 1964] B. N. Delone and D. K. Faddeev, *The theory of irrationalities of the third degree*, Translations of Mathematical Monographs **10**, Amer. Math. Soc., Providence, 1964. MR 28 #3955 Zbl 0133.30202

[Eisenstein 1844] G. Eisenstein, "Théorèmes sur les formes cubiques et solution d'une équation du quatrième degré à quatre indéterminées", *J. Reine Angew. Math.* **27** (1844), 75–79.

[Ennola and Turunen 1985] V. Ennola and R. Turunen, "On totally real cubic fields", *Math. Comp.* **44**:170 (1985), 495–518. MR 86e:11100 Zbl 0564.12006

[Gan et al. 2002] W. T. Gan, B. Gross, and G. Savin, "Fourier coefficients of modular forms on G_2", *Duke Math. J.* **115**:1 (2002), 105–169. MR 2004a:11036 Zbl 1165.11315

[Hasse 1930] H. Hasse, "Arithmetische Theorie der kubischen Zahlkörper auf klassenkörpertheoretischer Grundlage", *Math. Z.* **31**:1 (1930), 565–582. MR 1545136

[Hilbert 1900] D. Hilbert, "Theorie der algebraischen Zahlkörper", *Encykl. d. math. Wiss.* **1** (1900), 675–714. JFM 31.0207.01

[Hoffman and Morales 2000] J. W. Hoffman and J. Morales, "Arithmetic of binary cubic forms", *Enseign. Math.* (2) **46**:1-2 (2000), 61–94. MR 2001h:11048 Zbl 0999.11021

[Marcus 1977] D. A. Marcus, *Number fields*, Springer, New York, 1977. MR 56 #15601 Zbl 0383.12001 Zbl

[Scholz 1932] A. Scholz, "Über die Beziehung der Klassenzahlen quadratischer Körper zueinander", *J. Reine Angew. Math.* **166** (1932), 201–203. Zbl 0004.05104

Communicated by Raman Parimala
Received 2009-06-18 Revised 2009-12-05 Accepted 2010-05-15

mantilla@math.ubc.ca *Department of Mathematics, University of Wisconsin-Madison, 480 Lincoln Drive, Madison, WI 53705, United States*

ALGEBRA AND NUMBER THEORY 4:6(2010)

Parabolic induction and Hecke modules in characteristic p for p-adic GL_n

Rachel Ollivier

We classify the simple supersingular modules for the pro-p-Iwahori Hecke algebra $\mathcal{H}$ of p-adic GL_n by proving a conjecture by Vignéras about a *mod p numerical Langlands correspondence* on the side of the Hecke modules. We define a process of induction for $\mathcal{H}$-modules in characteristic p that reflects the parabolic induction for representations of the p-adic general linear group and explore the semisimplification of the standard nonsupersingular $\mathcal{H}$-modules in light of this process.

1. Introduction 701
2. Affine root system and Weyl groups 704
3. Hecke algebras and universal modules 711
4. Pro-p-Iwahori Hecke algebra relative to a Levi subgroup of G 712
5. Inducing Hecke modules 718
6. Parabolic induction and compact induction 730
7. Supersingular modules 733
8. Generic spherical Hecke algebra and Iwahori–Hecke algebra 735
Acknowledgments 741
References 741

1. Introduction

Let F be a p-adic field and let $n \geq 1$ be an integer. When exploring the category of smooth mod p representations of $\mathrm{GL}_n(F)$, it is natural to consider the functor that associates to such a representation its subspace of invariant vectors under the action of the pro-p-Iwahori subgroup of $\mathrm{GL}_n(F)$. It has values in the category of right modules in characteristic p over the pro-p Hecke algebra $\mathcal{H}$. The structure of this Hecke algebra has been studied by Vignéras [2005], and the classification of the simple modules in the case $n = 3$ is given in [Ollivier 2006b]. Three families of $\mathcal{H}$-modules appear, namely, the *regular, singular*, and *supersingular* ones. This

MSC2000: primary 20C08; secondary 20G05, 22E50.
Keywords: mod p representations of Hecke algebras and p-adic groups, parabolic induction, integral Bernstein presentation, integral Satake transform.

definition resonates with the idea that, just as the regular modules should be related to the principal series and the supersingular modules to the supersingular representations, likewise the singular modules should be related to the hybrid case where one induces a supersingular representation from a strict Levi subgroup of $\mathrm{GL}_n(F)$. The first link has been explored and proves fruitful [Ollivier 2006a; 2006c; Grosse-Klönne 2009; Vignéras 2008]. Except for the isolated case of $\mathrm{GL}_2(\mathbb{Q}_p)$, the link between supersingular modules and representations does not seem tight enough to give substantial information about the supersingular representations [Breuil and Paskunas 2007]. However, a striking numerical coincidence occurs: in this article (Section 7), we prove Conjecture 1 of [Vignéras 2005], which says that any nonzero simple supersingular module contains a character for the affine Hecke subalgebra of $\mathcal{H}$. It implies the following result, which can be seen as a numerical Langlands correspondence on the side of the Hecke modules.

Theorem 1.1. *The number of n-dimensional simple supersingular modules (with fixed action of the uniformizer) over the pro-p-Hecke algebra of $\mathrm{GL}_n(F)$ is equal to the number of smooth irreducible n-dimensional mod p representations of the absolute Galois group of F (with fixed determinant of a Frobenius).*

The aim of Sections 5 and 6 is to investigate the nonsupersingular Hecke modules. We define a process of *induction* for Hecke modules in characteristic p and relate it to the parabolic induction on the side of the representations of $\mathrm{GL}_n(F)$. In characteristic zero, one of the ingredients for the construction of types by covers consists in embedding a Hecke algebra relative to a Levi subgroup into a Hecke algebra relative to $\mathrm{GL}_n(F)$ using Iwahori decomposition and the notion of *positive subalgebra*. This allows a reading of the parabolic induction of representations in terms of induction on the side of the Hecke modules [Bushnell and Kutzko 1998, §6]. Some of these results can be adapted to the case of mod ℓ representations when $\ell \neq p$ [Vignéras 1998; Dat 1999]. In characteristic p, one cannot expect an injection of the pro-p Hecke algebra $\mathcal{H}(L)$ relative to a strict standard Levi subgroup L into the pro-p Hecke algebra of $\mathrm{GL}_n(F)$. Nevertheless, it is still true for the *positive* part $\mathcal{H}(L^+)$ of $\mathcal{H}(L)$. We now provide a summary of the results proved in this article, keeping in mind that all the modules have mod p coefficients.

Let $\mathfrak{M}$ be a right $\mathcal{H}(L)$-module with scalar action of the uniformizers. The $\mathcal{H}$-module *induced* from $\mathfrak{M}$ is defined in Section 5A by the tensor product over $\mathcal{H}(L^+)$ of $\mathfrak{M}$ by $\mathcal{H}$. This process of induction defines an exact functor from the category of $\mathcal{H}(L)$-modules with scalar action of the uniformizers into the category of right $\mathcal{H}$-modules.

In Section 5B, we recall the definition of a standard $\mathcal{H}$-module: a regular, singular or supersingular character (with values in a field with characteristic p) of the commutative part $\mathcal{A}$ of $\mathcal{H}$ gives rise to a standard module. This standard module

and any of its quotients are then called regular, singular or supersingular respectively. Any simple $\mathcal{H}$-module is a quotient of a standard module. We show in Section 5C that the standard modules relative to *L-adapted* characters of $\mathcal{A}$ are *induced* from $\mathcal{H}(L)$-modules in the sense defined above. These are a special case of nonsupersingular standard modules. Owing to intertwining operators defined in Section 5D, any nonsupersingular standard module can be related to a standard module of this kind. We then give sufficient conditions for these operators to be isomorphisms, from which we deduce:

• Assuming that Conjecture 5.20 is true, we bolster the definition of *nonsupersingular modules* with the proof that any simple nonsupersingular $\mathcal{H}$-module appears in the semisimplification of a standard module that is *induced* from a $\mathcal{H}(L)$-module, where L is a strict Levi subgroup of $\mathrm{GL}_n(F)$. We prove the conjecture and its consequence for the simple modules that are actually modules over the Iwahori–Hecke algebra. The key to this proof is a theorem by Rogawski [1985] which relies on the Kazhdan–Lusztig polynomials for the Iwahori–Hecke algebra in characteristic zero (Section 5E).

• We show that if an irreducible $\mathcal{H}(L)$-module $\mathfrak{M}$ satisfies Hypothesis ($\star$), it gives rise by induction to an irreducible $\mathcal{H}$-module (Section 5F).

• In Section 6B, we consider the compact induction $\mathcal{U}$ (resp. $\mathcal{U}_L$) of the trivial character of the pro-p-Iwahori subgroup of $\mathrm{GL}_n(F)$ (resp. L), and relate the representation $\mathfrak{M}\otimes_{\mathcal{H}(L^+)}\mathcal{U}$ to the one which is parabolically induced from $\mathfrak{M}\otimes_{\mathcal{H}(L)}\mathcal{U}_L$. Denote the latter representation of $\mathrm{GL}_n(F)$ by $\rho_{\mathfrak{M}}$.

We compare the $\mathcal{H}$-module induced from $\mathfrak{M}$ with the pro-p-invariant subspace of $\rho_{\mathfrak{M}}$. So far we have made no specific hypothesis about the p-adic field F, the Levi subgroup L, or the $\mathcal{H}(L)$-module $\mathfrak{M}$ with scalar action of the uniformizers.

In Section 6D we give some examples in the case where $F = \mathbb{Q}_p$ and the standard Levi subgroup L is isomorphic to a product of $\mathrm{GL}_1(\mathbb{Q}_p)$'s and $\mathrm{GL}_2(\mathbb{Q}_p)$'s. In these cases, the irreducible representations of L and the corresponding Hecke modules are thoroughly understood. Our process of induction describes explicitly the pro-p-invariant subspace of $\rho_{\mathfrak{M}}$, which is irreducible as a Hecke module in the chosen examples. After the first version of this article was written, however, Herzig announced that he could prove that these representations $\rho_{\mathfrak{M}}$ are actually irreducible.

While this article does not draw on Herzig's work [2010, Theorem 8.1], it is noticeable that Hypothesis ($\star$) reflects parallel conditions. Our approach, which focuses on the Hecke modules, does not require any further hypotheses on F and L. A barrier to further investigation of the pro-p-invariant subspace of the irreducible induced representations classified in [Herzig 2010] is the lack of knowledge of the (pro-p-invariants of) supersingular representations of L, for general L and F.

In Section 8, we work with the Iwahori–Hecke algebra. Using [Schneider and Teitelbaum 2006], which deals with p-adic Hecke algebras, we make an integral Satake transform for the generic Iwahori–Hecke algebra of $\mathrm{GL}_n(F)$ explicit. By analyzing the map (8-7), Barthel and Livné's method for producing unramified representations [1995] can then be related to the construction of representations arising from the natural left adjoint of the functor of the Iwahori-invariants.

2. Affine root system and Weyl groups

2A. We consider an affine root datum $(\Lambda, \check{\Lambda}, \Phi, \check{\Phi}, \Pi, \check{\Pi})$; for this notion and the facts in the subsequent review, see [Lusztig 1989, 1]. An element of the free abelian group Λ is called a *weight*. We will denote by $\langle\,.\,,\,.\,\rangle$ the perfect pairing on $\Lambda \times \check{\Lambda}$. The elements of $\check{\Lambda}$ are the *coweights*. The elements in $\Phi \subset \Lambda$ are the *coroots*, while those in $\check{\Phi} \subset \check{\Lambda}$ are the *roots*. There is a correspondence $\alpha \leftrightarrow \check{\alpha}$ between roots and coroots satisfying $\langle \alpha, \check{\alpha}\rangle = 2$. The set Π of simple coroots is a basis for Φ, and the corresponding set $\check{\Pi}$ of simple roots is a basis for $\check{\Phi}$. Let $\check{\Phi}^+$ and $\check{\Phi}^-$ denote, respectively the set of roots which are positive and negative with respect to $\check{\Pi}$. There is a partial order on $\check{\Phi}$ given by $\check{\alpha} \leq \check{\beta}$ if and only if $\check{\beta} - \check{\alpha}$ is a linear combination with (integral) nonnegative coefficients of elements in $\check{\Pi}$. Denote by Π_m the set of coroots such that the associated root is a minimal element in $\check{\Phi}$ for $\leq$.

To the (simple) root $\check{\alpha}$ corresponds the (simple) reflection $s_\alpha : \lambda \mapsto \lambda - \langle \lambda, \check{\alpha}\rangle\, \alpha$, which leaves Φ stable. Reciprocally, we will denote by $\check{\alpha}_s$ the simple root associated to the simple reflection s. The finite Weyl group W_0 is the subgroup of $\mathrm{GL}(\Lambda)$ generated by the simple reflections s_α for $\alpha \in \Pi$. It is a Coxeter system with generating set $S_0 = \{s_\alpha,\ \alpha \in \Pi\}$. We will denote by $(w_0, \lambda) \mapsto {}^{w_0}\lambda$ the natural action of W_0 on the set of weights and by $W_0(\lambda)$ the stabilizer of a weight λ under the action of W_0. This action induces a natural action of W_0 on the coweights which stabilizes the set of roots. The set Λ acts on itself by translations: for any weight λ, we denote by e^λ the associated translation. The Weyl group W is the semidirect product of W_0 and Λ. For $w_0 \in W_0$ and $\lambda \in \Lambda$, observe that $w_0 e^\lambda = e^{{}^{w_0}\lambda} w_0$. The affine Weyl group W_{aff} is the semidirect product of W_0 and Φ.

The Weyl group acts on $\check{\Phi} \times \mathbb{Z}$ by

$$w_0 e^\lambda : (\check{\alpha}, k) \mapsto (w_0\check{\alpha},\ k - \langle \lambda, \check{\alpha}\rangle),$$

where we denote by $(w_0, \check{\alpha}) \mapsto w_0\check{\alpha}$ the natural action of W_0 on the roots. Define the set of affine roots by $\check{\mathbf{\Phi}} = \check{\mathbf{\Phi}}^+ \cup \check{\mathbf{\Phi}}^- \subset \check{\Phi} \times \mathbb{Z}$, where

$$\check{\mathbf{\Phi}}^+ := \{(\check{\alpha}, k),\ \check{\alpha} \in \Phi,\ k > 0\} \cup \{(\check{\alpha}, 0),\ \check{\alpha} \in \Phi^+\},$$
$$\check{\mathbf{\Phi}}^- := \{(\check{\alpha}, k),\ \check{\alpha} \in \Phi,\ k < 0\} \cup \{(\check{\alpha}, 0),\ \check{\alpha} \in \Phi^-\},$$

and the set of simple affine roots by $\check{\mathbf{\Pi}} := \{(\check{\alpha}, 0),\ \alpha \in \Pi\} \cup \{(\check{\alpha}, 1),\ \check{\alpha} \in \Pi_m\}$. Identifying $\check{\alpha}$ with $(\check{\alpha}, 0)$, we will often consider Π a subset of $\check{\mathbf{\Pi}}$.

For $A \in \check{\mathbf{\Pi}}$, denote by s_A the associated reflection $s_A = s_\alpha$ if $A = (\check{\alpha}, 0)$ and $s_A = s_\alpha e^\alpha$ if $A = (\check{\alpha}, 1)$. The affine Weyl group is a Coxeter system with generating set

$$S_{\mathrm{aff}} = \{s_A,\ A \in \check{\mathbf{\Pi}}\}.$$

The length on the Coxeter group W_{aff} extends to W in such a way that, for any $w \in W$,

$$\ell(w) := \#\{A \in \check{\mathbf{\Phi}}^+,\ w(A) \in \check{\mathbf{\Phi}}^-\}.$$

The Weyl group is the semidirect product of W_{aff} by the subgroup $\mathbf{\Omega}$ of the elements with length zero. The Bruhat order $\leq$ inflates from W_{aff} to W [Vignéras 2005, Proposition 1].

2B. The length on W has the following properties [Lusztig 1989; Vignéras 2006, appendice]. Let $\lambda, \lambda' \in \Lambda$, $w_0, w_0' \in W_0$, $w \in W$, $A \in \check{\mathbf{\Phi}}$.

2B1. $\ell(ws_A) = \begin{cases} \ell(w) + 1 & \text{if } wA \in \check{\mathbf{\Phi}}^+, \\ \ell(w) - 1 & \text{if } wA \in \check{\mathbf{\Phi}}^-. \end{cases}$

2B2. The quantity $\ell(w_0) + \ell(w_0' e^\lambda) - \ell(w_0 w_0' e^\lambda)$ is twice the number of positive roots $\check{\alpha} \in \check{\Phi}^+$ satisfying

$$\begin{aligned} &w_0'\check{\alpha} \in \check{\Phi}^-, \quad w_0 w_0' \check{\alpha} \in \check{\Phi}^+, \quad \langle \lambda, \check{\alpha} \rangle \geq 0 \quad \text{or} \\ &w_0'\check{\alpha} \in \check{\Phi}^+, \quad w_0 w_0' \check{\alpha} \in \check{\Phi}^-, \quad \langle \lambda, \check{\alpha} \rangle < 0. \end{aligned}$$

2B3. Set $n(\check{\alpha}, w_0 e^\lambda) = \langle \lambda, \check{\alpha} \rangle$ if $w_0 \check{\alpha} \in \check{\Phi}^+$ and $n(\check{\alpha}, w_0 e^\lambda) = 1 + \langle \lambda, \check{\alpha} \rangle$ otherwise. If the integers $n(\check{\alpha}, w_0 e^\lambda)$ and $n(\check{\alpha}, e^{\lambda'})$ have the same sign (or one of them vanishes) for all $\check{\alpha} \in \check{\Phi}^+$, then

$$\ell(w_0 e^{\lambda + \lambda'}) = \ell(w_0 e^\lambda) + \ell(e^{\lambda'}).$$

2C. ***The root datum associated to p-adic*** $\mathbf{GL}_n$.

2C1. We denote by F a nonarchimedean locally compact field with ring of integers $\mathbb{O}$, maximal ideal $\mathcal{P}$ and residue field $\mathbb{F}_q$, where q is a power of p. We choose a uniformizer π and fix the valuation (denoted by val) normalized by $\mathrm{val}(\pi) = 1$ and the corresponding absolute value $|\,.\,|$ such that $|\pi| = q^{-1}$.

Let $n \in \mathbb{N}$, $n \geq 2$. Denote by G the group of F-valued points of the general linear group GL_n, by K_0 the maximal compact $\mathrm{GL}_n(\mathbb{O})$, by I the standard upper Iwahori subgroup of K_0 and by $I(1)$ its unique pro-p-Sylow. It contains the first congruent subgroup K_1 of the matrices in K_0 which are congruent to the identity

modulo π. The element

$$\varpi = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ \vdots & & \ddots & \ddots & \\ 0 & \cdots & & 0 & 1 \\ \pi & 0 & \cdots & \cdots & 0 \end{pmatrix}$$

normalizes the Iwahori subgroup and $\varpi^n = \pi.\mathrm{Id}$ is central in G. Let B denote the upper triangular Borel subgroup of G with Levi decomposition $B = UT$ and modulus character $\delta : B \to \mathbb{Z}[q^{\pm 1}]$.

Consider the affine root datum associated to (G, B, T). The set of cocharacters of T identifies with $\Lambda \simeq T/(T \cap K_0) \simeq \mathbb{Z}^n$. We will also consider it a multiplicative subgroup of G by lifting $T/(T \cap K_0)$ to the subgroup of diagonal matrices with coefficients in $\pi^{\mathbb{Z}}$. The simple positive roots are

$$\check{\alpha}_i : \mathrm{diag}(\pi^{x_1}, \pi^{x_2} \ldots, \pi^{x_n}) \mapsto x_{i+1} - x_i, \quad \text{for} \quad i = 1, \ldots, n-1.$$

Identifying the reflection s_i associated to $\check{\alpha}_i$ with the transposition $(i, i+1)$ gives an isomorphism between the finite Weyl group W_0 and the symmetric group $\mathfrak{S}_n$. We see $W = W_0\Lambda$ as a subgroup of G. It is a system of representatives of the double cosets $I\backslash G/I$.

There is a unique coroot in Π_m and the associated root is $-\check{\alpha}_0$, where $\check{\alpha}_0$ denotes the positive root

$$\check{\alpha}_0 = \check{\alpha}_1 + \cdots + \check{\alpha}_{n-1}.$$

The reflection associated to $(-\check{\alpha}_0, 1)$ is $s_0 = \varpi s_1 \varpi^{-1}$. A generating set for the affine Weyl group is $S_{\mathrm{aff}} = \{s_0, s_1, \ldots, s_{n-1}\}$. The subgroup Ω of W of the elements with length zero is generated by ϖ.

For $s \in S_{\mathrm{aff}}$, denote by $\Phi_s : \mathrm{GL}_2(F) \to G$ the associated morphism [Iwahori and Matsumoto 1965]. Recall that the cocharacter associated to s is the map $F^* \to T$, $x \mapsto \Phi_s\left(\begin{smallmatrix} x & 0 \\ 0 & x^{-1} \end{smallmatrix}\right)$. Denote by $\mathbb{T}_s$ the image of $\mathbb{F}_q^*$ by this cocharacter and set $\phi_s = \Phi_s\left(\begin{smallmatrix} -1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$.

Define the *dominant* and *antidominant* weights respectively by

$$\Lambda_{\mathrm{dom}} = \{\lambda \in \Lambda, \langle \lambda, \check{\alpha} \rangle \geq 0 \text{ for any } \check{\alpha} \in \check{\Phi}^+\},$$
$$\Lambda_{\mathrm{anti}} = \{\lambda \in \Lambda, \langle \lambda, \check{\alpha} \rangle \leq 0 \text{ for any } \check{\alpha} \in \check{\Phi}^+\}.$$

A weight $\mu \in \Lambda$ is said to be *minuscule* if $\langle \mu, \check{\alpha} \rangle \in \{0, \pm 1\}$ for any positive root $\check{\alpha} \in \check{\Phi}^+$. To any subset $J \subset \{1, \ldots, n\}$ corresponds a minuscule weight μ_J defined by $(\mu_J)_i = \pi$ if $i \in J$, $(\mu_J)_i = 1$ otherwise. The semigroup Λ_{anti} of the antidominant weights is generated by the minuscule antidominant weights

$$\{\mu_1, \ldots, \mu_{n-1}, \mu_n^{\pm 1}\},$$

where, for $i \in \{1, \dots, n\}$, we denote by μ_i the minuscule weight associated to $\{1, \dots, i\}$. Set $\mu_0 := \mu_\varnothing$.

2C2. The Weyl group W of G identifies with the quotient of the normalizer $N_G(T)$ of T in G by $T \cap K_0$. The extended Weyl group $W^{(1)}$ of G is defined to be the quotient $N_G(T)/(T \cap K_1)$. We have an exact canonically split sequence

$$0 \to \mathbb{T} \to W^{(1)} \to W \to 0,$$

where $\mathbb{T}$ denotes the finite diagonal torus of the Chevalley group $\mathrm{GL}_n(\mathbb{F}_q)$. For any subset X of W we will denote by $X^{(1)}$ its inverse image in $W^{(1)}$. In particular, the set of extended weights $\Lambda^{(1)}$, which identifies with the direct product of Λ by $\mathbb{T}$, is seen as the set of translations on itself. Again, for any extended weight λ, we denote by e^λ the associated translation. An extended weight is said to be dominant, antidominant, or minuscule if its component in Λ is so. The action of the extended Weyl group on $\Lambda^{(1)}$ and on $\check{\Phi} \times \mathbb{Z}$ is the one inflated from the action of W. By Teichmüller lifting, we identify $\Lambda^{(1)}$ and $W^{(1)} = W_0\Lambda^{(1)}$ with subgroups of G. The extended affine Weyl group $W^{(1)}_{\mathrm{aff}}$ is generated by $S^{(1)}_{\mathrm{aff}}$. The length function on W extends to $W^{(1)}$ in such a way that the elements of $\mathbb{T}$ have length zero.

The extended Weyl group $W^{(1)}$ is a system of representatives of the double cosets $I(1)\backslash G/I(1)$.

2C3. Throughout, we fix a standard Levi subgroup $L = L_1 \times \cdots \times L_m$ in G, where $L_j \simeq \mathrm{GL}_{n_j}(F)$ for $j \in \{1, \dots, m\}$ with $n_1 + \cdots + n_m = n$. Set $\Delta := \{1, \dots, n-1\}$ and define its subset Δ_L to be the set of i such that $s_i \in L$. Denote by $W_{0,L}$ the finite Weyl group of L. It is a Coxeter group generated by $\{s_i, i \in \Delta_L\}$. Denote by $\check{\Phi}_L \subset \check{\Phi}$ the set of associated roots, and by $\check{\Phi}^+_L = \check{\Phi}_L \cap \check{\Phi}^+$ the set of positive ones. The Weyl group W_L of L is the semidirect product of $W_{0,L}$ by Λ. The extended Weyl group $W^{(1)}_L$ of L is the semidirect product of $W_{0,L}$ by $\Lambda^{(1)}$.

Proposition 2.1. *There exists a system D_L of representatives of the right cosets $W_{0,L}\backslash W_0$ such that*

$$\ell(w_0 d) = \ell(w_0) + \ell(d) \quad \textit{for all } w_0 \in W_{0,L},\ d \in D_L. \tag{2-1}$$

Any $d \in D_L$ is the unique element with minimal length in $W_{0,L}d$.

Proof. The proposition is proved in [Carter 1985, 2.3.3], where D_L is explicitly given by

$$D_L := \{d \in W_0,\ d^{-1}\check{\Phi}^+_L \subset \check{\Phi}^+\}. \tag{2-2}$$

This concludes the proof. □

Proposition 2.2. *Let $d \in D_L$ and $s \in S_0$.*

(1) *If $\ell(ds) = \ell(d) - 1$ then $ds \in D_L$.*

(2) *If $\ell(ds) = \ell(d) + 1$ then either $ds \in D_L$ or $W_{0,L}ds = W_{0,L}d$.*

Proof. Suppose $ds \notin D_L$. Let $i \in \{1, \dots, n-1\}$ be such that $s = s_i$. Since $d \in D_L$ and $ds_i \notin D_L$, there is an element $\check{\beta} \in \check{\Phi}_L^+$ such that $d^{-1}\check{\beta} \in \check{\Phi}^+$ and $s_i d^{-1}\check{\beta} \notin \check{\Phi}^+$. But $\check{\alpha}_i$ is the only positive root made negative by s_i [Carter 1985, Proposition 2.2.6], so $d^{-1}\check{\beta} = \check{\alpha}_i$. This implies in particular that $d\check{\alpha}_i \in \check{\mathbf{\Phi}}^+$, and so $\ell(ds_i) = \ell(d) + 1$ by 2B1. The fact that $d\check{\alpha}_i$ belongs to $\check{\Phi}_L$ ensures that $ds_i d^{-1} \in W_{0,L}$. □

2C4. We denote the upper standard parabolic subgroup associated to L by P. It has Levi decomposition $P = LN$, and $\overline{N}$ will denote the opposite unipotent subgroup. The Iwahori subgroup decomposes into $I = I^+ I_L I^-$, where

$$I^+ = I \cap N, \quad I_L = I \cap L, \quad I^- = I \cap \overline{N}.$$

We also set $I_L(1) := I(1) \cap L$. As in [Vignéras 1998, II.4] and [Bushnell and Kutzko 1998, 6], we consider the semigroup L^+ of *L-positive elements*: an element $w \in L$ is called L-positive if it contracts I^+ and dilates I^-, that is,

$$wI^+w^{-1} \subset I^+ \quad \text{and} \quad w^{-1}I^-w \subset I^-.$$

The elements w in $W_L^{(1)}$ which are L-positive are the ones satisfying

$$w(\check{\Phi}^+ - \check{\Phi}_L^+) \subset \check{\mathbf{\Phi}}^+. \tag{2-3}$$

A weight $\lambda \in \Lambda^{(1)}$ is said to be L-positive if the associated translation in $W^{(1)}$ is L-positive. It means that $\langle \lambda, \check{\alpha} \rangle \leq 0$ for any $\check{\alpha} \in \check{\Phi}^+ - \check{\Phi}_L^+$. For example, if L is the diagonal torus, a weight λ is T-positive if and only if it is antidominant.

The set D_L is also a system of representatives of the right cosets $W_L \backslash W$, and we have a weak analog of (2-1):

Lemma 2.3. *For any $w \in W_L^{(1)}$ which is L-positive and any $d \in D_L$,*

$$\ell(wd) = \ell(w) + \ell(d). \tag{2-4}$$

Proof. Let $d \in D_L$, and let $w \in W_L^{(1)}$ be a L-positive element. Write $w = e^{\lambda} w_0$. Equality (2-4) is equivalent to $\ell(d^{-1}) + \ell(w_0^{-1}e^{-\lambda}) - \ell(d^{-1}w_0^{-1}e^{-\lambda}) = 0$.

Let $\check{\alpha} \in \check{\Phi}^+$ be a positive root. Suppose $w_0^{-1}\check{\alpha} \in \check{\Phi}^+$ and $d^{-1}w_0^{-1}\check{\alpha} \in \check{\Phi}^-$. Then by (2-2) and (2-3), one has $w_0^{-1}\check{\alpha} \in \check{\Phi}^+ - \check{\Phi}_L^+$ and $w(w_0^{-1}\check{\alpha}, 0) = (\check{\alpha}, -\langle \lambda, \check{\alpha} \rangle) \in \check{\mathbf{\Phi}}^+$, so $\langle -\lambda, \check{\alpha} \rangle \geq 0$. In the same way, one gets $\langle -\lambda, \check{\alpha} \rangle < 0$ if $w_0^{-1}\check{\alpha} \in \check{\Phi}^-$ and $d^{-1}w_0^{-1}\check{\alpha} \in \check{\Phi}^+$. Applying the length property 2B2 then gives the required equality. □

Lemma 2.4. *The set $I(1)L^+K_0$ is the disjoint union of the sets $I(1)L^+dI(1)$ where d runs over D_L.*

Proof. Lemma 2.3 implies that $I(1)w^+dI(1) = I(1)w^+I(1)dI(1)$ for any $d \in D_L$ and any L-positive $w^+ \in W_L^{(1)}$. So the set $I(1)L^+I(1)dI(1)$ is the disjoint union of the sets $I(1)w^+dI(1)$, where w^+ runs over the L-positive elements of $W_L^{(1)}$. It

is equal to $I(1)L^+dI(1)$. In particular, the sets $I(1)L^+dI(1)$ are pairwise disjoint for $d \in D_L$.

The set $I(1)L^+K_0$ is the union of the sets $I(1)w^+I(1)w_0dI(1)$, where d runs over D_L, w_0 over $W_{0,L}$ and w^+ over the L-positive elements in $W_L^{(1)}$. By Proposition 2.1, we have $I(1)w_0dI(1) = I(1)w_0I(1)dI(1)$, so $I(1)w^+I(1)w_0dI(1) = I(1)w^+I(1)w_0I(1)dI(1)$, and, since w^+ and w_0 are L-positive,

$$I(1)w^+I(1)w_0dI(1) = I(1)w^+I_L(1)w_0I(1)dI(1) \subset I(1)L^+I(1)dI(1). \qquad \square$$

Proposition 2.5. *There is a system $\mathfrak{D}$ of representatives of the right cosets $W_0\backslash W$ such that*

$$\ell(w_0d) = \ell(w_0) + \ell(d), \quad \textit{for all } w_0 \in W_0,\ d \in \mathfrak{D}. \tag{2-5}$$

Any $d \in \mathfrak{D}$ is the unique element with minimal length in W_0d.

Proof. Set

$$\mathfrak{D} := \{d \in W,\ d^{-1}\check{\Phi}^+ \subset \check{\mathbf{\Phi}}^+\}.$$

First check that the cosets W_0d are pairwise disjoint for $d \in \mathfrak{D}$. Let $d, d' \in \mathfrak{D}$, $w_0, w_0' \in W_0$ be such that $w_0d = w_0'd'$. If $d \neq d'$, then $w_0 \neq w_0'$ and there exists a simple root $\check{\beta} \in \check{\Pi}$ such that $\ell(s_\beta w_0^{-1}w_0') = \ell(w_0^{-1}w_0') - 1$, that is, $(w_0'^{-1}w_0)\check{\beta} = (d'd^{-1})\check{\beta} \in \check{\Phi}^-$. But $d' \in \mathfrak{D}$, and hence $d^{-1}\check{\beta} \in d'^{-1}(\check{\Phi}^-) \subset \check{\mathbf{\Phi}}^-$, which contradicts the fact that $d \in \mathfrak{D}$.

For $w \in W$, we prove by induction on the length of w that there exists an (obviously unique) $(w_0, d) \in W_0 \times \mathfrak{D}$ such that $w = w_0d$ and $\ell(w_0d) = \ell(w_0) + \ell(d)$.

By 2B1, saying that w does not belong to $\mathfrak{D}$ means that there exists a simple root $\check{\alpha} \in \check{\Pi}$ such that $\ell(s_\alpha w) = \ell(w) - 1$. In particular, if w has length 0, it belongs to $\mathfrak{D}$. Suppose now that $\ell(w) > 0$ and that it does not belong to $\mathfrak{D}$. Then, by induction, there exists $(w_0, d) \in W_0 \times \mathfrak{D}$ with $s_\alpha w = w_0d$ and $\ell(s_\alpha w) = \ell(w_0) + \ell(d)$, where α is chosen as before. So $w = s_\alpha w_0 d$ and

$$\ell(w) = \ell(s_\alpha w) + 1 = \ell(w_0) + \ell(d) + 1.$$

Verifying that $\ell(w) = \ell(s_\alpha w_0) + \ell(d)$ is just verifying that $\ell(s_\alpha w_0) = \ell(w_0) + 1$, which is true, since otherwise $\ell(s_\alpha w_0) < \ell(w_0)$ and $\ell(w) \leq \ell(s_\alpha w_0) + \ell(d) < \ell(w_0) + \ell(d) = \ell(w) - 1$.

We have proved that $\mathfrak{D}$ is a system of representatives of the right cosets $W_0\backslash W$ and that it satisfies (2-5). In particular, any $d \in \mathfrak{D}$ is the unique element with minimal length in W_0d, since $w_0 \in W_0$ has length zero if and only if $w_0 = 1$. $\square$

Lemma 2.6. *Any $d \in \mathfrak{D}$ can be written $d = e^\lambda w_0 \in W$, with $w_0 \in W_0$ and $\lambda \in \Lambda$ a dominant weight such that*

$$\ell(e^\lambda w_0) + \ell(w_0^{-1}) = \ell(e^\lambda).$$

Proof. By definition of the set $\mathfrak{D}$, we have $(w_0^{-1}\check{\alpha}, \langle\lambda, \check{\alpha}\rangle) \in \check{\mathbf{\Phi}}^+$ for every $\check{\alpha} \in \check{\Phi}^+$. Then λ is dominant and $w_0^{-1}\check{\alpha} \in \check{\Phi}^+$ if $\check{\alpha} \in \check{\Phi}^+$ satisfies $\langle\lambda, \check{\alpha}\rangle = 0$. Applying the length property 2B2, one gets the required equality. □

Proposition 2.7. *Let $d \in \mathfrak{D}$ and $s \in S_{\mathrm{aff}}$.*

(1) *If $\ell(ds) = \ell(d) - 1$ then $ds \in \mathfrak{D}$.*

(2) *If $\ell(ds) = \ell(d) + 1$ then either $ds \in \mathfrak{D}$ or $W_0 ds = W_0 d$.*

Proof. Write $d = e^\lambda w_0 \in W$.

(A) We first prove the proposition for s in the finite Weyl group; write $s = s_i$ with $1 \le i \le n-1$. Saying that $ds_i \notin \mathfrak{D}$ means that there exists $\check{\beta} \in \check{\Phi}^+$ such that $d^{-1}\check{\beta} = (\check{\alpha}_i, 0)$, since $(\check{\alpha}_i, 0)$ is the only positive affine root made negative by s_i. This implies in particular that $d\check{\alpha}_i \in \check{\mathbf{\Phi}}^+$, so $\ell(ds_i) = \ell(d) + 1$. We have

$$\check{\beta} = w_0\check{\alpha}_i, \quad \langle\lambda, w_0\check{\alpha}_i\rangle = 0.$$

The latter equality means that $w_0 s_i w_0^{-1}$ fixes λ, so

$$ds_i = e^\lambda w_0 s_i = w_0 s_i w_0^{-1} e^\lambda w_0 \in W_0 d.$$

(B) Now suppose $s = s_0$. Recall that the associated affine simple root is $(-\check{\alpha}_0, 1)$. The coroot α_0 can be seen as the diagonal matrix $(\pi^{-1}, 1, \dots, 1, \pi)$. Write $s_0 = \rho e^{-\alpha_0}$, where ρ denotes the reflection sending α_0 to its opposite. Saying that $\ell(ds_0) = \ell(d) + 1$ means that $d(-\check{\alpha}_0, 1) \in \check{\mathbf{\Phi}}^+$, that is, we are either in case (a) or in case (b):

(a) $\langle\lambda, w_0\check{\alpha}_0\rangle \ge 0$,

(b) $w_0\check{\alpha}_0 \in \check{\Phi}^-$ and $\langle\lambda, w_0\check{\alpha}_0\rangle = -1$.

Saying that $\ell(ds_0) = \ell(d) - 1$ means that $d(-\check{\alpha}_0, 1) \in \check{\mathbf{\Phi}}^-$, so we are in case (c) (note that since λ is dominant, it is impossible to simultaneously have the conditions $w_0\check{\alpha}_0 \in \Phi^+$ and $\langle\lambda, w_0\check{\alpha}_0\rangle = -1$):

(c) $\langle\lambda, w_0\check{\alpha}_0\rangle < -1$.

By definition of the reflection ρ, hypothesis (b) says that ${}^{w_0\rho w_0^{-1}}\lambda = \lambda + {}^{w_0}\alpha_0$, so that we have $ds_0 = e^\lambda w_0 \rho e^{-\alpha_0} = w_0 \rho w_0^{-1} e^\lambda w_0 \in W_0 d$.

Suppose that we are under hypothesis (a) or (c), that is, $\langle\lambda, w_0\check{\alpha}_0\rangle \ne -1$. Take $\check{\beta} \in \check{\Pi}$. Under the action of $s_0 d^{-1}$, it becomes the affine root

$$s_0 d^{-1}\check{\beta} = (\rho w_0^{-1}\check{\beta}, \langle\lambda, \check{\beta}\rangle + \langle\alpha_0, w_0^{-1}\check{\beta}\rangle).$$

Let us check that it belongs to $\check{\mathbf{\Phi}}^+$, which will prove that $ds_0 \in \mathfrak{D}$. Recall that $d \in \mathfrak{D}$, so

$$d^{-1}\check{\beta} = (w_0^{-1}\check{\beta}, \langle\lambda, \check{\beta}\rangle) \in \check{\mathbf{\Phi}}^+.$$

First we verify that $\langle \alpha_0, w_0^{-1}\check{\beta}\rangle + \langle \lambda, \check{\beta}\rangle \geq 0$. Since $\langle \alpha_0, w_0^{-1}\check{\beta}\rangle \in \{0, \pm 1, \pm 2\}$, the required inequality is true if $\langle \lambda, \check{\beta}\rangle \geq 2$. If $\langle \lambda, \check{\beta}\rangle = 0$, then $w_0^{-1}\check{\beta} \in \Phi^+$ and $\langle \alpha_0, w_0^{-1}\check{\beta}\rangle \geq 0$. If $\langle \lambda, \check{\beta}\rangle = 1$ then, by the chosen hypotheses, $w_0^{-1}\check{\beta} \neq -\check{\alpha}_0$, so $\langle \alpha_0, w_0^{-1}\check{\beta}\rangle \neq -2$.

Finally, we have to show that $\langle \alpha_0, w_0^{-1}\check{\beta}\rangle + \langle \lambda, \check{\beta}\rangle = 0$ implies $\rho w_0^{-1}\check{\beta} \in \check{\Phi}^+$. A positive root $\check{\gamma}$ becomes a positive root under the action of ρ if and only if it is fixed by the action of ρ, or in other words, if $\langle \alpha_0, \check{\gamma}\rangle = 0$. Suppose that

$$\langle \lambda, \check{\beta}\rangle = \langle \alpha_0, w_0^{-1}\check{\beta}\rangle = 0;$$

then $w_0^{-1}\check{\beta} \in \check{\Phi}^+$, and so, by the preceding remark, $\rho w_0^{-1}\check{\beta} \in \check{\Phi}^+$. Suppose that

$$\langle \lambda, \check{\beta}\rangle = -\langle \alpha_0, w_0^{-1}\check{\beta}\rangle > 0;$$

then $w_0^{-1}\check{\beta} \in \check{\Phi}^-$, and by the preceding remark, $\rho w_0 \check{\beta} \in \check{\Phi}^+$. □

3. Hecke algebras and universal modules

3A. Consider the Chevalley group $\mathbb{G} = \mathrm{GL}_n(\mathbb{F}_q)$ and its standard upper Borel subgroup $\mathbb{B}$ with Levi decomposition $\mathbb{B} = \mathbb{T}\mathbb{U}$. We denote by $\overline{\mathbb{U}}$ the opposite unipotent subgroup. The double cosets $\mathbb{U}\backslash\mathbb{G}/\mathbb{U}$ are represented by the extended Weyl group of $\mathbb{G}$, which is isomorphic to the extended finite Weyl group $W_0^{(1)}$ of G. The *finite universal module* $\mathbb{Z}[\mathbb{U}\backslash\mathbb{G}]$ of $\mathbb{Z}$-valued functions with support on the right cosets $\mathbb{U}\backslash\mathbb{G}$ is endowed with a natural action of $\mathbb{G}$. The ring $\mathcal{H}(\mathbb{G}, \mathbb{U})$ of its $\mathbb{Z}[\mathbb{G}]$-endomorphisms will be called the *finite Hecke ring*. By Frobenius reciprocity, a $\mathbb{Z}$-basis of the latter identifies with the characteristic functions of the double cosets $\mathbb{U}\backslash\mathbb{G}/\mathbb{U}$.

We call the space $\mathbb{Z}[I(1)\backslash G]$ of $\mathbb{Z}$-valued functions with finite support on the right cosets $I(1)\backslash G$ the *pro-p-universal module*. It is endowed with an action of G. The subspace of the functions that are actually left invariant under the Iwahori subgroup constitute a G-subspace that is isomorphic to the universal module $\mathbb{Z}[I\backslash G]$.

The $\mathbb{Z}$-ring of the $\mathbb{Z}[G]$-endomorphisms of $\mathbb{Z}[I(1)\backslash G]$ will be called the *pro-p-Hecke ring* and denoted by $\mathcal{H}(G, I(1))$. By Frobenius reciprocity, $\mathcal{H}(G, I(1))$ is seen as the convolution ring of the functions with finite support on the double cosets of G modulo $I(1)$. Among these functions, the ones that are actually biinvariant under the Iwahori subgroup constitute a ring that is isomorphic to the Iwahori–Hecke ring $\mathcal{H}(G, I)$ of the $\mathbb{Z}[G]$-endomorphisms of $\mathbb{Z}[I\backslash G]$.

A $\mathbb{Z}$-basis for $\mathcal{H}(G, I(1))$ (resp. $\mathcal{H}(G, I)$) is given by the characteristic functions of the double cosets $I(1)\backslash G/I(1)$ (resp. $I\backslash G/I$).

For $w \in W^{(1)}$, we denote by τ_w the element of $\mathcal{H}(G, I(1))$ corresponding to the associated double coset. The subalgebra generated by the elements τ_w for $w \in W^{(1)}_{\mathrm{aff}}$ is called the affine Hecke algebra.

The subspace of $\mathbb{Z}[I(1)\backslash G]$ of the functions with support in K_0 identifies with the finite universal module. Among the $\mathbb{Z}[G]$-endomorphisms of the pro-p-universal module, those stabilizing this subspace form a subring that identifies with the finite Hecke algebra. It is the subring generated by the elements τ_w for $w \in W^{(1)}_0$.

Fix k an algebraic closure of $\mathbb{F}_q$. The space $\mathbb{Z}[I(1)\backslash G] \otimes_{\mathbb{Z}} k$ is endowed with a smooth action of G and is isomorphic to the compact induction $\mathrm{ind}_{I(1)}^{G} \mathbf{1}_k$ of the trivial character with values in k of the pro-p-Iwahori subgroup. We will denote by $\mathcal{U}$ this representation of G.

3B. The pro-p-Hecke ring is the ring with $\mathbb{Z}$-basis $(\tau_w)_{w \in W^{(1)}}$ satisfying the *braid* and *quadratic* relations, namely

- $\tau_w \tau_{w'} = \tau_{ww'}$ for any $w,\ w' \in W^{(1)}$ such that $\ell(ww') = \ell(w) + \ell(w')$, and
- $\tau_s^2 = q + \left(\sum_{t \in \mathbb{T}_s} \tau_{\phi_s} \tau_t\right) \tau_s$ for $s \in S_{\mathrm{aff}}$,

in the notation of 2C1. From now on, we consider q an indeterminate and work with the $\mathbb{Z}[q]$-algebra $\mathcal{H}$ with generators $(\tau_w)_{w \in W^{(1)}}$ satisfying the relations above. It will be called the *generic pro-p-Hecke algebra*.

For $w \in W^{(1)}$, set

$$\tau_w^* := q^{\ell(w)} \tau_w^{-1}. \tag{3-1}$$

The map $\mu : \tau_w \mapsto (-1)^{\ell(w)} \tau^*_{w^{-1}}$ defines an involutive algebra endomorphism of $\mathcal{H}$ [Vignéras 2005, Corollary 2].

Remark 3.1. For $s \in S_{\mathrm{aff}}$, one checks that the following equalities hold in $\mathcal{H} \otimes_{\mathbb{Z}[q]} k$:

$$(\tau_s^*)^2 = (\tau_s + \nu_s)^2 = \tau_s^* \nu_s = \nu_s \tau_s^*,$$

where $\nu_s := -\sum_{t \in \mathbb{T}_s} \tau_{\phi_s} \tau_t$.

4. Pro-p-Iwahori Hecke algebra relative to a Levi subgroup of G

The generic pro-p-Hecke algebra $\mathcal{H}(L)$ of the Levi subgroup L is the tensor product of the generic pro-p-Hecke algebras of the L_j's, for $j \in \{1, \dots, m\}$. For any element $w = (w_1, \dots, w_m)$ in the extended Weyl group $W_L^{(1)}$ of L, we will denote by

$$\tau_w^{\otimes} := \bigotimes_{j=1}^{m} \tau_{w_j}$$

the corresponding element of $\mathcal{H}(L)$. Denote by $\mathcal{H}(L^+)$ the subspace of $\mathcal{H}(L)$ generated over $\mathbb{Z}[q]$ by the elements $\tau_w^{\otimes}$ corresponding to L-positive elements w in $W_L^{(1)}$. From [Bushnell and Kutzko 1998, 6.12] and [Vignéras 1998, II], we know that $\mathcal{H}(L^+)$ is a $\mathbb{Z}[q]$-algebra and the following holds.

Proposition 4.1. *The natural injective map* θ_L^+

$$\mathcal{H}(L^+) \to \mathcal{H}, \quad \tau_w^{\otimes} \mapsto \tau_w, \tag{4-1}$$

where $w \in W_L^{(1)}$ *is* L*-positive, respects the product. It extends uniquely into an injective morphism* θ_L *of* $\mathbb{Z}[q^{\pm 1}]$*-algebras*

$$\theta_L : \mathcal{H}(L) \otimes_{\mathbb{Z}[q]} \mathbb{Z}[q^{\pm 1}] \to \mathcal{H} \otimes_{\mathbb{Z}[q]} \mathbb{Z}[q^{\pm 1}].$$

The proof of the second assertion [Bushnell and Kutzko 1998; Vignéras 1998] makes use of the following (strongly) L-positive central element in L:

$$a_L = e^{\lambda_L}, \quad \text{where} \quad \lambda_L = \sum_{j \in \Delta - \Delta_L} \mu_j, \tag{4-2}$$

and the fact that for any $w \in W_L^{(1)}$ there exists $r \in \mathbb{N}$ such that $a_L^r w$ is L-positive. Then $\theta_L(\tau_w^{\otimes})$ is given by $\tau_{a_L}^{-r} \tau_{a_L^r w}$, which is well-defined in $\mathcal{H} \otimes_{\mathbb{Z}[q]} \mathbb{Z}[q^{\pm 1}]$ (and does not depend on the choice of r).

We will call $\mathcal{H}(L^+)$ the *positive subalgebra* of $\mathcal{H}(L)$. We will sometimes identify it with its image in $\mathcal{H}$ without further notice.

4A. ***Classical Bernstein presentation.*** In the case where the Levi subgroup L is the diagonal torus T, the map θ_T is simply denoted by θ and called the Bernstein embedding. It is more traditional to consider its renormalization

$$\begin{aligned} \tilde{\theta} : \mathbb{Z}[q^{\pm 1/2}][\Lambda^{(1)}] &\to \mathcal{H} \otimes_{\mathbb{Z}[q]} \mathbb{Z}[q^{\pm 1/2}], \\ \lambda &\mapsto \delta^{1/2}(\lambda)\theta(\lambda), \end{aligned} \tag{4-3}$$

whose image is denoted by $\mathcal{A}[q^{\pm 1/2}]$, where δ is the modulus character of the Borel subgroup defined in 2C1. The following well-known properties of this commutative subalgebra are proved in, for example, [Lusztig 1989, 3] (and [Vignéras 2005, 1.4] for the extension to the pro-p case). The center of $\mathcal{H} \otimes_{\mathbb{Z}[q]} \mathbb{Z}[q^{\pm 1/2}]$ is the image under $\tilde{\theta}$ of the subspace $\mathbb{Z}[q^{\pm 1/2}][\Lambda^{(1)}]^{W_0}$ of the invariants in $\mathbb{Z}[q^{\pm 1/2}][\Lambda^{(1)}]$ under the natural action of W_0. The Hecke algebra $\mathcal{H} \otimes_{\mathbb{Z}[q]} \mathbb{Z}[q^{\pm 1/2}]$ is a free right module over $\mathcal{A}[q^{\pm 1/2}]$ with basis $\{\tau_{w_0},\ w_0 \in W_0\}$.

4B. ***Integral Bernstein presentation.*** In this section, we recall the results obtained by Vignéras [2005] concerning an integral version of the previous Bernstein presentation. We present them in the light of [Schneider and Teitelbaum 2006].

4B1. Following [Schneider and Teitelbaum 2006, p. 10 and Example 2], we consider the action of W_0 on $\mathbb{Z}[q^{\pm 1/2}][\Lambda^{(1)}]$ twisted by the map

$$\gamma : W_0 \times \Lambda^{(1)} \to \mathbb{Z}[q^{\pm 1/2}], \quad (w_0, \lambda) \mapsto \frac{\delta^{1/2}({}^{w_0}\lambda)}{\delta^{1/2}(\lambda)}. \tag{4-4}$$

This map is a cocycle in the sense that it satisfies

(a) $\gamma(v_0 w_0, \lambda) = \gamma(v_0, {}^{w_0}\lambda)\gamma(w_0, \lambda)$, for $v_0, w_0 \in W_0$ and $\lambda \in \Lambda^{(1)}$,

so we have a well-defined action of W_0 on $\Lambda^{(1)}$ denoted by $(w_0, \lambda) \mapsto w_0 \centerdot \lambda$ and given by

$$w_0 \centerdot \lambda = \gamma(w_0, \lambda)^{w_0}\lambda. \tag{4-5}$$

The map γ also satisfies the following conditions:

(b) $\gamma(w_0, \lambda\mu) = \gamma(w_0, \lambda)\gamma(w_0, \mu)$, for $w_0 \in W_0$ and $\lambda, \mu \in \Lambda^{(1)}$,

(c) $\gamma(w_0, \lambda) = 1$ for $w_0 \in W_0$, $\lambda \in \Lambda^{(1)}$ such that ${}^{w_0}\lambda = \lambda$,

so the twisted action (4-5) extends into an action on $\mathbb{Z}[q^{\pm 1/2}][\Lambda^{(1)}]$, which is compatible with the structure of $\mathbb{Z}[q^{\pm 1/2}]$-algebra.

Lemma 4.2 [Schneider and Teitelbaum 2006, Example 2 and Lemma 4.2]. (1) *For $w_0 \in W_0$, $\lambda \in \Lambda^{(1)}$, one has*

$$\gamma(w_0, \lambda) = \prod_{\check{\alpha} \in \check{\Phi}^+ \cap w_0^{-1}(\check{\Phi}^-)} |\check{\alpha}(\lambda)|,$$

so γ actually takes values in $\mathbb{Z}[q^{\pm 1}]$.

(2) *Any $\lambda \in \Lambda^{(1)}$ can be written $\lambda_1 - \lambda_2$ with λ_1, λ_2 antidominant weights. Let $w_0 \in W_0$ such that ${}^{w_0}\lambda$ is antidominant. Then*

$$\gamma(w_0, \lambda) = q^{-(\ell(\lambda) - \ell(\lambda_1) + \ell(\lambda_2))/2}$$

and it does not depend on the choice of $w_0, \lambda_1, \lambda_2$.

4B2. Let $\lambda \in \Lambda^{(1)}$ and $w_0 \in W_0$ such that ${}^{w_0}\lambda$ is antidominant. Define the element $E(\lambda)$ in $\mathcal{H} \otimes_{\mathbb{Z}[q]} \mathbb{Z}[q^{\pm 1}]$ by

$$E(\lambda) := \gamma(w_0, \lambda)^{-1}\theta(\lambda) = \gamma(w_0^{-1}, {}^{w_0}\lambda)\theta(\lambda). \tag{4-6}$$

It is proved in [Vignéras 2005] that $E(\lambda)$ actually lies in $\mathcal{H}$ (see Theorem 4.5 below for the precise statement). Hence, we have an injective $\mathbb{Z}[q]$-equivariant map

$$E : \mathbb{Z}[q][\Lambda^{(1)}] \to \mathcal{H}, \tag{4-7}$$

but it does not respect the product. The natural action of W_0 on $\Lambda^{(1)}$ induces an action of W_0 on the image $\mathcal{A}$ of E.

Proposition 4.3 (integral Bernstein relations). *Let $\lambda \in \Lambda^{(1)}$ be a weight, $\check{\alpha} \in \check{\Pi}$ a simple root and s the associated reflection. The following holds in $\mathcal{H}$:*

(1) *If $\langle \lambda, \check{\alpha} \rangle = 0$, then $E(\lambda)$ and τ_s commute.*

(2) *If $\langle \lambda, \check{\alpha} \rangle = 1$, then $\tau_s E(\lambda) = E({}^s\lambda)\tau_s^*$ and $E(\lambda)\tau_s = \tau_s^* E({}^s\lambda)$.*

Proof. This is a direct corollary of the classical Bernstein relations proved in [Lusztig 1989, Proposition 3.6] and [Vignéras 2005, Proposition 5]. An integral version of these is proved in [Ollivier 2006a, 4.4.1] (use the involution μ defined in 3B to pass from the definition of the Bernstein map in the latter to the present situation). □

Lemma 4.4. *Let $\lambda \in \Lambda$ be antidominant. Then $E(\lambda) = \tau_{e^\lambda}$ and $E(\lambda^{-1}) = \tau^*_{e^\lambda}$. Suppose also that λ is minuscule. Let $d \in W_0$ with minimal length in $W_0(\lambda)d$. Then*

$$E({}^{d^{-1}}\lambda) = \tau_{d^{-1}e^\lambda}\tau^*_{d^{-1}} \quad \textit{and} \quad \tau_d E({}^{d^{-1}}\lambda) = E(\lambda)\tau^*_{d^{-1}}.$$

Proof. First recall that an element $\lambda \in \Lambda$ is T-positive if and only if it is antidominant. So $\theta(\lambda) = \theta^+_T(\lambda) = \tau_{e^\lambda}$. Then, by Lemma 4.2(2) and since θ respects the product, one has $E(\lambda^{-1}) = q^{\ell(\lambda)}\tau^{-1}_{e^\lambda} = \tau^*_{e^\lambda}$. We have proved the first statement, which gives the second one for the case $d = 1$. Suppose λ is minuscule and show the second one by induction on $\ell(d)$. Let $d \in W_0$ with minimal length in $W_0(\lambda)d$ and $\ell(d) > 0$. Let $s \in S_0$ such that $\ell(ds) = \ell(d) - 1$. Then $\tau^*_{sd^{-1}}\tau^*_s = \tau^*_{d^{-1}}$ and $d\check{\alpha}_s \in \check{\Phi}^-$. The stabilizer $W_0(\lambda)$ is a Coxeter subgroup of W_0, so Proposition 2.2 applies: ds has minimal length in $W_0(\lambda)ds$. In particular, this implies that dsd^{-1} does not stabilize λ, so $\langle\lambda, d\check{\alpha}_s\rangle > 0$. The length property 2B2 then gives $\ell(s) + \ell(d^{-1}e^\lambda) = \ell(sd^{-1}e^\lambda)$. By induction, $E({}^{sd^{-1}}\lambda) = \tau_{sd^{-1}e^\lambda}\tau^*_{sd^{-1}} = \tau_s\tau_{d^{-1}e^\lambda}\tau^*_{sd^{-1}}$. Now work in $\mathcal{H} \otimes_{\mathbb{Z}[q]} \mathbb{Z}[q^{\pm 1}]$ and apply the Bernstein relations (2) to ${}^{d^{-1}}\lambda$:

$$E({}^{d^{-1}}\lambda) = \tau_s^{-1}E({}^{sd^{-1}}\lambda)\tau^*_s = \tau_{d^{-1}e^\lambda}\tau^*_{d^{-1}}.$$

The last equality of the lemma easily follows using 2B2 and the fact that $\langle\lambda, \check{\alpha}\rangle = 0$ implies $d^{-1}\check{\alpha} \in \check{\Phi}^+$ for any $\check{\alpha} \in \check{\Phi}^+$. □

Theorem 4.5 [Vignéras 2005, Theorems 2, 3, and 4]. *The image $\mathcal{A}$ of E is a $\mathbb{Z}[q]$-algebra. It coincides with the intersection $\mathcal{A}[q^{\pm 1/2}] \cap \mathcal{H}$. The action of W_0 on $\mathcal{A}$ is compatible with the structure of $\mathbb{Z}[q]$-algebra.*

A $\mathbb{Z}[q]$-basis for $\mathcal{A}$ is given by $(E(\lambda))_{\lambda \in \Lambda^{(1)}}$.

As a $\mathbb{Z}[q]$-algebra, $\mathcal{A}$ is generated by elements corresponding to minuscule weights, that is, by the elements τ_t for $t \in \mathbb{T}$ and

$$(E(\mu_I))_{I \subsetneq \{1,\dots,n\}},\ E(\mu_{\{1,\dots,n\}})^{\pm 1}$$

with the relations

$$E(\mu_I)E(\mu_J) = q^{bc}E(\mu_{I\cup J})E(\mu_{I\cap J}) \tag{4-8}$$

for any $I, J \subset \{1, \dots, n\}$ with $|I \cap J| = a$, $|I| = a + b$, $|J| = a + c$.

The center of $\mathcal{H}$ is the space of W_0-invariants in $\mathcal{A}$.

As an $\mathcal{A}$-module, $\mathcal{H}$ is finitely generated; as a module over the center, $\mathcal{A}$ is finitely generated.

The proof of the theorem relies on the more general definition of an element $E(w) \in \mathcal{H}$ associated to any $w = e^{\lambda} w_0 \in W^{(1)}$:

$$E(w) := q^{(\ell(w)-\ell(w_0)-\ell(e^{\lambda}))/2} E(\lambda)\tau_{w_0} \tag{4-9}$$

in $\mathcal{H} \otimes_{\mathbb{Z}[q]} \mathbb{Z}[q^{\pm 1/2}]$, and the fact that the elements $(E(w))_{w \in W^{(1)}}$ constitute a $\mathbb{Z}[q]$-basis for $\mathcal{H}$ called the *integral Bernstein basis*.

Remark 4.6. Note that (4-8) implies that in $\mathcal{H} \otimes_{\mathbb{Z}[q]} k$, the product $E(\mu_I)E(\mu_J)$ is zero unless either $I \subset J$ or $J \subset I$.

4C. For $w = (w_1, \ldots, w_m) \in W_L^{(1)}$, we denote by $E^{\otimes}(w) \in \mathcal{H}(L)$ the tensor product of the Bernstein elements corresponding to the elements w_j in the generic pro-p-Hecke algebras of the L_is. The Hecke algebra $\mathcal{H}(L)$ contains the commutative subring $\mathcal{A}_L$ with $\mathbb{Z}[q]$-basis $(E^{\otimes}(\lambda))_{\lambda \in \Lambda^{(1)}}$.

Proposition 4.7. *A $\mathbb{Z}[q]$-basis for the positive subalgebra $\mathcal{H}(L^+)$ is given by*

$$(E^{\otimes}(w))_w,$$

where w runs over the L-positive elements in $W_L^{(1)}$. For any such w, one has

$$\theta_L^+(E^{\otimes}(w)) = E(w). \tag{4-10}$$

Proof. **(A)** We first check that $E^{\otimes}(\lambda)$ lies in the positive subalgebra $\mathcal{H}(L^+)$ for any L-positive weight $\lambda \in \Lambda^{(1)}$. It is enough to show the property for λ minuscule. In this case, using Lemma 4.4, one easily computes $E^{\otimes}(\lambda)$ and checks that the elements of the Iwahori–Matsumoto basis appearing in its decomposition correspond to L-positive elements in $W_L^{(1)}$.

Now consider $w = (w_1, \ldots, w_m) \in W_L^{(1)}$. Write $w = e^{\lambda} v$ with $\lambda \in \Lambda^{(1)}$, $v \in W_{0,L}$. Since $W_{0,L}$ normalizes I^- and I^+, the element w is L-positive if and only if λ is an L-positive weight. Decompose $\lambda = (\lambda_1, \ldots, \lambda_m)$ and $v = (v_1, \ldots, v_m)$ in the Levi L and recall that, after extending the scalars to $\mathbb{Z}[q^{\pm 1/2}]$,

$$E^{\otimes}(w) = \prod_{j=1}^{m} q^{(\ell(w_j)-\ell(v_j)-\ell(e^{\lambda_j}))/2} E^{\otimes}(\lambda)\tau_v^{\otimes}. \tag{4-11}$$

The element $\tau_v^{\otimes}$ lies in the positive subalgebra, and $E^{\otimes}(\lambda)$ does too if w is L-positive, so the property also holds for $E^{\otimes}(w)$.

Once we know that $E^{\otimes}(w)$ lies in the positive subalgebra $\mathcal{H}(L^+)$ for any L-positive element $w \in W_L^{(1)}$, it is clear that these elements constitute a $\mathbb{Z}[q]$-basis of $\mathcal{H}(L^+)$ by using [Vignéras 2006, 1.5].

(B) Let us show Equality (4-10) for L-positive elements of the form e^{λ} with $\lambda \in \Lambda^{(1)}$. The weight λ can be written $\lambda = \mu - \nu$, where $\mu, \nu \in \Lambda^{(1)}$ are anti-dominant weights which decompose into $\mu = (\mu_1, \ldots, \mu_m)$, $\nu = (\nu_1, \ldots, \nu_m)$, so

$\lambda = (\lambda_1, \dots, \lambda_m)$ with $\lambda_i = \mu_i - \nu_i$ for $i = 1, \dots, m$. By definition,

$$E^{\otimes}(\lambda) = \prod_{i=1}^{m} q^{(\ell(e^{\lambda_i}) + \ell(e^{\nu_i}) - \ell(e^{\mu_i}))/2} \theta^{\otimes}(\lambda)$$

and

$$E(\lambda) = q^{(\ell(e^{\lambda}) + \ell(e^{\nu}) - \ell(e^{\mu}))/2} \theta(\lambda).$$

Note that uniqueness in Proposition 4.1 gives $\theta_L \circ \theta^{\otimes} = \theta$, where $\theta^{\otimes}$ denotes the tensor product of the Bernstein maps, so the required equality will be proved once we have checked that

$$\ell(e^{\lambda}) + \ell(e^{\nu}) - \ell(e^{\mu}) = \sum_{i=1}^{m} (\ell(e^{\lambda_i}) + \ell(e^{\nu_i}) - \ell(e^{\mu_i})). \tag{4-12}$$

By the definition of the length on $\Lambda^{(1)}$,

$$\begin{aligned} \ell(e^{\lambda}) + \ell(e^{\nu}) - \ell(e^{\mu}) &= \sum_{\check{\alpha} \in \check{\Phi}^+} |\langle \mu - \nu, \check{\alpha} \rangle| + |\langle \nu, \check{\alpha} \rangle| - |\langle \mu, \check{\alpha} \rangle| \\ &= \sum_{\check{\alpha} \in \check{\Phi}^+} |\langle \mu - \nu, \check{\alpha} \rangle| - \langle \nu, \check{\alpha} \rangle + \langle \mu, \check{\alpha} \rangle. \end{aligned}$$

A positive root $\check{\alpha}$ will give a zero contribution to this sum if and only if $\langle \nu, \check{\alpha} \rangle \geq \langle \mu, \check{\alpha} \rangle$. According to (2-3), the fact that λ is L-positive ensures that it is the case for every $\check{\alpha} \in \check{\Phi}^+ - \check{\Phi}_L^+$. Hence the sum can be restricted to the roots $\check{\alpha} \in \check{\Phi}_L^+$, which proves that (4-12) holds.

We return to the general case of an L-positive element of the form $w = e^{\lambda} v$. By the previous case, applying θ_L to (4-11) gives

$$\theta_L(E^{\otimes}(w)) = \prod_{j=1}^{k} q^{(\ell(w_j) - \ell(v_j) - \ell(e^{\lambda_j}))/2} E(\lambda) \tau_v.$$

Since $E(w) = q^{(\ell(w) - \ell(v) - \ell(e^{\lambda}))/2} E(\lambda) \tau_v$, it remains to check that

$$\sum_{j=1}^{k} (\ell(e^{\lambda_j}) + \ell(v_j) - \ell(e^{\lambda_j} v_j)) = \ell(e^{\lambda}) + \ell(v) - \ell(e^{\lambda} v).$$

By 2B2, the right side of this equality is twice the number of roots $\check{\alpha} \in \check{\Phi}^+$ such that $v\check{\alpha} \in \check{\Phi}^-$ and $\langle \lambda, v\check{\alpha} \rangle < 0$. But $v \in W_{0,L}$, so any $\check{\alpha} \in \check{\Phi}^+$ satisfying $v\check{\alpha} \in \check{\Phi}^-$ belongs to $\check{\Phi}_L^+$. Now applying 2B2 to each summand of the left hand side, this remark ensures that the equality holds. □

Proposition 4.7 says in particular that the $\mathbb{Z}[q]$-algebra

$$\mathcal{A}_{L^+} := \mathcal{A}_L \cap \mathcal{H}(L^+) \tag{4-13}$$

has $\mathbb{Z}[q]$-basis $E^{\otimes}(\lambda)$, where λ runs over the L-positive weights λ in $\Lambda^{(1)}$.

Proposition 4.8. *For any $h \in \mathcal{H}$, there is $r \in \mathbb{N}$ such that*

$$\tau_{a_L}^r h \in \sum_{d \in D_L} \mathcal{H}(L^+)\tau_d.$$

Proof. Let $w \in W^{(1)}$. Write $w = e^\lambda w_0 d$ with $w_0 \in W_{0,L}$, $d \in D_L$ and $\lambda \in \Lambda^{(1)}$ a weight that decomposes into $\lambda = \mu - \nu$ where μ and ν are antidominant. There is $r \in \mathbb{N}$ such that $a_L^r e^\lambda w_0$ is a L-positive element and $\ell(a_L^r w) = \ell(a_L^r e^\lambda w_0) + \ell(d)$ by Property (2-4). Note that $a_L^r e^\lambda = e^{r\lambda_L + \mu - \nu}$ and that $r\lambda_L + \mu$ is antidominant. The elements $E(w)$ and $E(a_L^r e^\lambda w_0)$ of the integral Bernstein basis of $\mathcal{H}$ can be written respectively

$$E(w) = q^{(\ell(w) - \ell(w_0) - \ell(d) + \ell(e^\nu) - \ell(e^\mu))/2} \tau_\mu \tau_\nu^{-1} \tau_{w_0} \tau_d$$

and

$$E(a_L^r e^\lambda w_0) = q^{(\ell(a_L^r e^\lambda w_0) - \ell(w_0) + \ell(e^\nu) - \ell(a_L^r) - \ell(e^\mu))/2} \tau_{a_L}^r \tau_\mu \tau_\nu^{-1} \tau_{w_0},$$

so the element

$$\tau_{a_L}^r E(w) = q^{(\ell(w) + \ell(a_L^r) - \ell(a_L^r w))/2} E(a_L^r e^\lambda w_0)\tau_d$$

belongs to $\mathcal{H}(L^+)\tau_d$. □

5. Inducing Hecke modules

5A. We consider the category $\mathcal{C}_L$ of the k-vector spaces $\mathfrak{M}$ endowed with a structure of right $\mathcal{H}(L)$-module such that the central invertible elements $\tau_{\mu_j}^{\otimes}$, $j \in \Delta - \Delta_L$ act by multiplication by nonzero scalars. This category is closed relative to subquotients.

Proposition 5.1. *Let $\mathfrak{M}$ be a k-vector space endowed with a right action of the positive algebra $\mathcal{H}(L^+)$. Suppose that the central invertible elements $\tau_{\mu_j}^{\otimes}$, $j \in \Delta - \Delta_L$ act by multiplication by nonzero scalars. Then there is a unique structure of right module over $\mathcal{H}(L)$ on $\mathfrak{M}$ extending the action of $\mathcal{H}(L^+)$.*

Proof. The element $\tau_{a_L}^{\otimes}$ defined by (4-2) is the product of the $\tau_{\mu_j}^{\otimes}$, $j \in \Delta - \Delta_L$. Denote by ζ the scalar action of $\tau_{a_L}^{\otimes}$ on $\mathfrak{M}$. The Hecke algebra $\mathcal{H}(L)$ is generated by $\mathcal{H}(L^+)$ and by the central elements $(\tau_{a_L}^{\otimes})^{\pm 1}$. So, if $\mathfrak{M}$ is endowed with an action of $\mathcal{H}(L)$, it is unique and the natural map $\mathfrak{M} \to \mathfrak{M} \otimes_{\mathcal{H}(L^+)} \mathcal{H}(L)$ is surjective. Define the map $\mathfrak{M} \otimes_{\mathcal{H}(L^+)} \mathcal{H}(L) \to \mathfrak{M}$, $v \otimes \tau_h^{\otimes} \mapsto \zeta^{-r}\, v\, \tau_{a_L^r h}^{\otimes}$, where $h \in W_L^{(1)}$ and $r \in \mathbb{N}$ is chosen so that $a_L^r h$ is L-positive. One checks that this map is well-defined and factors into an inverse for the previous one. □

Proposition 5.2. *Let $\mathfrak{M}$ in $\mathcal{C}_L$. As a vector space, $\mathfrak{M} \otimes_{\mathcal{H}(L^+)} \mathcal{H}$ decomposes into the direct sums*

$$\mathfrak{M} \otimes_{\mathcal{H}(L^+)} \mathcal{H} = \bigoplus_{d \in D_L} \mathfrak{M} \otimes \tau_d \tag{5-1}$$

and

$$\mathfrak{M} \otimes_{\mathcal{H}(L^+)} \mathcal{H} = \bigoplus_{d \in D_L} \mathfrak{M} \otimes \tau_{d^{-1}}^*. \tag{5-2}$$

Each subspace in these decompositions is isomorphic to $\mathfrak{M}$ via the natural maps $\mathfrak{M} \to \mathfrak{M} \otimes \tau_d$ and $\mathfrak{M} \to \mathfrak{M} \otimes \tau_{d^{-1}}^*$.

The decomposition (5-2) is a decomposition into eigenspaces for the action of τ_{a_L}: it acts by zero on each $\mathfrak{M} \otimes \tau_{d^{-1}}^*$ with $d \neq 1$ and by ζ on $\mathfrak{M} \otimes \tau_1$.

Corollary 5.3. *Let $\mathfrak{L}, \mathfrak{M}, \mathfrak{N}$ in $\mathcal{C}_L$ be such that there is an exact sequence of right $\mathcal{H}(L)$-modules $0 \to \mathfrak{L} \to \mathfrak{M} \to \mathfrak{N} \to 0$. Then one has an exact sequence of $\mathcal{H}$-modules*

$$0 \to \mathfrak{L} \otimes_{\mathcal{H}(L^+)} \mathcal{H} \to \mathfrak{M} \otimes_{\mathcal{H}(L^+)} \mathcal{H} \to \mathfrak{N} \otimes_{\mathcal{H}(L^+)} \mathcal{H} \to 0.$$

Corollary 5.4. *Suppose that $\mathfrak{N}$ and $\mathfrak{L}$ in $\mathcal{C}_L$ are finite-dimensional over k and that they have the same semisimplification as $\mathcal{H}(L)$-modules. Then any irreducible quotient of the $\mathcal{H}$-module $\mathfrak{N} \otimes_{\mathcal{H}(L^+)} \mathcal{H}$ is also an irreducible subquotient of $\mathfrak{L} \otimes_{\mathcal{H}(L^+)} \mathcal{H}$.*

Corollary 5.5. *Let $\mathfrak{M}$ in $\mathcal{C}_L$ be such that $\mathfrak{M} \otimes_{\mathcal{H}(L^+)} \mathcal{H}$ is an irreducible $\mathcal{H}$-module. Then $\mathfrak{M}$ is an irreducible $\mathcal{H}(L)$-module.*

Corollaries 5.3 and 5.5 easily follow from Proposition 5.2.

Proof of Corollary 5.4. Let N be an irreducible quotient of $\mathfrak{N} \otimes_{\mathcal{H}(L^+)} \mathcal{H}$. Let $\mathfrak{N}_0$ be a subquotient of the $\mathcal{H}(L)$-module $\mathfrak{N}$ with minimal dimension over k such that N is a quotient of $\mathfrak{N}_0 \otimes_{\mathcal{H}(L^+)} \mathcal{H}$. Using Corollary 5.3 and the irreducibility of N, one sees that $\mathfrak{N}_0$ is irreducible as an $\mathcal{H}(L)$-module. Hence $\mathfrak{N}_0$ is an irreducible subquotient of $\mathfrak{L}$, so that N appears in the semisimplification of $\mathfrak{L} \otimes_{\mathcal{H}(L^+)} \mathcal{H}$. □

Proof of Proposition 5.2.
(A) Proposition 4.8 ensures that

$$\mathfrak{M} \otimes_{\mathcal{H}(L^+)} \mathcal{H} = \sum_{d \in D_L} \mathfrak{M} \otimes \tau_d. \tag{5-3}$$

Since $\tau_{d^{-1}}^*$ decomposes with respect to the Iwahori–Matsumoto basis into the sum of τ_d and of other terms corresponding to elements with strictly smaller length [Vignéras 2005, Lemma 13], we also have

$$\mathfrak{M} \otimes_{\mathcal{H}(L^+)} \mathcal{H} = \sum_{d \in D_L} \mathfrak{M} \otimes \tau_{d^{-1}}^*. \tag{5-4}$$

(B) Let $\mu \in \Lambda^{(1)}$ be a minuscule weight and $m \in \mathfrak{M}$. If μ is not L-positive, then $E(\mu)$ acts by zero on $m \otimes 1$ (Because of relations (4-8), there is $j \in \Delta - \Delta_L$ such

that $E(\mu)E(\mu_j)=0$; since $E(\mu_j)=\tau_{e^{\mu_j}}$ acts by a nonzero scalar on $m\otimes 1$, the element $E(\mu)$ acts by zero.) We show by induction on $\ell(d)$ that, for any $d\in D_L$,

$$m\otimes\tau_{d^{-1}}^{*}E(\mu)=m\otimes E({}^{d}\mu)\tau_{d^{-1}}^{*}. \tag{5-5}$$

Let $d\in D_L$ and let $s\in S_0$ be such that $ds\in D_L$ and $\ell(ds)=\ell(d)+1$. These hypotheses imply $d\check{\alpha}_s\in\check{\Phi}^{+}-\check{\Phi}_L^{+}$. Suppose that (5-5) holds. We have to show that

$$m\otimes\tau_{d^{-1}}^{*}\tau_s^{*}E(\mu)=m\otimes E({}^{ds}\mu)\tau_{d^{-1}}^{*}\tau_s^{*}. \tag{5-6}$$

If $\langle\mu,\check{\alpha}_s\rangle=0$, then $\mu={}^{s}\mu$ and $E(\mu)$ and τ_s^{*} commute by Proposition 4.3(1), so we have the required equality.

If $\langle\mu,\check{\alpha}_s\rangle>0$, then

$$\begin{aligned}m\otimes\tau_{d^{-1}}^{*}\tau_s^{*}E(\mu)&=m\otimes\tau_{d^{-1}}^{*}(\tau_s+\nu_s)E(\mu)\\&=m\otimes\tau_{d^{-1}}^{*}E({}^{s}\mu)\tau_s^{*}+m\otimes\tau_{d^{-1}}^{*}E(\mu)\nu_s\quad\text{by the Bernstein relations}\\&=m\otimes E({}^{ds}\mu)\tau_{d^{-1}}^{*}\tau_s^{*}+m\otimes E({}^{d}\mu)\tau_{d^{-1}}^{*}\nu_s\quad\text{by induction.}\end{aligned}$$

The hypothesis on μ implies that $\langle{}^{d}\mu,d\check{\alpha}_s\rangle>0$, so ${}^{d}\mu$ is not L-positive. Hence the second part of the preceding sum is zero, which gives the required equality.

If $\langle\mu,\check{\alpha}_s\rangle<0$, then

$$\begin{aligned}m\otimes\tau_{d^{-1}}^{*}\tau_s^{*}E(\mu)&=m\otimes\tau_{d^{-1}}^{*}E({}^{s}\mu)\tau_s\quad\text{by the Bernstein relations}\\&=m\otimes E({}^{ds}\mu)\tau_{d^{-1}}^{*}\tau_s\quad\text{by induction.}\end{aligned}$$

But $\langle{}^{ds}\mu,d\check{\alpha}_s\rangle>0$, so ${}^{ds}\mu$ is not L-positive. Hence we have proved that both sides of (5-6) are zero.

By Proposition 2.2, we have proved (5-6) by induction.

(C) Result (B) shows that the right action of $E({}^{d^{-1}}\lambda_L)$ on $\mathfrak{M}\otimes\tau_{d'^{-1}}^{*}$ is zero for any $d'\in D_L$ $d'\neq d$ and that it is a multiplication by ζ on $\mathfrak{M}\otimes\tau_{d^{-1}}^{*}$. Hence, the decomposition (5-4) is a direct sum.

(D) Let us prove that

$$\mathfrak{M}\to\mathfrak{M}\otimes\tau_{d_0^{-1}}^{*}$$

is injective for any $d_0\in D_L$. Let $m\in\mathfrak{M}$ such that

$$m\otimes\tau_{d_0^{-1}}^{*}=0. \tag{5-7}$$

Let $(m_\nu)_{\nu\in\mathcal{N}}$ be a family of generators of the $\mathcal{H}(L^{+})$-module $\mathfrak{M}$ that contains m, say $m_{\nu_0}=m$. By [Bourbaki 1961, Chapitre 1, §2, n° 11], (5-7) implies that there is a finite family $(k_\iota)_{\iota\in\mathcal{I}}$ of elements in $\mathcal{H}$ and a finitely supported family $(b_{\iota,\nu})_{\iota\in\mathcal{I},\nu\in\mathcal{N}}$ of elements in $\mathcal{H}(L^{+})$ such that

- $\sum_{\nu\in\mathcal{N}}m_\nu b_{\iota,\nu}=0$ for any $\iota\in\mathcal{I}$,

- $\sum_{\imath\in\mathcal{I}} b_{\imath,\nu_0}k_\imath = \tau^*_{d_0^{-1}}$,
- $\sum_{\imath\in\mathcal{I}} b_{\imath,\nu}k_\imath = 0$ for any $\nu \neq \nu_0$.

By Proposition 4.8, there exists $r \in \mathbb{N}$ such that $\tau^r_{a_L}k_\imath = \sum_{d\in\mathcal{D}_L} c_{\imath,d}\tau_d$ with $c_{\imath,d} \in \mathcal{H}(L^+)$ for any $\imath \in \mathcal{I}$. The component of

$$\tau^r_{a_L}\tau^*_{d_0^{-1}} = \sum_{d\in D_L}\sum_{\imath} b_{\imath,\nu_0}c_{\imath,d}\tau_d$$

with support in $I(1)L^+d_0I(1)$ is equal to $\tau^r_{a_L}\tau_{d_0}$ on one hand, and to $\sum_\imath b_{\imath,\nu_0}c_{\imath,d_0}\tau_{d_0}$ on the other hand. So, by Lemma 2.4, we get $\tau^r_{a_L}\tau_{d_0} = \sum_\imath b_{\imath,\nu_0}c_{\imath,d_0}\tau_{d_0}$ and then $\tau^r_{a_L} = \sum_\imath b_{\imath,\nu_0}c_{\imath,d_0}$.

The same argument applied to $0 = \sum_{\imath\in\mathcal{I}} b_{\imath,\nu}k_\imath$ shows that $0 = \sum_\imath b_{\imath,\nu}c_{\imath,d_0}$ for $\nu \neq \nu_0$.

Multiplying $0 = \sum_{\nu\in\mathcal{N}} m_\nu b_{\imath,\nu}$ by $c_{\imath,d_0}$ for any $\imath \in \mathcal{I}$, and then summing over $\imath$, gives $0 = m_{\nu_0}\tau^r_{a_L}$, and hence $m = 0$.

This proves the remaining assertions of Proposition 5.2, also using again the argument of [Vignéras 2005, Lemma 13] to deduce the direct sum (5-1) from the direct sum (5-2). □

5B. ***Standard modules.*** The field k is naturally a $\mathbb{Z}[q]$-module via the specialization $q \mapsto 0$. A k-character of $\mathcal{A}$ is a morphism of unitary rings $\chi : \mathcal{A} \to k$ which is compatible with the structures of $\mathbb{Z}[q]$-modules. The set of k-characters of $\mathcal{A}$ inherits a natural action of W_0 given by $(w_0, \chi) \mapsto {}^{w_0}\chi$.

Because of (4-8), one has $E(\mu_J)E(\mu_K) = 0$ for any $J, K \subset \{1, \dots, n\}$, unless either $J \subset K$ or $K \subset J$. So, a k-character χ of $\mathcal{A}$ is completely determined by its values on $\{\tau_t,\ t \in \mathbb{T}\}$, the flag

$$J_0 = \varnothing \subsetneq J_1 \subsetneq \cdots \subsetneq J_r = \{1, \dots, n\}$$

of the subsets $J_i \subset \{1, \dots, n\}$ such that $\chi(E(\mu_{J_i}))$ is nonzero, and these nonzero values. The standard module induced by χ is the right $\mathcal{H}$-module

$$\chi \otimes_{\mathcal{A}} \mathcal{H}.$$

The set of minuscule weights $(\mu_{J_i})_{i\in\{1,\dots,r\}}$ we call the *support* of χ. We say that χ has dominant or antidominant support if every weight in the support is so.

Recall that any k-vector space which is a simple $\mathcal{H}$-module is a quotient of a standard module [Vignéras 2005, 1.4].

Definition 5.6. The character χ, the associated standard module, and any quotient of the latter are said to be *regular* if the flag is maximal, that is, $r = n$; *supersingular* if the flag is minimal, that is, $r = 1$; and *singular* otherwise.

If $n = 1$, we make the convention that any character of $\mathcal{A}$ is supersingular.

5C. ***Inducing standard modules.***

5C1. A k-character $\chi : \mathscr{A} \to k$ is called *adapted* to L (or L-*adapted*) if $\chi(E(\lambda_L))$ is nonzero, where λ_L is defined by (4-2). This implies that χ has L-positive support, that is, any weight in its support is L-positive. A k-character $\chi_L : \mathscr{A}_L \to k$ of the integral Bernstein subalgebra of $\mathscr{H}(L)$ is the tensor product of k-characters of the integral Bernstein algebras corresponding to the L_j's, $j \in \{1, \dots, m\}$. The value of χ_L on the invertible element $\tau_{a_L}^{\otimes}$ being nonzero, χ_L is completely determined by its restriction to $\mathscr{A}_{L^+}$ and we have an isomorphism of $\mathscr{H}(L)$-modules:

$$\chi_L \otimes_{\mathscr{A}_L} \mathscr{H}(L) \simeq \chi_L \otimes_{\mathscr{A}_{L^+}} \mathscr{H}(L).$$

There is a one-to-one correspondence between the k-characters χ_L of $\mathscr{A}_L$ and the k-characters of $\mathscr{A}$ adapted to L: it associates the character $\chi : \mathscr{A} \to k$ adapted to L with the character χ_L given on $\mathscr{A}_{L^+}$ by

$$\chi_L(E^{\otimes}(\lambda)) := \chi \circ \theta_L(E^{\otimes}(\lambda)) = \chi(E(\lambda))$$

for any L-positive weight $\lambda \in \Lambda^{(1)}$.

The algebra $\mathscr{A}_L$ is endowed not only with an action of the finite Weyl group $W_{0,L}$, but also of the normalizer of $W_{0,L}$ in W_0. Nevertheless, the previous correspondence is only compatible with the action of $W_{0,L}$ which preserves the set of L-positive weights in $\Lambda^{(1)}$.

5C2. With Proposition 5.1, the previous paragraph gives the following result.

Proposition 5.7. *Given $\chi_L : \mathscr{A}_L \to k$, let $\chi : \mathscr{A} \to k$ be the associated L-adapted character of $\mathscr{A}$. The standard module relative to χ is induced by the standard module relative to χ_L in the sense that the following isomorphisms of $\mathscr{H}$-modules hold*:

$$\chi \otimes_{\mathscr{A}} \mathscr{H} \simeq \chi_L \otimes_{\mathscr{A}_L} \mathscr{H}(L) \otimes_{\mathscr{H}(L^+)} \mathscr{H} \simeq \chi_L \otimes_{\mathscr{A}_{L^+}} \mathscr{H}.$$

5D. ***Intertwining operators between standard modules.*** Let $\chi : \mathscr{A} \to k$ be a character. We assume that L is a strict Levi subgroup of G and that χ is adapted to L. Then its support contains at least $\{\mu_j,\ j \in \Delta - \Delta_L\}$.

Let $d \in W_0$ and $s \in S_0$ be a simple reflection such that $d, ds \in D_L$ and $\ell(ds) = \ell(d) + 1$. Let ξ be the k-character $\xi = {}^{d^{-1}}\chi$. Denote respectively by φ and φ_s the canonical generators of the standard modules induced by ξ and ${}^{s}\xi$.

5D1. *Definition of the intertwiners.*

Remark 5.8. The fact that $\ell(ds) = \ell(d) + 1$ implies that $d\check{\alpha}_s$ is a positive root.

That both ds and d belong to D_L implies that $ds \notin W_{0,L}d$, so there exists j in $\Delta - \Delta_L$ such that ${}^{dsd^{-1}}\mu_j \neq \mu_j$: the weight ${}^{d^{-1}}\mu_j$ lies in the support of ξ and satisfies $\langle {}^{d^{-1}}\mu_j, \check{\alpha}_s \rangle = \langle \mu_j, d\check{\alpha}_s \rangle < 0$. Because of relations (4-8), any other minuscule weight μ in the support of ξ will then satisfy $\langle \mu, \check{\alpha}_s \rangle \leq 0$.

Lemma 5.9. *The vector $\varphi\tau_s^*$ is an eigenvector for the character ${}^s\xi$ of $\mathscr{A}$.*

Proof. It is easy to check that $\varphi\tau_s^*\tau_t = {}^s\xi(\tau_t)\varphi\tau_s^*$ for any $t \in \mathbb{T}$ (or see [Ollivier 2006a, 4.4.2]). We have yet to show that

$$\varphi\tau_s^* E(\mu_J) = {}^s\xi(E(\mu_J))\,\varphi\tau_s^* \tag{5-8}$$

for any minuscule weight μ_J associated to $J \subset \{1, \dots, n-1\}$.

If μ_J is fixed by s, the Bernstein relations ensure that τ_s^* and $E(\mu_J)$ commute and (5-8) holds.

If $\langle \mu_J, \check{\alpha}_s\rangle > 0$, the Bernstein relations give

$$\varphi\tau_s^* E(\mu_J) = \varphi\tau_s E(\mu_J) + \varphi E(\mu_J)\nu_s = \varphi E({}^s\mu_J)\tau_s^* = {}^s\xi(E(\mu_J))\,\varphi\tau_s^*,$$

because μ_J is not in the support of ξ by Remark 5.8 .

If $\langle \mu_J, \check{\alpha}_s\rangle < 0$, the Bernstein relations give $\varphi\tau_s^* E(\mu_J) = \varphi E({}^s\mu_J)\tau_s = 0$, because ${}^s\mu_J$ is not in the support of ξ, and (5-8) holds. □

We choose a weight ${}^{d^{-1}}\mu_j$ as in Remark 5.8. It is a minuscule weight in the support of ξ. It can be denoted by μ_K for some $K \subset \{1, \dots, n\}$. Recall that $\langle \mu_K, \check{\alpha}_s\rangle < 0$. Set

$$\beta := \xi(E(\mu_{K\cup {}^sK}))\xi(E(\mu_{K\cap {}^sK}))\xi(E(\mu_K))^{-1},$$

where ${}^sK$ denotes the image of K under the natural action of s.

Remark 5.10. Because of the relations (4-8), this scalar β is zero as soon as there exists a minuscule weight μ_J different from μ_K in the support of ξ such that $\langle \mu_J, \check{\alpha}_s\rangle < 0$.

Lemma 5.11. *The vector $\varphi_s(E(se^{\mu_K}) - \beta\nu_s)$ is an eigenvector for the character ξ of $\mathscr{A}$.*

Proof. Note that ν_s lies in $\mathscr{A}$ and commutes with τ_s. See [Ollivier 2006a, 4.4.2] to check that $\varphi_s(E(se^{\mu_K}) - \beta\nu_s)\tau_t = \xi(\tau_t)\varphi_s(E(se^{\mu_K}) - \beta\nu_s)$ for any $t \in \mathbb{T}$. We have yet to prove that

$$\varphi_s(E(se^{\mu_K}) - \beta\nu_s)E(\mu_J) = \xi(E(\mu_J))\,\varphi_s(E(se^{\mu_K}) - \beta\nu_s) \tag{5-9}$$

for any minuscule weight μ_J associated to $J \subset \{1, \dots, n-1\}$.

We use the fact that after extending the scalars to $\mathbb{Z}[q^{\pm 1/2}]$, we have

$$E(se^{\mu_K}) = q^{-1}E({}^s\mu_K)\tau_s = q^{-1}\tau_s^* E(\mu_K). \tag{5-10}$$

If μ_J is fixed by s, then (5-9) holds.

If $\langle \mu_J, \check{\alpha}_s\rangle < 0$, then μ_J is not in the support of ${}^s\xi$ by Remark 5.8, and the left side of (5-9) is $\varphi_s E(se^{\mu_K})E(\mu_J)$. The Bernstein relations and (5-10) give

$$E(se^{\mu_K})E(\mu_J) = E({}^s\mu_J)E(se^{\mu_K}) - \nu_s q^{(|K|-|{}^sK\cap J|)(|J|-|{}^sK\cap J|)-1} E(\mu_{{}^sK\cup J})E(\mu_{{}^sK\cap J}). \quad (5\text{-}11)$$

- If $J \neq K$, the power of q in the preceding equality is at least 1, so
$$\varphi_s E(se^{\mu_K})E(\mu_J) = \xi(E(\mu_J))\,\varphi_s E(se^{\mu_K}).$$
If J is in the support of ξ, then $\beta = 0$ by Remark 5.10, and (5-9) holds. If J is not in the support of ξ, we have proved that both sides of (5-9) are zero.
- If $J = K$, then (5-11) gives equality (5-9).

If $\langle \mu_J, \check{\alpha}_s\rangle > 0$, then μ_J is not in the support of ξ and the right side of (5-9) is zero. The Bernstein relations give

$$E(se^{\mu_K})E(\mu_J) = E({}^s\mu_J)E(se^{\mu_K}) + \nu_s q^{(|K|-|K\cap J|)(|J|-|K\cap J|)-1} E({}^s\mu_{K\cup J})E({}^s\mu_{K\cap J}), \quad (5\text{-}12)$$

so $\varphi_s E(se^{\mu_K})E(\mu_J) = \varphi_s\nu_s q^{(|K|-|K\cap J|)(|J|-|K\cap J|)-1} E({}^s\mu_{K\cup J})E({}^s\mu_{K\cap J})$.

- If $J \neq {}^sK$, the latter power of q is at least 1, so the only remaining term in the left side of (5-9) is equal to $-{}^s\xi(E(\mu_J))\beta\,\varphi_s\nu_s$: if μ_J is in the support of ${}^s\xi$, then $\beta = 0$ by Remark 5.10; if μ_J is not in the support of ${}^s\xi$, then ${}^s\xi(E(\mu_J)) = 0$.
- If $J = {}^sK$, then $\varphi_s E(se^{\mu_K})E(\mu_J) = \xi(E(\mu_K))\beta\varphi_s\nu_s$, so the left side of (5-9) is zero. □

The preceding lemmas allow us to define an $\mathcal{H}$-equivariant morphism Φ from the standard module induced by ξ into the one induced by ${}^s\xi$, and another, Ψ, going the other way around. They are fully determined by $\Phi(\varphi) = \varphi_s(E(se^{\mu_K}) - \beta\nu_s)$ and $\Psi(\varphi_s) = \varphi\tau_s^*$.

Lemma 5.12. *The composition of Φ and Ψ is the homothety with ratio*

$$\xi(E(\mu_K) - \beta\nu_s^2).$$

Proof. Any $d_0 \in W_0$ such that ${}^s\mu_K = {}^{d_0^{-1}}\mu_j$ satisfies $\langle \mu_j, d_0\check{\alpha}_s\rangle = -\langle \mu_K, \check{\alpha}_s\rangle > 0$, so $d_0\check{\alpha}_s \in \check{\Phi}^-$ and $\ell(d_0 s) = \ell(d_0) - 1$. Hence

$$\tau^*_{d_0^{-1}}\tau_s = 0$$

in $\mathcal{H}\otimes_{\mathbb{Z}[q]}k$, and Lemma 4.4 ensures that $E({}^s\mu_K)\tau_s = 0$ in $\mathcal{H}\otimes_{\mathbb{Z}[q]}k$. Thus $\varphi_s\tau_s = 0$ and $\varphi_s(E(se^{\mu_K}) - \beta\nu_s)\tau_s^* = \xi(E(\mu_K) - \beta\nu_s^2)\varphi_s$, and $\Phi\circ\Psi$ is a homothety with ratio $\xi(E(\mu_K) - \beta\nu_s^2)$. Using the equalities $E(\mu_K)E(se^{\mu_K}) = \tau_s E(\mu_{K\cap{}^sK})E(\mu_{K\cup{}^sK})$ and $\tau_s E(se^{\mu_K}) = E(\mu_K)$, one checks that $\Psi\circ\Phi$ is a homothety with the same ratio. □

5D2. *Conditions of isomorphism.*

5D2.1. Suppose that χ_L is a tensor product of supersingular characters. Then the support of χ is exactly $\{\mu_j,\ j \in \Delta - \Delta_L\}$.

Recall that the standard Levi subgroup L decomposes into $L = L_1 \times \cdots \times L_m$, where L_i is isomorphic to $\mathrm{GL}_{n_i}(F)$ for $i \in \{1, \dots, m\}$. There exists a simple reflection not belonging to $W_{0,L}$ but normalizing $W_{0,L}$ if and only if one can find two consecutive L_i and L_{i+1} with $i \in \{1, \dots, m-1\}$ such that $n_i = n_{i+1} = 1$.

We will say that χ_L *satisfies Hypothesis* $(\star)$ if for any simple reflection s_j not belonging to $W_{0,L}$ but normalizing $W_{0,L}$, the characters ${}^{s_j}\chi_L$ and χ_L differ.

Lemma 5.13. *Let $j \in \Delta$ and suppose that the simple reflection s_j does not belong to $W_{0,L}$ but normalizes $W_{0,L}$. The k-character χ_L and its conjugate by s_j coincide if and only if two conditions are satisfied*:

- $\chi(\nu_{s_j}^2) \neq 0$, *that is*, $\chi(\nu_{s_j}^2) = 1$,
- $\chi(E(\mu_j))^2 = \chi(E(\mu_{j-1}))\chi(E(\mu_{j+1}))$.

Proof. First note that $\nu_{s_j}^2 = \sum_{t \in \mathbb{T}_{s_j}} \tau_t$. One then easily checks that $\chi(\nu_{s_j}^2) = 1$ if the characters χ_L and its conjugate by s_j coincide on the space generated by $\{\tau_t^{\otimes}, t \in \mathbb{T}\}$, and that $\chi(\nu_{s_j}^2) = 0$ otherwise (see also [Ollivier 2006a, Remarque 7]).

Saying that s_j does not belong to $W_{0,L}$ means that $e^{\mu_{j-1}}$, e^{μ_j}, $e^{\mu_{j+1}}$ are central elements in L, so $\chi(E(\mu_{j-1}))$, $\chi(E(\mu_j))$, $\chi(E(\mu_{j+1}))$ are nonzero elements in k. The characters χ_L and its conjugate by s_j coincide if and only if they coincide on the space generated by $\{\tau_t^{\otimes}, t \in \mathbb{T}\}$, and

$$\frac{\chi(E(\mu_j))}{\chi(E(\mu_{j-1}))} = \frac{\chi(E(\mu_{j+1}))}{\chi(E(\mu_j))}. \qquad \square$$

By Lemma 5.12, it is clear that if $\beta = 0$, then $I(\xi)$ and $I({}^s\xi)$ are isomorphic. Saying that β is nonzero means that $\mu_{K \cup {}^sK}$ and $\mu_{K \cap {}^sK}$ both belong to the support of ξ. Because of the hypothesis on the support of χ, this implies that $\mu_{j+1} = {}^d\mu_{K \cup {}^sK}$, $\mu_j = {}^d\mu_K$, $\mu_{j-1} = {}^d\mu_{K \cap {}^sK}$ belong to the support of χ and that $dsd^{-1} = s_j$ is a simple reflection not belonging to $W_{0,L}$ and normalizing $W_{0,L}$. By Hypothesis $(\star)$, Lemma 5.13 then proves that $\xi(E(\mu_K) - \beta\nu_s^2)$ is nonzero, so $I(\xi)$ and $I({}^s\xi)$ are isomorphic.

By induction and using Proposition 2.2, we get the following result.

Proposition 5.14. *Let $\chi : \mathcal{A} \to k$ be an L-adapted character.*

Suppose that the associated $\chi_L : \mathcal{A}_L \to k$ is a tensor product of supersingular characters and that it satisfies Hypothesis $(\star)$. Then the standard module induced by χ is isomorphic to the standard module induced by any conjugate ${}^{d^{-1}}\chi$ of χ under the action of the inverse of an element $d \in \mathcal{D}_L$.

5D2.2. Let $\chi_0 : \mathscr{A} \to k$ be a character with antidominant support, and L be the maximal Levi subgroup such that the associated character $\chi_{0,L} : \mathscr{A} \to k$ is a tensor product of supersingular or regular characters. This Levi subgroup can be described in the following way: any $j \in \Delta$ lies in $\Delta - \Delta_L$ if and only if μ_j lies in the support of χ_0 and at least one of μ_{j+1} or μ_{j-1} does not lie in the support of χ_0.

We suppose now that $\chi = {}^{w_0}\chi_0$, where $w_0 \in W_{0,L}$. It is adapted to L and we can apply the results of Section 5D1.

Consider as before the weight μ_K in the support of ξ and the element $j \in \Delta - \Delta_L$ such that $\mu_K = {}^{d^{-1}}\mu_j$. Then $\mu_{K \cup {}^sK}$ and $\mu_{K \cap {}^sK}$ cannot be simultaneously in the support of ξ; otherwise $\mu_{j-1}, \mu_j, \mu_{j+1}$ would be in the support of χ_0, which contradicts the definition of $\Delta - \Delta_L$. Hence $\beta = 0$ and $I(\xi)$ and $I({}^s\xi)$ are isomorphic. By induction (using Proposition 2.2), the following proposition is proved.

Proposition 5.15. *Let $\chi_0 : \mathscr{A} \to k$ be a character with antidominant support and L the maximal Levi subgroup such that the associated character $\chi_{0,L} : \mathscr{A}_L \to k$ is a tensor product of supersingular or regular characters. Let $w_0 \in W_{0,L}$. The standard module induced by $\chi := {}^{w_0}\chi_0$ is isomorphic to the standard module induced by any conjugate ${}^{d^{-1}}\chi$ of χ under the action of the inverse of an element $d \in \mathscr{D}_L$.*

5E. *Nonsupersingular Hecke modules.*

5E1. *Regular standard modules.*

Proposition 5.16. *The standard module induced by a character $\chi : \mathscr{A} \to k$ with regular support is a k-vector space with dimension $n!$.*

5E1.1. Our proof relies on further ingredients relative to root data and Coxeter systems. Let $R \subset \check{\Pi}$ be a set of simple roots and denote by $\langle R \rangle$ the subset of Φ^+ generated by R. Define $W_0(R)$ to be the subset of W_0 whose elements w satisfy $w(R) \subset \check{\Phi}^-$ and $w(\check{\Pi} - R) \subset \check{\Phi}^+$.

Lemma 5.17. *In $W_0(R)$ there is a unique element w_R with minimal length. It is an involution and its length is equal to the cardinality of $\langle R \rangle$.*

Proof. The length of an element w in W_0 being the number of positive roots $\alpha \in \Phi^+$ such that $w\alpha \in \Phi^-$ (Section 2A), any element in $W_0(R)$ has length larger than the cardinality of $\langle R \rangle$. The subgroup of W_0 generated by the simple reflections corresponding to the simple roots in R has a unique maximal length element w_R, with length the cardinality of $\langle R \rangle$. It is an involution satisfying $w_R(R) = -R$ and $w_R(\Phi^+ - \langle R \rangle) \subset \Phi^+$ [Bourbaki 1968, Chapitre VI, §1, n° 1.6, corollaire 3]. This element belongs to $W_0(R)$.

Let $w \in W_0(R)$. Suppose that $\ell(w) = \ell(w_R)$. Then the roots in $\langle R \rangle$ are the only positive ones made negative by w. Applying the length property 2B2 and the definition of $W_0(R)$, we then see that $\ell(w) = \ell(ww_R) + \ell(w_R)$, so $ww_R = 1$ and $w = w_R$. □

Lemma 5.18. *Let $w \in W_0(R)$. Suppose that $w \neq w_R$ and consider $s_j \in S_0$ such that $\ell(s_j w w_R) = \ell(w w_R) - 1$. Then $\ell(s_j w) = \ell(w) - 1$, the element $s_j w$ lies in $W_0(R)$ and the positive root $-w^{-1}\check{\alpha}_j$ is not a simple root.*

Proof. The hypothesis on the length ensures that $w_R w^{-1}\check{\alpha}_j \in \check{\Phi}^-$. Because of the properties of w and w_R, it implies $w^{-1}\check{\alpha}_j \in \check{\Phi}^-$ and $\ell(s_j w) = \ell(w) - 1$. More precisely, one checks that the only possibility is $-w^{-1}\check{\alpha}_j \in \check{\Phi}^+ - \langle R \rangle$. So, if $-w^{-1}\check{\alpha}_j$ were a simple root, it would be an element in $\check{\Pi} - R$, which would contradict $w(\check{\Pi} - R) \subset \check{\Phi}^+$. It remains to check that $s_j w$ lies in $W_0(R)$. Let $\check{\alpha} \in R$. Since $w\check{\alpha} \in \check{\Phi}^- - \{-\check{\alpha}_j\}$, we have $s_j w \check{\alpha} \in \check{\Phi}^-$. Let $\check{\alpha} \in \check{\Pi} - R$. Since $w\check{\alpha} \in \check{\Phi}^+ - \{\check{\alpha}_j\}$, we have $s_j w \check{\alpha} \in \check{\Phi}^+$. □

Lemma 5.19. *Denote by $\sigma \in W_0$ the cycle $(n, n-1, \dots, 1)$. Let $\check{\alpha} \in \check{\Pi} - R$. There exists $j \in \{1, \dots, n-1\}$ such that $\sigma^j w_R \in W(R \cup \{\check{\alpha}\})$.*

Proof. We first make some remarks.

(1) Let $\check{\beta} \in \check{\Pi} - R$ be a simple root. Then $w_R \check{\beta}$ is a positive root. Also, s_β appears in any reduced decomposition of the transposition $w_R s_\beta w_R$ according to the set S_0. From this, one easily deduces that $w_R \check{\beta} \geq \check{\beta}$, where $\geq$ denotes the partial order on $\check{\Phi}$ described in 2A. Conversely, let $\check{\alpha} \in \check{\Pi} - R$. If $w_R \check{\beta} \geq \check{\alpha}$, this means that s_α appears in any reduced decomposition of $w_R s_\beta w_R$, so $\check{\beta} = \check{\alpha}$.

(2) Let $j \in \{1, \dots, n-1\}$ and $\check{\beta} \in \check{\Phi}^+$. Then $\sigma^j \check{\beta} \in \check{\Phi}^-$ if and only if $\beta \geq \check{\alpha}_j$.

Let $\check{\alpha} \in \check{\Pi} - R$ as in the lemma and $j \in \{1, \dots, n-1\}$ such that $\check{\alpha} = \check{\alpha}_j$. We check that $\sigma^j w_R \in W_0(R \cup \{\check{\alpha}_j\})$. Any $\check{\beta} \in R$ is sent by w_R to an element in $-R$, which in turn is sent by σ^j to an element in $\check{\Phi}^-$ by (2). Let $\check{\beta} \in \check{\Pi} - R$. Then $w_R \check{\beta} \in \Phi^+$ and using (2), $\sigma^j w_R \check{\beta} \in \check{\Phi}^-$ if and only if $w_R \check{\beta} \geq \check{\alpha}_j$, which by (1) is equivalent to $\check{\beta} = \check{\alpha}_j$. □

Proof of Proposition 5.16. Let $\chi : \mathscr{A} \to k$ be a character with regular antidominant support.

(A) Let $R \subset \check{\Pi}$ be as in 5E1.1. We prove by induction on the length of $w \in W_0(R)$ that the standard modules induced by ${}^{w}\chi$ and ${}^{w_R}\chi$ are isomorphic as $\mathscr{H}$-modules.

Let $w \in W_0(R)$. Suppose $w \neq w_R$; then there is $s_j \in S_0$ such that $\ell(s_j w w_R) = \ell(w w_R) - 1$. By Lemma 5.18, this implies $\ell(s_j w) = \ell(w) - 1$ and the element $s_j w$ also lies in $W_0(R)$. Set $\xi = {}^{s_j w}\chi$. We prove that ξ and ${}^{s_j}\xi$ induce isomorphic standard modules. We are in the situation of Section 5D; the Levi subgroup here is simply the diagonal torus. So we have two well-defined intertwining operators between the standard modules in question. By Remark 5.10, there is an easy sufficient condition for these operators to be isomorphisms: it suffices to check that there is more than one minuscule weight μ in the support of ξ satisfying $\langle \mu, \check{\alpha}_j \rangle < 0$; that is, that there is more than one antidominant minuscule weight λ

such that $\langle \lambda, w^{-1}\check{\alpha}_j \rangle > 0$. This is true, because $w^{-1}\check{\alpha}_j \in \Phi^-$ and $-w^{-1}\check{\alpha}_j$ is not a simple root, by Lemma 5.18.

(B) For $w \in W_0$, the standard modules induced by ${}^{w}\chi$ and ${}^{\sigma w}\chi$ have the same dimension, as proved in [Ollivier 2006a, Proposition 2].

(C) Let $R \subset \check{\Pi}$ be a set of simple roots. We prove by induction on the cardinality of R that the standard module induced by ${}^{w}\chi$ is $n!$-dimensional for any $w \in W_0(R)$. If $R = \varnothing$, then $W_0(R) = \{1\}$, and the result is given by Propositions 5.2 and 5.7. Suppose that the property holds for some set of simple roots $R \subsetneq \check{\Pi}$. Let $\check{\alpha} \in \check{\Pi} - R$ and $w \in W_0(R \cup \{\check{\alpha}\})$. By Lemma 5.19, there is a power σ^j of the cycle σ such that $\sigma^j w_R \in W_0(R \cup \{\check{\alpha}\})$. We conclude using (A) and (B). □

5E1.2. The motivation for Proposition 5.16 is this:

Conjecture 5.20. *Let $\chi : \mathscr{A} \to k$ be a character with regular support and $w_0 \in W_0$. The standard modules induced by ${}^{w_0}\chi$ and χ have the same semisimplification as modules over $\mathscr{H}$.*

We can prove the conjecture if we consider characters of $\mathscr{A}$ which are totally degenerate on the finite torus, that is, for $t \in \mathbb{T}$, the value $\chi(\tau_t)$ only depends on the orbit of t under the action of W_0. By twisting, we can consider that χ is trivial on the finite torus. Then the standard module induced by χ can be seen as a module over the Iwahori–Hecke algebra (see for example Section 8). One can then apply the arguments listed in [Ollivier 2006b, 2.4] (for the case of GL_3) to show that χ and its conjugates induce standard modules which have the same semisimplification. The first argument comes from [Vignéras 2006, théorème 6]: the character χ can be lifted to a character χ_0 with values in $\overline{\mathbb{Z}}_p$, and we see the latter as a character with values in $\overline{\mathbb{Q}}_p$. Since the standard module induced by χ is $n!$-dimensional over $\overline{\mathbb{F}}_p$, [Vignéras 2006, théorème 5] says that it is isomorphic to the reduction of the canonical integral structure of the $\mathscr{H} \otimes_{\mathbb{Z}[q]} \overline{\mathbb{Q}}_p$ standard module induced by χ_0. To conclude, we recall Proposition 2.3 of [Rogawski 1985]: two standard modules for the Iwahori–Hecke algebra in characteristic zero have the same semisimplification if they are induced by conjugate characters. The proof is based on the description of an explicit basis for the standard modules owing to the Kazhdan–Lusztig polynomials for the Iwahori–Hecke algebra.

Proposition 5.21. *Conjecture 5.20 is true for the standard modules over the Iwahori–Hecke algebra, that is, for characters χ that are trivial on the finite torus.*

5E2. *Nonsupersingular simple modules and induction.* Recall that a nonsupersingular character $\chi : \mathscr{A} \to k$ with antidominant support is adapted to some strict Levi subgroup L of G. So the associated standard module is *induced* from a $\mathscr{H}(L)$-module by Proposition 5.7. In the light of this, the following proposition bolsters the definition of a *nonsupersingular* module.

Proposition 5.22. *Assume that Conjecture 5.20 is true. Any simple nonsupersingular $\mathcal{H}$-module appears in the semisimplification of a standard module for $\mathcal{H}$ relative to a nonsupersingular character with antidominant support.*

Proof. Let M be a simple nonsupersingular module: it is a quotient of a standard module induced by some nonsupersingular character $\xi : \mathcal{A} \to k$. Let $w \in W_0$ with minimal length such that $\chi := {}^{w}\xi$ has antidominant support. We want to prove that M appears in the semisimplification of the standard module induced by χ. Let L be the standard Levi subgroup associated to χ as in Proposition 5.15. Let $(w_0, d) \in W_{0,L} \times D_L$ be such that $w = w_0 d$. Recall that $\ell(w) = \ell(w_0) + \ell(d)$. By Proposition 5.15, the standard modules induced by ${}^{w_0^{-1}}\chi$ and ξ are isomorphic. So M is an irreducible quotient of the standard module induced by ${}^{w_0^{-1}}\chi$. We have yet to check that it is a subquotient of the standard module induced by χ.

- If $L = G$, then χ is a regular character and the claim comes from Conjecture 5.20.
- Suppose $L \neq G$. Decompose $L \simeq L_1 \times \cdots \times L_m$ and $w_0^{-1} = (w_1, \ldots, w_m) \in L_1 \times \cdots \times L_m$. Both χ and ${}^{w_0^{-1}}\chi$ are L-adapted: denote by $\chi_L = \chi_{L_1} \otimes \cdots \otimes \chi_{L_m}$ the character of $\mathcal{A}_L$ corresponding to χ. Then ${}^{w_0^{-1}}\chi_L = {}^{w_1}\chi_{L_1} \otimes \cdots \otimes {}^{w_m}\chi_{L_m}$ corresponds to ${}^{w_0^{-1}}\chi$. If χ_{L_i} is a supersingular character for an $i \in \{1, \ldots, m\}$, then ${}^{w_i}\chi$ and χ have the same support, so by minimality of the length of w, we must have $w_i = 1$. In other words, if $w_i \neq 1$, then χ_{L_i} is a regular character of $\mathcal{A}_{L_i}$. So Conjecture 5.20 says that the standard modules for $\mathcal{H}(L)$ induced by χ_L and ${}^{w_0^{-1}}\chi_L$ have the same semisimplification. Then applying Proposition 5.7 and Corollary 5.4, one gets that M is an irreducible subquotient of the standard module induced by χ. □

Proposition 5.23. *The statement of Proposition 5.22 holds without further hypothesis for modules over the Iwahori–Hecke algebra.*

5F. *Irreducible induced modules.* Let $\mathfrak{M}$ be a k-vector space endowed with a structure of right $\mathcal{H}(L)$-module. Let $\mathfrak{M}$ be irreducible as an $\mathcal{H}(L)$-module. Then it is finite-dimensional and has a central character [Vignéras 2007, 5.3], so $\mathfrak{M}$ is a quotient of some standard module for $\mathcal{H}(L)$ induced by a character $\chi_L : \mathcal{A}_L \to k$. In particular, $\mathfrak{M}$ belongs to the category $\mathcal{C}_L$ defined in 5A. Suppose that χ_L is the tensor product of supersingular characters and consider as before its associated L-adapted character $\chi : \mathcal{A} \to k$.

Proposition 5.24. *Let χ' be a k-character for $\mathcal{A}$ contained in $\mathfrak{M} \otimes_{\mathcal{H}(L^+)} \mathcal{H}$. There is $d \in D_L$ such that ${}^{d}\chi'$ is the L-adapted character associated to some $W_{0,L}$-conjugate of χ_L.*

Proof. First note, using the braid relations in $\mathcal{H}(L)$ and the fact that χ_L is a product of supersingular characters, that any k-character for $\mathcal{A}_L$ contained in $\mathfrak{M}$ is a $W_{0,L}$-conjugate of χ_L. Then, using Proposition 5.7, note that $\mathfrak{M} \otimes_{\mathcal{H}(L^+)} \mathcal{H}$ is a quotient of the standard module for $\mathcal{H}$ induced by χ. So it has a central character given by the restriction of χ to the center of $\mathcal{H}$. Any k-character χ' for $\mathcal{A}$ contained in $\mathfrak{M} \otimes_{\mathcal{H}(L^+)} \mathcal{H}$ has the same restriction to the center, which ensures that the supports of χ' and χ are conjugate, and more precisely, that there is an element $d \in D_L$ such that χ and ${}^d\chi'$ coincide on $(E(\lambda))_{\lambda \in \Lambda}$. In particular, $\chi'(E({}^{d^{-1}}\lambda_L)) \neq 0$, so the character χ' is supported by an element in $\mathfrak{M} \otimes \tau^*_{d^{-1}}$ by Proposition 5.2 and its proof. With the braid relations in $\mathcal{H}$, our first remark then shows that ${}^d\chi'$ is the L-adapted character associated to some $W_{0,L}$-conjugate of χ_L. □

Corollary 5.25. *Suppose that χ_L satisfies Hypothesis ($\star$). Then $\mathfrak{M} \otimes_{\mathcal{H}(L^+)} \mathcal{H}$ is an irreducible $\mathcal{H}$-module.*

Proof. A nontrivial irreducible submodule of $\mathfrak{M} \otimes_{\mathcal{H}(L^+)} \mathcal{H}$ is a quotient of a standard module for $\mathcal{H}$. By Proposition 5.24, the latter is induced by a k-character χ' such that ${}^d\chi'$ is the L-adapted character associated to ${}^{w_0}\chi_L$ for some $d \in D_L$ and $w_0 \in W_{0,L}$. It is clear that ${}^{w_0}\chi_L$ satisfies Hypothesis ($\star$) since χ_L does, so Proposition 5.14 ensures that the standard module induced by χ' is isomorphic to the one induced by ${}^d\chi'$. In particular, any nonzero submodule of $\mathfrak{M} \otimes_{\mathcal{H}(L^+)} \mathcal{H}$ contains an L-adapted character, and hence a nonzero eigenvector for τ_{a_L} and the value ζ. By Proposition 5.2 and by the irreducibility of $\mathfrak{M}$, any nonzero submodule contains $\mathfrak{M} \otimes \tau_1$, and hence it is the whole $\mathfrak{M} \otimes_{\mathcal{H}(L^+)} \mathcal{H}$. □

6. Parabolic induction and compact induction

Recall that the universal module $\mathcal{U}$ is the compact induction to G of the trivial character of $I(1)$ with values in k. We will denote by $\mathcal{U}_L$ the compact induction to L of the trivial character of $I_L(1)$ with values in k. These representations of G and L are respectively generated by the characteristic functions of the pro-p-Iwahori subgroups $I(1)$ and $I_L(1)$. We will denote both of these by $\mathbf{1}$ when there is no possible ambiguity.

We consider a module $\mathfrak{M}$ in the category $\mathcal{C}_L$ defined in 5A. Let $(\pi(\mathfrak{M}), V)$ be the representation of G on $\mathfrak{M} \otimes_{\mathcal{H}(L^+)} \mathcal{U}$ and $(\pi_L(\mathfrak{M}), V_L)$ the representation of L on $\mathfrak{M} \otimes_{\mathcal{H}(L)} \mathcal{U}_L$.

6A. The parabolic induction $\mathrm{Ind}_P^G\, \pi_L(\mathfrak{M})$ is the smooth part of the space of functions $f : G \to V_L$ satisfying $f(lng) = l.f(g)$ for $g \in G$, $(l, n) \in L \times N$, endowed with the action of G by right translation.

6A1. The set D_L is a system of representatives of the double cosets $\mathbb{P}\backslash\mathbb{G}/\mathbb{U}$ in the Chevalley group. For $d \in D_L$, set $\mathbb{U}_d = \mathbb{U} \cap d^{-1}\mathbb{U}d$ and $\overline{\mathbb{U}}_d = \mathbb{U} \cap d^{-1}\overline{\mathbb{U}}d$. Any

element in $\mathbb{U}$ can be written as a product of an element of $\mathbb{U}_d$ and of an element of $\overline{\mathbb{U}}_d$ and this decomposition is unique. From this, one deduces that any element in $\mathbb{P}d\mathbb{U}$ decomposes uniquely in $\mathbb{P}d\overline{\mathbb{U}}_d$ [Carter 1985, 2.5.12].

The set D_L is also a system of representatives of the double cosets $P\backslash G/I(1)$. For any $d \in D_L$, one has

$$PdI(1) = \coprod_y PI(1)dy, \tag{6-1}$$

where dy runs over a system of representatives of $I(1)\backslash I(1)dI(1)$.

For any $d \in D_L$ and any $I_L(1)$-invariant element v in $\pi_L(\mathfrak{M})$, the $I(1)$-invariant function $f_{PdI(1),v}$ with support $PdI(1)$ and value v at d is a well-defined element of $\mathrm{Ind}_P^G\, \pi_L(\mathfrak{M})$. Any $I(1)$-invariant function in the latter representation is a linear combination of such functions.

6A2. The right action of τ_d maps $f_{PI(1),v}$ to an $I(1)$-invariant element with support $PdI(1)$, which is completely determined by its value at d. Using (6-1), one easily checks that this value is v, so

$$(f_{PI(1),v})\tau_d = f_{PdI(1),v} \tag{6-2}$$

6A3. Let $w \in W_L^{(1)}$. Suppose it is a L-positive element.

According to [Vignéras 1998, II.4], there is a system of representatives of the right cosets $I(1)\backslash I(1)wI(1)$ respecting the decomposition of $I_L(1)wI_L(1)$ into right cosets mod $I_L(1)$. Explicitly, from the decomposition

$$I_L(1)wI_L(1) = \coprod_x I_L(1)wx$$

one gets

$$I(1)wI(1) = \coprod_x I(1)wxI(1)^-$$

and a decomposition $I(1)wI(1) = \coprod_{x,u_x} I(1)wxu_x$, where u_x belong to $I(1)^-$.

From arguments analogous to [Schneider and Stuhler 1991, Proposition 7], one shows that $PI(1)wx \cap PI(1)wxu_x \neq \varnothing$ implies $I(1)wxu_x = I(1)wx$: the hypothesis can be written $Pw^{-1}I(1)^-wx \cap Pw^{-1}I(1)^-wxu_x \neq \varnothing$, and we recall that $I(1)^-$ is normalized by $x \in I_L(1)$. So there exists an element $\kappa_1 x u_x x^{-1}\kappa_2$ in P with $\kappa_1, \kappa_2 \in w^{-1}I(1)^-w \subset I(1)^-$. Since $P \cap I(1)^- = \{1\}$, one deduces that $xu_xx^{-1} \in w^{-1}I(1)w$ and $I(1)wxu_x = I(1)wx$.

The right action of $\tau_w \in \mathscr{H}$ on $f_{PI(1),v}$ gives the $I(1)$-invariant function with support $PI(1)$ and value at 1_G given by $\sum_{x,u_x} f_{PI(1),v}((wxu_x)^{-1})$. But $(wxu_x)^{-1} \in PI(1)$ implies $1 \in PI(1)wx \cap PI(1)wxu_x$; therefore this value is $\sum_x (wx)^{-1}v = v\tau_w^{\otimes}$, and

$$(f_{PI(1),v})\tau_w = f_{PI(1),v\tau_w^{\otimes}}. \tag{6-3}$$

6B. For any $m \in \mathfrak{M}$, there is a well-defined G-equivariant map

$$\mathcal{F}_m : \mathcal{U} \to \mathrm{Ind}_P^G \ \pi_L(\mathfrak{M})$$

sending the characteristic function of $I(1)$ on $f_{PI(1),m\otimes\mathbf{1}}$. The computation of 6A3 shows that we then have a G-equivariant morphism

$$\mathcal{F} : \pi(\mathfrak{M}) \to \mathrm{Ind}_P^G \, \pi_L(\mathfrak{M}), \quad m \otimes u \mapsto \mathcal{F}_m(u). \tag{6-4}$$

Remark 6.1. In the case where L is the diagonal torus T and $\mathfrak{M}$ is a character of $\mathcal{A}_T$, the map $\mathcal{F}$ is an isomorphism [Schneider and Stuhler 1991; Vignéras 2004].

6C. In the tensor product $\mathfrak{M} \otimes_{\mathcal{H}(L)} \mathcal{U}_L$, the group L only acts on $\mathcal{U}_L$, so there is a natural morphism of $\mathcal{H}(L)$-modules

$$\mathfrak{M} \to (\mathfrak{M} \otimes_{\mathcal{H}(L)} \mathcal{U}_L)^{I_L(1)}, \tag{6-5}$$

and a natural morphism of $\mathcal{H}$-modules

$$\mathfrak{M} \otimes_{\mathcal{H}(L^+)} \mathcal{H} \to (\mathfrak{M} \otimes_{\mathcal{H}(L^+)} \mathcal{U})^{I(1)}, \tag{6-6}$$

which composes with $\mathcal{F}$ to give the morphism of $\mathcal{H}$-modules

$$\mathfrak{M} \otimes_{\mathcal{H}(L^+)} \mathcal{H} \to (\mathrm{Ind}_P^G \pi_L(\mathfrak{M}))^{I(1)}. \tag{6-7}$$

6C1. If (6-5) is not trivial, then (6-7) is not trivial and neither is (6-6). By adjunction, if there exists a representation (π_L, V_L) of L and a nonzero $\mathcal{H}(L)$-equivariant map $\mathfrak{M} \to V_L^{I_L(1)}$, then (6-5) is not trivial.

6C2. Suppose (6-5) is surjective. Then (6-7) is surjective.

6C3. Using Proposition 5.2, one sees that (6-7) is injective if (6-5) is injective. In this case, (6-6) is also injective.

In 5F, we gave sufficient conditions for certain irreducible $\mathcal{H}(L)$-modules $\mathfrak{M}$ to induce irreducible $\mathcal{H}$-modules. Under these conditions, and if (6-5) is nonzero, then (6-7) allows us to describe an irreducible subspace $\mathfrak{M} \otimes_{\mathcal{H}(L)} \mathcal{H}$ of the pro-p-invariants of $\mathrm{Ind}_P^G \pi_L(\mathfrak{M})$.

If $\mathcal{H}(L)$ is a direct factor of $\mathcal{U}_L$ as a left $\mathcal{H}(L)$-module, then (6-5) is injective for any $\mathfrak{M}$ in $\mathcal{C}_L$. This is the case if F has residue field $\mathbb{F}_p$ and L is isomorphic to a product of $\mathrm{GL}_1(F)$'s and $\mathrm{GL}_2(F)$'s [Ollivier 2007, 2.1.3].

6D. ***Examples.***

6D1. If L is the diagonal torus T, then $\mathfrak{M}$ identifies with a character $\chi_T : \mathcal{A}_T \to k$. By Remark 6.1 and previous results, the representation $\mathfrak{M} \otimes_{\mathcal{H}(L^+)} \mathcal{U}$ is isomorphic to the principal series induced by the character $T \to k^*$, $t \mapsto \chi_T(t^{-1})$. The semisimplification of this representation and of its space of pro-p-invariants is well-understood [Grosse-Klönne 2009; Ollivier 2006a; Ollivier 2006c; Vignéras 2008].

6D2. We consider the case where $F = \mathbb{Q}_p$. Suppose that L is isomorphic to a product of $\mathrm{GL}_1(\mathbb{Q}_p)$'s and $\mathrm{GL}_2(\mathbb{Q}_p)$'s. There is an equivalence of categories between the right $\mathcal{H}(L)$-modules (with scalar action of the uniformizers) and the representations of L generated by their $I_L(1)$-invariants (with scalar action of the uniformizers). In particular, (6-5) is an isomorphism for any $\mathfrak{M}$. If L is the diagonal torus, it is clear. Otherwise, the result is given by [Ollivier 2009]. So, for any $\mathfrak{M}$ in $\mathscr{C}_L$, the map (6-7) is an isomorphism.

6D2.1. Suppose that $G = \mathrm{GL}_3(\mathbb{Q}_p)$ and L is isomorphic to $\mathrm{GL}_2(\mathbb{Q}_p) \times \mathrm{GL}_1(\mathbb{Q}_p)$.

Let $\chi_L : \mathscr{A}_L \to k$ be the tensor product of two supersingular characters. It satisfies Hypothesis $(\star)$. Denote by $\mathfrak{M}$ the standard module for $\mathcal{H}(L)$ induced by χ_L. It is irreducible and 2-dimensional. Because of the above-mentioned equivalence of categories, the representation $\mathfrak{M} \otimes_{\mathcal{H}(L^+)} \mathcal{U}_L$ is the tensor product of a supersingular representation of $\mathrm{GL}_2(\mathbb{Q}_p)$ by a character of $\mathrm{GL}_1(\mathbb{Q}_p)$.

By Corollary 5.25, the $\mathcal{H}$-module $\mathfrak{M} \otimes_{\mathcal{H}(L^+)} \mathcal{H}$ is irreducible. By the remarks of 6C, it is isomorphic to the subspace of $I(1)$-invariants of the representation which is parabolically induced from $\mathfrak{M} \otimes_{\mathcal{H}(L^+)} \mathcal{U}_L$. Hence, this subspace generates an irreducible subrepresentation for $\mathrm{GL}_3(\mathbb{Q}_p)$. By the results of Herzig, this subrepresentation is actually the whole $\mathrm{Ind}_P^G \pi_L(\mathfrak{M})$.

6D2.2. Suppose that G is $\mathrm{GL}_4(\mathbb{Q}_p)$ and L is isomorphic to $\mathrm{GL}_2(\mathbb{Q}_p) \times \mathrm{GL}_2(\mathbb{Q}_p)$. Let $\chi_L : \mathscr{A}_L \to k$ be the tensor product of two supersingular characters. It satisfies Hypothesis $(\star)$. Denote by $\mathfrak{M}$ the standard module for $\mathcal{H}(L)$ induced by χ_L. It is irreducible and 4-dimensional. The same arguments as before ensure that $\mathfrak{M} \otimes_{\mathcal{H}(L^+)} \mathcal{U}_L$ is the tensor product of two supersingular representations of $\mathrm{GL}_2(\mathbb{Q}_p)$, and that the $\mathcal{H}$-module $\mathfrak{M} \otimes_{\mathcal{H}(L^+)} \mathcal{H}$ is irreducible and isomorphic to the space of $I(1)$-invariants of the representation which is parabolically induced from $\mathfrak{M} \otimes_{\mathcal{H}(L^+)} \mathcal{U}_L$. The latter is an irreducible representation by the results of Herzig.

7. Supersingular modules

Fix a supersingular character $\chi : \mathscr{A} \to k$. It is defined by its restriction to $\{\tau_t,\ t \in \mathbb{T}\}$, its value $\zeta \in k^*$ on $E(\mu_{\{1,\dots,n\}})$ and by the fact that for any $\lambda \in \Lambda^{(1)}$ such that $\ell(e^\lambda) > 0$, the scalar $\chi(E(\lambda))$ is zero.

Let M be a nonzero quotient of the standard module for $\mathcal{H}$ induced by χ. Denote by M_χ the sum of the equivariant subspaces in M for $\mathscr{A}$ and the W_0-conjugates of χ (it is nonzero).

Proposition 7.1. *M_χ is stable under the action of the finite Hecke algebra.*

Proof. This is a direct consequence of the integral Bernstein relations. □

Recall that the simple modules for the finite Hecke algebra are the characters [Cabanes and Enguehard 2004, Theorem 6.12]. So the proposition says in particular that M_χ contains a character for the finite Hecke algebra. Denote by $m \in M_\chi$ its support. The set $\mathfrak{D}$ was introduced by Proposition 2.5 and one has the following result.

Proposition 7.2. *The set of the lengths $\ell(d)$, where d runs over the elements of $\mathfrak{D}$ such that $mE(d) \neq 0$, is bounded.*

Proof. Let $d \in \mathfrak{D}$. Write $d = e^\lambda w_0 \in W$. According to Lemma 2.6, the weight λ is dominant, so (after a suitable twist of d by a power of the central element ϖ^n) it decomposes into a linear combination

$$\lambda = \sum_{1 \le i \le n-1} -n_i \mu_i$$

with nonnegative integral coefficients. Suppose that one of the coefficients, say n_j, is at least 2. Then $\lambda + \mu_j$ is still dominant and we show that

(1) $d' := e^{\lambda+\mu_j} w_0 \in \mathfrak{D}$,

(2) $\ell(d) = \ell(e^{-\mu_j}) + \ell(d')$, which easily implies that $E(d) = E(-\mu_j)E(d')$ and $mE(d) = 0$.

Since $\lambda+\mu_j$ is dominant, the only thing one has to check to make sure that $d' \in \mathfrak{D}$ is the following: for any $\check{\alpha} \in \check{\Phi}^+$, if $\langle \lambda+\mu_j, \check{\alpha}\rangle = 0$ then $w_0^{-1}\check{\alpha} \in \check{\Phi}^+$. Since $d = e^\lambda w_0$ is already in $\mathfrak{D}$, the only tricky case is $\langle \lambda, \check{\alpha}\rangle = -\langle \mu_j, \check{\alpha}\rangle > 0$. By definition of the weight μ_j, this condition implies that $\check{\alpha} \ge \check{\alpha}_j$ and $2 \le n_j = \langle \lambda, \check{\alpha}_j\rangle \le \langle \lambda, \check{\alpha}\rangle$, which contradicts the fact $\langle \lambda, \check{\alpha}\rangle = -\langle \mu_j, \check{\alpha}\rangle = 1$, since μ_j is minuscule.

Now for the second assertion, recall from 2B3 that this equality holds if and only if, for any $\check{\alpha} \in \check{\Phi}^+$,

$$\langle \mu_j, \check{\alpha}\rangle\, n(\check{\alpha}, w_0^{-1} e^{-\lambda-\mu_j}) \ge 0, \tag{7-1}$$

where the integer $n(\check{\alpha}, w_0^{-1}e^{-\lambda-\mu_j})$ is $\langle -\lambda-\mu_j, \check{\alpha}\rangle$ in the case $w_0^{-1}\check{\alpha} \in \check{\Phi}^+$ and $1 - \langle \lambda+\mu_j, \check{\alpha}\rangle$ if $w_0^{-1}\check{\alpha} \in \check{\Phi}^-$. In the case $w_0^{-1}\check{\alpha} \in \check{\Phi}^+$, inequality (7-1) obviously holds. Suppose now that $w_0^{-1}\check{\alpha} \in \check{\Phi}^-$ and that $\langle \mu_j, \check{\alpha}\rangle = -1$. Then $\check{\alpha} \ge \check{\alpha}_j$, so again, $2 \le \langle \lambda, \check{\alpha}\rangle$ and $n(\check{\alpha}, w_0^{-1}e^{-\lambda-\mu_j}) \le 0$. □

Choose $d \in \mathfrak{D}$ an element with maximal length such that $mE(d) \neq 0$.

Theorem 7.3. *The element $mE(d)$ is an eigenvector for the action of the affine Hecke algebra.*

Proof. With Lemmas 2.6 and 4.4 we compute $E(d) = \tau^*_{d^{-1}}$ for any $d \in \mathfrak{D}$. First note that the braid relations in $\mathcal{H}$ ensure that $mE(d)$ is an eigenvector for the elements of the form τ_t with $t \in \mathbb{T}$. Let $s \in S_{\mathrm{aff}}$. We have to show that $mE(d)\tau_s^*$ is proportional to $mE(d)$.

- If $\ell(ds) = \ell(d) - 1$, then $\tau^*_{d^{-1}} = \tau^*_{(sd)^{-1}}\tau^*_s$. In $\mathcal{H} \otimes_{\mathbb{Z}} k$, where $(\tau^*_s)^2 = \tau^*_s \nu_s$ (Remark 3.1), we have $\tau^*_{d^{-1}}\tau^*_s = \tau^*_{d^{-1}}\nu_s$, so $mE(d)\tau^*_s = mE(d)\nu_s$, which is proportional to $mE(d)$ by our first remark.
- If $\ell(ds) = \ell(d) + 1$, then $\tau^*_{(ds)^{-1}} = \tau^*_{d^{-1}}\tau^*_s$. If $ds \in \mathcal{D}$, then $0 = mE(ds) = m\tau^*_{(ds)^{-1}} = mE(d)\tau^*_s$ by the maximal property of $\ell(d)$. If $ds \notin \mathcal{D}$, then Proposition 2.7 says that there exists $w_0 \in W_0$ such that $ds = w_0 d$ with $\ell(w_0) + \ell(d) = \ell(ds)$. So
$$E(d)\tau^*_s = \tau^*_{w_0^{-1}} E(d).$$
Since m is a character for the finite Hecke algebra, $m\tau^*_{w_0^{-1}}$ is proportional to m, so $mE(d)\tau^*_s$ is proportional to $mE(d)$. □

The statement of the theorem is exactly the claim of [Vignéras 2005, Conjecture 1], where it is proven that it implies the numerical correspondence described by Theorem 1.1 in our introduction.

8. Generic spherical Hecke algebra and Iwahori–Hecke algebra

8A. Denote by $*$ the convolution operator in the generic pro-p-Hecke algebra $\mathcal{H}$ and by $e_I \in \mathcal{H}$ the characteristic function of the Iwahori subgroup. The generic Iwahori–Hecke algebra $\mathbf{H}$ coincides with the algebra $e_I * \mathcal{H} * e_I$ with unit e_I, so all the results of Sections 3 and 4 have (well-known) analogs in the Iwahori case. The generic Iwahori–Hecke algebra $\mathbf{H}$ has $\mathbb{Z}[q]$-basis $(\mathbf{T}_w)_{w\in W}$, where $\mathbf{T}_w = e_I * \tau_w * e_I$ corresponds to the double coset IwI, satisfying the following *braid* and *quadratic* relations.

- $\mathbf{T}_w\mathbf{T}_{w'} = \mathbf{T}_{ww'}$ for any $w,\ w' \in W$ such that $\ell(ww') = \ell(w) + \ell(w')$,
- $\mathbf{T}_s^2 = q + (q-1)\mathbf{T}_s$ for $s \in S_{\mathrm{aff}}$.

Denote by Θ the classic Bernstein embedding
$$\Theta : \mathbb{Z}[q^{\pm 1/2}][\Lambda] \to \mathbf{H} \otimes_{\mathbb{Z}[q]} \mathbb{Z}[q^{\pm 1/2}]$$
naturally arising from the Bernstein map θ of Section 4A and satisfying $\Theta(\lambda) = \mathbf{T}_{e^\lambda}$ for any antidominant weight $\lambda \in \Lambda$. For $w \in W$, define $\mathbf{E}(w) := e_I * E(w) * e_I$. It is explicitly given by the formula
$$\mathbf{E}(w) = q^{(\ell(w)-\ell(w_0)-\ell(e^{\lambda_1})+\ell(e^{\lambda_2}))/2}\Theta(\lambda)\mathbf{T}_{w_0}$$
for $\lambda \in \Lambda$ and $w_0 \in W_0$ such that $w = e^\lambda w_0$ and $\lambda_1, \lambda_2 \in \Lambda$ are antidominant weights satisfying $\lambda = \lambda_1 - \lambda_2$. Theorem 4.5, translated to the Iwahori case, gives the following results (see also [Vignéras 2006, Chapitre 3]). The image $\mathbf{A}$ of $\mathbf{E} : \mathbb{Z}[q][\Lambda] \to \mathbf{H}$ coincides with the intersection of $\mathbf{H}$ with the image of Θ. It has

$\mathbb{Z}[q]$-basis $(\mathbf{E}(\lambda))_{\lambda\in\Lambda}$. As a $\mathbb{Z}[q]$-algebra, it is generated by the elements

$$(\mathbf{E}(\lambda_I))_{I\subsetneq\{1,\dots,n\}},\ \mathbf{E}(\lambda_{\{1,\dots,n\}})^{\pm 1}$$

with the relations

$$\mathbf{E}(\lambda_I)\mathbf{E}(\lambda_J) = q^{bc}\mathbf{E}(\lambda_{I\cup J})\mathbf{E}(\lambda_{I\cap J}) \tag{8-1}$$

for any $I, J \subset \{1,\dots,n\}$ with $|I\cap J| = a$, $|I| = a+b$, $|J| = a+c$. The center of $\mathbf{H}$ is the space of W_0-invariants in $\mathbf{A}$. It is equal to the $\mathbb{Z}[q]$-algebra of polynomials in the variables

$$Z_1,\dots,Z_{n-1},Z_n^{\pm 1},$$

where, for $i\in\{1,\dots,n\}$, we denote by Z_i the central element

$$Z_i = \sum_{w_0\in W_0/W_0(\mu_i)} \mathbf{E}({}^{w_0}\mu_i).$$

8B. ***Integral Satake isomorphism.*** We closely follow the work of Schneider and Teitelbaum [2006], who introduce a renormalized version of the classic Satake map in order to get a p-adic Satake isomorphism, and check that their description provides us in addition with an *integral Satake isomorphism*.

8B1. In Section 4B, we defined a twisted action of W_0 on the weights. Denote by $\mathbb{Z}[q^{\pm 1/2}][\Lambda]^{W_0,\gamma}$ the space of invariants of $\mathbb{Z}[q^{\pm 1/2}][\Lambda]$ under this action. It has $\mathbb{Z}[q^{\pm 1/2}]$-basis $\{\sigma_\lambda\}_\lambda$ with

$$\sigma_\lambda = \sum_{w_0\in W_0/W_0(\lambda)} w_0\centerdot\lambda = \sum_{w_0\in W_0/W_0(\lambda)} \gamma(w_0,\lambda)^{w_0}\lambda,$$

where λ runs over the set Λ_{anti} of antidominant weights. Note that σ_λ is well-defined for any weight λ thanks to property (c) (of Section 4B) of the cocycle γ.

We call *the generic spherical Hecke algebra* and denote by $\mathcal{H}_{\mathbb{Z}[q]}(G,K_0)$ the $\mathbb{Z}[q]$-algebra $\mathbb{Z}[q][K_0\backslash G/K_0]$ of the functions with finite support on the double cosets of G modulo K_0, with the usual convolution product. The $\mathbb{Z}[q^{\pm 1/2}]$-algebra $\mathcal{H}_{\mathbb{Z}[q]}(G,K_0)\otimes_{\mathbb{Z}[q]}\mathbb{Z}[q^{\pm 1/2}]$ will be denoted by $\mathcal{H}_{\mathbb{Z}[q^{\pm 1/2}]}(G,K_0)$.

A system of representatives for the double cosets $K_0\backslash G/K_0$ is given by the set Λ_{anti} of antidominant weights. For $\lambda\in\Lambda$, denote by ψ_λ the characteristic function of $K_0e^\lambda K_0$. The results of [Schneider and Teitelbaum 2006, p. 23] with $\xi = 1$ give the next theorem, the proof of which involves the subsequent lemma.

Theorem 8.1. *There is an injective morphism of $\mathbb{Z}[q^{\pm 1/2}]$-algebras*

$$\begin{aligned} \mathrm{S} : \mathcal{H}_{\mathbb{Z}[q^{\pm 1/2}]}(G,K_0) &\to \mathbb{Z}[q^{\pm 1/2}][\Lambda], \\ \psi_\lambda,\ \lambda\in\Lambda_{\text{anti}} &\mapsto \sum_{\eta\in\Lambda_{\text{anti}}} c(\eta,\lambda)\sigma_\eta, \end{aligned} \tag{8-2}$$

where $c(\eta,\lambda) = [(Ue^\eta K_0\cap K_0e^\lambda K_0)/K_0]$. Its image is equal to $\mathbb{Z}[q^{\pm 1/2}][\Lambda]^{W_0,\gamma}$.

Lemma 8.2. *If $\eta, \lambda \in \Lambda$ are antidominant weights, then*

(1) $c(\lambda, \lambda) = 1$, *and*

(2) $c(\eta, \lambda) = 0$ *unless $\lambda - \eta$ is an antidominant weight.*

Note that, the coefficient $c(\eta, \lambda)$ being integral, the image of $\mathcal{H}_{\mathbb{Z}[q]}(G, K_0)$ by the map S lies in $\mathbb{Z}[q][\sigma_\lambda, \lambda \in \Lambda_{\mathrm{anti}}]$. From this lemma, one also deduces the following result.

Lemma 8.3. *The image of $\mathcal{H}_{\mathbb{Z}[q]}(G, K_0)$ by the map* S *is $\mathbb{Z}[q][\sigma_\lambda, \lambda \in \Lambda_{\mathrm{anti}}]$.*

Proof. One has to check that any σ_λ with $\lambda \in \Lambda_{\mathrm{anti}}$ lies in the image of $\mathcal{H}_{\mathbb{Z}[q]}(G, K_0)$ by the map S. Recall that the element $\varpi^n = e^{\mu_n}$ is central in G, so if the weight λ has the form $k\mu_n$ with $k \in \mathbb{Z}$, then σ_λ is the image by S of $\psi_{k\mu_n}$, which is invertible in $\mathcal{H}_{\mathbb{Z}[q]}(G, K_0)$. So it remains to prove the property for nontrivial weights λ that can be written $\lambda = \sum_{i=1}^{n-1} k_i \mu_i$, with $k_i \in \mathbb{N}$, and we do it by induction on $\sum_{i=1}^{n-1} k_i$. The only antidominant weights η such that $\lambda - \eta$ is antidominant are the $\sum_{i=1}^{n-1} m_i \mu_i$ with $0 \le m_i \le k_i$. By induction, if such an η satisfies $\eta \neq \lambda$, then σ_η is in the image of $\mathcal{H}_{\mathbb{Z}[q]}(G, K_0)$ by S. Lemma 8.2(1) then ensures that it is also true for σ_λ. □

We have checked that the map in Theorem 8.1 actually defines an integral version of a Satake isomorphism: the restriction of S to the generic spherical algebra $\mathcal{H}_{\mathbb{Z}[q]}(G, K_0)$ defines an isomorphism

$$\mathrm{S} : \mathcal{H}_{\mathbb{Z}[q]}(G, K_0) \xrightarrow{\sim} \mathbb{Z}[q][\sigma_\lambda, \lambda \in \Lambda_{\mathrm{anti}}]. \tag{8-3}$$

An important consequence of Lemma 4.2 and property (a) of the cocycle γ is the fact that for any $w_0 \in W_0$, the coefficient $\gamma(w_0, \lambda)$ belongs to $\mathbb{Z}[q]$ if λ is antidominant. So σ_λ actually lies in $\mathbb{Z}[q][\Lambda]$. The supports of the elements σ_λ being disjoint for $\lambda \in \Lambda_{\mathrm{anti}}$ and each coefficient $\gamma(1, \lambda)$ being 1, one obtains

$$\mathbb{Z}[q^{\pm 1/2}][\Lambda]^{W_0, \gamma} \cap \mathbb{Z}[q][\Lambda] = \mathbb{Z}[q][\sigma_\lambda, \lambda \in \Lambda_{\mathrm{anti}}]. \tag{8-4}$$

8C. *Compatibility of Bernstein and Satake transforms.* Note that for any antidominant weight λ, the element

$$\Theta(\sigma_\lambda) = \sum_{w_0 \in W_0 / W_0(\lambda)} \mathbf{E}({}^{w_0}\lambda) \tag{8-5}$$

belongs to the center of $\mathbf{H}$. The description of the center of $\mathbf{H}$ in Section 8A implies the following.

Proposition 8.4. *Composing Θ with the isomorphism (8-3) gives an isomorphism between $\mathcal{H}_{\mathbb{Z}[q]}(G, K_0)$ and the center of* $\mathbf{H}$.

For $1 \le i \le n$, denote by T_i the element ψ_{μ_i}. The generic spherical algebra $\mathcal{H}_{\mathbb{Z}[q]}(G, K_0)$ is an algebra of polynomials in the variables $T_1, \dots, T_{n-1}, T_n^{\pm 1}$. Consider the G-equivariant map

$$\begin{aligned} \mathbb{Z}[q][I\backslash G] &\to \mathbb{Z}[q][K_0\backslash G], \\ f &\mapsto e_{K_0} * f, \end{aligned} \tag{8-6}$$

where e_{K_0} denotes the characteristic function of K_0 and the convolution product is given by

$$e_{K_0} * f(x) = \sum_{t \in G/I} e_{K_0}(t)\, f(t^{-1}x) = \sum_{t \in I\backslash K_0} f(tx) \quad \text{for} \quad x \in G.$$

Proposition 8.5. *Composing the maps*

$$\mathcal{H}_{\mathbb{Z}[q]}(G, K_0) \xrightarrow{\Theta \circ \mathrm{S}} \mathbf{A} \xrightarrow{e_{K_0} * \cdot} \mathbb{Z}[q][K_0\backslash G]$$

gives the identity on $\mathcal{H}_{\mathbb{Z}[q]}(\mathrm{G}, K_0)$.

Proof. See [Schneider and Stuhler 1991, p. 32]. □

Note that the compatibility refers to the *classic* Bernstein map and the *integral* Satake transform.

8D. Denote by $\mathbf{R}$ the mod p reduction of the map (8-6), that is, the G-equivariant map

$$\begin{aligned} \mathrm{ind}_I^G \mathbf{1}_k &\to \mathrm{ind}_{K_0}^G \mathbf{1}_k, \\ f &\mapsto e_{K_0} * f, \end{aligned} \tag{8-7}$$

where $\mathrm{ind}_I^G \mathbf{1}_k$ and $\mathrm{ind}_{K_0}^G \mathbf{1}_k$ denote respectively the compact induction of the trivial character with values in k of the Iwahori subgroup I and of the maximal compact subgroup K_0.

Proposition 8.6. *Let $\mu \in \Lambda$ be a minuscule weight. The image by $\mathbf{R}$ of $\mathbf{E}(\mu) \in \mathbf{A} \otimes_{\mathbb{Z}[q]} k$ is equal to $\psi_\mu \in \mathcal{H}_{\mathbb{Z}[q]}(G, K_0) \otimes_{\mathbb{Z}[q]} k$ if μ is a dominant weight, and to zero otherwise.*

The proof will be a consequence of the following lemmas.

Lemma 8.7. *For $\mu \in \Lambda$ dominant and minuscule,*

$$K_0 e^\mu K_0 = \coprod_{d \in \mathcal{D},\, d \preceq e^\mu} K_0 d I,$$

where $\preceq$ denotes the extended Bruhat order on W.

Proof. We have to prove that for $\mu \in \Lambda$ dominant and minuscule, $\mathcal{D} \cap K_0 e^\mu K_0 = \{d \in \mathcal{D},\ d \preceq e^\mu\}$. For any such weight μ, the corresponding translation can be written $e^\mu = \varpi^k w_0$ with $k \in \{0, \dots, n\}$ and $w_0 \in W_0$. By definition of the extended

Bruhat order, an element $d \in W$ satisfies $d \preceq e^\mu$ if and only if it has the form $d = \varpi^k w$ with $w \in W_0$ such that $w \preceq w_0$. So $\{d \in \mathfrak{D},\, d \preceq e^\mu\} \subset \mathfrak{D} \cap K_0 e^\mu K_0$. Let $d \in \mathfrak{D}$. Lemma 2.6 says that d can be written $d = e^\lambda w$ with $w \in W_0$ and $\lambda \in \Lambda$ a dominant weight such that $\ell(e^\lambda) = \ell(d) + \ell(w^{-1})$. If $d \in K_0 e^\mu K_0$, then $K_0 e^\lambda K_0 = K_0 e^\mu K_0$ and $\lambda = \mu$. Since ϖ has length zero, one then has $\ell(w_0 w) + \ell(w^{-1}) = \ell(w_0)$, so $w_0 w \preceq w_0$ and $d = \varpi^k w_0 w \preceq e^\mu$. □

Lemma 8.8. *For $w \in W$,*

$$\mathbf{R}(\mathbf{T}_w) = |I \backslash (K_0 \cap I w I w^{-1})|\, \mathbf{1}_{K_0 w I} = \begin{cases} \mathbf{1}_{K_0 w I} & \text{if } w \in \mathfrak{D}, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. By definition, the map (8-6) sends the characteristic function $\mathbf{T}_w$ of IwI onto $|I \backslash (K_0 \cap I w I w^{-1})|\, \mathbf{1}_{K_0 w I}$. We have to show that the index $|I \backslash (K_0 \cap I w I w^{-1})|$ is equal to 1 if $w \in \mathfrak{D}$ and is equal to a nontrivial power of q otherwise. If $w \in \mathfrak{D}$, then by length property, one easily checks that $K_0 \cap I w I w^{-1} = I$. Suppose now that w is not an element of $\mathfrak{D}$, that is, that it is not the minimal length element in $W_0 w$: there exists $s \in S_0$ such that $IwI = IsIswI$. Hence $IwIw^{-1} \cap K_0$ contains $IsIs$, which has q right cosets modulo I. □

Lemma 8.9. *For any dominant weight $\lambda \in \Lambda$, the following holds in $\mathbf{H} \otimes_{\mathbb{Z}[q]} k$:*

$$\mathbf{E}(\lambda) = \sum_{w \in W,\, w \preceq e^\lambda} \mathbf{T}_w.$$

Proof. Let us show that for any $x \in W$, one has

$$\mathbf{T}^*_{x^{-1}} = \sum_{w \in W,\, w \preceq x} \mathbf{T}_w \in \mathbf{H} \otimes_{\mathbb{Z}[q]} k.$$

This proves the lemma because $\mathbf{E}(\lambda) = \mathbf{T}^*_{e^{-\lambda}}$ for a dominant weight λ. It is enough to show the equality for $x \in W_{\mathrm{aff}}$, and we do it by induction on $\ell(x)$. If $x = s \in S_{\mathrm{aff}}$, then $\mathbf{T}^*_{x^{-1}} = \mathbf{T}^*_s = \mathbf{T}_s + 1 - q = \mathbf{T}_s + 1$ in $\mathbf{H} \otimes_{\mathbb{Z}[q]} k$. Now suppose $x \in W_{\mathrm{aff}}$ and $s \in S$ is such that $\ell(sx) = \ell(x) + 1$. In $\mathbf{H} \otimes_{\mathbb{Z}[q]} k$, one has by induction

$$\mathbf{T}^*_{(sx)^{-1}} = \mathbf{T}^*_s \mathbf{T}^*_{x^{-1}} = (\mathbf{T}_s + 1) \sum_{y \preceq x} \mathbf{T}_y = \sum_{y \preceq x} \mathbf{T}_s \mathbf{T}_y + \sum_{y \preceq x} \mathbf{T}_y.$$

Let $y \preceq x$. If $\ell(sy) = \ell(y) + 1$, then $\mathbf{T}_s \mathbf{T}_y = \mathbf{T}_{sy}$ and $sy \preceq sx$. Otherwise $\mathbf{T}_s \mathbf{T}_y = -\mathbf{T}_y$, so $\mathbf{T}^*_{(sx)^{-1}} = \sum_{sy' \preceq y' \preceq sx} \mathbf{T}_{y'} + \sum_{y \preceq sx,\, y \preceq sy} \mathbf{T}_y = \sum_{y \preceq sx} \mathbf{T}_y$. □

Lemma 8.10. *If μ is minuscule and not dominant, then $\mathbf{R}(\mathbf{E}(\mu)) = 0$.*

Proof. Let λ be the unique antidominant weight in the orbit of μ and $d \in W_0$ with minimal length in $W_0(\lambda)d$ such that $\mu = {}^{d^{-1}}\lambda$. Lemma 4.4 says that $\mathbf{E}(\mu) = \mathbf{T}_{e^\mu d^{-1}} \mathbf{T}^*_{d^{-1}}$. For any $w_0 \in W_0$, we have $\ell(e^\mu d^{-1}) + \ell(w_0) = \ell(e^\mu d^{-1} w_0)$, which can be seen by applying 2B2 and recalling that for any $\check{\alpha} \in \check{\Phi}^+$, if $\langle \lambda, \check{\alpha} \rangle = 0$

then $d^{-1}\check{\alpha} \in \check{\Phi}^+$. This implies that the elements of the Iwahori–Matsumoto basis appearing in the decomposition of $\mathbf{E}(\mu)$ have the form $\tau_{e^\mu d^{-1} w_0}$, with $w_0 \in W_0$. In particular, if μ is not dominant, then $e^\mu d^{-1} w_0$ is not an element of $\mathfrak{D}$, by Lemma 2.6, and $\mathbf{E}(\mu)$ is sent by $\mathbf{R}$ on zero, by Lemma 8.8. □

Proof of Proposition 8.6. Let $\mu \in \Lambda$ be a minuscule weight. If it is not dominant, Lemma 8.10 says that its image by $\mathbf{R}$ is zero. If it is dominant, Lemmas 8.8 and 8.9 together say that $\mathbf{R}(\mathbf{E}(\mu))$ is the sum of the characteristic functions of $K_0 w I$, where $w \in \mathfrak{D}$, $w \preceq e^\mu$, which, by Lemma 8.7, is the characteristic function of $K_0 e^\mu K_0$. □

8E. ***On Barthel–Livné's unramified representations for*** $\mathbf{GL}_n$**.** For $i \in \{1, \dots, n\}$, choose $\alpha_i \in k$ with $\alpha_n \neq 0$. Set $\alpha_0 = 1$. Define χ_0 to be the k-character of $\mathbf{A}$ with dominant support given by $\mathbf{E}(\mu_{\{n-i+1,\dots,n\}}) \mapsto \alpha_i$ for $i \in \{1, \dots, n\}$.

Define the associated character of $\mathcal{H}(G, K_0)_{\mathbb{Z}[q]} \otimes k$ by $T_i \mapsto \alpha_i$ for $i \in \{1, \dots, n\}$ and denote by

$$\frac{\operatorname{ind}_{K_0}^G \mathbf{1}_k}{\sum_i (T_i - \alpha_i)}$$

the quotient of the universal representation $\operatorname{ind}_{K_0}^G \mathbf{1}_k$ by $\sum_i (T_i - \alpha_i) \operatorname{ind}_{K_0}^G \mathbf{1}_k$.

By the results of 8D, the G-equivariant surjective morphism

$$\begin{aligned} \operatorname{ind}_I^G \mathbf{1}_k &\to \frac{\operatorname{ind}_{K_0}^G \mathbf{1}_k}{\sum_i (T_i - \alpha_i)}, \\ f &\mapsto \mathbf{R}(f) \mod \sum_i (T_i - \alpha_i) \end{aligned} \tag{8-8}$$

factors into a surjective G-equivariant morphism

$$\chi_0 \otimes_{\mathbf{A}} \operatorname{ind}_I^G \mathbf{1}_k \to \frac{\operatorname{ind}_{K_0}^G \mathbf{1}_k}{\sum_i (T_i - \alpha_i)}. \tag{8-9}$$

Example 8.11. Suppose that one of the α_i, $i \in \{1, \dots, n-1\}$ is nonzero. The unique character of $\mathbf{A}$ with antidominant support in the W_0-orbit of χ_0 satisfies Hypothesis ($\star$) of Section 5D if and only if $\alpha_{i-1}\alpha_i\alpha_{i+1} \neq 0$ implies ${\alpha_i}^2 \neq \alpha_{i-1}\alpha_{i+1}$ for any $i \in \{1, \dots, n-1\}$.

Under this hypothesis and if none of the elements α_i is zero, then, by the results of [Ollivier 2006a], the representation $\chi_0 \otimes_{\mathbf{A}} \operatorname{ind}_I^G \mathbf{1}_k$ is irreducible and isomorphic to the principal series induced by the unramified character

$$T \to k^*, \quad \mu_i \mapsto \alpha_i^{-1},$$

and (8-9) is an isomorphism.

Acknowledgments

I would like to thank Marie-France Vignéras for introducing me to the subject of mod p representations a few years ago, and for enthusiastic and fruitful discussions since then. I am grateful to the referee of this article for a thorough and helpful report, and to Emily Dolan for stylistic comments. This article was conceived during the fall semester of 2008, while I was a fellow at the Radcliffe Institute for Advanced Study at Harvard University, enjoying exceptional work conditions and extremely stimulating interactions, in Cambridge and in Somerville, MA.

References

[Barthel and Livné 1995] L. Barthel and R. Livné, "Modular representations of GL_2 of a local field: the ordinary, unramified case", *J. Number Theory* **55**:1 (1995), 1–27. MR 96m:22036 Zbl 0841.11026

[Bourbaki 1961] N. Bourbaki, *Algèbre commutative, I–II: Modules plats, localisation*, Actualités Scientifiques et Industrielles **1290**, Herman, Paris, 1961. MR 36 #146 Zbl 0108.04002

[Bourbaki 1968] N. Bourbaki, *Groupes et algèbres de Lie, IV–VI: Groupes de Coxeter et systèmes de Tits, groupes engendrés par des réflexions, systèmes de racines*, Actualités Scientifiques et Industrielles **1337**, Hermann, Paris, 1968. MR 39 #1590 Zbl 0186.33001

[Breuil and Paskunas 2007] C. Breuil and V. Paskunas, "Towards a modulo p Langlands correspondence for GL_2", preprint, Institut des Hautes Études Scientifiques, 2007, Available at http://inc.web.ihes.fr/prepub/PREPRINTS/2007/M/M-07-25.pdf.

[Bushnell and Kutzko 1998] C. J. Bushnell and P. C. Kutzko, "Smooth representations of reductive p-adic groups: structure theory via types", *Proc. London Math. Soc.* (3) **77**:3 (1998), 582–634. MR 2000c:22014 Zbl 0911.22014

[Cabanes and Enguehard 2004] M. Cabanes and M. Enguehard, *Representation theory of finite reductive groups*, New Mathematical Monographs **1**, Cambridge University Press, 2004. MR 2005g:20067 Zbl 1069.20032

[Carter 1985] R. W. Carter, *Finite groups of Lie type*, Wiley, New York, 1985. MR 87d:20060 Zbl 0567.20023

[Dat 1999] J.-F. Dat, "Types et inductions pour les représentations modulaires des groupes p-adiques", *Ann. Sci. École Norm. Sup.* (4) **32**:1 (1999), 1–38. MR 99m:22018 Zbl 0935.22013

[Grosse-Klönne 2009] E. Grosse-Klönne, "On special representations of p-adic reductive groups", preprint, The Erwin Schrödinger International Institute for Mathematical Physics, 2009.

[Herzig 2010] F. Herzig, "The classification of irreducible admissible mod p representations of a p-adic GL_n", preprint, 2010. arXiv 1005.1713v1

[Iwahori and Matsumoto 1965] N. Iwahori and H. Matsumoto, "On some Bruhat decomposition and the structure of the Hecke rings of p-adic Chevalley groups", *Inst. Hautes Études Sci. Publ. Math.* 25 (1965), 5–48. MR 32 #2486 Zbl 0228.20015

[Lusztig 1989] G. Lusztig, "Affine Hecke algebras and their graded version", *J. Amer. Math. Soc.* **2**:3 (1989), 599–635. MR 90e:16049 Zbl 0715.22020

[Ollivier 2006a] R. Ollivier, "Critère d'irréductibilité pour les séries principales de $\mathrm{GL}_n(F)$ en caractéristique p", *J. Algebra* **304**:1 (2006), 39–72. MR 2007k:22017 Zbl 1108.22012

[Ollivier 2006b] R. Ollivier, "Modules simples en caractéristique p de l'algèbre de Hecke du pro-p-Iwahori de $\mathrm{GL}_3(F)$", *J. Algebra* **304**:1 (2006), 1–38. MR 2007j:22030 Zbl 1107.22010

[Ollivier 2006c] R. Ollivier, "Pro-p-invariants des séries principales de $\mathrm{GL}_n(F)$ en caractéristique p", preprint, 2006.

[Ollivier 2007] R. Ollivier, "Platitude du pro-p-module universel de $\mathrm{GL}_2(F)$ en caractéristique p", *Compos. Math.* **143**:3 (2007), 703–720. MR 2008e:22017 Zbl 1170.22007

[Ollivier 2009] R. Ollivier, "Le foncteur des invariants sous l'action du pro-p-Iwahori de $\mathrm{GL}_2(F)$", *J. Reine Angew. Math.* **635** (2009), 149–185. MR 2572257 Zbl pre05633739

[Rogawski 1985] J. D. Rogawski, "On modules over the Hecke algebra of a p-adic group", *Invent. Math.* **79**:3 (1985), 443–465. MR 86j:22028 Zbl 0579.20037

[Schneider and Stuhler 1991] P. Schneider and U. Stuhler, "The cohomology of p-adic symmetric spaces", *Invent. Math.* **105**:1 (1991), 47–122. MR 92k:11057 Zbl 0751.14016

[Schneider and Teitelbaum 2006] P. Schneider and J. Teitelbaum, "Banach–Hecke algebras and p-adic Galois representations", *Doc. Math.* Extra Vol. (2006), 631–684. MR 2008b:11126 Zbl 1140.11026

[Vignéras 1998] M.-F. Vignéras, "Induced R-representations of p-adic reductive groups", *Selecta Math.* (*N.S.*) **4**:4 (1998), 549–623. MR 99k:22026 Zbl 0943.22017

[Vignéras 2004] M.-F. Vignéras, "Representations modulo p of the p-adic group $\mathrm{GL}(2, F)$", *Compos. Math.* **140**:2 (2004), 333–358. MR 2004m:22028 Zbl 1049.22010

[Vignéras 2005] M.-F. Vignéras, "Pro-p-Iwahori Hecke ring and supersingular $\overline{\mathbf{F}}_p$-representations", *Math. Ann.* **331**:3 (2005), 523–556. Erratum in **333**:2, 699–701. MR 2005m:22020 Zbl 1107.22011

[Vignéras 2006] M.-F. Vignéras, "Algèbres de Hecke affines génériques", *Represent. Theory* **10** (2006), 1–20. MR 2006i:20005 Zbl 1134.22014

[Vignéras 2007] M.-F. Vignéras, "Représentations irréductibles de $\mathrm{GL}(2, F)$ modulo p", pp. 548–563 in *L-functions and Galois representations*, edited by D. Burns et al., London Math. Soc. Lecture Note Ser. **320**, Cambridge University Press, 2007. MR 2009h:11084 Zbl 1172.11017

[Vignéras 2008] M.-F. Vignéras, "Série principale modulo p de groupes réductifs p-adiques", *Geom. Funct. Anal.* **17**:6 (2008), 2090–2112. MR 2009a:22015 Zbl 05275302

Communicated by Barry Mazur
Received 2009-07-02 Revised 2010-04-21 Accepted 2010-06-06

ollivier@math.columbia.edu *Université de Versailles Saint-Quentin, Laboratoire de Mathématiques de Versailles, 45 avenue des États-Unis, 78035 Versailles Cedex, France*

Current address: *Columbia University, Mathematics Department, MC 4445, 2990 Broadway, New York, NY 10027, United States*

ALGEBRA AND NUMBER THEORY 4:6(2010)

Patching and admissibility over two-dimensional complete local domains

Danny Neftin and Elad Paran

We develop a patching machinery over the field $E = K((X, Y))$ of formal power series in two variables over an infinite field K. We apply this machinery to prove that if K is separably closed and G is a finite group of order not divisible by $\operatorname{char} E$, then there exists a G-crossed product algebra with center E if and only if the Sylow subgroups of G are abelian of rank at most 2.

Introduction

Complete local domains play an important role in commutative algebra and algebraic geometry, and their algebraic properties were already described in 1946 by Cohen's structure theorem. The Galois theoretic properties of their quotient fields were extensively studied over the past two decades. The pioneering work in this line of research is due to Harbater [1987], who introduced the method of patching to prove that if R is a complete local domain with quotient field K, then every finite group occurs as a Galois group over $K(x)$. This result was strengthened by Pop [1996] and, in a different language, by Haran and Jarden [1998], who showed that if moreover R is of dimension 1, then every finite split embedding problem over $K(x)$ is solvable.

The first step towards higher dimension was made by Harbater and Stevenson [2005], who essentially showed that if R is a complete local domain of dimension 2, then every finite split embedding problem over $\operatorname{Quot} R$ has $|R|$ independent solutions. That is, the absolute Galois group of $\operatorname{Quot} R$ is semifree of rank $|R|$ (see [Bary-Soroker et al. 2008] for details on this notion). This result was generalized in [Pop 2010; Paran 2010], where it is shown that if K is the quotient field of a complete local domain of dimension exceeding 1, then $\operatorname{Gal} K$ is semifree.

Despite the major progress made in the study of Galois theory over these fields, little is known about the structure of division algebras over them. A step in that

MSC2000: primary 12E30; secondary 16S35.
Keywords: patching, admissible groups, division algebras, complete local domains.
The second author was partially supported by the Israel Science Foundation (grant No. 343/07), and by an ERC grant.

direction was recently made by Harbater et al. [2009], who consider a question relating Galois theory and Brauer theory over a field E: *Which groups are admissible over E*? That is, which finite groups occur as a Galois group of an *adequate* Galois extension F/E (recall that an extension F/E is called adequate if F is a maximal subfield in an E-central division algebra). Equivalently, for which groups G there is a G-crossed product division algebra with center E. Note that for E as above and a finite extension F/E, the above maximality requirement can be omitted since any F which is a subfield of an E-division algebra is also a maximal subfield of some E-division algebra (see Remark 3.9).

This question was first considered by Schacher [1968] over global fields. He proved that any $\mathbb{Q}$-admissible group has metacyclic Sylow subgroups and conjectured the converse. Admissibility has been studied extensively over global fields [Stern 1982; Sonn 1983; Schacher and Sonn 1992; Liedahl 1996; Feit 2004; Neftin 2009], function fields, and fields of Laurent series [Fein et al. 1992; Fein and Schacher 1995a; 1995b].

The main theorem of [Harbater et al. 2009] asserts that if E is a function field in one variable over a complete discretely valued field with an algebraically closed residue field, then a finite group G of order not divisible by $\operatorname{char} E$ is admissible over E if and only if each of the Sylow subgroups of G is abelian of rank at most 2 (i.e., generated by two elements).

In this work, we take the next natural step, and determine the admissible groups over quotient fields of equicharacteristic (that is, having the same characteristic as their residue field) two-dimensional complete local domains, with a separably closed residue field. In particular, we determine the admissible groups over $C((X, Y))$, whenever C is a separably closed field. This problem was posed to the first author by David Harbater. We show that the result of [Harbater et al. 2009] holds over these fields as well.

Main Theorem. *Let R be an equicharacteristic complete local domain of dimension* 2, *with a separably closed residue field. Let G be a finite group of order not divisible by* $\operatorname{char} R$. *Then G is admissible over* $\operatorname{Quot} R$ *if and only if each of its Sylow subgroups is abelian of rank at most* 2.

The "only if" part of this result is essentially proven in [Harbater et al. 2009], using results of [Colliot-Thélène et al. 2002].

The "if" part actually holds in greater generality — the residue field need not be separably closed, it suffices that it contains a primitive root of unity of order k, for each $k \in \mathbb{N}$ not divisible by $\operatorname{char} R$ (Proposition 3.7). Our proof strategy for the backward direction is as follows. We first use Cohen's structure theorem to reduce the problem from $\operatorname{Quot} R$ to a field E of the form $K((X, Y))$. We then apply a patching argument as in [Harbater et al. 2009]; we explicitly realize each

Sylow subgroup of G by a Galois extension of E which is a maximal subfield in some E-central division algebra. We then patch these realizations into a Galois extension F/E with group G, in a way that also patches the division algebras into an E-central division algebra D containing F as a maximal subfield.

A key ingredient in our proof is a patching machinery over fields of the form $K((X, Y))$, where K may be an arbitrary infinite field. In [Pop 2010; Paran 2010], problems over $K((X, Y))$ are lifted to $K((X, Y))(Z)$, solved there (via different methods), and then the solutions are specialized into solutions over $K((X, Y))$ using Hilbertianity. This approach seems inapplicable to our current goal, since adequate extensions usually do not remain adequate under specialization. Instead we patch groups directly over $K((X, Y))$. To this end we define "analytic fields" over $K((X, Y))$, satisfying the axioms of algebraic patching (i.e., matrix factorization and intersection), as presented in [Haran and Jarden 1998] (a slightly different axiomatization from the "field patching" axiomatization of [Harbater and Hartmann 2007]). The construction of these analytic rings is of a rigid-geometric nature. In recent communication with David Harbater, we learned that a formal-geometric analogue of this form of patching was carried out by him in [2003, Theorem 5.3.9] in order to solve split embedding problems over the field $\mathbb{C}((X, Y))$ of formal power series in two variables over the complex numbers. The core patching arguments in the proof of [Harbater 2003, Theorem 5.3.9] can be extended to replace $\mathbb{C}$ by an arbitrary field and used to study admissible groups, in a similar fashion to our development here.

1. Analytic fields

In this section we establish our patching machinery. Fix an infinite field K, and let $E = K((X, Y)) = \mathrm{Quot}(K[\![X, Y]\!])$ be the field of formal power series over K in the variables X and Y. Denote by v the order function of the maximal ideal $\langle X, Y\rangle$ in $K[\![X, Y]\!]$. Then v extends uniquely to a discrete rank-1 valuation of E. Note that $K[\![X, Y]\!]$ is strictly contained in the valuation ring of v in E.

Construction 1.1 (analytic rings over E). Let I be a finite set. For each $i \in I$ let $c_i \in K$, such that $c_i \neq c_j$ for $i \neq j$ (such a choice is possible since K is infinite). For each $i \in I$ denote $z_i = Y/(X - c_i Y) \in E$. For each $J \subseteq I$, consider the subring $K[z_j \mid j \in J][X, Y]$ of E, and let D_J be the completion of this ring with respect to v. Note that for each $J \subseteq I$, $D_J \subseteq D_I$, and that $D_\varnothing = K[\![X, Y]\!]$, since $K[\![X, Y]\!]$ is complete. Let $Q = \mathrm{Quot}\, D_I$, and for each $i \in I$ let $Q_i = E \cdot D_{I \smallsetminus \{i\}}$ and $Q'_i = \bigcap_{j \neq i} Q_j$.

For the rest of this section, we fix the notation of Construction 1.1. A geometric interpretation of the rings defined in Construction 1.1 appears in Remark 1.12. In order to present this interpretation, we need several lemmas.

Lemma 1.2. *Let $i \in I$. Then v is trivial on $K(z_i)$.*

Proof. It suffices to prove that v is trivial on $K[z_i]$. Let $0 \neq f = \sum_{n=0}^{d} a_n z_i^n \in K[z_i]$, with $a_0, \dots, a_d \in K$. Without loss of generality, $a_0 \neq 0$. We have

$$\sum_{n=0}^{d} a_n z_i^n = \frac{\sum_{n=0}^{d} a_n Y^n (X - c_i Y)^{d-n}}{(X - c_i Y)^d}.$$

By opening parentheses, the numerator in this expression can be written as a sum of monomials of degree d. For $n = 0$ we get the summand $a_0 X^d$, while all other monomials in this presentation have a positive power of Y, so they do not cancel $a_0 X^d$. Thus the numerator has value d, and clearly so does the denominator, so $v(f) = 0$. □

Corollary 1.3. *The valuation v is trivial on $K[z_i \mid i \in I]$.*

Proof. Note that for each $i, j \in I$ we have $K(z_i) = K(z_j)$. Thus by the preceding lemma, v is trivial on $K(z_i) = K(z_j \mid j \in I)$, and in particular v is trivial on the subring $K[z_i \mid i \in I]$. □

Lemma 1.4. *Let $J \subseteq I$ and $j \in J$. Then the ring $K[z_l \mid l \in J][X - c_j Y]$ is isomorphic to the ring of polynomials in one variable over $K[z_l \mid l \in J]$.*

Proof. To prove the claim we show that if $\sum_{n=0}^{d} a_n (X - c_j Y)^n = 0$, for $a_0, \dots, a_d \in K[z_l \mid l \in J]$, then $a_0 = \cdots = a_d = 0$. If not, suppose (without loss of generality) that $a_0 \neq 0$. By Corollary 1.3, $v(a_0) = 0$ while $v(a_k (X - c_j Y)^k) = v(a_k) + k = k > 0$ for each $k > 0$. Hence $\infty = v(0) = v\left(\sum_{n=0}^{d} a_n (X - c_j Y)^n\right) = 0$, a contradiction. □

By Lemma 1.4, for each $J \subseteq I$, $j \in J$, each element of $K[z_l \mid l \in J][X - c_j Y]$ has a unique presentation as a polynomial in $X - c_j Y$. Thus we have a natural valuation on this ring, given by $v'\left(\sum_{n=0}^{d} a_n (X - c_j Y)^n\right) = \min(n \mid a_n \neq 0)$, and we may form the completion $K[z_l \mid l \in J][\![X - c_j Y]\!]$ of this ring with respect to v'.

Proposition 1.5. *Let $J \subseteq I$ and $j \in J$. Then $D_J = K[z_l \mid l \in J][\![X - c_j Y]\!]$, and v is given on D_J by $v\left(\sum_{n=0}^{\infty} a_n (X - c_j Y)^n\right) = \min(n \mid a_n \neq 0)$.*

Proof. By Lemma 1.4, v coincides with v' (given in the paragraph preceding this proposition) on $K[z_l \mid l \in J][X - c_j Y]$; hence $K[z_l \mid l \in J][\![X - c_j Y]\!]$ is the completion of $K[z_l \mid l \in J][X - c_j Y]$ with respect to v, and v coincides with v' on the completion. Note that $K[z_l \mid l \in J][X - c_j Y] = K[z_l \mid l \in J][X, Y]$ (since $Y = z_j (X - c_j Y)$ and $X = (1 + c_j z_j)(X - c_j Y)$); hence by the definition of D_J we are done. □

Lemma 1.6. *Let $J \subseteq I$. Then $K[z_j \mid j \in J] = \sum_{j \in J} K[z_j]$.*

Proof. For each $i \neq j \in J$ we have

$$z_i \cdot z_j = \frac{Y^2}{(X - c_i Y) \cdot (X - c_j Y)} = \frac{1}{c_i - c_j} \cdot z_i + \frac{1}{c_j - c_i} z_j.$$

The claim now follows by induction on $|I|$. □

Proposition 1.7. *Let $J, J' \subseteq I$. Then for each $f \in D_{J \cup J'}$ there exist $f_1 \in D_J$ and $f_2 \in D_{J'}$ with $v(f_1), v(f_2) \geq v(f)$, such that $f = f_1 + f_2$.*

Proof. Replace J with $J \smallsetminus (J \cap J')$ to assume that $J \cap J' = \varnothing$. Moreover, without loss of generality J, J' are nonempty. Choose $j \in J$ and $j' \in J'$, and let

$$A_J = K[z_l \mid l \in J], \quad A_{J'} = K[z_l \mid l \in J'], \quad A = K[z_l \mid l \in J \cup J'].$$

By Proposition 1.5, $D_J = A_J[\![X - c_j Y]\!]$, $D_{J'} = A_J[\![X - c_{j'} Y]\!]$, and $D_{J \cup J'} = A[\![X - c_j Y]\!]$. Let $f = \sum_{n=m}^{\infty} a_n (X - c_j Y)^n \in D_{J \cup J'}$, with $a_m \neq 0$. Then $v(f) = m$ by Proposition 1.5. By Lemma 1.6, $A = A_J + A_{J'}$. For each $n \geq m$, let $b_n \in A_J, b'_n \in A_{J'}$ such that $a_n = b_n + b'_n$. Let $f_1 = \sum_{n=m}^{\infty} b_n (X - c_j Y)^n$ and $f_2 = f - f_1 = \sum_{n=m}^{\infty} b'_n (X - c_j Y)^n$. Then $f_1 \in D_J$ and $v(f_1) \geq m$. It remains to prove that $f_2 \in D_{J'}$ and that $v(f_2) \geq m$. This follows by the equality:

$$\begin{aligned} f_2 = \sum_{n=m}^{\infty} b'_n (X - c_j Y)^n &= \sum_{n=m}^{\infty} b'_n ((X - c_{j'} Y) + (c_{j'} - c_j) Y)^n \\ &= \sum_{n=m}^{\infty} (b'_n (1 + (c_{j'} - c_j) z_{j'})^n)(X - c_{j'} Y)^n. \end{aligned}$$ □

The next lemma is a variant of [Harbater and Hartmann 2007, Lemma 3.3], allowing nonprincipal ideals.

Lemma 1.8. *Let $R \subseteq R_1$ and $R_2 \subseteq R_0$ be domains such that $R_0 = R_1 + R_2$. Let w be a nontrivial discrete valuation on* Quot R_0 *such that R is complete with respect to w and $w(x) \geq 0$ for all $x \in R_0$. Let $\mathfrak{p}$, $\mathfrak{p}_1$, $\mathfrak{p}_2$, and $\mathfrak{p}_0$ be the centers of w in R, R_1, R_2, and R_0, respectively. Suppose that $\mathfrak{p} R_0 = \mathfrak{p}_0$ and $R/\mathfrak{p} = R_1/\mathfrak{p}_1 \cap R_2/\mathfrak{p}_2$ (inside $R_0/\mathfrak{p}_0$). Then $R_1 \cap R_2 = R$.*

Proof. First, note that $\mathfrak{p}_0 = \mathfrak{p}_1 + \mathfrak{p}_2$. Indeed, suppose $x \in \mathfrak{p}_0 = \mathfrak{p} R_0$. Then $x = \sum_{i=1}^{n} a_i x_i$ for some $a_1, \ldots, a_n \in R_0$ and $x_1, \ldots, x_n \in \mathfrak{p}$. For each $1 \leq i \leq n$, write $a_i = b_i + b'_i$ with $b_i \in R_1$ and $b'_i \in R_2$. Then $\sum a_i x_i = \sum b_i x_i + \sum b'_i x_i \in \mathfrak{p}_1 + \mathfrak{p}_2$, since $\mathfrak{p} \subseteq \mathfrak{p}_1, \mathfrak{p}_2$.

Let $S = R_1 \cap R_2$ and $\mathfrak{q}$ be the center of w at S. Then the sequence $0 \to S \to R_1 \times R_2 \to R_0 \to 0$ is exact (where the second map is the diagonal map and the third map is substraction). This sequence induces an exact sequence $0 \to S/\mathfrak{q} \to (R_1/\mathfrak{p}_1) \times (R_2/\mathfrak{p}_2) \to R_0/\mathfrak{p}_0 \to 0$. Indeed, the only nontrivial part in showing this

is to check that the kernel of the substraction map is contained in the image of the diagonal map. Suppose $(x_1+\mathfrak{p}_1, x_2+\mathfrak{p}_2) \in (R_1/\mathfrak{p}_1)\times(R_2/\mathfrak{p}_2)$ satisfies $x_1-x_2 \in \mathfrak{p}_0$. Since $\mathfrak{p}_0 = \mathfrak{p}_1+\mathfrak{p}_2$ we may choose $y_1 \in \mathfrak{p}_1$, $y_2 \in \mathfrak{p}_2$ such that $x_1 - y_1 = x_2 - y_2$. Then $(x_1+\mathfrak{p}_1, x_2+\mathfrak{p}_2) = (x_1-y_1+\mathfrak{p}_1, x_2-y_2+\mathfrak{p}_2)$ belongs to the image of the diagonal map. Thus the sequence is exact.

Since $\mathfrak{p}$ is the center of w on R and $\mathfrak{p}_1$ the center of w on R_1, we have $R\cap\mathfrak{p}_1 = \mathfrak{p}$. In particular, the diagonal map $R/\mathfrak{p} \to (R_1/\mathfrak{p}_1)\times(R_2/\mathfrak{p}_2)$ is injective. Since $R_0 = R_1+R_2$, the substraction map $(R_1/\mathfrak{p}_1)\times(R_2/\mathfrak{p}_2) \to R_0/\mathfrak{p}_0$ is surjective. Thus, since $R/\mathfrak{p} = R_1/\mathfrak{p}_1 \cap R_2/\mathfrak{p}_2$, the sequence $0 \to R/\mathfrak{p} \to (R_1/\mathfrak{p}_1)\times(R_2/\mathfrak{p}_2) \to R_0/\mathfrak{p}_0 \to 0$ is also exact. It follows that the natural map $R/\mathfrak{p} \to S/\mathfrak{q}$ is an isomorphism. In particular, $S = R+\mathfrak{p}S$. By induction we have $S = R+\mathfrak{p}^k S$ for each $k \in \mathbb{N}$. Since $\mathfrak{p}_0 = \mathfrak{p}R_0$, $\mathfrak{p} \neq 0$. Since w is discrete, there exists an integer m such that $v(x) \geq m$ for each $x \in \mathfrak{p}$. Thus $v(x) \geq mk$ for each $x \in \mathfrak{p}^k$; hence R is w-dense in S and therefore the completion of R with respect to w contains S. By our assumptions, R is complete; hence $R = S$. □

Lemma 1.9. *The set* $\{z_i^n \mid i \in I, n \in \mathbb{N}\cup\{0\}\}$ *is K-linearly independent.*

Proof. Suppose $a_0 + \sum_{i\in I}\sum_{n=1}^{d_i} a_{i,n}z_i^n = 0$, where $d_i \in \mathbb{N}$ and $a_0, a_{i,n} \in K$ for each i and n. We wish to show that $a_0 = a_{i,n} = 0$ for each i and n. Suppose there exist $i \in I$ and $n \in \mathbb{N}$ such that $a_{i,n} \neq 0$. Without loss of generality, $n = d_i$. Since $X - c_iY$ is a prime element of $K[X,Y]$, it defines a discrete valuation on $K(X,Y)$, which we denote by w. We have $w(Y) = w(Y-c_jX) = 0$ for each $j \neq i$ in I. Thus $w\big(a_0 + \sum_{j\neq i}\sum_{n=1}^{d_j} a_{j,n}z_j^n\big) \geq 0$, while $w\big(\sum_{n=1}^{d_i} a_{i,n}z_i^n\big) = -d_i$. Thus $w(0) = w\big(a_0 + \sum_{j\in I}\sum_{n=1}^{d_j} a_{j,n}z_j^n\big) = -d_i$, a contradiction. □

Proposition 1.10. *Suppose* $J, J' \subseteq I$. *Then* $D_J \cap D_{J'} = D_{J\cap J'}$.

Proof. Clearly, $D_{J\cap J'} \subseteq D_J \cap D_{J'}$. For the converse inclusion, we distinguish between two cases. First suppose that $J \cap J' \neq \varnothing$, and fix $j \in J\cap J'$. Then $D_J = K[z_k \mid k \in J][\![X - c_jY]\!]$ and $D_{J'} = K[z_k \mid k \in J'][\![X - c_jY]\!]$; hence $D_J \cap D_{J'} = (K[z_k \mid k \in J] \cap K[z_k \mid k \in J'])[\![X - c_jY]\!]$. By Lemma 1.9 $K[z_k \mid k \in J] \cap K[z_k \mid k \in J'] = K[z_k \mid k \in J\cap J']$; hence $y \in D_{J\cap J'}$.

Now suppose that $J \cap J' = \varnothing$ and let $R = K[\![X,Y]\!] = D_\varnothing$, $R_1 = D_J$, $R_2 = D_{J'}$, and $R_0 = D_{J\cup J'}$. Since $v(f) \geq 0$ for each $f \in K[z_j \mid j \in J\cup J'][X,Y]$, we also have $v(f) \geq 0$ for each f in the completion R_0. The ring R is complete with respect to v, and $R = R_1 + R_2$ by Proposition 1.7. Let $\mathfrak{p}$, $\mathfrak{p}_1$, $\mathfrak{p}_2$, and $\mathfrak{p}_0$ be the centers of v at R, R_1, R_2, and R_0, respectively. Then $\mathfrak{p}$ is generated by X and Y, and $\mathfrak{p}_0$ is generated by $X - c_jY$ for any $j \in J$, by Proposition 1.5. It follows that $\mathfrak{p}R_0 = \mathfrak{p}_0$. In order to apply Lemma 1.8, it remains to check that $R_1/\mathfrak{p}_1 \cap R_2/\mathfrak{p}_2 = R/\mathfrak{p}$ in $R_0/\mathfrak{p}_0$. Indeed, we have $R_1/\mathfrak{p}_1 = K[z_j \mid j \in J]$, $R_2/\mathfrak{p}_2 = K[z_j \mid j \in J']$, and $R_0/\mathfrak{p}_0 = K$. By Lemma 1.9, we are done. □

Proposition 1.11. *For each $i \in I$, $Q'_i \cap Q_i = E$.*

Proof. Since $Q'_i = \bigcap_{j \neq i} Q_j$, the assertion is that $\bigcap_{j \in I} Q_j = E$. Indeed, let $y \in \bigcap_{j \in I} Q_j$. For each $j \in J$ write $y = f_j/q_j$ with $f_j \in D_{I \smallsetminus \{j\}}$, $q_j \in K[\![X, Y]\!]$. Taking a common denominator we may assume that q_j is independent of j, and denote $q = q_j$ (for all $j \in J$). It suffices to prove that $qy \in K[\![X, Y]\!] \subseteq E$. But $qy = q_j y = f_j \in D_{I \smallsetminus \{j\}}$ for all $j \in I$; hence, by Proposition 1.10,

$$qy \in \bigcap_{j \in I} D_{I \smallsetminus \{j\}} = D_\varnothing = K[\![X, Y]\!]. \qquad \square$$

The next remark gives a rigid-geometric and a formal-geometric interpretation of the rings D_J.

Remark 1.12. Let $J \subseteq I$, $j \in J$, and $t = X - c_j Y$. By Proposition 1.5, $D_J = K[z_l \mid l \in J][\![t]\!]$ is the t-adic completion of $K[z_l \mid l \in J][t]$; thus $D_J[t^{-1}]$ is the t-adic completion of $K[z_l \mid l \in J][t, t^{-1}]$. We have $K[z_l \mid l \in J][t, t^{-1}] \subseteq K((t))[z_l \mid l \in J] \subseteq D_J[t^{-1}]$; hence $D_J[t^{-1}]$ is the t-adic completion of $A := K((t))[z_l \mid l \in J]$. Let $T = K[\![t]\!]$, $F = K((t))$, and $s = X/Y$. Then s is a free variable over F. Let v_t be the t-adic valuation on F, and extend it to $F(s)$ by $v_t(s) = 0$. Note that $z_k = 1/(s - c_k)$ and $v_t(c_l - c_k) = 0$ for all distinct $l, k \in J$. By [Haran and Jarden 1998, Lemma 3.1(c)] (with w_k, K, and x there replaced by z_k, F, and s here), each element $0 \neq f \in A$ can be uniquely written as

$$f = f_0 + \sum_{k \in J} \sum_{n=1}^{\infty} f_{kn} z_k^n, \tag{1}$$

where $f_0, f_{kn} \in F$ are almost all zero. Uniqueness in the presentation (1) implies that $v_t(f) = \min_{kn}\{v_t(f_0), v_t(f_{kn})\}$.

By [Haran and Jarden 1998, Lemma 3.3] the completion $D_J[t^{-1}]$ of A is the ring of holomorphic functions on the affinoid $U = \mathbb{P} \smallsetminus \big(\bigcup_{l \in J} B(c_l)\big)$, where $\mathbb{P}$ is the projective s-line and $B(c_l)$ is a disc of radius 1 with center c_l for each $l \in J$ [Fresnel and van der Put 2004, §2.2]. Moreover, each element $f \in D_J[t^{-1}]$ can be uniquely presented as in (1), where $f_0 \in F$ and $\{f_{ln}\}_{n=1}^{\infty}$ is a null sequence in F (with respect to v_t) for each $l \in J$. Thus, in the rigid-geometric language, D_J is the ring of holomorphic functions on U having no pole at t. Its elements are of the form (1), where the coefficients are now in T (and $\{f_{kn}\}_{n=1}^{\infty}$ is a null sequence for each $k \in J$). In particular, $T[z_l \mid l \in J]$ is dense in D_J.

Let $\hat{X}$ be the projective s-line over T, and let X be its closed fibre. Put

$$U = X \smallsetminus \{c_l \mid l \in J\}.$$

Then $R_U = T[1/(s - c_l) \mid l \in J] = T[z_l \mid l \in J]$ is the set of functions on $\hat{X}$ which are regular on U. Since R_U is t-adically dense in $D_J = K[1/(s - c_l) \mid l \in J][\![t]\!]$,

D_J is the t-adic completion of R_U. In formal-geometric language, this means that $D_J = \hat{R}_U$ is the ring of regular functions on the t-adic thickening of U [Harbater and Hartmann 2007, Notation 4.3].

Corollary 1.13. *Let $J \subseteq I$, $j \in J$.*

(a) *For each $0 \neq g \in D_J$,*

$$D_J[(X - c_j Y)^{-1}] = K((X - c_j Y))[z_k \mid k \in J] + g D_J[(X - c_j Y)^{-1}].$$

(b) *For each $f \in D_J$ there are $h \in K[\![X - c_j Y]\!][z_j]$ and $u \in D_J^\times$ such that $f = hu$.*

(c) *The ring Q_j is a field.*

Proof. In the notation of Remark 1.12, each element $f \in D_J[t^{-1}]$ can be written in the form $u \cdot h$ with $u \in D_J[t^{-1}]^\times$ and $h \in F[z_j]$, by [Haran and Jarden 1998, Lemma 3.7]. If $f \in D_J$ then we can multiply u and h by a power of t to assume that $u \in D_J^\times$ and $h \in K[\![t]\!][z_j]$. This proves (b). Part (a) is given by [Haran and Jarden 1998, Corollary 3.8]. By [Harbater and Hartmann 2007, Corollary 4.8] (now viewing the rings D_J in the formal-geometric context) Quot D_J is the compositum of $K((t))(X/Y)$ and D_J. Since $K((t))(X/Y) \subseteq E$, we have $\operatorname{Quot} D_J = E D_J$. Applying this argument for $J \smallsetminus \{j\}$ instead of J we have $Q_j = \operatorname{Quot}(D_{I \smallsetminus \{j\}}) = E D_{I \smallsetminus \{j\}}$; hence $\operatorname{Quot} Q_j = \operatorname{Quot}(E D_{I \smallsetminus \{j\}}) = E \operatorname{Quot}(D_{I \smallsetminus \{j\}}) = E D_{I \smallsetminus \{j\}} = Q_j$ is a field. □

The proof of the following proposition is based on that of Corollary 4.4 of [Haran and Jarden 1998]. (We cannot use that corollary as it is, since condition (e′) of that claim does not hold for D_I itself.)

Proposition 1.14. *Let $i \in I$, $n \in \mathbb{N}$, and $b \in \mathrm{GL}_n(Q)$. There exist $b_1 \in \mathrm{GL}_n(Q_i)$ and $b_2 \in \mathrm{GL}_n(Q_i')$ such that $b = b_1 \cdot b_2$.*

Proof. Denote by $|\cdot|$ the absolute value on Q that corresponds to v. Each of the rings $A = D_I$, $A_1 = D_{I \smallsetminus \{i\}}$, and $A_2 = D_{\{i\}}$ is complete with respect to $|\cdot|$ and Proposition 1.7 asserts that condition (d′) of Example 4.3 of [Haran and Jarden 1998] holds for these rings. We extend $|\cdot|$ to the maximum norm $\|\cdot\|$ on $M_n(Q)$, as in the same example. Then $M_n(A)$, $M_n(A_1)$, and $M_n(A_2)$ are complete with respect to $\|\cdot\|$ and condition (d) of [Haran and Jarden 1998, §4] holds. By Cartan's lemma [Haran and Jarden 1998, Lemma 4.2], for each $a \in \mathrm{GL}_n(A)$ with $\|a - 1\| < 1$ there exist $a_1 \in \mathrm{GL}_1(A_1)$ and $a_2 \in \mathrm{GL}_1(A_2)$ such that $a = a_1 \cdot a_2$.

Let $E_1 = \operatorname{Quot} A_1 = Q_i$ and $E_2 = \operatorname{Quot} A_2 = Q_i'$. In order to factor b (which need not be in $\mathrm{GL}_n(A)$), let $t = X - c_i Y$, $T = k[\![t]\!]$. By Remark 1.12 (for $J = I$) $A_0 = T[z_k \mid k \in I]$ is a dense subring of A, and by Corollary 1.13(b) there exists $h \in A_0$ such that $hb \in \mathrm{M}_n(A)$. If $hb = b_1 b_2'$ with $b_1 \in \mathrm{GL}_n(E_1)$ and $b_2' \in \mathrm{GL}_n(E_2)$, then $b = b_1 b_2$ with $b_2 = b_2'/h \in \mathrm{GL}_n(E_2)$. So we may assume that $b \in \mathrm{M}_n(A)$.

Let $0 \neq d = \det(b) \in A$. By Corollary 1.13(b) there are $0 \neq g \in A_0$ and $u \in A^\times$ such that $d = gu$. Let $b'' \in \mathrm{M}_n(A)$ be the adjoint matrix of b, so that $bb'' = d1$. Let $b' = u^{-1}b''$. Then $b' \in \mathrm{M}_n(A)$ and $bb' = g1$. Put

$$V = \{a' \in \mathrm{M}_n(A[t^{-1}]) \mid ba' \in gM_n(A[t^{-1}])\}, \qquad V_0 = V \cap M_n(A_0[t^{-1}]).$$

V is an additive subgroup of $M_n(A[t^{-1}])$ and $gM_n(A[t^{-1}]) \leq V$. By Corollary 1.13(a), $M_n(A[t^{-1}]) = M_n(A_0[t^{-1}]) + gM_n(A[t^{-1}])$; hence $V = V_0 + gM_n(A[t^{-1}])$.

Since A_0 is dense in A, $gM_n(A_0[t^{-1}])$ is dense in $gM_n(A[t^{-1}])$. It follows that $V_0 = V_0 + gM_n(A_0[t^{-1}])$ is dense in $V = V_0 + gM_n(A[t^{-1}])$. As $b' \in V$, there exists $a_0 \in V_0$ such that $\|b' - a_0\| < |g|/\|b\|$. Put $a = a_0/g \in M_n(Q)$. Then $ba \in M_n(A[t^{-1}])$ and $\|1 - ba\| = \|(1/g)b(b' - a_0)\| \leq (1/|g|)\|b\| \cdot \|b' - a_0\| < 1$. Hence $\|ba\| = 1$, so each entry in ba has a nonnegative value at v. By Remark 1.12, v coincides with the t-adic valuation on A; hence all the entries of ba belong to A. Thus $ba \in M_n(A)$, and since $\|1 - ba\| < 1$ and $M_n(A)$ is complete, $ba \in \mathrm{GL}_n(A)$. In particular, $\det a \neq 0$ and hence $a \in \mathrm{GL}_n(\mathrm{Quot}\, A_0) \subseteq \mathrm{GL}_n(E_2)$. By the first paragraph of this proof, there exist $b_1 \in \mathrm{GL}_n(A_1) \subseteq \mathrm{GL}_n(E_1)$ and $b_2' \in \mathrm{GL}_n(A_2)$ such that $ba = b_1 b_2'$. Then $b_2 = b_2' a^{-1} \in \mathrm{GL}_n(E_2)$ satisfies $b = b_1 b_2$. □

Corollary 1.15. *Suppose G is a finite group. For each $i \in I$ let F_i be a Galois extension of E with group G_i contained in G, such that $F_i \subseteq Q_i'$. If $G = \langle G_i \mid i \in I\rangle$ then $\mathscr{E} = (E, F_i, Q_i, Q; G_i, G)_{i \in I}$ is a patching datum* [Haran and Jarden 1998, Definition 1.1]. *In particular, G occurs as a Galois group over E.*

Proof. By Corollary 1.13(c), Q_i is a field for each $i \in I$. Conditions (2a), (2b), and (2d) of [Haran and Jarden 1998, Definition 1.1] are given in the hypothesis. Conditions (2c) and (2e) are given by Propositions 1.11 and 1.14, respectively. Thus $\mathscr{E}$ is a patching datum. By [Haran and Jarden 1998, Lemma 1.3(a)], there exists a Galois extension F of E with group G. □

2. *p*-groups

Fix the notation of Section 1, including that of Construction 1.1, and let p denote a prime number. In this section we realize p-groups of rank at most 2 by adequate extensions of E, and embed these extensions into the analytic fields.

Lemma 2.1. *Let $J \subseteq I$, $j \in J$, and $t = X - c_j Y$.*

(a) *Suppose $f = \sum_{l=0}^{d} f_l z_j^l \in K[\![t]\!][z_j]$ is a polynomial such that $v(f_1) = 0$ and $v(f_l) > 0$ for each $l > 1$. Then f is prime in $D_J[t^{-1}]$.*

(b) *The ring $D_J[t^{-1}]$ is a unique factorization domain.*

(c) *For each $a, b, c \in K^\times$ with $a \neq -b$ and $2 \leq m \in \mathbb{N}$, the elements $1 + az_j + t^{m-1}z_j^m$, $1 + bz_j - t^{m-1}z_j^m$, $1 + cz_j$ are nonassociate primes of $D_J[t^{-1}]$.*

Proof. Set $F = K((t))$. Then $D_J[t^{-1}] = F\{z_k \mid k \in J\}$ (see Remark 1.12). Viewed as an element of $F\{z_j\}$, f is regular of pseudodegree 1; see Definition 1.4 of [Haran and Völklein 1996]. Hence, by Corollary 1.7 of the same work, we have $f = u \cdot q$, where $u \in F\{z_j\}^\times \subseteq D_J[t^{-1}]^\times$ and $q = q_0 + z_j \in F[z_j]$ is a linear polynomial with $v(q_0) \geq 0$. Thus to prove (a), it suffices to show that q is prime in $D_J[t^{-1}]$. Without loss of generality $q_0 \neq 0$, and we set $c = c_j - 1/q_0$. Then $q = z_j - 1/(c - c_j)$; hence, by [Paran 2008, Lemma 6.4(a)] (with D, r, and 1 there replaced by F, 1, and j here), q generates the kernel of an epimorphism from $D_J[t^{-1}]$ onto a domain (actually a field here); hence q is prime. This proves (a).

Since $D_J[t^{-1}]$ is a principal ideal domain by [Haran and Jarden 1998, Proposition 3.9], part (b) follows.

By part (a), $r = 1 + az_j + t^{m-1}z_j^m$, $r' = 1 + bz_j - t^{m-1}z_j^m$, and $s = 1 + cz_j$ are primes of $D_J[t^{-1}]$. If $s \mid r$, then $-1/c$ is a root of r, a contradiction. Thus r and s (and similarly, r' and s) are nonassociates.

If $r \mid r'$ then $r \mid r + r'$. By the argument of the preceding paragraph, $r + r' = 2 + (a+b)z_j$ is a prime, nonassociate to r, a contradiction. This proves (c). □

Lemma 2.2. *Let K be a field that contains a primitive q-th root of unity, for some $q \in \mathbb{N}$. Let v be a discrete valuation on K which is trivial on the prime field of K, and let $a \in K$ with $v(a) = 0$. Suppose $L = K(a^{1/q})$ is a Kummer extension of K, and that L/K is unramified at v. Then $v(x^\sigma) = v(x)$ for each $x \in L$ and $\sigma \in \mathrm{Gal}(L/K)$.*

Proof. Extend v arbitrarily to L, let O be the valuation ring of v in K, and O' the valuation ring of v in L. Since K contains a primitive q-th root of unity, q is not divisible by $p = \operatorname{char} K$. Thus $d = \operatorname{disc}(T^q - a, K) = ka^{q-1}$, where $k \in \mathbb{Z}$ is not divisible by p. Hence $v(d) = 0$, and by [Fried and Jarden 2005, Lemma 6.1.2] we have $O' = O[a^{1/q}]$. Put $\alpha = a^{1/q}$ and let $x = \sum_{i=0}^{q-1} b_i \alpha^n \in K$, with $b_0, \ldots, b_{q-1} \in K$. We claim that $v(x) = \min_i v(b_i)$. Indeed, since L/K is unramified at v, we may multiply x by a power of a uniformizer of v in K, to assume that $v(x) = 0$. Since $O' = O[\alpha]$, $v(b_i) \geq 0$ for each $0 \leq i \leq q-1$. On the other hand $v(x) \geq \min_i v(b_i \alpha^i) = \min_i v(b_i)$, since $v(\alpha) = (1/n)v(a) = 0$. Thus $v(b_i) = 0$ for some $0 \leq i \leq q-1$; hence $v(x) = \min_i v(b_i)$.

Now, let $\sigma \in \mathrm{Gal}(L/K)$ and let $x = \sum_{i=0}^{q-1} b_i \alpha^n \in K$, with $b_0, \ldots, b_{q-1} \in K$, be an arbitrary element. We have $\alpha^\sigma = \zeta\alpha$, where ζ is some q-th root of unity. Then $v(x^\sigma) = v\left(\sum_{i=0}^{q-1} b_i \zeta^i \alpha^i\right) = \min_i v(b_i \zeta^i) = \min_i v(b_i) = v(x)$. □

Recall that given a field K, any K-central simple algebra A is of the form $\mathrm{M}_n(D)$ for some K-division algebra D. The index of A is defined to be $\operatorname{ind} A = \sqrt{\dim_K D}$. So, A is a division algebra if and only if $\operatorname{ind} A = \sqrt{\dim_K A}$. Let us denote Brauer equivalence by $\sim$ and the exponent of A (its order in the Brauer group) by $\exp A$. A subfield F of A is a maximal subfield of A if and only if $\dim_K A = [F : K]^2$.

Furthermore, a field F is a maximal subfield of A if and only if $\dim_K A = [F:K]^2$ and F splits A [Reiner 1975, Theorem 28.4 and Corollary 28.11].

The proof of the next proposition is partially based on that of [Harbater et al. 2009, Proposition 4.4].

Proposition 2.3. *Fix $i \in I$, and let H be an abelian p-group of rank at most 2, where $p \neq \operatorname{char} K$. Suppose K contains an $|H|$-th primitive root of unity. Let E' be a finite extension of E. Then there exists an H-Galois extension F_i/E such that $F_i \subseteq Q_i'$, F_i is contained as a maximal subfield in an E-division algebra D_i', and $D_i' \otimes_E E'Q_i$ remains a division algebra (where $E'Q_i$ is the compositum of E' and Q_i in an algebraic closure of Q).*

Proof. Let us start by constructing F_i. Write $H = C_q \times C_{q'}$, where q and q' are powers of p. For each $k \in \mathbb{N}$, the elements $X - c_iY + Y^k$ and $X + c_iY - Y^k$ are irreducible and hence prime in the unique factorization domain $K[\![X, Y]\!]$. Only finitely many primes of $K[\![X, Y]\!]$ are ramified at E'/E; hence for a sufficiently large $2 \leq k \in \mathbb{N}$, $f = X - c_iY + Y^k$, and $g = X + c_iY - Y^k$ are unramified at E'/E. That is, the corresponding valuations v_f and v_g are unramified. Let $a = f/(X - c_iY)$ and $b = g/(X - c_iY)$. Clearly $v_f(X - c_iY) = v_f(g) = 0$; hence $v_f(a) = 1$ and $v_f(b) = 0$. Similarly, $v_g(a) = 0$ and $v_g(b) = 1$. Consider the polynomial $h(U) = U^q - a$ over $D_{\{i\}} = K[z_i][\![X - c_iY]\!]$. Note that $a = 1 + z_i^k(X - c_iY)^{k-1}$, hence $h(1) \in (X - c_iY)D_{\{i\}}$ and $h'(1) = q \in K^\times \subseteq D_{\{i\}}^\times$. By the ring version of Hensel's lemma (for the ideal $(X - c_iY)D_{\{i\}}$) $h(U)$ has a root $s \in D_{\{i\}}$. Note that $v_f(s) = 1/q \notin \mathbb{Z}$; hence $s \notin E$. Since K contains a primitive $|H|$-th root of unity, it contains a primitive q-th root of unity. By Kummer theory $E(s)/E$ is a Galois extension with group C_q. Similarly, there exists $s' \in D_{\{i\}}$ satisfying $(s')^{q'} = b$, and $E(s')/E$ is Galois with group $C_{q'}$. Let $F_i = E(s, s') \subseteq Q_i'$.

Since $v_f(a) = 1$, $h(U)$ is irreducible over E, by Eisenstein's criterion. Denoting the reduction modulo g by $\bar{\cdot}$, $\bar{h}(U) = U^q - \bar{a}$ is separable, since $\bar{a} \neq 0$. Thus by [Fried and Jarden 2005, Lemma 2.3.4], $E(s)/E$ is unramified at v_g. Clearly, $E(s')/E$ is totally ramified at v_g. Thus $E(s)$ and $E(s')$ are linearly disjoint over E; hence $\operatorname{Gal}(F_i/E) = H$.

Let D_i' be the quaternion algebra $(a, b)_{qq'}$ [Pierce 1982, Section 15.4]. Note that D_i' can be also viewed as the cyclic algebra $(E(a^{1/qq'})/E, \sigma, b)$, for some generator σ of $\operatorname{Gal}(E(a^{1/qq'})/E)$. We claim that F_i splits D_i'. By [Reiner 1975, Theorem 30.8], we have

$$D_i' \otimes_E E(s) \sim (E(s^{1/q'})/E(s), \sigma^q, b);$$

thus $D_i' \otimes_E F_i \sim (F_i(s^{1/q'})/F_i, \sigma^q, b)$. The cyclic algebra $(F_i(s^{1/q'})/F_i, \sigma^q, b)$ is split if and only if b is a norm from $F_i(s^{1/q'})$ (see for example [Reiner 1975, Theorem 30.4]), i.e., if and only if $b \in N_{F_i(s^{1/q'})/F_i}(F_i(s^{1/q'}))$. This holds since

$b = N_{F_i(s^{1/q'})/F_i}(s')$. Thus F_i splits D_i'. As $[F_i : E] = qq'$, F_i is a maximal subfield of the E-central simple algebra D_i'. We shall show that $D_i' \otimes_E E'Q_i$ is a division algebra and from this it will follow that D_i' is a division algebra.

In order to show that $D_i' \otimes_E E'Q_i$ is a division algebra we construct auxiliary valuations. Choose $j \in I \smallsetminus \{i\}$, and let

$$t = X - c_j Y, \quad r = 1 + (c_j + c_i)z_j - t^{k-1}z_j^k, \quad r' = 1 + (c_j - c_i)z_j + t^{k-1}z_j^k.$$

By Lemma 2.1(c) r and r' are nonassociate prime elements in $D_{I\smallsetminus\{i\}}[t^{-1}]$, so they define discrete valuations v_r and $v_{r'}$ on $Q_i = \mathrm{Quot}(D_{I\smallsetminus\{i\}}) = \mathrm{Quot}(D_{I\smallsetminus\{i\}}[t^{-1}])$ such that $v_r(r') = v_{r'}(r) = 0$. By Lemma 2.1(c) we also have

$$v_{r'}(1 + (c_j - c_i)z_j) = v_r(1 + (c_j - c_i)z_j) = 0.$$

Note that

$$b = \frac{X - c_jY + (c_j + c_i)Y - Y^k}{X - c_jY + (c_j - c_i)Y} = \frac{t + (c_j + c_i)tz_j - t^j z_j^k}{t + (c_j - c_i)tz_j} = \frac{r}{1 + (c_j - c_i)z_j}.$$

Similarly, $a = r'/(1 + (c_j - c_i)z_j)$. Thus $v_r(b) = 1$, $v_{r'}(b) = 0$, $v_r(a) = 0$, and $v_{r'}(a) = 1$. Then the polynomial $U^{qq'} - a$ is irreducible over $D_{I\smallsetminus\{i\}}$, by Eisenstein's criterion (using $v_{r'}$). Thus $Q_i(a^{1/qq'})/Q_i$ is unramified at v_r (again by [Fried and Jarden 2005, Lemma 2.3.4]); hence so is $E'Q_i(a^{1/qq'})/E'Q_i$.

Only finitely many primes of the unique factorization domain $D_{I\smallsetminus\{i\}}[t^{-1}]$ (see Lemma 2.1(b)) are ramified at the finite extension $E'Q_i/Q_i$; hence, without loss of generality, we may assume that $E'Q_i/Q_i$ is unramified at $v_{r'}$ (by possibly choosing an even larger k beforehand). On the other hand, $Q_i(a^{1/qq'})/Q_i$ is totally ramified at $v_{r'}$; hence

$$[E'Q_i(a^{1/qq'}) : E'Q_i] = [Q_i(a^{1/qq'}) : Q_i] = qq'.$$

We can now show that $D_i' \otimes_E E'Q_i$ is a division algebra. A sufficient condition for this to hold is that $\exp(D_i' \otimes_E E'Q_i) = qq'$. This happens if and only if for every $1 \le m \le qq' - 1$ the algebra $(E'Q_i(a^{1/qq'})/E'Q_i, \sigma, b^m) \sim (D_i' \otimes_E E'Q_i)^m$ does not split. Let N denote the norm $N_{E'Q_i(a^{1/qq'})/E'Q_i}$. For any $1 \le m \le qq' - 1$, the algebra $(D_i' \otimes_E E'Q_i)^m$ splits if and only if $b^m \in N(E'Q_i(a^{1/qq'})^\times)$ [Reiner 1975, Theorem 30.4].

Since $E'Q_i(a^{1/qq'})/E'Q_i$ is unramified at v_r, we have $v_r(x) = v_r(x^\sigma)$ for each $x \in E'Q_i(a^{1/qq'})$, by Lemma 2.2. Hence

$$v_r(N(x)) = \sum_{l=0}^{qq'-1} v_r(x^{\sigma^l}) = qq' v_r(x)$$

for all $x \in E'Q_i(a^{1/qq'})$. Since $v_r(b) = 1$, $b^m \notin N(E'Q_i(a^{1/qq'})^\times)$ for all $1 \le m \le qq' - 1$. Thus, $\exp(D_i' \otimes_E E'Q_i) = qq'$ and $D_i' \otimes_E E'Q_i$ is a division algebra. $\square$

3. Patching and admissibility

We have established the patching machinery needed to prove our Main Theorem. We first recall some general properties of induced algebras and Frobenius algebras.

Remark 3.1 (induced algebras). Let G be a finite group and $H \leq G$. Let P/Q be a finite field extension with $H = \mathrm{Gal}(P/Q)$. Let

$$N = \mathrm{Ind}_H^G P = \Big\{ \sum_{\theta \in G} a_\theta \theta \mid a_\theta \in P,\ a_\theta^\tau = a_{\theta\tau} \text{ for all } \theta \in G,\ \tau \in H \Big\}$$

be a ring with respect to point-wise addition and multiplication. Then P can be embedded as a subring of N by choosing representatives $\theta_1 = 1, \dots, \theta_k$ of $H \backslash G$ and sending an element $x \in P$ to $\sum_{i=1, \tau \in H}^{k} x^\tau \theta_i \tau$. Furthermore, by choosing different representatives N can be presented as a direct sum of copies of P.

If P splits a central simple Q-algebra A then

$$\mathrm{Ind}_H^G P \otimes_Q A \cong \mathrm{Ind}_H^G P \otimes_P (P \otimes_Q A) \sim \mathrm{Ind}_H^G P \otimes_P P \cong \mathrm{Ind}_H^G P;$$

hence A splits over $\mathrm{Ind}_H^G P$. For a definition of a split separable (Azumaya) algebra over a ring, see [DeMeyer and Ingraham 1971, §5].

The next definition, remark, and proposition appear in Section 2.1 of [Jacobson 1996].

Definition 3.2 (Frobenius algebras). Let F be a field. An F-algebra A is a Frobenius algebra if A contains a hyperplane that does not contain any nonzero one sided ideal of A.

Remark 3.3. An algebra $A_1 \oplus \cdots \oplus A_s$ is Frobenius if and only if A_i is Frobenius for each $1 \leq i \leq s$. Any algebra $F[a]$ (with a single generator) is Frobenius. Let L/K be an H-extension. By Remark 3.1, $\mathrm{Ind}_H^G L$ can be decomposed into a sum of copies of L and it follows that $\mathrm{Ind}_H^G L$ is a Frobenius algebra.

Proposition 3.4 [Jacobson 1996, Theorem 2.2.3]. *Let A be an F-central simple algebra and K a commutative Frobenius subalgebra of A such that* $\dim_F A = [K : F]^2$. *Then any embedding of K into A can be extended to an inner automorphism of A.*

Lemma 3.5. *Let R be an equicharacteristic complete local domain of dimension r. Suppose that the residue field of R contains a primitive root of unity of order k, for each $k \in \mathbb{N}$ with* $\mathrm{char}\, R \nmid k$. *Then R is a finite module over a subring of the form $K[\![X_1, \dots, X_r]\!]$, where K is a field containing a primitive root of unity of order k, for each $k \in \mathbb{N}$ with* $\mathrm{char}\, K \nmid k$.

Proof. By Cohen's structure theorem [Matsumura 1986, §29], R is finitely generated over a subring of the form $B = K_0[\![X_1, \dots, X_n]\!]$, for some field K_0. Since $\dim B = \dim R = r$, we have $n = r$.

Let K be the field obtained by adjoining all primitive roots of unity of order not divisible by char R to K_0. By our assumptions, K is contained in the residue field of R; hence K/K_0 is a finite (separable) extension. By Hensel's lemma, R contains K. Since $[K : K_0]$ is finite, $K(K_0[\![X_1, \dots, X_r]\!]) = K[\![X_1, \dots, X_r]\!]$. Thus R is finite over $K[\![X_1, \dots, X_r]\!]$ (since it is finite over $K_0[\![X_1, \dots, X_r]\!]$). □

The final ingredient we need in order to prove our Main Theorem is patching of central simple algebras. The content of the next proposition is essentially given in [Harbater et al. 2009], but for specific fields Q_i, while here we present it for general fields satisfying a matrix factorization property. We note that [Harbater et al. 2009, Theorem 4.1] uses the terminology of categories and equivalence of categories. Here we prefer a more explicit presentation, working with vector spaces and bases, as in [Haran and Jarden 1998]. The proof of the proposition combines the proof of [Haran and Jarden 1998, Lemma 1.2] (where a more restricted assertion is made for specific types of algebras), and the proof of [Harbater and Hartmann 2007, Theorem 7.1(vi)] (where the assertion is made for specific types of fields).

Proposition 3.6. *Let I be a finite set. For each $i \in I$ let Q_i be a field contained in a Q_i-algebra A_i. Let Q be a field containing Q_i for each $i \in I$, and contained in a Q-algebra A_Q which contains A_i for each $i \in I$. Moreover, suppose that $A_i Q = A_Q$ and $\dim_{Q_i} A_i = \dim_Q A_Q$ for each $i \in I$. Finally, suppose that*

(∗) *for each $B \in \mathrm{GL}_n(Q)$ there exist $B_i \in \mathrm{GL}_n(Q_i)$ and $B_i' \in \mathrm{GL}_n\big(\bigcap_{j \neq i} Q_j\big)$ such that $B = B_i B_i'$.*

Then, letting $E = \bigcap_{i \in I} Q_i$, $A = \bigcap_{i \in I} A_i$ is an E-algebra satisfying $AQ_i = A_i$ for each $i \in I$. Moreover, if each A_i is central simple, then so is A.

Proof. For each $i \in I$, let $\mathscr{C}_i$ be a basis for A_i over Q_i. Since $A_i Q = A_Q$, $\mathrm{Span}_Q(\mathscr{C}_i) = A_Q$, and since $\dim_{Q_i} A_i = \dim_Q A_Q$, $\mathscr{C}_i$ is a basis for A_Q over Q, for each $i \in I$. We now construct a basis $\mathscr{C}$ for A_Q over Q, which is also a basis for A_i over Q_i, for *all* $i \in I$.

For each subset J of I we find, by induction on $|J|$, a basis $\mathscr{V}_J$ for A_Q over Q which is also a basis for A_j over Q_j, for each $j \in J$. Then for $I = J$ we will get the basis $\mathscr{C}$.

If $J = \varnothing$ there is nothing to prove. Suppose that $|J| \geq 1$, choose $k \in J$ and let $J' = J \smallsetminus \{k\}$. By assumption there is a basis $\mathscr{V}_J$ for A_Q over Q which is a basis for A_i over Q_i for each $i \in J'$. Since $\mathscr{C}_i$ is a common basis for A_Q and A_i, there is a matrix $B \in \mathrm{GL}_n(Q)$ such that $\mathscr{C}_k B = \mathscr{V}_{J'}$. By condition (∗) in the statement of the proposition, there exist $B_k \in \mathrm{GL}_n(Q_k)$ and $M \in \mathrm{GL}_n\big(\bigcap_{k \neq j \in I} Q_j\big) \subseteq \bigcap_{j \in J'} \mathrm{GL}_n(Q_j)$ such that $B = B_k M$. Put $\mathscr{V}_J = \mathscr{V}_{J'} M^{-1}$. Then $\mathscr{V}_J$ is a basis for A_Q over Q which is also a basis for A_j over Q_j for each $j \in J'$. Moreover, $\mathscr{V}_J$ is also a basis for A_k over Q_k, since $\mathscr{V}_J = \mathscr{V}_k B M^{-1} = \mathscr{V}_k B_k$. This completes the induction.

The existence of the common basis C implies that $AQ_i = A_i$ for each $i \in I$. As A_i is a Q_i-central simple algebra for any $i \in I$ (a single i suffices), A is an E-central simple algebra. This follows, for example, from [Saltman 1999, Theorem 2.2.c]. □

Proposition 3.7. *Let R be an equicharacteristic complete local domain of dimension 2, with residue field containing a primitive root of unity of order k for each $k \in \mathbb{N}$ with* char $R \nmid k$. *Let G be a finite group of order not divisible by* char R, *whose Sylow subgroups are abelian of rank at most 2. Then G is admissible over* Quot R.

Proof. By Lemma 3.5, R is a finite module over a subring of the form $B = K[\![X, Y]\!]$, where K contains a primitive root of unity of order k for each $k \in \mathbb{N}$ not divisible by $p = \operatorname{char} R$. Let $E = \operatorname{Quot} B = K((X, Y))$ and $E' = \operatorname{Quot} R$. Then E' is a finite extension of E.

Part A: A patching datum. Let $(p_i)_{i\in I}$ be the prime factors of $n = |G|$, for some index set I. For each $i \in I$, let G_i be a p_i-Sylow subgroup of G. Apply Construction 1.1 to obtain rings Q_i, $i \in I$, contained in the common field Q. For each $i \in I$, we may apply Proposition 2.3 to obtain a Galois extension F_i/E with group G_i, such that $F_i \subseteq Q_i'$ and F_i is contained as a maximal subfield in a division E-algebra D_i'. Moreover, $D_i' \otimes_E E'Q_i$ remains a division algebra. Thus $D_i := D_i' \otimes_E Q_i$ is also a division algebra. By Corollary 1.13(c), Q_i is a field for each $i \in I$. Put $P_i = F_i Q_i$. Since F_i splits D_i', P_i splits D_i, and since $[P_i : Q_i] = [F_i : E] = \operatorname{ind}(D_i') = \operatorname{ind}(D_i)$, P_i is a maximal subfield of D_i. Let $\mathscr{E} = (E, F_i, Q_i, Q; G_i, G)_{i\in I}$. By Corollary 1.15, $\mathscr{E}$ is a patching datum.

Part B: Induced algebras [Haran and Jarden 1998, §1]. Consider the induced algebra $N = \operatorname{Ind}_1^G Q$ of dimension n over Q, and the Q_i-subalgebra $N_i = \operatorname{Ind}_{G_i}^G P_i$ for each $i \in I$ (see Remark 3.1). Then G acts on N by $\left(\sum_{\theta\in G} a_\theta \theta\right)^\sigma = \sum_{\theta\in G} a_\theta \sigma^{-1}\theta = \sum_{\theta\in G} a_{\sigma\theta}\theta$ for each $\sigma \in G$. The field Q is embedded diagonally in N, which induces an embedding of Q_i in N_i, for each $i \in I$. We view these embeddings as containments. By Lemma 1.2 of [Haran and Jarden 1998] there is a basis for N over Q, which is also a basis for N_i over Q_i, for each $i \in I$. In particular, we have $N_i Q = N$ for each $i \in I$. By Lemma 1.3 of the same paper, $F = \bigcap_{i\in I} N_i$ is a Galois field extension of E with group G, and there exists an E-embedding of F into Q. Denote the image of F under this embedding by F'.

Part C: Division algebras. It remains to prove that the extension F'/E is adequate. Let $A_Q = \mathrm{M}_n(Q)$, and for each $i \in I$ let $n_i = [G : G_i]$. As A_Q is split of dimension n^2 and N is of dimension n over Q, we also have an embedding of N into A_Q. We view N as a subalgebra of A_Q via this embedding.

Fix $i \in I$. Since $P_i = F_i Q_i$ splits D_i, it follows by Remark 3.1 that N_i also splits D_i. Moreover, by [DeMeyer and Ingraham 1971, Theorem 5.5] there is a central

simple Q_i-algebra A_i which is Brauer equivalent to D_i, in which N_i embeds as a maximal commutative separable Q_i-subalgebra so that $\dim_{Q_i}(A_i) = \dim_{Q_i}(N_i)^2 = n^2$. We view N_i as contained in A_i via this embedding.

Since P_i splits D_i, we have

$$D_i \otimes_{Q_i} Q \cong (D_i \otimes_{Q_i} P_i) \otimes_{P_i} Q \cong \mathrm{M}_{n/n_i}(P_i) \otimes_{P_i} Q \cong \mathrm{M}_{n/n_i}(Q).$$

Since $\mathrm{ind}(D_i) = n/n_i$ and $\dim_Q A_Q = n^2$ we get that $A_i \cong M_{n_i}(D_i)$. Thus we have $A_i \otimes_{Q_i} Q \cong M_{n_i}(D_i) \otimes_{Q_i} Q \cong M_n(Q) = A_Q$, and we denote the induced Q-algebras isomorphism $A_i \otimes_{Q_i} Q \to A_Q$ by ψ_i. We cannot identify these two algebras via this isomorphism, since it might not be compatible with the containments $N_i \subseteq A_i$ and $N \subseteq A_Q$. This compatibility problem can be settled similarly to [Harbater et al. 2009, Lemma 4.2]:

By Part B we have $N = N_i Q$ and $\dim_{Q_i} N_i = \dim_Q N = n$. Thus we have an isomorphism $\delta_i : N = N_i Q \to N_i \otimes_{Q_i} Q$ for which the following diagram commutes:

$$\begin{tikzcd} A_i \arrow[r, dash] \arrow[d, dash] & A_i \otimes_{Q_i} Q \\ N_i \arrow[r, dash] & N \arrow[r, "\delta_i"] & N_i \otimes_{Q_i} Q. \arrow[u, "\mathrm{id} \otimes_{Q_i} Q"'] \end{tikzcd} \tag{2}$$

By Remark 3.3, $N = \mathrm{Ind}_1^G Q$ is a Frobenius (commutative) subalgebra of A_Q. By Proposition 3.4, the embedding $\psi_i(\mathrm{id} \otimes_{Q_i} Q)\delta_i : N \to A_Q$ extends to an inner automorphism α_i of A_Q. Let $\psi_i' = \alpha_i^{-1}\psi_i$. Then $\alpha_i^{-1}\psi_i(\mathrm{id} \otimes_{Q_i} Q)\delta_i$ is the identity map on $N = N_i Q$, so we have the commutative diagram

$$\begin{tikzcd} A_i \otimes_{Q_i} Q \arrow[r, "\psi_i'"] & A_Q \\ N_i Q \arrow[u, "(\mathrm{id} \otimes_{Q_i} Q)\delta_i"] \arrow[r, equal] & N. \arrow[u] \end{tikzcd} \tag{3}$$

Combining (2) and (3), we get the following commutative diagram:

$$\begin{tikzcd} A_i \arrow[r, dash] \arrow[d, dash] & A_i \otimes_{Q_i} Q \arrow[r, "\psi_i'"] & A_Q \\ N_i \arrow[r, dash] & N_i Q \arrow[u, "(\mathrm{id} \otimes_{Q_i} Q)\delta_i"] \arrow[r, equal] & N. \arrow[u] \end{tikzcd}$$

This diagram gives an embedding $A_i \to A_Q$ which is compatible with the containments $N_i \subseteq A_i$ and $N \subseteq A_Q$, so we may now identify A_i as a subring of A_Q, via this embedding. Moreover, since ψ_i' is an isomorphism, we have $A_i \otimes_{Q_i} Q = A_i Q = A_Q$

by this identification. The following diagram explains the containment relations:

$$\begin{tikzcd}
 & & A_i \arrow[rr, dash] \arrow[dd, dash] & & A_Q \\
N_i \arrow[urr, dash] \arrow[rrr, dash] \arrow[dd, dash] & & & N \arrow[ur, dash] \arrow[dd, dash] & \\
 & & D_i & & \\
Q_i \arrow[r, dash] \arrow[dd, dash] & P_i \arrow[ur, dash] \arrow[rr, dash] \arrow[dd, dash] & & Q \arrow[dd, dash] & \\
 & & D'_i & & \\
E \arrow[r, dash] & F_i \arrow[ur, dash] \arrow[rr, dash] & & Q'_i. &
\end{tikzcd}$$

Let $A = \bigcap_{i \in I} A_i$. By Proposition 3.6, A is a central simple E-algebra for which $AQ_i = A_i$ for each $i \in I$. In particular, $A = M_k(D)$ for some division algebra D of index n/k.

Now, $D \otimes_E Q_i$ is Brauer equivalent to $A \otimes_E Q_i \cong A_i$, which is Brauer equivalent to D_i. Thus, $n/n_i = \operatorname{ind}(D_i) \mid \operatorname{ind} D$ for each $i \in I$ and $n = \operatorname{lcm}_i(n/n_i) \mid \operatorname{ind} D$. It follows that $k = 1$ and A is a division algebra. Naturally, F is a subfield of A and $\operatorname{ind} A = [F : E]$. It follows that F is a maximal subfield of the division algebra A.

By choosing a basis for A/F and considering the corresponding structure constants one can form an E-division algebra A' which is E-isomorphic to A such that F' is a maximal subfield of A'.

We will show that $A' \otimes_E E'$ is an E'-division algebra, but first let us show that this implies that $F'E'/E'$ is an adequate G-extension (and hence G is E'-admissible). Indeed, if $A' \otimes_E E'$ is a division algebra, then $F' \otimes_E E'$ is a field. It follows that $F' \otimes_E E' \cong F'E'$, since $F' \otimes_E E'$ is G-Galois over E' [Saltman 1999, Theorem 6.3]. Thus, $[F'E' : E'] = [F' : E]$ and $F' \cap E' = E$. Since $F'E'$ splits $A' \otimes_E E'$ and as $\operatorname{ind}(A' \otimes_E E') = [F'E' : E']$, $F'E'$ is a maximal subfield of $A' \otimes_E E'$ and hence an adequate G-extension.

In order to show that $A' \otimes_E E'$ is an E'-division algebra, we first note that for each $i \in I$, $P_i = F_i Q_i = F'Q_i$, by [Haran and Völklein 1996, Lemma 3.6(b)]. Thus, we have the diagram

$$\begin{tikzcd}
 & A' \otimes_E Q_i \arrow[rr, dash] \arrow[dd, dash] & & A' \otimes_{E'} Q_i E' \arrow[dd, dash] \\
A' \arrow[ur, dash] \arrow[rr, dash] \arrow[dd, dash] & & A' \otimes_E E' \arrow[ur, dash] \arrow[dd, dash] & \\
 & P_i \arrow[rr, dash] \arrow[dd, dash] & & P_i E' \arrow[dd, dash] \\
F' \arrow[ur, dash] \arrow[rr, dash] \arrow[dd, dash] & & F'E' \arrow[ur, dash] \arrow[dd, dash] & \\
 & Q_i \arrow[rr, dash] & & Q_i E' \\
E \arrow[ur, dash] \arrow[rr, dash] & & E'. \arrow[ur, dash] &
\end{tikzcd}$$

As mentioned above, $A' \otimes_E Q_i$ is Brauer equivalent to $D_i = D'_i \otimes_E Q_i$. Thus $A' \otimes_E Q_i E'$ is Brauer equivalent to $D'_i \otimes_{Q_i} Q_i E'$, which by the choice of D'_i is a

division algebra. Then, for all $i \in I$,

$$\frac{n}{n_i} = \operatorname{ind}(D_i') = \operatorname{ind}(D_i' \otimes_{Q_i} Q_i E') \mid \operatorname{ind}(A' \otimes_E Q_i E') \mid \operatorname{ind}(A' \otimes_E E').$$

It follows that $n = \operatorname{lcm}_{i \in I}(n/n_i) \mid \operatorname{ind}(A' \otimes_E E')$. Hence $n = \operatorname{ind}(A' \otimes E')$, which shows that $A' \otimes_E E'$ is a division algebra. □

As a corollary, we get our Main Theorem, which we restate for convenience:

Theorem 3.8. *Let R be an equicharacteristic complete local domain of dimension* 2, *with a separably closed residue field. Let $E = \operatorname{Quot} R$ and let G be a finite group of order not divisible by* $\operatorname{char} E$. *Then G is E-admissible if and only if all the Sylow subgroups of G are abelian of rank at most* 2.

Proof. By Proposition 3.7, if the Sylow subgroups of G are abelian of rank at most 2 then G is E-admissible. For the converse, assume G is E-admissible. For a prime v of E, let ram_v denote the ramification map $\operatorname{ram}_v : \operatorname{Br} E \to \mathrm{H}^1(G_{E_v}, Q/Z)$ [Saltman 1999]. Following [Harbater et al. 2009], we say that an $\alpha \in \operatorname{Br} E$ is determined by ramification with respect to a set of primes Ω if there is a prime $v \in \Omega$ for which $\exp(\alpha) = \exp(\operatorname{ram}_v(\alpha))$. Let D be an E-division algebra with maximal subfield L that has Galois group $G = \operatorname{Gal}(L/E)$. Let $p = \operatorname{char} E$ (possibly $p = 0$). By [Harbater et al. 2009, Theorem 3.3], if D satisfies:

(1) the order of D is prime to p and $\operatorname{ind} D = \exp D$, and

(2) D is determined by ramification with respect to some set of discrete valuations whose residue characteristic is prime to $|G|$,

then G has Sylow subgroups that are abelian of rank at most 2. Condition (1) is satisfied for any α of order prime to p by Theorem 2.1 of [Colliot-Thélène et al. 2002], while condition (2) is satisfied by Corollary 1.9(c) of the same paper with respect to the set of codimension 1 primes of R. □

Remark 3.9. Let E be as above. By [Colliot-Thélène et al. 2002, Theorem 2.1], any Brauer class $\alpha \in \operatorname{Br} E$ of order prime to $\operatorname{char} E$ has $\operatorname{ind}(\alpha) = \exp(\alpha)$. Thus by [Schacher 1968, Proposition 2.2], a subfield of an E-division algebra is also a maximal subfield of some E-division algebra.

Acknowledgements

The authors wish to thank David Harbater for useful comments, and the referee, for many comments and corrections that significantly improved the presentation of this work.

References

[Bary-Soroker et al. 2008] L. Bary-Soroker, D. Haran, and D. Harbater, "Permanence criteria for semi-free profinite groups", preprint, 2008. To appear in *Math. Ann.* arXiv 0810.0845v2

[Colliot-Thélène et al. 2002] J.-L. Colliot-Thélène, M. Ojanguren, and R. Parimala, "Quadratic forms over fraction fields of two-dimensional Henselian rings and Brauer groups of related schemes", pp. 185–217 in *Algebra, arithmetic and geometry* (Mumbai, 2000), vol. 1, edited by R. Parimala, Tata Inst. Fund. Res. Stud. Math. **16**, Tata Inst. Fund. Res., Bombay, 2002. MR 2004c:14031 Zbl 1055.14019

[DeMeyer and Ingraham 1971] F. DeMeyer and E. Ingraham, *Separable algebras over commutative rings*, Lecture Notes in Math. **181**, Springer, Berlin, 1971. MR 43 #6199 Zbl 0215.36602

[Fein and Schacher 1995a] B. Fein and M. Schacher, "$\boldsymbol{Q}(t)$ and $\boldsymbol{Q}((t))$-admissibility of groups of odd order", *Proc. Amer. Math. Soc.* **123**:6 (1995), 1639–1645. MR 95g:12003 Zbl 0974.12003

[Fein and Schacher 1995b] B. Fein and M. Schacher, "Crossed products over algebraic function fields", *J. Algebra* **171**:2 (1995), 531–540. MR 96b:12003 Zbl 0827.12005

[Fein et al. 1992] B. Fein, D. J. Saltman, and M. Schacher, "Brauer–Hilbertian fields", *Trans. Amer. Math. Soc.* **334**:2 (1992), 915–928. MR 93b:12006 Zbl 0767.12003

[Feit 2004] W. Feit, "$\mathrm{PSL}_2(11)$ is admissible for all number fields", pp. 295–299 in *Algebra, arithmetic and geometry with applications* (West Lafayette, IN, 2000), edited by C. Christensen et al., Springer, Berlin, 2004. MR 2004k:12006 Zbl 1121.12004

[Fresnel and van der Put 2004] J. Fresnel and M. van der Put, *Rigid analytic geometry and its applications*, Progress in Mathematics **218**, Birkhäuser, Boston, MA, 2004. MR 2004i:14023 Zbl 1096.14014

[Fried and Jarden 2005] M. D. Fried and M. Jarden, *Field arithmetic*, 2nd ed., Ergebnisse der Math. (3) **11**, Springer, Berlin, 2005. MR 2005k:12003 Zbl 1055.12003

[Haran and Jarden 1998] D. Haran and M. Jarden, "Regular split embedding problems over complete valued fields", *Forum Math.* **10**:3 (1998), 329–351. MR 99e:12007 Zbl 0903.12003

[Haran and Völklein 1996] D. Haran and H. Völklein, "Galois groups over complete valued fields", *Israel J. Math.* **93** (1996), 9–27. MR 97c:12002 Zbl 0869.12006

[Harbater 1987] D. Harbater, "Galois coverings of the arithmetic line", pp. 165–195 in *Number theory* (New York, 1984–1985), edited by D. V. Chudnovsky et al., Lecture Notes in Math. **1240**, Springer, Berlin, 1987. MR 88h:14020 Zbl 0627.12015

[Harbater 2003] D. Harbater, "Patching and Galois theory", pp. 313–424 in *Galois groups and fundamental groups*, edited by L. Schneps, Math. Sci. Res. Inst. Publ. **41**, Cambridge Univ. Press, 2003. MR 2004j:14030 Zbl 1071.14029

[Harbater and Hartmann 2007] D. Harbater and J. Hartmann, "Patching over fields", preprint, 2007. To appear in *Israel Journal of Mathematics*. arXiv 0710.1392v3

[Harbater and Stevenson 2005] D. Harbater and K. F. Stevenson, "Local Galois theory in dimension two", *Adv. Math.* **198**:2 (2005), 623–653. MR 2007e:12002 Zbl 1104.12003

[Harbater et al. 2009] D. Harbater, J. Hartmann, and D. Krashen, "Patching subfields of division algebras", preprint, 2009. To appear in *Transactions of the AMS*. arXiv 0904.1594v2

[Jacobson 1996] N. Jacobson, *Finite-dimensional division algebras over fields*, Springer, Berlin, 1996. MR 98a:16024 Zbl 0874.16002

[Liedahl 1996] S. Liedahl, "K-admissibility of wreath products of cyclic p-groups", *J. Number Theory* **60**:2 (1996), 211–232. MR 97j:11059 Zbl 0861.12002

[Matsumura 1986] H. Matsumura, *Commutative ring theory*, Cambridge Studies in Advanced Math. **8**, Cambridge University Press, 1986. MR 88h:13001 Zbl 0603.13001

[Neftin 2009] D. Neftin, "Admissibility and realizability over number fields", preprint, 2009. arXiv 0904.3772v1

[Paran 2008] E. Paran, "Algebraic patching over complete domains", *Israel J. Math.* **166** (2008), 185–219. MR 2009e:12006 Zbl 1159.12002

[Paran 2010] E. Paran, "Galois theory over complete local domains", preprint, 2010, Available at www.tau.ac.il/~paranela/LocalGal.pdf. To appear in *Math. Annalen*.

[Pierce 1982] R. S. Pierce, *Associative algebras*, Graduate Texts in Math. **88**, Springer, New York, 1982. MR 84c:16001 Zbl 0497.16001

[Pop 1996] F. Pop, "Embedding problems over large fields", *Ann. of Math.* (2) **144**:1 (1996), 1–34. MR 97h:12013 Zbl 0862.12003

[Pop 2010] F. Pop, "Henselian implies large", preprint, 2010, Available at http://tinyurl.com/29dhoxl. To appear in *Ann. Math.*

[Reiner 1975] I. Reiner, *Maximal orders*, London Mathematical Society Monographs **5**, Academic Press, London, 1975. MR 52 #13910 Zbl 0305.16001

[Saltman 1999] D. J. Saltman, *Lectures on division algebras*, CBMS Regional Conference Series in Mathematics **94**, Amer. Math. Soc., Providence, RI, 1999. MR 2000f:16023 Zbl 0934.16013

[Schacher 1968] M. M. Schacher, "Subfields of division rings, I", *J. Algebra* **9** (1968), 451–477. MR 37 #2809 Zbl 0174.34103

[Schacher and Sonn 1992] M. Schacher and J. Sonn, "K-admissibility of A_6 and A_7", *J. Algebra* **145**:2 (1992), 333–338. MR 93a:12003 Zbl 0739.12004

[Sonn 1983] J. Sonn, "$\boldsymbol{Q}$-admissibility of solvable groups", *J. Algebra* **84**:2 (1983), 411–419. MR 86b:20022 Zbl 0524.16007

[Stern 1982] L. Stern, *Admissibility of finite groups over global fields*, Ph.D. thesis, Technion, Haifa, 1982.

Communicated by Jean-Louis Colliot-Thélène
Received 2009-10-09 Revised 2010-02-15 Accepted 2010-03-21

neftind@tx.technion.ac.il *Department of Mathematics, Technion – Institute of Technology, Haifa 32000, Israel*

paranela@post.tau.ac.il *Einstein Institute of Mathematics, Edmond J. Safra Campus, Givat Ram, The Hebrew University of Jerusalem, Jerusalem 91904, Israel*
http://www.tau.ac.il/~paranela/

ALGEBRA AND NUMBER THEORY 4:6(2010)

Minimisation and reduction of 2-, 3- and 4-coverings of elliptic curves

John E. Cremona, Tom A. Fisher and Michael Stoll

We consider models for genus-one curves of degree n for $n = 2$, 3 and 4, which arise in explicit n-descent on elliptic curves. We prove theorems on the existence of minimal models with the same invariants as the minimal model of the Jacobian elliptic curve and provide simple algorithms for minimising a given model, valid over general number fields. Finally, for genus-one models defined over $\mathbb{Q}$, we develop a theory of reduction and again give explicit algorithms for $n = 2$, 3 and 4.

1.	Introduction	763
2.	Genus-one models	765
3.	Minimisation theorems	770
4.	Minimisation algorithms	778
5.	Minimisation of insoluble genus-one models	793
6.	Reduction	801
7.	Examples	810
References		817

1. Introduction

Let E be an elliptic curve defined over a number field K. An n-descent on E computes the n-Selmer group $\mathrm{Sel}^{(n)}(K, E)$ of E, which parametrises the everywhere locally soluble n-coverings of E up to isomorphism. An *n-covering of E* is a principal homogeneous space C for E, together with a map $\pi : C \to E$ that fits into a commutative diagram

$$\begin{tikzcd} C \arrow[d, dashed, "\psi"'] \arrow[dr, "\pi"] & \\ E \arrow[r, "\cdot n"] & E \end{tikzcd}$$

where $\psi : C \to E$ is an isomorphism defined over the algebraic closure $\overline{K}$, compatible with the structure of C as a principal homogeneous space. In a series of papers

MSC2000: primary 11G05; secondary 11G07, 11G05, 14H52, 14H25.
Keywords: elliptic curves, genus-one curves, minimisation, reduction, descent.

[Cremona et al. 2008; 2009; n.d.], it is shown how to produce explicit equations of covering curves from a more abstract representation of the Selmer group. (The latter is computed, at least for n prime, in [Schaefer and Stoll 2004].)

In general, an n-covering C can be realised as a smooth curve of degree n inside a Severi–Brauer variety S of dimension $n-1$ (when $n=2$, we obtain a double cover of a conic instead of an embedding). If C has points everywhere locally, as will be the case when C represents an element of the n-Selmer group of E, then the same statement is true of S, and hence $S \cong \mathbb{P}^{n-1}$, so that C has a degree-n model in projective space. Thus, for $n=2$, we get a double cover of $\mathbb{P}^1$ ramified in four points, for $n=3$, we get a plane cubic curve, and for $n=4$, we get an intersection of two quadrics in $\mathbb{P}^3$. For larger n, these models are no longer complete intersections, but can be given by a number of quadratic equations.

In this paper, we will focus on the problem of how to produce nice models of the covering curves, i.e., models given by equations with small integral coefficients, in the cases $n=2$, 3 and 4. The advantage of having such a nice model is two-fold. On the one hand, rational points on the covering curve can be expected to be of smaller height on a model with small coefficients, and therefore will be found more easily. On the other hand, if no rational points are found, one would like to use the covering curve as the basis for a further descent, and the necessary computations are greatly facilitated when the given model is nice.

This problem naturally splits into two parts: *minimisation* and *reduction*. Minimisation makes the invariants of the model smaller by eliminating spurious bad primes and reducing the exponents of primes of bad reduction, to obtain a so-called minimal model. We will prove the following theorem. (See Section 2 for the definitions of models for n-coverings and their invariants.)

Theorem 1.1. *Let $n=2$, 3 or 4. Let K be a number field of class number one, and E an elliptic curve defined over K. If $\mathscr{C}$ is an n-covering of E which is everywhere locally soluble (i.e., $\mathscr{C}$ has points over all completions of K) then $\mathscr{C}$ has a model with integral coefficients and the same discriminant as a global minimal Weierstrass equation for E.*

By contrast, reduction attempts to reduce the size of the coefficients by an invertible integral (i.e., unimodular) linear change of coordinates, which leaves the invariants unchanged. Both processes are necessary to obtain a nice model: minimisation without reduction will provide a model with small invariants, but most likely rather large coefficients, whereas reduction without minimisation will not be able to make the coefficients really small, since the invariants will still be large.

After introducing the kinds of models we will be using and their invariants in Section 2, we state our main results on minimisation over local fields in Section 3A, and discuss how they relate to earlier work. The most important of these results

(the Minimisation Theorem, Theorem 3.4) is proved in Section 3B. The proof is short and transparent, but is not algorithmic. We remedy this in Section 4 where we give practical algorithms for computing minimal models, that may be seen as generalising Tate's algorithm [1975]. In Section 4E we deduce Theorem 1.1 from our local results, and explain how it may be generalised to arbitrary number fields. Moreover, as our local minimisation results make no restriction on the characteristic of the local field, they have more general global applications; in particular, one obtains results over function fields as well as number fields.

The algorithms of Section 4 may be combined with the Minimisation Theorem to prove the Strong Minimisation Theorem, Theorem 3.5(i). This states that if an n-covering of E (defined over a local field, and represented by a degree-n model) is soluble over the maximal unramified extension, then it has a model with integral coefficients and the same discriminant as a minimal Weierstrass equation for E. In Section 5 we prove the converse, Theorem 3.5(ii), thereby showing that the Strong Minimisation Theorem is best possible.

In Section 6, we discuss reduction for general n-coverings, and more specifically for $n = 2$, 3 and 4. Our results for reduction only cover the case where the ground field is $\mathbb{Q}$. A comparable theory of reduction over a general number field would be very useful in practice, but has not yet been sufficiently developed.

Finally we give examples of both minimisation and reduction (over $K = \mathbb{Q}$) in Section 7. Our algorithms have all been implemented over $\mathbb{Q}$ in MAGMA [Bosma et al. 1997] for $n = 2$, 3 and 4.

As stated earlier, the main application of our results is in explicit n-descent on elliptic curves over number fields. Minimisation and reduction of binary quartics is also used in the invariant theory method for 2-descent [Birch and Swinnerton-Dyer 1963; Cremona 1997]. For $n = 3$, Djabri and Smart [1998] consider the possibility of carrying out 3-descent using invariant theory in a similar way; one stumbling-block there was the inability to minimise plane cubic models for 3-coverings.

2. Genus-one models

In this section, we specify the models of the covering curves that we will use, together with their invariants c_4, c_6 and Δ. For completeness and later reference we include the case $n = 1$. Note that we use the term *genus-one model* to include singular models, which do not define curves of genus one.

Definition 2.1. A *Weierstrass equation*, or *genus-one model of degree* 1, is an equation of the form

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6. \tag{2-1}$$

The space of all Weierstrass equations with coefficients $a_1, \dots, a_6$ in a ring R will be denoted $X_1(R)$. We say that two such models are *R-equivalent* if they are related by substitutions

$$x \leftarrow u^2 x + r, \qquad y \leftarrow u^3 y + u^2 s x + t, \tag{2-2}$$

for some $u \in R^\times$ and $r, s, t \in R$. We write $\mathcal{G}_1(R)$ for the group of all transformations $[u; r, s, t]$ and define $\det([u; r, s, t]) = u^{-1}$. The invariants c_4, c_6 and Δ are certain primitive polynomials in $a_1, \dots, a_6$ with integer coefficients, satisfying

$$c_4^3 - c_6^2 = 1728\Delta.$$

Definition 2.2. A *genus-one model of degree* 2, or *generalised binary quartic*, is an equation of the form

$$y^2 + P(x, z)y = Q(x, z)$$

where P and Q are homogeneous polynomials of degrees 2 and 4. We sometimes abbreviate this as (P, Q). The space of all such models with coefficients in a ring R is denoted $X_2(R)$. Two such models are *R-equivalent* if they are related by substitutions $x \leftarrow m_{11}x + m_{21}z$, $z \leftarrow m_{12}x + m_{22}z$ and $y \leftarrow \mu^{-1}y + r_0x^2 + r_1xz + r_2z^2$ for some $\mu \in R^\times$, $r = (r_0, r_1, r_2) \in R^3$ and $M = (m_{ij}) \in \mathrm{GL}_2(R)$. We write $\mathcal{G}_2(R)$ for the group of all such transformations $[\mu, r, M]$, and define

$$\det([\mu, r, M]) = \mu \det(M).$$

A generalised binary quartic $y^2 + P(x_1, x_2)y = Q(x_1, x_2)$ over a field K defines a subscheme $\mathcal{C}_{(P,Q)} \subset \mathbb{P}(1, 1, 2)$, the ambient space being a weighted projective space with coordinates x_1, x_2, y. We say that the model $\Phi = (P, Q)$ is *K-soluble* if $\mathcal{C}_\Phi(K) \neq \varnothing$.

The binary quartic $F(x, z) = ax^4 + bx^3z + cx^2z^2 + dxz^3 + ez^4$ has invariants $c_4(F) = 2^4 I$ and $c_6(F) = 2^5 J$, where I and J are given by

$$I = 12ae - 3bd + c^2, \qquad J = 72ace - 27ad^2 - 27b^2e + 9bcd - 2c^3.$$

The discriminant $\Delta = (c_4^3 - c_6^2)/1728$ is 16 times the usual discriminant of a quartic polynomial. The invariants of a generalised binary quartic are obtained by completing the square, i.e., $c_4(P, Q) = c_4(\frac{1}{4}P^2 + Q)$ and so on. We find that c_4, c_6 and Δ are primitive integer coefficient polynomials in the coefficients of P and Q, again satisfying $c_4^3 - c_6^2 = 1728\Delta$.

Earlier work on 2-coverings, including [Birch and Swinnerton-Dyer 1963] and [Stoll and Cremona 2002], used the more restrictive binary quartic models with $P = 0$. We use generalised binary quartics here, in order to obtain more uniform local results at places with residue characteristic 2.

Definition 2.3. A *genus-one model of degree* 3 is a ternary cubic. We write $X_3(R)$ for the space of all ternary cubics with coefficients in a ring R. Two such models are *R-equivalent* if they are related by multiplying by $\mu \in R^\times$ and then substituting $x_j \leftarrow \sum_{i=1}^3 m_{ij}x_i$ for some $M = (m_{ij}) \in \mathrm{GL}_3(R)$. We write $\mathcal{G}_3(R) = R^\times \times \mathrm{GL}_3(R)$ for the group of all such transformations $[\mu, M]$, and we define

$$\det([\mu, M]) = \mu \det(M).$$

A ternary cubic $F(x, y, z)$ over a field K defines a subscheme $\mathcal{C}_F \subset \mathbb{P}^2$. The model F is *K-soluble* if $\mathcal{C}_F(K) \neq \varnothing$.

The invariants c_4 and c_6 may be defined as follows. Let

$$H(F) = \det \begin{pmatrix} F_{xx} & F_{xy} & F_{xz} \\ F_{yx} & F_{yy} & F_{yz} \\ F_{zx} & F_{zy} & F_{zz} \end{pmatrix}$$

be the *Hessian* of F, which is again a ternary cubic. Then we have

$$H(H(F)) = 48c_4(F)^2F + 16c_6(F)H(F);$$

the sign of $c_4(F)$ is fixed by requiring that $\Delta = (c_4^3 - c_6^2)/1728$ has integer coefficients. Then c_4, c_6 and Δ are primitive integer coefficient polynomials in the coefficients of F and satisfy $c_4^3 - c_6^2 = 1728\Delta$.

Definition 2.4. A *genus-one model of degree* 4, or *quadric intersection*, is an ordered pair (Q_1, Q_2) of quadrics (homogeneous polynomials of degree 2) in 4 variables. The space of all such models with coefficients in a ring R is denoted $X_4(R)$. Quadric intersections (Q_1, Q_2) and (Q_1', Q_2') are *R-equivalent* if they are related by putting $Q_1' = m_{11}Q_1 + m_{12}Q_2$ and $Q_2' = m_{21}Q_1 + m_{22}Q_2$ for some $M = (m_{ij}) \in \mathrm{GL}_2(R)$ and then substituting $x_j \leftarrow \sum_{i=1}^4 n_{ij}x_i$ for some $N = (n_{ij}) \in \mathrm{GL}_4(R)$. We write $\mathcal{G}_4(R) = \mathrm{GL}_2(R) \times \mathrm{GL}_4(R)$ for the group of all such transformations $[M, N]$, and define $\det([M, N]) = \det(M)\det(N)$.

A quadric intersection $\Phi = (Q_1, Q_2)$ over a field K defines a subscheme $\mathcal{C}_\Phi$ of $\mathbb{P}^3$. The model Φ is *K-soluble* if $\mathcal{C}_\Phi(K) \neq \varnothing$.

The invariants c_4 and c_6 may be defined as follows. Let A and B be the matrices of second partial derivatives of Q_1 and Q_2. Then $F(x, z) = \det(Ax + Bz)$ is a binary quartic. We define $c_4(Q_1, Q_2) = 2^{-4}c_4(F)$, $c_6(Q_1, Q_2) = 2^{-6}c_6(F)$ and $\Delta(Q_1, Q_2) = 2^{-12}\Delta(F)$. These scalings are chosen so that c_4, c_6 and Δ are primitive integer coefficient polynomials in the coefficients of Q_1 and Q_2. They satisfy $c_4^3 - c_6^2 = 1728\Delta$.

Earlier work on 4-coverings, including [Siksek 1995] and [Womack 2003], used pairs of symmetric matrices rather the pairs of quadrics. We use quadrics here, in order to obtain more uniform local results at places with residue characteristic 2.

Remark 2.5. There is also a definition of *genus-one model of degree* 5; see [Fisher 2008a]. The minimisation and reduction of these models (and possible extensions to larger degrees) will be the subject of future investigations.

Remark 2.6. There is a natural way in which we can rewrite a Weierstrass equation (a genus-one model of degree 1) as a genus-one model of degree 2, 3 or 4 (see Lemma 3.11). We have normalised the invariants c_4, c_6 and Δ so that they agree with the usual formulae (see [Silverman 1986, Chapter III], for example) when specialised to one of these Weierstrass models.

Definition 2.7. Let K be a field and $\overline{K}$ its algebraic closure. Let $K[X_n]$ be the polynomial ring in the coefficients of a genus-one model of degree n. A polynomial $F \in K[X_n]$ is an *invariant* of *weight* k if $F \circ g = \det(g)^k F$ for all $g \in \mathcal{G}_n(\overline{K})$.

For $n = 1, 2, 3, 4$ we defined polynomials $c_4, c_6, \Delta \in \mathbb{Z}[X_n]$ with

$$c_4^3 - c_6^2 = 1728\Delta.$$

Theorem 2.8. *Let* $n = 1, 2, 3$ *or* 4.

(i) *The polynomials* $c_4, c_6, \Delta \in K[X_n]$ *are invariants of weights* 4, 6 *and* 12.

(ii) *A genus-one model* $\Phi \in X_n(K)$ *defines a smooth curve* $\mathcal{C}_\Phi$ *of genus one* (*over* $\overline{K}$) *if and only if* $\Delta(\Phi) \neq 0$.

(iii) *If* $\mathrm{char}(K) \neq 2, 3$ *then* c_4 *and* c_6 *generate the ring of invariants. Moreover if* $\Phi \in X_n(K)$ *with* $\Delta(\Phi) \neq 0$ *then the Jacobian of the curve* $\mathcal{C}_\Phi$ *has Weierstrass equation* $y^2 = x^3 - 27c_4(\Phi)x - 54c_6(\Phi)$.

Proof. The invariants c_4, c_6 and Δ were known to the nineteenth century invariant theorists. The observation that they give a formula for the Jacobian is due to Weil [1954; 1983]. See [An et al. 2001] for a brief survey, or [Fisher 2008a] for a proof of the theorem exactly as it is stated here. □

As was first pointed out to us by Rodriguez-Villegas, it is possible to work back through Tate's formulaire [Silverman 1986, Chapter III] to write the invariants c_4 and c_6 in terms of polynomials $a_1, \dots, a_6$.

Lemma 2.9. *There exist* $a_1, a_2, a_3, a_4, a_6 \in \mathbb{Z}[X_n]$ *and* $b_2, b_4, b_6 \in \mathbb{Z}[X_n]$ *with*

$$b_2 = a_1^2 + 4a_2, \qquad b_4 = a_1a_3 + 2a_4, \qquad b_6 = a_3^2 + 4a_6,$$
$$c_4 = b_2^2 - 24b_4, \qquad c_6 = -b_2^3 + 36b_2b_4 - 216b_6.$$

Proof. The lemma is proved by splitting into the cases $n = 2, 3, 4$ and giving explicit formulae for the a-invariants. (The case $n = 1$ is a tautology.)

Case $n = 2$. The a-invariants of the generalised binary quartic,

$$y^2 + (lx^2 + mxz + nz^2)y = ax^4 + bx^3z + cx^2z^2 + dxz^3 + ez^4,$$

are

$$\begin{aligned} &a_1 = m, \qquad a_2 = c - ln, \qquad a_3 = ld + nb, \\ &a_4 = -4ae + bd - (l^2 e + lnc + n^2 a), \\ &a_6 = -4ace + ad^2 + b^2 e - (l^2 ce + m^2 ae + n^2 ac + lnbd) + lmbe + mnad. \end{aligned} \tag{2-3}$$

Case $n = 3$. The a-invariants of the ternary cubic,

$$ax^3 + by^3 + cz^3 + a_2 x^2 y + a_3 x^2 z + b_1 xy^2 + b_3 y^2 z + c_1 xz^2 + c_2 yz^2 + mxyz,$$

are[1]

$$\begin{aligned} &a_1 = m, \\ &a_2 = -(a_2 c_2 + a_3 b_3 + b_1 c_1), \\ &a_3 = 9abc - (ab_3 c_2 + ba_3 c_1 + ca_2 b_1) - (a_2 b_3 c_1 + a_3 b_1 c_2), \\ &a_4 = -3(abc_1 c_2 + acb_1 b_3 + bca_2 a_3) + a(b_1 c_2^2 + b_3^2 c_1) + b(a_2 c_1^2 + a_3^2 c_2) \\ &\qquad\qquad + c(a_2^2 b_3 + a_3 b_1^2) + a_2 c_2 a_3 b_3 + b_1 c_1 a_2 c_2 + a_3 b_3 b_1 c_1, \\ &a_6 = -27a^2 b^2 c^2 + 9abc(ab_3 c_2 + ca_2 b_1 + ba_3 c_1) + \cdots + abcm^3. \end{aligned} \tag{2-4}$$

These formulae in the case $n = 3$ were first given in [Artin et al. 2005].

Case $n = 4$. Let $Q = \sum_{i \le j} c_{ij} x_i x_j$ be a quadric in 4 variables. Then

$$\det\left(\frac{\partial^2 Q}{\partial x_i \partial x_j}\right) = \mathrm{pf}(Q)^2 + 4\,\mathrm{rd}(Q)$$

where $\mathrm{pf}(Q) = c_{12}c_{34} + c_{13}c_{24} + c_{14}c_{23}$ and $\mathrm{rd}(Q) \in \mathbb{Z}[c_{11}, c_{12}, \ldots, c_{44}]$. We define the a-invariants of the quadric intersection (Q_1, Q_2) to be the a-invariants of the generalised binary quartic

$$y^2 + \mathrm{pf}(xQ_1 + zQ_2)y = \mathrm{rd}(xQ_1 + zQ_2). \qquad \square$$

The polynomials a_i of Lemma 2.9 are far from unique. They can be modified by any transformation of the form $[\pm 1; r, s, t]$ with $r, s, t \in \mathbb{Z}[X_n]$. We next extend Theorem 2.8(iii) to fields of arbitrary characteristic. (The reader only interested in applications over number fields and their completions may safely skip this result.)

Theorem 2.10. *Let K be any field, and $n = 1, 2, 3$ or 4. For all $\Phi \in X_n(K)$ with $\Delta(\Phi) \neq 0$, the Jacobian of the curve $\mathscr{C}_\Phi$ has Weierstrass equation*

$$y^2 + a_1 xy + a_3 y = x^3 + a_2 x^2 + a_4 x + a_6, \tag{2-5}$$

where $a_i = a_i(\Phi)$.

[1] We follow classical notation for the coefficient labels, but warn the reader that the symbols a_2, a_3 have different meanings on the left and right sides of (2-4).

Proof. For $n = 3$ this is a special case of a theorem of Artin, Rodriguez-Villegas and Tate [2005]. The cases $n = 2, 4$ may be proved using similar techniques. We sketch a simplified form of the proof, covering the cases $n = 2$, 3 and 4. (The case $n = 1$ is of course a tautology.)

Let C/S be the universal family over[2] $S = \mathrm{Spec}(\mathbb{Z}[X_n][\Delta^{-1}])$. By Theorem 2.8(ii) the fibres are smooth projective curves of genus one. Let J/S be the Jacobian of C/S, in the sense that J is the S-scheme representing the relative Picard functor $\mathrm{Pic}^0_{C/S}$; see [Bosch et al. 1990, §9.3, Theorem 1]. Each fibre of J/S is the Jacobian of the corresponding fibre of C/S and hence an elliptic curve. By a generalisation of the usual procedure for putting an elliptic curve in Weierstrass form (see [Deligne 1975] or [Artin et al. 2005, Theorem 2] for a further generalisation) J is defined as a subscheme of $\mathbb{P}^2_S$ by the homogenisation of

$$y^2 + a_1' xy + a_3' y = x^3 + a_2' x^2 + a_4' x + a_6' \tag{2-6}$$

for some $a_1', \dots, a_6' \in \mathbb{Z}[X_n][\Delta^{-1}]$. Thus for every field K, and every $\Phi \in X_n(K)$ with $\Delta(\Phi) \neq 0$, the Weierstrass equation (2-6) gives a model for the Jacobian of $\mathscr{C}_\Phi$.

It only remains to show that (2-5) and (2-6) are related by a transformation in $\mathscr{G}_1(R)$ where $R = \mathbb{Z}[X_n][\Delta^{-1}]$. By Theorem 2.8(iii) they are related by some $[u; r, s, t] \in \mathscr{G}_1(K)$ where $K = \mathbb{Q}(X_n)$. Since for any genus-one model with $\Delta \neq 0$, (2-5) and (2-6) both specialise to a nonsingular Weierstrass equation, it follows that $u \in R^\times$. Then, since R is integrally closed, a standard argument (see [Silverman 1986, Chapter VII, Proposition 1.3]) shows that $r, s, t \in R$. □

We note that $a_1, \dots, a_6$ are not invariants in the sense of Definition 2.7. The ring of invariants when $\mathrm{char}(K) = 2$ or 3 is described in [Fisher 2008a, §10]. As is noted there, these do not give a formula for the Jacobian.

3. Minimisation theorems

3A. ***Statement of results.*** Let K be a field with normalised discrete valuation $v : K^\times \to \mathbb{Z}$. We write $\mathcal{O}_K$ for the valuation ring (or ring of integers) of K and fix a uniformiser $\pi \in K$. We assume throughout that the residue field $k = \mathcal{O}_K/\pi\mathcal{O}_K$ is perfect. A field extension L/K is *unramified* if there is a (normalised) discrete valuation $w : L^\times \to \mathbb{Z}$ extending v. The strict Henselisation K^{sh} of K is an unramified extension of K, that satisfies the conclusions of Hensel's lemma and has residue field $\bar{k}$, the algebraic closure of k. (See [Milne 2008, Definition 4.8] or [Raynaud 1970, Chapter VIII] for the precise definition.) If K is complete (with

[2] In [Artin et al. 2005] the authors work over $S = \mathrm{Spec}(\mathbb{Z}[X_3]) \setminus \{0\}$. This gives a more general result, but also makes the proof more difficult.

respect to v) then K^{sh} is the maximal unramified extension K^{nr} of K as defined in [Serre 1979, Chapter III, §5].

We work with genus-one models of degree $n = 1, 2, 3$ or 4. The invariants c_4, c_6 and Δ of a genus-one model were defined in Section 2.

Definition 3.1. (i) A genus-one model $\Phi \in X_n(K)$ is *nonsingular* if $\Delta(\Phi) \neq 0$.

(ii) A genus-one model $\Phi \in X_n(K)$ is *integral* if it has coefficients in $\mathcal{O}_K$.

(iii) A nonsingular model $\Phi \in X_n(\mathcal{O}_K)$ is *minimal* if $v(\Delta(\Phi))$ is minimal among all integral models K-equivalent to Φ, otherwise Φ is *nonminimal*.

Algorithms for computing minimal models in the case $n = 1$ have been given by Tate [1975] (see also [Silverman 1994, Chapter IV, §9]) and Laska [1982]. The latter can be refined using Kraus' conditions [Kraus 1989] as described in [Connell 1996, Chapter V] or [Cremona 1997, §3.2]. (Laska's algorithm and its refinements are simpler than Tate's algorithm, but are only applicable when $\mathrm{char}(K) \neq 2, 3$.) In Section 4 we give algorithms for computing minimal models in the cases $n = 2, 3, 4$.

In the following lemma we define the *level* of a genus-one model.

Lemma 3.2. *Let $\Phi \in X_n(K)$ be a nonsingular model of degree n. Let Δ_E be the minimal discriminant of $E = \mathrm{Jac}(\mathscr{C}_\Phi)$.*

(i) $v(\Delta(\Phi)) = v(\Delta_E) + 12l$ *for some integer l, which we call the level of Φ.*

(ii) *If* $\mathrm{char}(k) \neq 2, 3$ *then* $l = \min\{\lfloor v(c_4(\Phi))/4 \rfloor, \lfloor v(c_6(\Phi))/6 \rfloor\}$.

(iii) *The level of an integral model is always nonnegative.*

Proof. If $\mathrm{char}(k) \neq 2, 3$ then this is clear by Theorem 2.8 and the standard formulae for transforming Weierstrass equations. In general (that is, to prove (iii) when $\mathrm{char}(k) = 2$ or 3, or even to define the level when $\mathrm{char}(K) = 2$ or 3) we use Lemma 2.9 and Theorem 2.10 instead. □

The level of $\Phi \in X_n(K)$ may be computed as $v(u)$ where $[u; r, s, t] \in \mathscr{G}_1(K)$ is a transformation that minimises the Weierstrass equation (2-5).

Definition 3.3. The *minimal level* of $\Phi \in X_n(K)$ is the minimum of the levels of all integral models K-equivalent to Φ. Thus an integral model Φ is minimal (see Definition 3.1) if and only if it has level equal to this minimal level.

If $n = 1$ then the minimal level is 0, for trivial reasons. So from now on we take $n = 2$, 3 or 4. The most important result on minimisation states that every K-soluble model has minimal level 0, or in other words, that every K-soluble model is K-equivalent to an integral model whose discriminant has the same valuation as that of the discriminant of the minimal model of the Jacobian elliptic curve.

Theorem 3.4 (Minimisation Theorem). *Let $\Phi \in X_n(K)$ be a nonsingular genus-one model. If $\mathscr{C}_\Phi(K) \neq \varnothing$ then Φ has minimal level* 0.

The following strengthening of the Minimisation Theorem shows that a nonsingular model is K-equivalent to an integral model of level 0 if and only if it is K^{sh}-soluble.

Theorem 3.5. *Let $\Phi \in X_n(K)$ be nonsingular.*

(i) (Strong Minimisation Theorem). *If $\mathscr{C}_\Phi(K^{\mathrm{sh}}) \neq \varnothing$ then Φ has minimal level* 0.

(ii) (Converse Theorem). *If $\mathscr{C}_\Phi(K^{\mathrm{sh}}) = \varnothing$ then the minimal level is at least* 1, *and is equal to* 1 *if* $\mathrm{char}(k) \nmid n$.

Algorithms for minimising K-soluble binary quartics over $K = \mathbb{Q}_p$ are sketched by Birch and Swinnerton-Dyer [1963, Lemmas 3–5], with details in the case where p is neither 2 nor 3. Their algorithms give a proof of the Minimisation Theorem for $n = 2$, except when $p = 2$ (in which case further work is required to handle the cross terms). As pointed out in [Stoll and Cremona 2002] this generalises immediately to any local field K with $\mathrm{char}(k) \neq 2, 3$. These calculations were extended to the case $n = 3$, in conjunction with the authors' work on 3-descent [Cremona et al. 2008; 2009; n.d.]. The case $n = 4$ was treated by Womack [2003, Section 2.5], using a method that goes via the results for $n = 2$.

In each case, the approach taken is to start with a K^{sh}-soluble model $\Phi \in X_n(\mathcal{O}_K)$ with $v(c_4(\Phi)) \geq 4$ and $v(c_6(\Phi)) \geq 6$, and then by a series of substitutions to show that Φ is K-equivalent to an integral model of smaller level. This leads to both a proof of the Strong Minimisation Theorem and a practical algorithm for minimising. However, this traditional approach suffers from the following drawbacks.

- It is necessary to split into a large number of (elementary yet tedious) cases, and the number of cases grows rapidly with n.
- The modifications required if $\mathrm{char}(k) = 2$ or 3 are somewhat involved. (The hypothesis that Φ has positive level has to be made explicit using either Kraus' conditions [1989] or the "a-invariants" defined in Lemma 2.9.)

We take a different approach, in which the tasks of proving the Minimisation Theorem and finding a practical algorithm for minimising are treated separately. A proof of the Minimisation Theorem for $n = 2, 3$ (in all residue characteristics) is given in [Fisher 2007]. In Section 3B we simplify the proof and extend it to the case $n = 4$. Unfortunately this approach does not lead to any readily implementable algorithm, nor does it prove the Strong Minimisation Theorem.

In Sections 4A (case $n = 2$) and 4B (case $n = 3$) we specify a rather simple-minded procedure and show that, given any nonminimal integral model, iterating this procedure will eventually decrease the level. This gives an algorithm for computing minimal models. In Section 4C we give an algorithm in the case $n = 4$ based

on the treatment in [Womack 2003]. The algorithms for $n = 2$ and $n = 4$ must be modified when $\mathrm{char}(k) = 2$ as described in Section 4D. These modifications are required since, as noted in Section 2, our models for n-coverings differ slightly from those used previously in the literature. We have also defined the level, not in an absolute way, but by comparison with a minimal model for the Jacobian elliptic curve. The combined effect of these changes is that our results are much cleaner to state, in particular for residue characteristic 2, and can be proved uniformly, without assumptions on the ramification index.

As is the case for Tate's algorithm, it is clear from the form of our algorithms (for $n = 2, 3, 4$) that their success or otherwise is unchanged by an unramified field extension. We deduce the following.

Theorem 3.6. *The minimal level of a nonsingular genus-one model of degree* 2, 3 *or* 4 *is unchanged by an unramified field extension.*

The Strong Minimisation Theorem is then an immediate consequence of Theorem 3.6 and the Minimisation Theorem.

In Section 5 we show how to write down examples of minimal genus-one models of positive level. We call the models arising in our construction *critical models*; see Definition 5.1. We show (for $n = 2, 3$) that any K^{sh}-insoluble model is K-equivalent to a critical model. There is a corresponding result for models of degree $n = 4$. The proof of the Converse Theorem, 3.5(ii), is then reduced to a statement about the possible levels of a critical model (see Lemma 5.4).

Theorem 3.5 in the case $n = 2$ may already be found in [Liu 1996, remarque 21]. We claim that our proof is much simpler, and in any case serves as a template for our generalisations to $n = 3, 4$. Liu also gives an algorithm for minimising [Liu 1996, p. 4594, remarque 11] (still for $n = 2$), which although not made explicit appears to be the same as ours.

We remark that minimisations are not unique, in the sense that there can be more than one $\mathcal{O}_K$-equivalence class of minimal models K-equivalent to a given genus-one model. Following on from our work and that of Liu, Sadek [2009] explains how to compute the number of such classes.

For a more general, but necessarily less explicit, discussion of the problem of minimising homogeneous polynomials (of degree d in n variables) see [Kollár 1997].

3B. *Proof of the Minimisation Theorem.* In this subsection only we relax our assumptions on $\mathcal{O}_K$ and K. It will only be necessary to assume that $\mathcal{O}_K$ is a principal ideal domain and K is its field of fractions. The definitions of a nonsingular model and an integral model (see Definition 3.1) carry over as before. We consider models of degree $n = 2, 3$ or 4.

Let E be an elliptic curve over K, with identity $\mathcal{O}_E \in E(K)$, and let D be a K-rational divisor on E of degree n. We write $[D]$ for the linear equivalence class of D. We pick a basis $f_1, \ldots, f_n$ for the Riemann–Roch space $\mathcal{L}(D)$, and let $E \to \mathbb{P}^{n-1}$ be the morphism given by $P \mapsto (f_1(P) : \ldots : f_n(P))$. Then according as $n = 2, 3$ or 4, we find that E may be written as either a double cover of $\mathbb{P}^1$, a plane cubic, or an intersection of two quadrics in $\mathbb{P}^3$. It is therefore defined by a suitable genus-one model $\Phi \in X_n(K)$. Moreover this model is uniquely determined up to K-equivalence by the pair $(E, [D])$: replacing D by an equivalent divisor or changing basis for the space $\mathcal{L}(D)$ only has the effect of a linear change of coordinates on $\mathbb{P}^{n-1}$, so only changes the genus-one model by a K-equivalence. In this situation we say that the genus-one model Φ *represents* the pair $(E, [D])$.

Similarly, we obtain a genus-one model $\Phi \in X_n(K)$, well-defined up to K-equivalence, representing every pair $(\mathcal{C}, [D])$ where $\mathcal{C}$ is a genus-one curve and D a divisor of degree n on $\mathcal{C}$; we have $\mathcal{C} \cong \mathcal{C}_\Phi$ (over K), and in particular, Φ is K-soluble if and only if $\mathcal{C}(K) \neq \varnothing$. Under this isomorphism, the divisor class $[D]$ on $\mathcal{C}$ maps to a distinguished divisor class $[D_\Phi]$ of degree n on $\mathcal{C}_\Phi$, namely the class of the fibres of the map $\mathcal{C}_\Phi \to \mathbb{P}^1$ if $n = 2$, or the hyperplane section if $n = 3, 4$. It is a tautology that Φ represents $(\mathcal{C}_\Phi, [D_\Phi])$.

Lemma 3.7. *Every K-soluble nonsingular genus-one model arises from a pair $(E, [D])$ in the manner described above.*

Proof. If $\Phi \in X_n(K)$ is a K-soluble nonsingular model then $\mathcal{C}_\Phi$ is a smooth curve of genus one with a rational point, hence is an elliptic curve. Now it is obvious that the genus-one model determined by the pair $(\mathcal{C}_\Phi, [D_\Phi])$ is just Φ. □

The aim of this section is to prove the following theorem. The Minimisation Theorem, 3.4, is then an immediate consequence by Lemma 3.7.

Theorem 3.8. *Let E/K be an elliptic curve with integral Weierstrass equation*

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6 \tag{3-1}$$

and let $D \in \mathrm{Div}_K(E)$ be a divisor on E of degree $n = 2, 3$ or 4. Then $(E, [D])$ can be represented by an integral genus-one model with the same discriminant as (3-1).

This theorem states that, in the K-equivalence class of genus-one models representing $(E, [D])$, there is one which is integral and has the same discriminant as any given integral Weierstrass model for E. Our strategy for proving this starts with two observations.

Firstly, the claim really does only depend on the divisor class $[D]$ and not the given specific divisor D in that class, since the K-equivalence class of genus-one models representing $(E, [D])$ only depends on the divisor class.

Secondly, if $\tau_Q : E \to E$ is translation by some point $Q \in E(K)$, then the pairs $(E, [D])$ and $(E, [\tau_Q^* D])$ determine K-equivalent genus-one models. This follows from the fact that the map $E \to \mathbb{P}^{n-1}$ determined by $[\tau_Q^* D]$ is the composite of τ_Q and the map determined by $[D]$.

Using the classical facts that every K-rational divisor D of degree n is linearly equivalent to a unique divisor of the form $(n-1).\mathcal{O}_E + P$ for some $P \in E(K)$, and that divisors on an elliptic curve are linearly equivalent if and only if they have the same degree and the same sum, it suffices to prove Theorem 3.8 for such divisors as P runs over a set of coset representatives for $E(K)/nE(K)$.

In Lemmas 3.11 and 3.12, we show using explicit formulae that Theorem 3.8 holds in the cases $D = n.\mathcal{O}_E$ and $D = (n-1).\mathcal{O}_E + P$ where $P \in E(K)$ is an integral point, that is, a point with coordinates in $\mathcal{O}_K$. This is already enough to prove Theorem 3.8 in the case $\mathcal{O}_K$ is a complete discrete valuation ring with residue characteristic prime to n. Indeed, by the theory of formal groups, every nonzero element of $E(K)/nE(K)$ may then be represented by an integral point.

In general we rely on the following two lemmas, proved later in this subsection.

Lemma 3.9 (unprojection lemma). *Let $D \in \mathrm{Div}_K(E)$ have degree 2 or 3, and let $P \in E(K)$. If Theorem 3.8 holds for D then it holds for $D + P$.*

Lemma 3.10 (projection lemma). *Let $D \in \mathrm{Div}_K(E)$ have degree 3 or 4, and let $P \in E(K)$. If Theorem 3.8 holds for D then it holds for $D - P$.*

Theorem 3.8 may be deduced from these lemmas in more than one way. For example, if $n = 3$ or 4 then $D \sim (n-1).\mathcal{O}_E + P$ for some $P \in E(K)$. Then we quote the result for $D' = (n-1).\mathcal{O}_E$ and use the unprojection lemma. Likewise if $n = 2$ or 3 then $D \sim (n+1).\mathcal{O}_E - P$ for some $P \in E(K)$. Then we quote the result for $D' = (n+1).\mathcal{O}_E$ and apply the projection lemma to D'.

Theorem 3.8 in the case $D = n.\mathcal{O}_E$ follows from the formulae we used to normalise the invariants c_4, c_6 and Δ: see Remark 2.6.

Lemma 3.11. *Let E be an elliptic curve with Weierstrass equation*

$$Y^2 + a_1XY + a_3Y = X^3 + a_2X^2 + a_4X + a_6. \tag{3-2}$$

Then the pair $(E, [n.\mathcal{O}_E])$ determines genus-one models as follows:

$$\begin{aligned} n = 2: &\quad y^2 + (a_1x_1x_2 + a_3x_2^2)y = x_1^3x_2 + a_2x_1^2x_2^2 + a_4x_1x_2^3 + a_6x_2^4; \\ n = 3: &\quad y^2z + a_1xyz + a_3yz^2 - x^3 - a_2x^2z - a_4xz^2 - a_6z^3 = 0; \\ n = 4: &\quad x^2 - zt = 0, \quad y^2 + a_1xy + a_3yz - xt - a_2x^2 - a_4xz - a_6z^2 = 0. \end{aligned}$$

Moreover, each of these models has the same invariants c_4, c_6 and Δ as (3-2).

Proof. In the case $n = 2$ we embed E in $\mathbb{P}(1, 1, 2)$ via $(x_1 : x_2 : y) = (X : 1 : Y)$. In the cases $n = 3, 4$ we embed E in $\mathbb{P}^{n-1}$ via $(z : x : y) = (1 : X : Y)$ and

$(z : x : y : t) = (1 : X : Y : X^2)$ respectively. The statement about the invariants follows by direct calculation. □

Next we prove Theorem 3.8 in the case $D = (n-1).\mathcal{O}_E + P$ where $P \in E(K)$ is an integral point. By a substitution $X \leftarrow X + X(P)$, $Y \leftarrow Y + Y(P)$ we may assume that P is the point $(0, 0)$.

Lemma 3.12. *Let E be an elliptic curve with Weierstrass equation*

$$Y^2 + a_1XY + a_3Y = X^3 + a_2X^2 + a_4X \tag{3-3}$$

and let $P = (0, 0)$. Then the pair $(E, [(n-1).\mathcal{O}_E + P])$ determines genus-one models as follows:

$$\begin{aligned} n = 2: &\quad y^2 + (-x_1^2 + a_1x_1x_2 + a_2x_2^2)y = -a_3x_1x_2^3 - a_4x_2^4; \\ n = 3: &\quad y^2z - x^2y + a_1xyz + a_2yz^2 + a_3xz^2 + a_4z^3 = 0; \\ n = 4: &\quad zt - xy + a_1yz + a_3z^2 = 0, \quad y^2 - xt + a_2yz + a_4z^2 = 0. \end{aligned}$$

Moreover, each of these models has the same invariants c_4, c_6 and Δ as (3-3).

Proof. The rational function

$$F = \frac{Y + a_1X + a_3}{X} = \frac{X^2 + a_2X + a_4}{Y}$$

belongs to the Riemann–Roch space $\mathcal{L}(\mathcal{O}_E + P)$. In the case $n = 2$ we embed E in $\mathbb{P}(1, 1, 2)$ via $(x_1 : x_2 : y) = (F : 1 : X)$. In the cases $n = 3, 4$ we embed E in $\mathbb{P}^{n-1}$ via $(z : x : y) = (1 : F : X)$ and $(z : x : y : t) = (1 : F : X : Y)$ respectively. The statement about the invariants follows by direct calculation. □

It remains to prove Lemmas 3.9 and 3.10. One observation that we use in the proofs is the following.

Lemma 3.13. *The group $\mathrm{SL}_n(\mathcal{O}_K)$ acts transitively on $\mathbb{P}^{n-1}(K)$.*

Proof. Since $\mathcal{O}_K$ is a principal ideal domain this is standard. See for example [Jacobson 1985, Exercise 6, p. 186]. □

We now explain how we pass between results for generalised binary quartics (case $n = 2$) and ternary cubics (case $n = 3$).

Lemma 3.14. *Let $D \in \mathrm{Div}_K(E)$ be a divisor of degree 2 and let $P \in E(K)$. Let f_1, f_2 and f_3 be binary forms over K with $\deg f_i = i$. The following statements are equivalent.*

(i) *The pair $(E, [D])$ is represented by the generalised binary quartic*

$$y^2 + f_2(x_1, x_2)y = f_1(x_1, x_2)f_3(x_1, x_2) \tag{3-4}$$

and P is the point defined by $f_1 = y = 0$.

(ii) *The pair* $(E, [D+P])$ *is represented by the ternary cubic*

$$f_1(X,Z)Y^2 - f_2(X,Z)Y - f_3(X,Z) = 0 \tag{3-5}$$

and P *is the point* $(X:Y:Z) = (0:1:0)$.

Proof. We first show that the curves C_2 and C_3 defined by (3-4) and (3-5) are isomorphic. An isomorphism $\phi : C_2 \to C_3$ is given by

$$\begin{aligned}\phi : (x_1 : x_2 : y) \mapsto (X:Y:Z) &= (x_1 f_1(x_1,x_2) : y + f_2(x_1,x_2) : x_2 f_1(x_1,x_2)) \\ &= (x_1 y : f_3(x_1,x_2) : x_2 y),\end{aligned}$$

with inverse

$$\phi^{-1} : (X:Y:Z) \mapsto (x_1 : x_2 : y) = (X : Z : f_1(X,Z)Y - f_2(X,Z)).$$

The isomorphism identifies the points $\{f_1 = y = 0\} \in C_2(K)$ and $(0:1:0) \in C_3(K)$. To prove the equivalence of (i) and (ii) we note that if $D = P_1 + P_2$ is a fibre of the map $C_2 \to \mathbb{P}^1$; $(x_1 : x_2 : y) \mapsto (x_1 : x_2)$ then the points $\phi(P_1)$, $\phi(P_2)$ and $(0:1:0)$ are collinear on $C_3 \subset \mathbb{P}^2$. □

There is an entirely analogous result for passing between ternary cubics (case $n = 3$) and quadric intersections (case $n = 4$).

Lemma 3.15. *Let* $D \in \mathrm{Div}_K(E)$ *be a divisor of degree* 3 *and let* $P \in E(K)$. *Let* l_1, l_2, q_1 *and* q_2 *be ternary forms over* K *with* $\deg l_i = 1$ *and* $\deg q_i = 2$. *The following statements are equivalent.*

(i) *The pair* $(E, [D])$ *is represented by the ternary cubic*

$$l_1(x_1,x_2,x_3)q_2(x_1,x_2,x_3) - l_2(x_1,x_2,x_3)q_1(x_1,x_2,x_3) = 0, \tag{3-6}$$

and P *is the point defined by* $l_1 = l_2 = 0$.

(ii) *The pair* $(E, [D+P])$ *is represented by the quadric intersection*

$$l_1(x_1,x_2,x_3)x_4 + q_1(x_1,x_2,x_3) = 0, \qquad l_2(x_1,x_2,x_3)x_4 + q_2(x_1,x_2,x_3) = 0, \tag{3-7}$$

and P *is the point* $(x_1 : x_2 : x_3 : x_4) = (0:0:0:1)$.

Proof. We first show that the curves C_3 and C_4 defined by (3-6) and (3-7) are isomorphic. An isomorphism $\phi : C_3 \to C_4$ is given by

$$\phi : (x_1 : x_2 : x_3) \mapsto (x_1 l_1 : x_2 l_1 : x_3 l_1 : -q_1) = (x_1 l_2 : x_2 l_2 : x_3 l_2 : -q_2)$$

with inverse $\phi^{-1} : (x_1 : x_2 : x_3 : x_4) \mapsto (x_1 : x_2 : x_3)$. This isomorphism identifies the points $\{l_1 = l_2 = 0\} \in C_3(K)$ and $(0:0:0:1) \in C_4(K)$. To prove the equivalence of (i) and (ii) we note that if $C_3 \subset \mathbb{P}^2$ meets some line in the divisor $D = P_1 + P_2 + P_3$ then the points $\phi(P_1)$, $\phi(P_2)$, $\phi(P_3)$ and $(0:0:0:1)$ are coplanar on $C_4 \subset \mathbb{P}^3$. □

A generic computation shows that the genus-one models (3-4) and (3-5) in Lemma 3.14 have the same discriminant. Likewise the models (3-6) and (3-7) in Lemma 3.15 have the same discriminant.

Proof of Lemma 3.9. (i) Let $D \in \mathrm{Div}_K(E)$ be a divisor of degree 2, and suppose the pair $(E, [D])$ is represented by an integral generalised binary quartic of discriminant Δ. By Lemma 3.13 (with $n = 2$) we may assume that P is the point $(x_1 : x_2 : y) = (1 : 0 : \eta)$ for some $\eta \in K$. Since $\mathcal{O}_K$ is integrally closed it follows that $\eta \in \mathcal{O}_K$. By making a substitution $y \leftarrow y + \eta x_1^2$ we may assume that $\eta = 0$. Our model is now of the form (3-4) with $f_1(x_1, x_2) = x_2$. Then the ternary cubic (3-5) is an integral model of discriminant Δ representing the pair $(E, [D + P])$.

(ii) Let $D \in \mathrm{Div}_K(E)$ be a divisor of degree 3, and suppose the pair $(E, [D])$ is represented by an integral ternary cubic of discriminant Δ. By Lemma 3.13 (with $n = 3$) we may assume that P is the point $(x_1 : x_2 : x_3) = (0 : 0 : 1)$. Our model is now of the form (3-6) with $l_1 = x_1$ and $l_2 = x_2$. We may choose the quadratic forms q_1 and q_2 to have coefficients in $\mathcal{O}_K$. Then the quadric intersection (3-7) is an integral model of discriminant Δ representing the pair $(E, [D + P])$. □

Proof of Lemma 3.10. (i) Let $D \in \mathrm{Div}_K(E)$ be a divisor of degree 3, and suppose the pair $(E, [D])$ is represented by an integral ternary cubic of discriminant Δ. By Lemma 3.13 (with $n = 3$) we may assume that P is the point $(X : Y : Z) = (0 : 1 : 0)$. Our model is now of the form (3-5). Then the generalised binary quartic (3-4) is an integral model of discriminant Δ representing the pair $(E, [D - P])$.

(ii) Let $D \in \mathrm{Div}_K(E)$ be a divisor of degree 4, and suppose the pair $(E, [D])$ is represented by an integral quadric intersection of discriminant Δ. By Lemma 3.13 (with $n = 4$) we may assume that P is the point $(x_1 : x_2 : x_3 : x_4) = (0 : 0 : 0 : 1)$. Our model is now of the form (3-7) for some forms l_1, l_2, q_1 and q_2 with coefficients in $\mathcal{O}_K$. Then the ternary cubic (3-6) is an integral model of discriminant Δ representing the pair $(E, [D - P])$. □

Remark 3.16. In principle these proofs give an algorithm for minimising K-soluble models, but only once a K-rational point is explicitly known. (Although it is easy to decide solubility over local fields, if the model is far from minimal then we would need to know a local point to very high precision. Hence our comment that this is not a readily implementable algorithm.)

4. Minimisation algorithms

In this section we give algorithms for minimising binary quartics (case $n = 2$), ternary cubics (case $n = 3$) and quadric intersections (case $n = 4$). As in Section 3A we work over a field K which is the field of fractions of a discrete valuation ring $\mathcal{O}_K$. There is no need to assume that K is complete (or even Henselian). We

fix a uniformiser π and write $k = \mathbb{O}_K/\pi\mathbb{O}_K$ for the residue field. In the cases $n = 2, 4$ we initially assume that $\mathrm{char}(k) \neq 2$, leaving the case $\mathrm{char}(k) = 2$ to Section 4D.

Our algorithms for $n = 2, 3$ share some common features which we now elucidate. In these cases we specify a procedure that takes as input an integral genus-one model of positive level, and returns a K-equivalent integral model of the same or smaller level. We then show that if the model is nonminimal then the level must decrease after finitely many iterations, and give a bound N on the number of iterations required. This also gives a test for minimality: if N iterations of the procedure fail to decrease the level, then the model must be minimal.

The proofs are by induction on the *slope*, which we define as the least valuation of the determinant of a matrix $M \in \mathrm{GL}_n(K)$ with entries in $\mathbb{O}_K$ that can be used to decrease the level. The slope of a minimal model is undefined. The arguments we use are incapable of proving the Minimisation Theorem, since we assume at the outset that the given model has a slope, i.e., is nonminimal.

The following lemma is used to show that our procedure gives a well-defined map on $\mathbb{O}_K$-equivalence classes. This is useful, since it means we are free to replace our model by an $\mathbb{O}_K$-equivalent one at any stage of the proof. We write I_m for the $m \times m$ identity matrix.

Lemma 4.1. *Let* $\mathrm{GL}_n(K)$ *act on* $\mathbb{P}^{n-1}$ *in the natural way* (*via left multiplication of column vectors by matrices*). *Let* $\alpha = \mathrm{Diag}(I_r, \pi I_{n-r})$ *for some* $0 < r < n$. *Then the subgroup of* $\mathrm{GL}_n(\mathbb{O}_K)$ *consisting of transformations whose reduction modulo* π *preserves the subspace* $\{x_{r+1} = \ldots = x_n = 0\}$ *is*

$$\mathrm{GL}_n(\mathbb{O}_K) \cap \alpha\, \mathrm{GL}_n(\mathbb{O}_K)\alpha^{-1}.$$

Proof. Identifying $\mathbb{P}^{n-1}(K)$ with the nonzero elements of K^n modular scalars, $\mathrm{GL}_n(\mathbb{O}_K)$ is the subgroup preserving $\mathbb{O}_K^n$ and we are interested in the subgroup which also preserves $\mathbb{O}_K^r \oplus (\pi\mathbb{O}_K)^{n-r} = \alpha(\mathbb{O}_K^n)$. The statement is now clear. □

This lemma is used as follows. Suppose that Φ and Ψ are $\mathrm{GL}_n(\mathbb{O}_K)$-equivalent models, and the matrix relating them (or its transpose, depending on conventions) is one whose reduction modulo π preserves the subspace $\{x_{r+1} = \ldots = x_n = 0\}$. Then the models Φ' and Ψ' obtained by applying $\alpha = \mathrm{Diag}(I_r, \pi I_{n-r})$ to both Φ and Ψ will again be $\mathrm{GL}_n(\mathbb{O}_K)$-equivalent.

4A. ***Minimisation of 2-coverings.*** Let $F \in K[x, z]$ be a binary quartic, say

$$F(x, z) = ax^4 + bx^3z + cx^2z^2 + dxz^3 + ez^4.$$

Viewing the set of these as a subset of $X_2(K)$, the group of K-equivalences between binary quartics is $K^\times \times \mathrm{GL}_2(K)$, where $[\mu, M]$ acts via $F \mapsto \mu^2(F \circ M)$. Note that $[\pi^{-2}, \mathrm{Diag}(\pi, \pi)]$ acts trivially, so we may if convenient assume that M has entries in $\mathbb{O}_K$, not all in $\pi\mathbb{O}_K$.

We say that an integral binary quartic F is *minimal* if $v(\Delta(F))$ is minimal among all integral binary quartics K-equivalent to F. If $\mathrm{char}(k) = 2$ then this need not be the same as being minimal as a generalised binary quartic. We define the valuation $v(F)$ to be the minimum of the valuations of the coefficients. If $v(F) \geq 2$, then F is not minimal, and indeed dividing through by π^2 gives a K-equivalent integral model of smaller level. The algorithm for minimising binary quartics is described in the following theorem.

Theorem 4.2. *Let $F \in \mathcal{O}_K[x, z]$ be a nonsingular binary quartic. Suppose that $v(F) = 0$ or 1, but F has positive level. If $\mathrm{char}(k) = 2$ then further assume that F is nonminimal. Then*

(i) *The reduction $\bmod \pi$ of $F_1(x, z) = \pi^{-v(F)} F(x, z)$ has either a triple or quadruple root defined over k.*

(ii) *The following procedure replaces F by a K-equivalent integral model of the same level.*

- *Move the repeated root of $F_1(x, z) \bmod \pi$ to $(x : z) = (0 : 1)$.*
- *Replace $F(x, z)$ by $\pi^{-2} F(\pi x, z)$.*

(iii) *If F is nonminimal then the procedure in (ii) gives $v(F) \geq 2$ after at most 2 iterations.*

Proof. We first prove the theorem in the case $F \in \mathcal{O}_K[x, z]$ is nonminimal. By hypothesis there exists $[\mu, M] \in K^\times \times \mathrm{GL}_2(K)$ with

$$v(\mu \det(M)) \leq -1 \quad \text{and} \quad v(\mu^2 (F \circ M)) \geq 0,$$

i.e., the transform of F by $[\mu, M]$ has smaller level and is still integral. Hence $v(F \circ M) \geq 2v(\det M) + 2$. The slope s of F is defined to be the least possible valuation of $\det M$, for M such a matrix with entries in $\mathcal{O}_K$. By Lemma 4.1 we are free to replace F by any $\mathcal{O}_K$-equivalent binary quartic. So, putting M in Smith normal form, we may assume that $F(\pi^s x, z) \equiv 0 \pmod{\pi^{2s+2}}$ where s is the slope. For $s \geq 2$, this condition works out as $\pi^2 \mid c$, $\pi^{s+2} \mid d$ and $\pi^{2s+2} \mid e$. So the only possible slopes are $s = 0, 1, 2$ (as if these conditions hold for some $s > 2$, then they also hold for $s = 2$, and s was defined to be minimal). If $s = 0$, then $v(F) \geq 2$ contrary to hypothesis. If $s = 1$, then the coefficients of F have valuations satisfying ≥ 0, ≥ 1, ≥ 2, ≥ 3 and ≥ 4. So either $v(F) = 0$ and $F(x, z) \bmod \pi$ has a quadruple root at $(x : z) = (0 : 1)$, or $v(F) = 1$ and $\pi^{-1} F(x, z) \bmod \pi$ has a triple or quadruple root at $(x : z) = (0 : 1)$. If $s = 2$, then the coefficients of F have valuations satisfying ≥ 0, $= 0$, ≥ 2, ≥ 4 and ≥ 6. Then $F(x, z) \bmod \pi$ has a triple root at $(0 : 1)$. In each of these cases ($s = 1$ and $s = 2$) statements (i) and (ii) of the theorem are now immediate. Moreover the procedure in (ii) returns a K-equivalent

integral model of smaller slope. Hence at most 2 iterations are required to give $v(F) \geq 2$, establishing (iii).

It remains to prove (i) and (ii) in the case $\mathrm{char}(k) \neq 2$ and F has positive level (but could be minimal). Statement (i) follows from the fact that $F_1 \bmod \pi$ is a null form, i.e., both the invariants I and J vanish. (Since k is perfect the multiple root is defined over k.) For (ii) we must show that if $v(F) = 0$ and the reduction of $F \bmod \pi$ has a repeated root at $(x : z) = (0 : 1)$ then $\pi^2 \mid e$. But in this case there are smooth $\bar{k}$-points on the reduction of $\mathscr{C} \bmod \pi$ where $\mathscr{C} = \{y^2 = F(x, z)\}$. We then have $\mathscr{C}(K^{\mathrm{sh}}) \neq \varnothing$, and Theorem 3.4 shows that F is nonminimal over K^{sh}. Our earlier argument now applies, keeping in mind that π is still a uniformizer for K^{sh}. □

To give a satisfactory analogue of this algorithm when $\mathrm{char}(k) = 2$ we must work with generalised binary quartics. We give details in Section 4D.

4B. ***Minimisation of* 3*-coverings.*** The valuation $v(F)$ of a ternary cubic

$$F(x, y, z) = ax^3 + by^3 + cz^3 + a_2x^2y + a_3x^2z + b_1xy^2 + b_3y^2z + c_1xz^2 + c_2yz^2 + mxyz$$

is the minimum valuation of a coefficient. If $v(F) \geq 1$ then F is nonminimal, and indeed dividing through by π gives a K-equivalent integral model of smaller level. The algorithm for minimising ternary cubics is described in the following theorem.

Theorem 4.3. *Let $F \in X_3(\mathbb{O}_K)$ be a nonsingular ternary cubic. Suppose $v(F) = 0$, but F has positive level. Then*

(i) *The singular locus of the reduction*

$$\mathscr{S} = \left\{ (x : y : z) \in \mathbb{P}^2 \;\middle|\; F \equiv \frac{\partial F}{\partial x} \equiv \frac{\partial F}{\partial y} \equiv \frac{\partial F}{\partial z} \equiv 0 \pmod{\pi} \right\}$$

is either a point or a line, and is defined over k.

(ii) *The following procedure replaces F by a K-equivalent integral ternary cubic of the same level.*

- *Make a $\mathrm{GL}_3(\mathbb{O}_K)$-transformation to move the singular locus $\mathscr{S}$ to the point $(1 : 0 : 0)$, respectively the line $\{z = 0\}$.*
- *Replace $F(x, y, z)$ by $\pi F(\pi^{-1}x, y, z)$, respectively $\pi^{-1}F(x, y, \pi z)$.*

(iii) *If F is nonminimal then the procedure in (ii) gives $v(F) \geq 1$ after at most 4 iterations.*

Proof. We are given that F has positive level. It follows that its reduction $\bmod \pi$ is a null-form, i.e., the invariants c_4, c_6 and Δ all vanish. The classification of singular ternary cubics (up to equivalence over an algebraically closed field) is well known. See for example [Dolgachev 2003, §10.3] or [Poonen 2001]. The possible null-forms are either a cuspidal cubic, a line touching a conic, three lines

through a common point, a double line and a line, or a triple line. So over $\bar{k}$ the singular locus of the reduction is either a point or a line. Since k is perfect, this point or line is already defined over k. This proves (i).

Next we prove (ii) and (iii) in the case F is nonminimal. By hypothesis there exists $[\mu, M] \in K^\times \times \mathrm{GL}_3(K)$ with $v(\mu \det(M)) \leq -1$ such that the transform of F by $[\mu, M]$ is still integral. The slope s of F is the least possible valuation of $\det M$, for M such a matrix with entries in $\mathbb{O}_K$. By Lemma 4.1 we are free to replace F by any $\mathbb{O}_K$-equivalent ternary cubic. So putting the transformation matrix in Smith normal form we may assume that F satisfies

$$F(x, \pi^a y, \pi^b z) \equiv 0 \pmod{\pi^{a+b+1}} \tag{4-1}$$

for some $0 \leq a \leq b$ with $a+b=s$. If $a=b=0$ then $v(F) \geq 1$, contrary to hypothesis. If $a=0$ and $b \geq 1$ then the reduction of F mod π only involves the monomials xz^2, yz^2 and z^3. Hence $\mathcal{S}$ is the line $\{z=0\}$. If $a \geq 1$ then the coefficients of x^3, x^2y and x^2z all vanish mod π. Hence $\mathcal{S}$ is either the point $(1:0:0)$ or a line through this point. In each of these cases it is clear that the procedure in (ii) returns an integral model of the same level and smaller slope. Moreover it gives $v(F) \geq 1$ after a finite number of iterations (bounded by the initial slope). The next lemma shows that the only possible slopes are 0, 1, 2, 3 and 5. Hence at most 4 iterations are required, establishing (iii).

It remains to prove (ii) in the case F has positive level, but could be minimal. We must show that if $(1:0:0)$ is the only singular point on the reduction then $F(1,0,0) \equiv 0 \pmod{\pi^2}$. But in this case there are smooth $\bar{k}$-points on the reduction. The proof is completed exactly as in Theorem 4.2. □

We say that a pair (a, b) is *admissible* for F if (4-1) holds.

Lemma 4.4. *If some pair (a, b) with $0 \leq a \leq b$ is admissible for F then at least one of the pairs $(0,0)$, $(0,1)$, $(1,1)$, $(1,2)$ or $(2,3)$ is admissible for F.*

Proof. Suppose (a, b) is admissible for F. We make the observations:

- If $a=0$ and $b \geq 1$ then $(0,1)$ is admissible.
- If $a=b \geq 1$ then $(1,1)$ is admissible.
- If $a \geq 1$ and $b \geq 2a$ then $(1,2)$ is admissible.
- If $a \geq 2$ and $b \geq a+1$ then $(2,3)$ is admissible.

The only remaining possibility is $(a, b) = (0, 0)$. □

Example 4.5. We apply our algorithm to a cuspidal cubic (although this is singular, there are π-adically close smooth ternary cubics that are treated in the same way by our algorithm). An arrow labelled $(0, a, b)$ indicates that we make the

transformation $[\pi^{-a-b}, \mathrm{Diag}(1, \pi^a, \pi^b)]$.

$$\begin{aligned} xz^2 - y^3 &\xrightarrow{(0,1,1)} xz^2 - \pi y^3 \\ &\xrightarrow{(0,0,1)} \pi xz^2 - y^3 \\ &\xrightarrow{(0,1,0)} xz^2 - \pi^2 y^3 \\ &\xrightarrow{(0,0,1)} \pi(xz^2 - y^3). \end{aligned}$$

So this is an example where our algorithm takes the maximum possible of 4 iterations to give $v(F) \geq 1$.

4C. ***Minimisation of* 4*-coverings.*** In this subsection we prove the Strong Minimisation Theorem, 3.5(i), and Theorem 3.6 in the case $n = 4$, assuming that $\mathrm{char}(k) \neq 2$. The proofs are constructive and give an algorithm for minimising quadric intersections. The modifications required when $\mathrm{char}(k) = 2$ are described in the next subsection.

We define a map

$$\mathfrak{d} : X_4(K) \to X_2(K), \quad (Q_1, Q_2) \mapsto F(x, z) = \det(Ax + Bz), \tag{4-2}$$

where A and B are the matrices of second partial derivatives of Q_1 and Q_2. As noted in Definition 2.4 we have $\Delta(Q_1, Q_2) = 2^{-12}\Delta(F)$.

Lemma 4.6. *Let $(Q_1, Q_2) \in X_4(K)$ be a nonsingular quadric intersection. Then $F = \mathfrak{d}(Q_1, Q_2)$ is nonsingular, and there is a morphism of genus-one curves $\mathscr{C}_{(Q_1,Q_2)} \to \mathscr{C}_F$ defined over K.*

Proof. A formula for this morphism is given by classical invariant theory [An et al. 2001; Merriman et al. 1996]. We write the binary quartic $F = \mathfrak{d}(Q_1, Q_2)$ as $F(x, z) = ax^4 + bx^3z + cx^2z^2 + dxz^3 + ez^4$, and let T_1 and T_2 be the quadrics whose matrices of second partial derivatives M_1 and M_2 are determined by

$$\mathrm{adj}\big(\mathrm{adj}(A)x + \mathrm{adj}(B)z\big) = a^2Ax^3 + aM_1x^2z + eM_2xz^2 + e^2Bz^3. \tag{4-3}$$

Then $J^2 \equiv F(T_1, -T_2) \bmod (Q_1, Q_2)$ where $J = \dfrac{1}{4}\dfrac{\partial(Q_1, Q_2, T_1, T_2)}{\partial(x_1, x_2, x_3, x_4)}$. □

Lemma 4.7. *If $[M, N] \in \mathscr{G}_4(K)$ then there is a commutative diagram*

$$\begin{array}{ccc} X_4(K) & \xrightarrow{[M,N]} & X_4(K) \\ \downarrow{\scriptstyle \mathfrak{d}} & & \downarrow{\scriptstyle \mathfrak{d}} \\ X_2(K) & \xrightarrow{[\det N, M]} & X_2(K). \end{array}$$

In particular $\mathfrak{d}$ induces a well-defined map on K-equivalence classes.

Proof. This is clear. □

Following [Womack 2003], we deduce the Minimisation Theorem for $n=4$ from the $n=2$ version. The modifications required to prove Theorems 3.5(i) and 3.6 are given at the end of this subsection (see Proposition 4.12).

Proposition 4.8. *If $(Q_1, Q_2) \in X_4(K)$ is nonsingular and K-soluble then it is K-equivalent to an integral model of level* 0.

Proof. Since (Q_1, Q_2) is K-soluble, it follows by Lemma 4.6 that $\mathfrak{d}(Q_1, Q_2)$ is K-soluble. So by the Minimisation Theorem for $n=2$ we know that $\mathfrak{d}(Q_1, Q_2)$ is K-equivalent to an integral binary quartic $F(x, z)$ of level 0. It is clear by Lemma 4.7 that (Q_1, Q_2) is K-equivalent to a quadric intersection (Q_1', Q_2') with $\mathfrak{d}(Q_1', Q_2') = F$. The following lemma shows we may take (Q_1', Q_2') integral. This is then the required integral model of level 0. □

Notice that the next three lemmas are false when $\mathrm{char}(k)=2$, as we could otherwise use the above proof to find integral models of level $-v(2)$.

Lemma 4.9. *Let $(Q_1, Q_2) \in X_4(K)$ be a K-soluble nonsingular quadric intersection. If $\mathfrak{d}(Q_1, Q_2)$ is integral then (Q_1, Q_2) is K-equivalent to an integral quadric intersection (Q_1', Q_2') with $\mathfrak{d}(Q_1', Q_2') = \mathfrak{d}(Q_1, Q_2)$.*

Proof. By a transformation $[\mu I_2, I_4]$ for suitable $\mu \in \mathcal{O}_K$ we obtain an integral quadric intersection (Q_1', Q_2') with $\mathfrak{d}(Q_1', Q_2') = \mu^4 \mathfrak{d}(Q_1, Q_2)$. We now apply the following lemma, as many times as required, at each stage preserving the integrality of (Q_1', Q_2') while dividing $\mathfrak{d}(Q_1', Q_2')$ by a square in $\pi\mathcal{O}_K$. □

Recall that we write $v(F)$ for the minimum of the valuations of the coefficients of the binary quartic F. The following is Womack's "main reduction lemma".

Lemma 4.10. *Let $(Q_1, Q_2) \in X_4(\mathcal{O}_K)$ be a nonsingular K-soluble integral quadric intersection. If $F = \mathfrak{d}(Q_1, Q_2)$ satisfies $v(F) \geq 2$ then (Q_1, Q_2) is K-equivalent to an integral quadric intersection of smaller level by means of a transformation $[\lambda I_2, N] \in \mathcal{G}_4(K)$ with $\lambda \in K^\times$ and $N \in \mathrm{GL}_4(K)$.*

The following geometric lemma prepares for the proof of Lemma 4.10. We say that two pairs of quadratic forms in m variables are k-equivalent if they are in the same orbit for the natural action of $\mathrm{GL}_2(k) \times \mathrm{GL}_m(k)$. (This extends our earlier definition in the case $m=4$.) Over an algebraically closed field, the lemma may alternatively be deduced from the classification of pairs of quadrics using the Segre symbol, as given in [Hodge and Pedoe 1952, Chapter XIII, §11].

Lemma 4.11. *Let Q_1 and Q_2 be quadratic forms in $m = 3$ or 4 variables over a field k with* $\mathrm{char}(k) \neq 2$. *Let A and B be the matrices of second partial derivatives of Q_1 and Q_2. Assume that*

- $\{Q_1 = Q_2 = 0\} \subset \mathbb{P}^{m-1}$ *is not a cone, that is,* $\ker A \cap \ker B = 0$, *and*
- *the binary form $F(x, z) = \det(Ax + Bz)$ is identically zero.*

Then the k-equivalence class of (Q_1, Q_2) is uniquely determined:

(i) *If $m = 3$ then (Q_1, Q_2) is k-equivalent to (x_1x_2, x_2x_3).*

(ii) *If $m = 4$ then (Q_1, Q_2) is k-equivalent to $(x_1x_2, x_2x_3 - x_4^2)$.*

Proof. (i) We must show that the GCD of Q_1 and Q_2 is a linear form, and for this we may assume that k is algebraically closed. Since some quadric in the pencil has rank 2, we may assume that $Q_1 = x_1x_2$. Then the condition $\det(Ax + Bz) = 0$ works out as $b_{33} = b_{13}b_{23} = \det B = 0$. Swapping x_1 and x_2 if necessary, we may assume that $b_{13} = b_{33} = 0$. Then $b_{23} \neq 0$ (otherwise we would have a cone) and the condition $\det B = 0$ forces $b_{11} = 0$. Making a substitution for x_3 now puts (Q_1, Q_2) in the required form.

(ii) Suppose $\{Q_1 = Q_2 = 0\} \subset \mathbb{P}^3$ has a singular point defined over k. Moving this point to $(1:0:0:0)$, it is easy to reduce to the case

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & \multicolumn{2}{c}{\multirow{2}{*}{A'}} \\ 0 & 0 & & \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & * & * & * \\ 0 & * & \multicolumn{2}{c}{\multirow{2}{*}{B'}} \\ 0 & * & & \end{pmatrix}.$$

The condition $\det(Ax + Bz) = 0$ now becomes $\det(A'x + B'z) = 0$. Hence we may assume that A' and B' are scalar multiples of

$$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Then $b_{23} \neq 0$ (otherwise we have a cone) and a substitution in x_3 brings us to the case

$$(Q_1, Q_2) = (x_1x_2 + \lambda x_4^2, x_2x_3 + \mu x_4^2)$$

for some $\lambda, \mu \in k$. Replacing one of these quadrics by a suitable linear combination, and then making a substitution in x_1 and x_3 to compensate, we may assume that $\lambda = 0$. Then $\mu \neq 0$ (otherwise we have a cone) and we rescale to get $\mu = -1$.

By Theorem 2.8(ii) there is a singular point defined over $\bar{k}$. So running the above proof over $\bar{k}$ shows that $\{Q_1 = Q_2 = 0\} \subset \mathbb{P}^3$ is the union of a conic and a line, meeting at a unique point. This point of intersection is a k-rational singular point. Our earlier proof now applies. □

Proof of Lemma 4.10. We write $\overline{Q}_1, \overline{Q}_2$ for the reductions of Q_1, Q_2 modulo π. In the proof we often arrive at one of the following three special situations.

Situation 1: The reduction $\mathscr{C}_{(\overline{Q}_1, \overline{Q}_2)}$ contains a plane defined over k. By means of a $\mathrm{GL}_4(\mathbb{O}_K)$-transformation we may move the plane to $\{x_1 = 0\}$. We apply the transformation $[\pi^{-1}I_2, \mathrm{Diag}(\pi, 1, 1, 1)]$ to give an integral model of smaller level.

Situation 2: The reduction $\mathscr{C}_{(\overline{Q}_1,\overline{Q}_2)}$ is a cone over a point $\mathbf{x} \in \mathbb{P}^3(k)$, and moreover $Q_1(\mathbf{x}) \equiv Q_2(\mathbf{x}) \equiv 0 \pmod{\pi^2}$. By a $\mathrm{GL}_4(\mathbb{O}_K)$-transformation we may move the point to $(1:0:0:0)$. We apply the transformation $[I_2, \mathrm{Diag}(\pi^{-1}, 1, 1, 1)]$ to give an integral model of smaller level.

Situation 3: The reduction $\mathscr{C}_{(\overline{Q}_1,\overline{Q}_2)}$ contains a line defined over k. By a $\mathrm{GL}_4(\mathbb{O}_K)$-transformation we may move the line to $\{x_1 = x_2 = 0\}$. The flip-flop transformation $[\pi^{-1} I_2, \mathrm{Diag}(\pi, \pi, 1, 1)]$ gives an integral model of the same level.

Let A and B be the matrices of second partial derivatives of Q_1 and Q_2. Let $\bar{A}$ and $\bar{B}$ be their reductions mod π. We split into cases according to the value of the *common nullity*, defined as $s = \dim(\ker \bar{A} \cap \ker \bar{B})$.

If $s = 0$ then by Lemma 4.11(ii) we are in Situation 3. Applying the flip-flop transformation brings us to the case $s \geq 1$.

If $s = 1$ we may assume that $\overline{Q}_1$ and $\overline{Q}_2$ are quadratic forms in x_2, x_3 and x_4 only. Let A' and B' be the 3 by 3 matrices of second partial derivatives. Then

$$F(x, z) \equiv (a_{11}x + b_{11}z) \det(A'x + B'z) \pmod{\pi^2}. \tag{4-4}$$

Since $v(F) \geq 2$ we have either $a_{11} \equiv b_{11} \equiv 0 \pmod{\pi^2}$, in which case we are in Situation 2, or $\det(A'x + B'z) = 0$, in which case Lemma 4.11(i) shows we are in Situation 1.

If $s \geq 2$ we may assume that $\overline{Q}_1$ and $\overline{Q}_2$ are binary quadratic forms in x_1 and x_2. If $\overline{Q}_1$ and $\overline{Q}_2$ simultaneously represent 0 over k, then we are in Situation 1. Otherwise we apply the flip-flop transformation $[\pi^{-1} I_2, \mathrm{Diag}(\pi, \pi, 1, 1)]$ to give an integral model (R_1, R_2) of the same level. Then $\bar{R}_1$ and $\bar{R}_2$ are binary quadratic forms in x_3 and x_4. If $\bar{R}_1$ and $\bar{R}_2$ simultaneously represent 0 over k then we are in Situation 1. Otherwise we obtain a contradiction to our hypothesis that (Q_1, Q_2) is K-soluble. Indeed if $(x_1 : x_2 : x_3 : x_4)$ were a K-point with $\min\{v(x_i) : 1 \leq i \leq 4\} = 0$ then from $Q_1(\mathbf{x}) \equiv Q_2(\mathbf{x}) \equiv 0 \pmod{\pi}$ we deduce $x_1 \equiv x_2 \equiv 0 \pmod{\pi}$ and from $Q_1(\mathbf{x}) \equiv Q_2(\mathbf{x}) \equiv 0 \pmod{\pi^2}$ we deduce $x_3 \equiv x_4 \equiv 0 \pmod{\pi}$. □

This completes the proof of Proposition 4.8. We now modify the proof so that we can deduce Theorems 3.5(i) and 3.6 in the case $n = 4$ from the corresponding results for $n = 2$. The situation considered at the end of the last paragraph motivates the definition of a critical model, see Definition 5.1(iii) below.

Proposition 4.12. *If $(Q_1, Q_2) \in X_4(K)$ is nonsingular then it is K-equivalent to either an integral model $\Phi \in X_4(\mathbb{O}_K)$ with $\mathfrak{d}(\Phi)$ minimal (and hence Φ minimal), or a critical model* (as specified in Definition 5.1(iii), page 793).

Proof. By Lemma 4.7 we may assume that $\mathfrak{d}(Q_1, Q_2)$ is a minimal binary quartic. We then follow the proof of Lemma 4.9, but without the hypothesis of K-solubility. This hypothesis was only used at the end of the proof of Lemma 4.10. We may

assume that one of the pairs, say $\overline{Q}_1$ and $\overline{Q}_2$, simultaneously represents 0 over $\bar{k}$. (Otherwise we would have a critical model.) If they do not simultaneously represent 0 over k, then they must be linearly dependent. So it is clear we can reduce the level, but not necessarily using a transformation of the specified form. In the proof of Lemma 4.9 we repeatedly applied Lemma 4.10. For the final application it does not matter what transformation we use. In all earlier applications we have $v(F) \geq 3$. If A_1, B_1 and A_2, B_2 are the 2 by 2 matrices representing the pairs of binary quadratic forms $\overline{Q}_1$, $\overline{Q}_2$ and $\bar{R}_1$, $\bar{R}_2$ then

$$F(x, z) \equiv \pi^2 \det(A_1 x + B_1 z) \det(A_2 x + B_2 z) \pmod{\pi^3}.$$

The hypothesis $v(F) \geq 3$ therefore ensures that one of the pairs simultaneously represents 0 over k. We are then in Situation 1 and can proceed as before. □

In Lemma 5.3 (see below) we show that critical models are minimal. Hence following the proof of Proposition 4.12 gives an algorithm for minimising quadric intersections, even in the case they are not K-soluble.

Proposition 4.12 also allows us to deduce the case $n = 4$ of Theorems 3.5(i) and 3.6 from the case $n = 2$. Here we use the easy facts that critical models are K^{sh}-insoluble, and remain critical after any unramified field extension.

4D. ***Minimisation in residue characteristic* 2.** We describe how to modify our algorithms in the cases $n = 2, 4$ when $\mathrm{char}(k) = 2$. In the case $n = 2$ the issue is that we must work with generalised binary quartics instead of just binary quartics. Recall that a generalised binary quartic, or genus-one model of degree 2, is an equation of the form

$$y^2 + P(x, z)y = Q(x, z),$$

where P and Q are homogeneous polynomials of degrees 2 and 4. The coefficients of P and Q are labelled l, m, n and a, b, c, d, e. Notice that in characteristic 2 the binary quadratic form $\partial^2 Q/\partial x \partial z = bx^2 + dz^2$ is a covariant of the quartic Q. Moreover this covariant vanishes if and only if Q is a square. (Recall that k is perfect, and so every element of k is a square.)

We say that two models are *y-equivalent* if they are related by a *y-substitution*, that is, a substitution of the form $x \leftarrow x$, $z \leftarrow z$, $y \leftarrow y + r_0 x^2 + r_1 xz + r_2 z^2$. The *valuation* of $(P, Q) \in X_2(\mathbb{O}_K)$ is

$$v(P, Q) = \max\{\min(2v(P'), v(Q')) : (P', Q') \text{ is } y\text{-equivalent to } (P, Q)\}.$$

It is easy to check that $v(P, Q)$ only depends on the $\mathbb{O}_K$-equivalence class of (P, Q).

If $v(P) = 0$ or $v(P) \geq 1$ and $Q(x, z)$ is not a square mod π then $v(P, Q) = 0$. Otherwise we can make a y-substitution so that $v(Q) \geq 1$. Then either $v(Q) = 1$, in which case $v(P, Q) = 1$, or $v(Q) \geq 2$, in which case (P, Q) is nonminimal, and

indeed dividing P and Q through by π and π^2 gives a K-equivalent integral model of smaller level. Theorem 4.2 has the following analogue.

Theorem 4.13. *Let $(P, Q) \in X_2(\mathbb{O}_K)$ be a nonsingular generalised binary quartic. Suppose that $v(P, Q) = 0$ or 1, but (P, Q) has positive level.*

(i) *The reduction* mod π *of*

$$Q_1(x, z) = \begin{cases} P(x, z) & \text{if } v(P) = 0, \\ \dfrac{\partial^2 Q}{\partial x \partial z} & \text{if } v(P) \geq 1 \text{ and } v(P, Q) = 0, \\ \pi^{-1} Q(x, z) & \text{if } v(P) \geq 1 \text{ and } v(Q) = 1 \end{cases}$$

has a unique repeated root defined over k.

(ii) *The following procedure replaces (P, Q) by a K-equivalent integral model of the same level.*

- *If $v(P, Q) = 1$ then make a y-substitution so that $v(Q) \geq 1$.*
- *Move the repeated root of $Q_1(x, z)$* mod π *to $(x : z) = (0 : 1)$.*
- *Make a y-substitution so that $\pi \mid e$. (This is possible since $\pi \mid n$ and every element of k is a square.)*
- *Replace $P(x, z)$ by $\pi^{-1} P(\pi x, z)$ and $Q(x, z)$ by $\pi^{-2} Q(\pi x, z)$.*

(iii) *If (P, Q) is nonminimal then the procedure in (ii) gives $v(P, Q) \geq 2$ after at most 2 iterations.*

Proof. We first show that if (i) holds for (P, Q) then it holds for any $\mathbb{O}_K$-equivalent model (P', Q'). We say that forms $f, g \in k[x, z]$ are *k-equivalent* if

$$f(x, z) = \lambda g(\alpha x + \beta z, \gamma x + \delta z)$$

for some $\lambda, \alpha, \beta, \gamma, \delta \in k$ with $\lambda(\alpha\delta - \beta\gamma) \neq 0$. Each of the following claims is an easy consequence of the definition of $\mathbb{O}_K$-equivalence (as given in Section 2) and our assumption that $\mathrm{char}(k) = 2$.

- The reductions mod π of $P(x, z)$ and $P'(x, z)$ are k-equivalent; in particular, $v(P) = 0 \iff v(P') = 0$.
- If $v(P) \geq 1$ then the reductions mod π of $\partial^2 Q/\partial x \partial z$ and $\partial^2 Q'/\partial x \partial z$ are k-equivalent; note that $v(P, Q) = v(P', Q')$.
- If $v(P) \geq 1$ and $v(Q) = v(Q') = 1$ then the reductions mod π of $\pi^{-1} Q(x, z)$ and $\pi^{-1} Q'(x, z)$ are k-equivalent.

It is now clear that if (i) holds for (P, Q) then it holds for (P', Q'). Next we show that the procedure in (ii) gives a well-defined map on $\mathbb{O}_K$-equivalence classes. This does not automatically follow from Lemma 4.1, as we must also consider y-substitutions. Suppose we start with some model satisfying (i), and carry out the

first three steps of the procedure in (ii) in two different ways. The result is a pair of $\mathbb{O}_K$-equivalent models (P, Q) and (P', Q') related by some $[1, r, M] \in \mathcal{G}_2(\mathbb{O}_K)$. Since the reduction of M mod π fixes the repeated root $(0:1)$ we have $\pi \mid m_{21}$. Labelling the coefficients of (P, Q) in the usual way, and likewise for (P', Q'), we have $\pi \mid n, e$ and $\pi \mid n', e'$. Therefore $\pi \mid r_2$. It is now routine to check that if (ii) holds for (P, Q), i.e. $\pi \mid n, d$ and $\pi^2 \mid e$, then (ii) holds for (P', Q'), i.e. $\pi \mid n', d'$ and $\pi^2 \mid e'$. Moreover the transformed models are related by

$$[1, (\pi r_0, r_1, \pi^{-1} r_2), \mathrm{Diag}(\pi, 1) M \,\mathrm{Diag}(\pi^{-1}, 1)] \in \mathcal{G}_2(\mathbb{O}_K).$$

Thus the procedure gives a well-defined map on $\mathbb{O}_K$-equivalence classes.

We are now free in the proof to replace (P, Q) by any $\mathbb{O}_K$-equivalent model. So if (P, Q) is nonminimal we may assume that $P(\pi^s x, z) \equiv 0 \pmod{\pi^{s+1}}$ and $Q(\pi^s x, z) \equiv 0 \pmod{\pi^{2s+2}}$ for some integer $s \geq 0$. We call the least such integer s the *slope*. As happened for binary quartics, the only possible slopes are $s = 0, 1, 2$. If $s = 0$ then $v(P, Q) \geq 2$ contrary to hypothesis. If $s = 1$ then the coefficients of (P, Q) have valuations satisfying

$$\geq 0, \quad \geq 1, \quad \geq 2; \qquad \geq 0, \quad \geq 1, \quad \geq 2, \quad \geq 3, \quad \geq 4.$$

If $v(P) = 0$ then $P(x, z)$ mod π has a double root at $(x:z) = (0:1)$. Otherwise, since every element of k is a square (recall k is perfect), we can make a y-substitution $y \leftarrow y + r_0 x^2$ so that $v(Q) \geq 1$. Then $\pi^{-1} Q(x, z)$ mod π has either a triple or quadruple root at $(x : z) = (0 : 1)$. If $s = 2$ then the coefficients of (P, Q) have valuations satisfying

$$\geq 0, \quad \geq 1, \quad \geq 3; \qquad \geq 0, \quad = 0, \quad \geq 2, \quad \geq 4, \quad \geq 6.$$

So in this case $v(P, Q) = 0$. If $v(P) = 0$ then $P(x, z)$ mod π has a double root at $(x : z) = (0 : 1)$. Otherwise $bx^2 + dz^2$ mod π has a double root at $(x : z) = (0 : 1)$. In each of the cases $s = 1, 2$ it is now clear that the procedure in (ii) returns a K-equivalent integral model of smaller slope. Hence at most 2 iterations are required to give $v(P, Q) \geq 2$.

It remains to prove (i) and (ii) in the case (P, Q) has positive level (but could be minimal). If (P, Q) is K^{sh}-soluble then after an unramified extension we have $\mathcal{C}_{(P,Q)}(K) \neq \varnothing$. Then Theorem 3.4 shows that (P, Q) is non-minimal, and our earlier argument applies. Otherwise, we show in Proposition 5.6 below, that (P, Q) is $\mathbb{O}_K$-equivalent to a model whose coefficients have valuations satisfying

$$\geq 1, \quad \geq 1, \quad \geq 2; \qquad = 1, \quad \geq 2, \quad \geq 2, \quad \geq 3, \quad = 3.$$

Statements (i) and (ii) are then clear. □

Next we modify the algorithm for minimising quadric intersections, as presented in Section 4C. First we replace $\mathfrak{d}$ by the map

$$\begin{aligned} &\mathfrak{d}' : X_4(K) \to X_2(K), \\ &(Q_1, Q_2) \mapsto (P, Q) = (\mathrm{pf}(xQ_1 + zQ_2), \mathrm{rd}(xQ_1 + zQ_2)), \end{aligned} \tag{4-5}$$

where pf and rd were defined in the proof of Lemma 2.9. Then

$$\Delta(Q_1, Q_2) = \Delta(P, Q).$$

We call (P, Q) the *doubling* of (Q_1, Q_2). (The reason for this name is that $\mathfrak{d}'$ acts as multiplication-by-2 on the Weil–Châtelet group.) The analogue of Lemma 4.6 (using $\mathfrak{d}'$ instead of $\mathfrak{d}$) is immediate if $\mathrm{char}(K) \neq 2$. Indeed, the covering map $\mathscr{C}_{(Q_1,Q_2)} \to \mathscr{C}_{(P,Q)}$ is given by

$$(x_1 : x_2 : x_3 : x_4) \mapsto (T_1 : -T_2 : J'),$$

where $J' = \frac{1}{2}(J - lT_1^2 + mT_1T_2 - nT_2^2)$, and l, m, n are the coefficients of P. If $\mathrm{char}(K) = 2$, then the role of J' is taken by

$$\begin{aligned} J'' = \tfrac{1}{2}\big(J - lT_1^2 + mT_1T_2 - nT_2^2 + mn(lT_1 + mT_2)Q_1 + lm(nT_2 + mT_1)Q_2 \\ + l^2n^3Q_1^2 + lmn(ln + m^2)Q_1Q_2 + l^3n^2Q_2^2\big). \end{aligned}$$

It may be verified by direct calculation that T_1, T_2 and J'' have coefficients in $\mathbb{Z}[X_4]$. Moreover T_1 and T_2 cannot both vanish identically on $\mathscr{C}_{(Q_1,Q_2)}$. (We checked this for the models specified in Lemma 3.11, and then used the covariance of T_1 and T_2.) Hence in all characteristics there is a morphism $\mathscr{C}_{(Q_1,Q_2)} \to \mathscr{C}_{(P,Q)}$ given by $(x_1 : x_2 : x_3 : x_4) \mapsto (T_1 : -T_2 : J'')$.

The diagram in Lemma 4.7 (using $\mathfrak{d}'$ instead of $\mathfrak{d}$) no longer commutes, but it does commute up to y-equivalence, and this is sufficient for our purposes.

Definition 4.14. Let $Q \in k[x_1, \dots, x_m]$ be a quadratic form in m variables.

(i) The *kernel* $\ker Q$ of Q is the subspace of k^m defined by the vanishing of Q and all its partial derivatives. (Recall that k is perfect, so the restriction of Q to the subspace where all the partial derivatives vanish is the square of a linear form.) The *rank* of Q is $m - \dim \ker Q$.

(ii) The *discriminant* of Q is

$$\Delta_m(Q) = \begin{cases} \det(\partial^2 Q/\partial x_i \partial x_j) & \text{if } m \text{ is even,} \\ \frac{1}{2}\det(\partial^2 Q/\partial x_i \partial x_j) & \text{if } m \text{ is odd.} \end{cases}$$

The discriminant Δ_m is a polynomial in the coefficients of Q with integer coefficients. Therefore Definition 4.14(ii) is valid in all characteristics. Recall that we defined pf and rd so that $\Delta_4(Q) = \mathrm{pf}(Q)^2 + 4\,\mathrm{rd}(Q)$.

Lemma 4.15. *Let Q_1 and Q_2 be quadratic forms in $m = 3$ or 4 variables over a field k with* $\mathrm{char}(k) = 2$. *Assume that*

- $\{Q_1 = Q_2 = 0\} \subset \mathbb{P}^{m-1}$ *is not a cone, i.e.,* $\ker Q_1 \cap \ker Q_2 = 0$, *and*
- *if $m = 3$ then $\Delta_3(xQ_1 + zQ_2) = 0$, whereas if $m = 4$ then* $\mathrm{pf}(xQ_1 + zQ_2) = 0$ *and* $\mathrm{rd}(xQ_1 + zQ_2)$ *is a square.*

Then the k-equivalence class of (Q_1, Q_2) is uniquely determined, and is as given in Lemma 4.11.

Proof. This is similar to the proof of Lemma 4.11. □

In Lemma 4.10 we made the hypothesis that $v(F) \geq 2$ where $F = \mathfrak{d}(Q_1, Q_2)$. This should now be replaced by the hypothesis that $\mathfrak{d}'(Q_1, Q_2)$ is y-equivalent to a model (P, Q) with $v(P) \geq 1$ and $v(Q) \geq 2$. Then

$$\begin{aligned} P(x, z) &= \mathrm{pf}(xQ_1 + zQ_2) + 2h(x, z), \\ Q(x, z) &= \mathrm{rd}(xQ_1 + zQ_2) - \mathrm{pf}(xQ_1 + zQ_2)h(x, z) - h(x, z)^2, \end{aligned} \tag{4-6}$$

for some $h \in K[x, z]$. Since (Q_1, Q_2) is integral it follows that $h \in \mathcal{O}_K[x, z]$. Then $\mathrm{pf}(x\overline{Q}_1 + z\overline{Q}_2) = 0$ and $\mathrm{rd}(x\overline{Q}_1 + z\overline{Q}_2)$ is a square. Moreover if $\mathrm{rd}(xQ_1 + zQ_2)$ vanishes mod π then it vanishes mod π^2.

The common nullity is $s = \dim(\ker \overline{Q}_1 \cap \ker \overline{Q}_2)$. In the case $s = 1$ we may assume that Q_1 and Q_2 reduce to quadratic forms in x_2, x_3 and x_4 only. Call these Q_1' and Q_2'. The analogue of (4-4) is

$$\mathrm{rd}(xQ_1 + zQ_2) \equiv (\alpha x + \beta z)\Delta_3(xQ_1' + zQ_2') \pmod{\pi^2}$$

where α and β are the coefficients of x_1^2 in Q_1 and Q_2. In all other respects, the proof of the Lemma 4.10 goes through as before. By repeated application of this lemma we obtain the following analogue of Lemma 4.9.

Lemma 4.16. *Let $(Q_1, Q_2) \in X_4(K)$ be a K-soluble nonsingular quadric intersection. If $\mathfrak{d}'(Q_1, Q_2)$ is y-equivalent to an integral generalised binary quartic then (Q_1, Q_2) is K-equivalent to an integral quadric intersection (Q_1', Q_2') such that $\mathfrak{d}'(Q_1', Q_2')$ is y-equivalent to $\mathfrak{d}'(Q_1, Q_2)$.*

The Minimisation Theorem for $n = 4$ now follows from the Minimisation Theorem for $n = 2$ exactly as before.

The proof of Proposition 4.12 (with $\mathfrak{d}$ replaced by $\mathfrak{d}'$) is modified as follows. We follow the proof of Lemma 4.16 but without the hypothesis of K-solubility. This hypothesis is only used when $s \geq 2$. In this case

$$(\overline{Q}_1, \overline{Q}_2) = (\alpha_{11}x_1^2 + \alpha_{12}x_1x_2 + \alpha_{22}x_2^2, \beta_{11}x_1^2 + \beta_{12}x_1x_2 + \beta_{22}x_2^2)$$

and applying the transformation $[\pi^{-1}I_2, \mathrm{Diag}(\pi, \pi, 1, 1)]$ gives (R_1, R_2) with

$$(\bar{R}_1, \bar{R}_2) = (\gamma_{33}x_3^2 + \gamma_{34}x_3x_4 + \gamma_{44}x_4^2, \delta_{33}x_3^2 + \delta_{34}x_3x_4 + \delta_{44}x_4^2).$$

We must show that if $\bar{Q}_1$ and $\bar{Q}_2$ are linearly dependent and $\mathfrak{d}'(Q_1, Q_2)$ is y-equivalent to a model (P, Q) with $v(P) \geq 2$ and $v(Q) \geq 3$ then one of the pairs $\bar{Q}_1$, $\bar{Q}_2$ or $\bar{R}_1$, $\bar{R}_2$ simultaneously represents 0 over k. Since $s \geq 2$ we already know that $\mathrm{pf}(xQ_1 + zQ_2)$ vanishes mod π and $\mathrm{rd}(xQ_1 + zQ_2)$ vanishes mod π^2. It follows by (4-6) that $\mathrm{pf}(xQ_1 + zQ_2)$ vanishes mod π^2 and $\pi^{-2}\,\mathrm{rd}(xQ_1 + zQ_2)$ is a square mod π. Hence

$$\alpha_{12}\gamma_{34} = \beta_{12}\delta_{34} = \alpha_{12}\delta_{34} + \beta_{12}\gamma_{34} = 0$$

and

$$\alpha_{12}^2(\gamma_{33}\delta_{44} + \gamma_{44}\delta_{33}) + \gamma_{34}^2(\alpha_{11}\beta_{22} + \alpha_{22}\beta_{11}) = 0$$
$$\beta_{12}^2(\gamma_{33}\delta_{44} + \gamma_{44}\delta_{33}) + \delta_{34}^2(\alpha_{11}\beta_{22} + \alpha_{22}\beta_{11}) = 0.$$

Since $\bar{Q}_1$ and $\bar{Q}_2$ are linearly dependent we have $\alpha_{11}\beta_{22} + \alpha_{22}\beta_{11} = 0$. So either $\alpha_{12} = \beta_{12} = 0$, in which case $\bar{Q}_1$ and $\bar{Q}_2$ simultaneously represent 0 over k, or $\gamma_{34} = \delta_{34} = \gamma_{33}\delta_{44} + \gamma_{44}\delta_{33} = 0$ in which case $\bar{R}_1$ and $\bar{R}_2$ simultaneously represent 0 over k.

4E. ***Minimisation over global fields.*** Our theorems and algorithms for minimisation, as given above, were stated for genus-one models defined over a local field. We now discuss the global situation, and in particular prove Theorem 1.1. We restate that here in a more precise form. A genus-one model defined over a number field is called *integral* if its coefficients are algebraic integers.

Theorem 4.17. *Let $n = 2, 3$ or 4. Let K be a number field of class number one. Let Φ be a nonsingular genus-one model of degree n defined over K, and let E be the Jacobian of $\mathscr{C}_\Phi$. Suppose that $\mathscr{C}_\Phi$ is locally soluble at all finite places. Then there is a K-equivalent integral genus-one model with discriminant equal to the discriminant of a global minimal Weierstrass equation for E.*

Proof. To deduce this result directly from the statement of the Minimisation Theorem, 3.4, one would be led to using a version of strong approximation. See [Fisher 2007] for details in the cases $n = 2, 3$. The case $n = 4$ is similar. Although these proofs are not difficult, it is a notable advantage of the algorithmic approach taken in this section that the passage from local to global becomes a triviality.

Indeed, suppose K is a number field with class number one. Let $\mathfrak{p}$ be a prime of K and put $k = \mathbb{O}_K/\mathfrak{p}$. Then for any pair of m-dimensional subspaces $U, V \subset k^n$ there exists $M \in \mathrm{SL}_n(\mathbb{O}_K)$ whose reduction mod $\mathfrak{p}$ takes U to V. (Indeed, the case $\dim U = \dim V = 1$ is Lemma 3.13, and the general case is similar.) Since $\mathfrak{p}$ is

principal, say $\mathfrak{p} = \pi\mathbb{O}_K$, we can therefore follow the algorithms for minimising at $\mathfrak{p}$ using π as the uniformiser, without changing the level (or integrality) at other primes.

After first scaling the given model to be integral at all primes, we apply this procedure to the finite number of primes at which the resulting model has positive level. This gives an integral model which has level zero at all primes of K. By the definition of level, the discriminant of this model is equal to that of any global minimal model for E, up to a unit factor. Since this unit must be a 12th power, a final scaling by a suitable global unit gives the result. □

Theorem 1.1 is an immediate corollary, since every n-covering which is locally soluble at all places of K can be represented by a genus-one model defined over K.

To extend this theorem to a general number field K, we may replace integrality by S-integrality, where S is a (finite) set of primes generating the class group, so that the ring of S-integers is a principal ideal domain. The minimal model may then only be S-integral rather than integral. Just as with Weierstrass models for elliptic curves, there may be no global minimal model when the class number is greater than 1. In practice, we can alternatively find models which are simultaneously minimal at all primes in any given finite set, while being at least integral at all other primes.

Similar results may be deduced from our local results in the case where K is a function field, i.e., a finite extension of $\mathbb{F}_q(t)$.

5. Minimisation of insoluble genus-one models

We return to working over a discrete valuation field K as specified in Section 3A. In this section we prove the Converse Theorem, 3.5(ii), which shows that the Strong Minimisation Theorem, 3.5(i), is best possible.

Definition 5.1. (i) A generalised binary quartic $(P, Q) \in X_2(\mathbb{O}_K)$ is *critical* if the valuations of its coefficients l, m, n, a, b, c, d and e satisfy

$$\geq 1, \quad \geq 1, \quad \geq 2; \qquad = 1, \quad \geq 2, \quad \geq 2, \quad \geq 3, \quad = 3.$$

(ii) A ternary cubic $F \in X_3(\mathbb{O}_K)$ is *critical* if the valuations of its coefficients satisfy the inequalities indicated in the following diagram.

$$\begin{array}{ccccccccc}
 & & & z^3 & & & & & \\
 & & xz^2 & & yz^2 & & & & \\
 & x^2z & & xyz & & y^2z & & & \\
x^3 & & x^2y & & xy^2 & & y^3 & & \\
 & & & & & & & & \\
 & & & & & & & & \\
 & & & & & &

and on putting

$$(R_1, R_2) = [\pi^{-1} I_2, \mathrm{Diag}(\pi, \pi, 1, 1)](Q_1, Q_2)$$

the reductions of R_1 and R_2 mod π are quadratic forms in x_3 and x_4 with no common root in $\mathbb{P}^1(\bar{k})$.

We show in the next three lemmas that critical models are insoluble, minimal and of positive level. We then show (for $n = 2, 3$) that every K^{sh}-insoluble model is K-equivalent to a critical model. There is a corresponding result for models of degree $n = 4$.

Lemma 5.2. *Critical models are insoluble over K.*

Proof. We give details in the case $n = 2$. Suppose $(x, y, z) \in K^3$ is a nonzero solution of $y^2 + P(x, z)y = Q(x, z)$. Clearing denominators we may assume that $\min\{v(x), v(z)\} = 0$. It follows that $y \in \mathbb{O}_K$. Then reducing the equation mod π^i for $i = 1, 2, 3, 4$ we successively deduce $\pi \mid y$, $\pi \mid x$, $\pi^2 \mid y$ and $\pi \mid z$. In particular $\min\{v(x), v(z)\} > 0$. This is the required contradiction. The cases $n = 3, 4$ are similar. □

Since the definition of a critical model is unchanged by an unramified field extension, it follows immediately that critical models are insoluble over K^{sh}.

Lemma 5.3. *Critical models are minimal.*

Proof. In the cases $n = 2, 3$ we give a very quick proof. Indeed, if Φ were nonminimal, then our algorithms in Sections 4A, 4B and 4D would succeed in reducing the level. But on the contrary, when given a critical model, these algorithms endlessly cycle between two or three $\mathbb{O}_K$-equivalence classes. (Treating the case $n = 4$ in the same way would give a circular argument, as the current lemma was cited at the end of Section 4C.)

Alternatively we can imitate the proof of Lemma 5.2. We give details in the case $n = 4$. We define

$$s(Q_1, Q_2) = \max\{-v(\det M) : [M, I_4](Q_1, Q_2) \in X_4(\mathbb{O}_K)\}.$$

Suppose $[M, N] \in \mathcal{G}_4(K)$ is a transformation taking the critical model $\Phi = (Q_1, Q_2)$ to an integral model of smaller level. We may assume that N has entries in $\mathbb{O}_K$, not all in $\pi\mathbb{O}_K$. Let $\xi_j(x_1, \ldots, x_4) = \sum_{i=1}^4 n_{ij} x_i$. For $i = 1, 2$ we put

$$Q_i \circ N = Q_i(\xi_1, \ldots, \xi_4) \in \mathbb{O}_K[x_1, \ldots, x_4].$$

Our hypothesis is that $s(Q_1 \circ N, Q_2 \circ N) > v(\det N)$.

If $v(Q_1 \circ N) = 0$ then replacing Q_2 by $Q_2 + \lambda Q_1$ for suitable $\lambda \in \mathbb{O}_K$ we may assume that $v(Q_2 \circ N) > v(\det N)$. To understand this last condition, we put N in Smith normal form. Explicitly we write $N = U\,\mathrm{Diag}(\pi^a, \pi^b, \pi^c, 1)V$

for some $U, V \in \mathrm{GL}_4(\mathbb{O}_K)$ and $a \geq b \geq c \geq 0$. Since $v(Q_2) = 0$ we must have $2a > v(\det N) = a + b + c$ and therefore $a - b + c \geq 1$. It follows that $Q_2 \circ V \equiv x_1\big(\sum_{i=1}^4 \epsilon_i x_i\big) \pmod{\pi^2}$ for some $\epsilon_i \in \mathbb{O}_K$ with $\epsilon_2 \equiv \epsilon_3 \equiv \epsilon_4 \equiv 0 \pmod{\pi}$. In other words, $Q_2 \equiv \mu l_1 l_2 \pmod{\pi^2}$ for some $\mu \in \mathbb{O}_K$ and linear forms l_1, l_2 in $\mathbb{O}_K[x_1, \ldots, x_4]$ with $l_1 \equiv l_2 \pmod{\pi}$. This contradicts the definition of a critical model (as it would follow that R_2 vanishes mod π). Hence $v(Q_1 \circ N) \geq 1$. Similarly $v(Q_2 \circ N) \geq 1$. Since $\overline{Q}_1$ and $\overline{Q}_2$ are binary quadratic forms with no common root we deduce $\xi_1 \equiv \xi_2 \equiv 0 \pmod{\pi}$. Let $\xi_i' = \pi^{-1}\xi_i$ for $i = 1, 2$. We put

$$(R_1, R_2) = [\pi^{-1} I_2, \mathrm{Diag}(\pi, \pi, 1, 1)](Q_1, Q_2).$$

Let N' be the matrix with columns the coefficients of ξ_3, ξ_4, ξ_1' and ξ_2'. Then (R_1, R_2) is a critical model and $s(R_1 \circ N', R_2 \circ N') > v(\det N')$. Repeating the same arguments we deduce $\xi_3 \equiv \xi_4 \equiv 0 \pmod{\pi}$. This contradicts our scaling of the matrix N. □

The next lemma describes the possible levels of a critical model. For this we need to work explicitly with the a-invariants defined in the proof of Lemma 2.9. Although $a_1, \ldots, a_6$ are not invariants (in the sense of Definition 2.7), they are isobaric in the sense that

$$\begin{aligned}
&n = 2, && a_i \circ [\mu, 0, \mathrm{Diag}(\xi_1, \xi_2)] = (\mu\xi_1\xi_2)^i a_i,\\
&n = 3, && a_i \circ [\mu, \mathrm{Diag}(\xi_1, \xi_2, \xi_3)] = (\mu\xi_1\xi_2\xi_3)^i a_i,\\
&n = 4, && a_i \circ [\mathrm{Diag}(\mu_1, \mu_2), \mathrm{Diag}(\xi_1, \xi_2, \xi_3, \xi_4)] = (\mu_1\mu_2\xi_1\xi_2\xi_3\xi_4)^i a_i,
\end{aligned}$$

for all i. (We use the notation for transformations of genus-one models introduced in Section 2.) In the following we write $t^{(m)}$ as a shorthand for $\pi^{-m}t$.

Lemma 5.4. *The level of a critical model is at least* 1 *and equal to* 1 *if* $\mathrm{char}(k) \nmid n$.

Proof. Case $n = 2$. By (2-3) we have $\pi^i \mid a_i$ for all i. A convenient way to check this is to note that $\pi^{-3/2}P(\pi^{1/2}x, z)$ and $\pi^{-3}Q(\pi^{1/2}x, z)$ have coefficients in $\mathbb{O}_K[\pi^{1/2}]$, and then to use the isobaric property. It follows that (P, Q) has positive level. Now suppose that $\mathrm{char}(k) \neq 2$ and (P, Q) has level greater than 1. Completing the square we may assume that $l = m = n = 0$. Then $a_1 = a_3 = 0$ and $y^2 = x^3 + a_2^{(2)}x^2 + a_4^{(4)}x + a_6^{(6)}$ is an integral Weierstrass equation of positive level. According to Tate's algorithm the cubic polynomial

$$x^3 + a_2^{(2)}x^2 + a_4^{(4)}x + a_6^{(6)} \equiv (x + c^{(2)})(x^2 - 4a^{(1)}e^{(3)}) \pmod{\pi}$$

has a triple root defined over k. This contradicts the definition of a critical model.

Case $n = 3$. By (2-4) we have $\pi^i \mid a_i$ for all i. A convenient way to check this is to note that $\pi^{-2}F(\pi^{2/3}x, \pi^{1/3}y, z)$ has coefficients in $\mathbb{O}_K[\pi^{1/3}]$, and then to use the

isobaric property. It follows that F has positive level. Now suppose that $\mathrm{char}(k) \neq 3$ and F has level greater than 1. Then

$$y^2 + a_1^{(1)}xy + a_3^{(3)}y = x^3 + a_2^{(2)}x^2 + a_4^{(4)}x + a_6^{(6)}$$

is an integral Weierstrass equation of positive level. By (2-4) we find $a_2^{(2)} \equiv a_4^{(4)} \equiv 0 \pmod{\pi}$ and

$$\begin{aligned} a_1^{(1)} &\equiv m^{(1)} \pmod{\pi}, \\ a_3^{(3)} &\equiv 9ab^{(1)}c^{(2)} \pmod{\pi}, \\ a_6^{(6)} &\equiv -27(ab^{(1)}c^{(2)})^2 + ab^{(1)}c^{(2)}(m^{(1)})^3 \pmod{\pi}. \end{aligned}$$

So it suffices to show that if there is a Weierstrass equation over k of the form

$$y^2 + \alpha xy + 9\beta y = x^3 + (\alpha^3 - 27\beta)\beta$$

with $c_4 = \Delta = 0$, then $\beta = 0$. We have $c_4 = \alpha(\alpha^3 - 216\beta)$ and $\Delta = -\beta(\alpha^3 + 27\beta)^3$. Since $216 + 27 = 3^5$ is nonzero in k, it follows that $\beta = 0$ as required.

Case $n = 4$. The quadric intersection $[\pi^{-1}I_2, \mathrm{Diag}(\pi^{1/2}, \pi^{1/2}, 1, 1)](Q_1, Q_2)$ has coefficients in $\mathbb{O}_K[\pi^{1/2}]$. It follows by the isobaric property of the a-invariants that $\pi^i \mid a_i$ for all i and hence that (Q_1, Q_2) has positive level. Now suppose that $\mathrm{char}(k) \neq 2$. Then $F = \mathfrak{d}(Q_1, Q_2)$ satisfies $F(x, z) \equiv \pi^2 f_1(x, z) f_2(x, z) \pmod{\pi^3}$ where $f_1, f_2 \in \mathbb{O}_K[x, z]$ are binary quadratic forms, neither having a repeated root mod π. (So their product cannot have a triple or quadruple root.) It follows by Theorem 4.2(i) that F and hence (Q_1, Q_2) has level 1. □

Example 5.5. The following examples of critical models, all with level 2, show that the hypothesis $\mathrm{char}(k) \nmid n$ cannot be removed from Lemma 5.4.

$$\begin{aligned} K &= \mathbb{Q}_2 \qquad & y^2 &= 2x^4 + 24x^2z^2 + 8z^4, \\ K &= \mathbb{Q}_3 \qquad & x^3 + 3y^3 + 9z^3 + 18xyz &= 0, \\ K &= \mathbb{Q}_2 \qquad & x_1^2 + 2x_3^2 + 4x_2x_4 = x_2^2 + 2x_4^2 + 4x_1x_3 &= 0, \end{aligned}$$

To complete the proof of Theorem 3.5(ii) we prove the following. The doubling map $\mathfrak{d}'$ was defined in Section 4D. (If $\mathrm{char}(k) \neq 2$ then we can work with $\mathfrak{d}$ instead.)

Proposition 5.6. *Let $\Phi \in X_n(\mathbb{O}_K)$ be a K^{sh}-insoluble minimal genus-one model.*

(i) *If $n = 2$ or 3 then Φ is $\mathbb{O}_K$-equivalent to a critical model.*

(ii) *If $n = 4$ then Φ is K-equivalent to either a critical model or an integral model (Q_1, Q_2) with $\mathfrak{d}'(Q_1, Q_2)$ critical.*

First we need three lemmas.

Lemma 5.7. *Let k be an algebraically closed field. Suppose that either*

(a) $\Phi = (P, Q) \in X_2(k)$ *and* $P^2 + 4Q$ *is not identically zero,*

(b) $\Phi = (F) \in X_3(k)$ *is nonzero and is not the cube of a linear form,*

(c) $\Phi = (Q_1, Q_2) \in X_4(k)$ *and every quadric in the pencil spanned by* Q_1 *and* Q_2 *has rank at least* 2.

Then $\mathscr{C}_\Phi$ *has a smooth k-point* (*on some* 1*-dimensional component*).

Proof. For $n = 2, 3$ this is clear. In the case $n = 4$ we are looking for a transverse point of intersection of Q_1 and Q_2, i.e., a point where the Jacobian matrix has rank 2. We prove the result more generally for intersections of two quadrics in m variables. This enables us to reduce to the case $\ker Q_1 \cap \ker Q_2 = 0$. Now let P be a singular point on the quadric intersection. (If there is no such point there is nothing to prove.) Then moving this point to $(1 : 0 : \ldots : 0)$ we may assume that $Q_1 = x_1x_2 + g_1(x_2, \ldots, x_m)$ and $Q_2 = g_2(x_2, \ldots, x_m)$ for some g_1 and g_2. Since $\operatorname{rank}(Q_2) \geq 2$ we can pick a smooth point $(x_2 : \ldots : x_m)$ on $\{Q_2 = 0\} \subset \mathbb{P}^{m-2}$ with $x_2 \neq 0$. Then solving the equation $Q_1 = 0$ for x_1 gives the required transverse point of intersection on $\{Q_1 = Q_2 = 0\}$. □

Lemma 5.8. *Let $\Phi \in X_n(\mathbb{O}_K)$ be a K^{sh}-insoluble minimal genus-one model.*

(a) *If $n = 2$ then $\Phi = (P, Q)$ with $v(P, Q) = 1$. Moreover if $v(Q) = 1$ then the reduction of $\pi^{-1}Q(x, z)$* mod π *has either two double roots or a quadruple root* (*over $\bar{k}$*).

(b) *If $n = 3$ then Φ is a ternary cubic whose reduction* mod π *is* (*a constant times*) *the cube of a linear form.*

(c) *If $n = 4$ then there is a quadric of rank* 1 *in the reduced pencil*; *that is, if* $\Phi = (Q_1, Q_2)$ *then* $\operatorname{rank}(\lambda \overline{Q}_1 + \mu \overline{Q}_2) = 1$ *for some* $(\lambda : \mu) \in \mathbb{P}^1(\bar{k})$.

Proof. We recall that K^{sh} has residue field $\bar{k}$. The idea of the proof is that if Φ is not of the form listed, then we can use Lemma 5.7 to find a smooth $\bar{k}$-point on the reduction, and use the Henselian property to lift it to a K^{sh}-point, thereby obtaining a contradiction.

A little more needs to be said in the case $n = 2$. If $\operatorname{char}(k) \neq 2$ then completing the square gives $v(P) \geq 1$ and Lemma 5.7 shows that $v(Q) \geq 1$. If $\operatorname{char}(k) = 2$ then Lemma 5.7 shows that $v(P) \geq 1$. If $Q(x, z)$ mod π had a simple root over $\bar{k}$ then we could lift to a K^{sh}-point on $\mathscr{C}_{(P,Q)}$ with $y = 0$. It follows that $Q(x, z)$ is a square mod π. So by a y-substitution we may suppose $v(Q) \geq 1$. In all residue characteristics we now have $v(P) \geq 1$ and $v(Q) \geq 1$. We cannot have $v(Q) \geq 2$ since (P, Q) is minimal. If $\pi^{-1}Q(x, z)$ mod π had a simple root over $\bar{k}$ then we could lift to a K^{sh}-point on $\mathscr{C}_{(P,Q)}$ with $y = 0$. It follows that this polynomial has either two double roots or a quadruple root. □

Lemma 5.9. *Suppose* $(P, Q), (P', Q') \in X_2(\mathbb{O}_K)$ *are K-equivalent models of the same level related by a substitution* $[\mu, r, M] \in \mathcal{G}_2(K)$ *where* $M \in \mathrm{GL}_2(K)$ *has Smith normal form* $\mathrm{Diag}(1, \pi^s)$. *Then* $v(\Delta(P, Q)) \geq 2s$.

Proof. Let (P, Q) have coefficients l, m, n, a, b, c, d and e. Replacing our models by $\mathbb{O}_K$-equivalent ones we may assume $\mu = \pi^{-s}$ and $M = \mathrm{Diag}(\pi^s, 1)$. If we assume for simplicity that $r = 0$, then we have $\pi^s \mid n, d$ and $\pi^{2s} \mid e$. Since the discriminant $\Delta \in \mathbb{Z}[X_2]$ belongs to the ideal (n^2, nd, d^2, e) it follows that $v(\Delta(P, Q)) \geq 2s$.

For general r we can write the transformation $[\pi^{-s}, r, \mathrm{Diag}(\pi^s, 1)]$ either as

$$y \leftarrow \pi^s y + r_0 x^2 + r_1 xz + r_2 z^2 \quad \text{followed by} \quad x \leftarrow \pi^s x,$$

or as

$$x \leftarrow \pi^s x \quad \text{followed by} \quad y \leftarrow \pi^s (y + \pi^s r_0 x^2 + r_1 xz + \pi^{-s} r_2 z^2).$$

Since Q' has coefficients in $\mathbb{O}_K$ we have

$$v(r_0^2 + r_0 l - a) \geq -2s \quad \text{and} \quad v(r_2^2 + r_2 n - e) \geq 2s.$$

Hence $\pi^s r_0, r_2 \in \mathbb{O}_K$. So replacing our models by $\mathbb{O}_K$-equivalent ones we may assume that $r_0 = r_2 = 0$. Then the middle coefficient of Q' gives $v(r_1^2 + r_1 m - c) \geq 0$ and hence $r_1 \in \mathbb{O}_K$. Once more replacing (P, Q) by an $\mathbb{O}_K$-equivalent model we may assume that $r_0 = r_1 = r_2 = 0$. Our earlier proof now applies. □

Proof of Proposition 5.6. We split into the cases $n = 2, 3, 4$.

Case $n = 2$. Applying Lemma 5.8 to $\Phi = (P, Q)$ we may assume that $v(P) \geq 1$, $v(Q) = 1$ and $\pi^{-1} Q(x, z) \bmod \pi$ has either two double roots or a quadruple root.

We first rule out the possibility of two double roots. After an unramified field extension we may assume that these roots are defined over k. So without loss of generality $Q(x, z) \equiv \pi x^2 z^2 \pmod{\pi^2}$. We replace $P(x, z)$ by $\pi^{-1} P(\pi x, z)$ and $Q(x, z)$ by $\pi^{-2} Q(\pi x, z)$. By Lemma 5.8 we again have $v(P, Q) \geq 1$. We make a substitution $y \leftarrow y + r_2 z^2$ so that $v(P) \geq 1$ and $v(Q) \geq 1$. Now $\pi^{-1} Q(x, z) \bmod \pi$ has a double root at $(x : z) = (1 : 0)$. By Lemma 5.8 it has a second double root, say at $(\lambda : 1)$. We make the substitution $x \leftarrow x + \lambda z$. Then $Q(x, z) \equiv \pi x^2 z^2 \pmod{\pi^2}$. We can now repeat this process indefinitely. It follows by Lemma 5.9 that $\Delta(P, Q) = 0$. This is the required contradiction.

It remains to consider the case of a quadruple root, say $Q(x, z) \equiv \pi x^4 \pmod{\pi^2}$. Let $l_1, m_1, n_1, a_1, b_1, c_1, d_1$ and e_1 be the coefficients of $P_1(x, z) = \pi^{-1} P(\pi x, z)$ and $Q_1(x, z) = \pi^{-2} Q(\pi x, z)$. By Lemma 5.8 we can make a substitution

$$y \leftarrow y + r_2 z^2$$

so that π divides n_1 and e_1. Then $\pi^{-1} Q_1(x, z) \bmod \pi$ has at least a triple root at

$(x : z) = (1 : 0)$. So by Lemma 5.8 we have $\pi^2 \mid d_1$ and $v(e_1) = 1$. The coefficients of (P, Q) now satisfy the definition of a critical model.

Case $n = 3$. By Lemma 5.8 our ternary cubic F must reduce mod π to the cube of a linear form. So without loss of generality, we have

$$F = \pi f_3(y, z) + \pi f_2(y, z)x + \pi f_1(y, z)x^2 + ax^3.$$

with $\pi \nmid a$. Then $F_1(x, y, z) = \pi^{-1} F(\pi x, y, z)$ is a minimal ternary cubic and by Lemma 5.8 its reduction mod π is the cube of a linear form in y and z. After a suitable transformation of y and z, we may assume that $f_3(y, z) \equiv by^3 \pmod{\pi}$ with $\pi \nmid b$ (otherwise F would not be minimal). Now

$$F_2(x, y, z) = \pi^{-1} F_1(x, \pi y, z)$$

is again a minimal ternary cubic, and its reduction mod π is $(c'x + cz)z^2$. Again this must be a nonzero cube. So $c' = 0$ and c is a unit. The coefficients of F now satisfy the definition of a critical model.

Case $n = 4$. We divide the proof into the following two lemmas.

Lemma 5.10. *Let $(Q_1, Q_2) \in X_4(\mathbb{O}_K)$ be a K^{sh}-insoluble minimal quadric intersection. Let*

$$s = \dim(\ker \overline{Q}_1 \cap \ker \overline{Q}_2)$$

be the common nullity of the reduced pencil.

(i) *If $s \leq 1$ then the reduced pencil contains a unique rank 1 quadric, and the following procedure replaces (Q_1, Q_2) by a K-equivalent minimal quadric intersection with $s \geq 1$.*

- *Make a $\mathrm{GL}_2(\mathbb{O}_K) \times \mathrm{GL}_4(\mathbb{O}_K)$-transformation so that $Q_2 \equiv x_1^2 \pmod{\pi}$.*
- *Apply the transformation* $[\mathrm{Diag}(1, \pi^{-1}), \mathrm{Diag}(\pi, 1, 1, 1)]$.

(ii) *If $s \geq 2$ then (Q_1, Q_2) is $\mathbb{O}_K$-equivalent to a critical model.*

Proof. (i) By Lemma 5.8 there is a rank 1 quadric in the reduced pencil. It is unique (and therefore defined over k) as we would otherwise have $s \geq 2$. The remaining statements are clear.

(ii) We may assume that $\overline{Q}_1$ and $\overline{Q}_2$ are binary quadratic forms in x_1 and x_2. Since the model is minimal, these forms have no common root in $\mathbb{P}^1(\bar{k})$. We put

$$(R_1, R_2) = [\pi^{-1} I_2, \mathrm{Diag}(\pi, \pi, 1, 1)](Q_1, Q_2).$$

Then R_1 and R_2 reduce to binary quadratic forms in x_3 and x_4. Again, since the model is minimal, these forms have no common root in $\mathbb{P}^1(\bar{k})$. Hence (Q_1, Q_2) is critical. $\square$

Lemma 5.11. *Let $\Phi \in X_4(\mathbb{O}_K)$ satisfy the hypotheses of Lemma 5.10 with $s = 1$. If the procedure in Lemma 5.10(i) may be iterated indefinitely, then Φ is $\mathbb{O}_K$-equivalent to a quadric intersection (Q_1, Q_2) where the valuations of the coefficients of Q_1 and Q_2 satisfy the inequalities indicated in the following diagram*:

$$\begin{array}{cccc}
x_1^2 & x_1x_2 & x_1x_3 & x_1x_4 \\
 & x_2^2 & x_2x_3 & x_2x_4 \\
 & & x_3^2 & x_3x_4 \\
 & & & x_4^2
\end{array}
\qquad
\begin{array}{cccc}
\geq 0 & \geq 0 & = 0 & \geq 1 \\
 & = 0 & \geq 1 & \geq 1 \\
 & & \geq 1 & \geq 1 \\
 & & & = 1
\end{array}
\qquad
\begin{array}{cccc}
= 0 & \geq 1 & \geq 1 & \geq 1 \\
 & \geq 1 & \geq 1 & = 1 \\
 & & = 1 & \geq 2 \\
 & & & \geq 2
\end{array}$$

Proof. We may assume that $\Phi = (Q_1, Q_2)$ has reduction

$$(\overline{Q}_1, \overline{Q}_2) = (x_1 l(x_2, x_3) + f(x_2, x_3), cx_1^2) \tag{5-1}$$

for some $c \in k$ and $l, f \in k[x_2, x_3]$. Since (Q_1, Q_2) is minimal we have $cf \neq 0$. So the reduction is (set-theoretically) either a line or a pair of lines. We show in the case of a pair of lines that the procedure in Lemma 5.10(i) must give $s \geq 2$ after a finite number of iterations (bounded in terms of the valuation of the discriminant). The first iteration gives (R_1, R_2) with $(\bar{R}_1, \bar{R}_2) = (f(x_2, x_3), g(x_2, x_3, x_4))$, for some $g \in k[x_2, x_3, x_4]$. Since f has rank 2 we may assume on replacing R_2 by $R_2 + \lambda R_1$ for suitable $\lambda \in \mathbb{O}_K$ that g has rank 1. If g has no coefficient of x_4^2 then $s \geq 2$. Otherwise a $\mathrm{GL}_4(\mathbb{O}_K)$-transformation puts $(\bar{R}_1, \bar{R}_2)$ in the form (5-1) with $l = 0$ (and the same f as before). The process is then repeated. By considering the effect on the doubling it follows by Lemma 5.9 that only finitely many iterations are possible.

It remains to consider the case where the reduction is (set-theoretically) a line. We may assume that $\Phi = (Q_1, Q_2)$ and its transforms,

$$(R_1, R_2) = [\mathrm{Diag}(1, \pi^{-1}), \mathrm{Diag}(\pi, 1, 1, 1)](Q_1, Q_2),$$
$$(S_1, S_2) = [\mathrm{Diag}(\pi^{-1}, 1), \mathrm{Diag}(1, \pi, 1, 1)](R_1, R_2),$$

under the first two iterations have reductions

$$(\overline{Q}_1, \overline{Q}_2) = (x_1(\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3) + x_2^2, x_1^2), \tag{5-2}$$
$$(\bar{R}_1, \bar{R}_2) = (x_2^2, x_2(\beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4) + g(x_3, x_4)), \tag{5-3}$$
$$(\bar{S}_1, \bar{S}_2) = (\alpha_3 x_1 x_3 + \lambda x_3^3 + \mu x_3 x_4 + \nu x_4^2, g(x_3, x_4)), \tag{5-4}$$

for some $\alpha_i, \beta_i, \lambda, \mu, \nu \in k$ and $g \in k[x_3, x_4]$. By (5-2) we have $\alpha_3 \neq 0$ (otherwise $s \geq 2$). Since the reduction cannot be a pair of lines, we see first by (5-3) that g has rank 1, and then by (5-4) that $g = \gamma x_3^2$ for some $\gamma \neq 0$. Finally (5-3) and (5-4) show that $\beta_4 \neq 0$ and $\nu \neq 0$ (otherwise $s \geq 2$). The valuations of the coefficients of Q_1 and Q_2 now satisfy the inequalities indicated in the statement of the lemma. □

Proposition 5.6(ii) follows from the last two lemmas and the observation that if (Q_1, Q_2) satisfies the conclusions of Lemma 5.11 then its doubling is critical. □

6. Reduction

In this section, we assume that the ground field is $\mathbb{Q}$. The main reason for this is that a comparable theory of reduction over a general number field has not yet been sufficiently developed.

Let $\mathscr{C} \subset \mathbb{P}^{n-1}$ be a genus-one normal curve defined over $\mathbb{Q}$ of degree n (or, if $n = 2$, let $\mathscr{C} \to \mathbb{P}^1$ be a double cover) with points everywhere locally, so that $\mathscr{C}$ represents an element of the n-Selmer group of its Jacobian elliptic curve E. If $n \in \{2, 3, 4\}$, we can, by the results and algorithms of the previous sections, assume that $\mathscr{C} = \mathscr{C}_\Phi$ where Φ is a genus-one model which is both integral and minimal, so that its invariants c_4, c_6 and Δ coincide with those of a minimal model of E. This means that the invariants are as small as possible (in absolute value). However, it does not necessarily mean that the equations defining $\mathscr{C}$ will have small coefficients. To achieve this, we will employ *reduction*. Leaving aside the aesthetic value of equations with small coefficients, the main benefit of a reduced model is that further computations like searching for rational points on $\mathscr{C}$ or performing further descents on $\mathscr{C}$ are greatly facilitated.

The idea of reduction is to find a unimodular transformation (i.e., an invertible integral linear change of coordinates on $\mathbb{P}^{n-1}$) that makes the equations defining $\mathscr{C}$ smaller. Unimodular transformations have the property of preserving the integrality and invariants of the model, so they will not destroy its minimality. In the language of Section 2, a unimodular transformation is just a $\mathbb{Z}$-equivalence.

If we were allowed to make a coordinate change from $\mathrm{SL}_n(\mathbb{C})$ instead, then we could always bring our model into one of the following standard forms, where in general $a, b \in \mathbb{C}$ (see for example [Hulek 1986]). When $n = 3$, we can achieve this normal form even by a transformation from $\mathrm{SL}_3(\mathbb{R})$, so in this case, we can take $a, b \in \mathbb{R}$. We will call these forms *Hesse forms*, generalising the classical terminology for $n = 3$. They are as follows:

$$\begin{aligned}
n = 2: &\quad y^2 = a(x_0^4 + x_1^4) + bx_0^2x_1^2,\\
n = 3: &\quad a(x_0^3 + x_1^3 + x_2^3) + bx_0x_1x_2 = 0,\\
n = 4: &\quad a(x_0^2 + x_2^2) + bx_1x_3 = 0, \quad a(x_1^2 + x_3^2) + bx_0x_2 = 0.
\end{aligned}$$

In these forms, the coefficients a and b are bounded in terms of the invariants, so we can expect them to be small. Therefore, we would like to come close to a model of this kind, but using a unimodular transformation.

We need some way of measuring how close two models are. On the standard Hesse models, the action of the n-torsion of the Jacobian, $E[n]$, is given by the so-call standard representation, where one generator multiplies each x_j by ζ_n^j and the other generator does a cyclic shift of the coordinates. (Here ζ_n denotes a primitive n-th root of unity.) To this representation, we can associate an invariant inner product on $\mathbb{C}^n$, which is unique up to scaling. It is easy to check that this invariant inner product is just the standard one on $\mathbb{C}^n$. Now our approach is to associate an inner product to a given model $\mathscr{C}$, and consider the model to be close to a standard model when the associated inner product is close to the standard one, which means that it is reduced in an appropriate sense. This is explained in some detail in the following subsection.

6A. ***The reduction covariant.*** Let $K = \mathbb{R}$ or $\mathbb{C}$. We write $\mathcal{Y}_n(K)$ for the set of all genus-one normal curves of degree n defined over K, inside a fixed copy of $\mathbb{P}^{n-1}$. (If $n = 2$ we consider double covers of $\mathbb{P}^1$ instead.) The difference between $\mathcal{Y}_n(K)$ and $X_n(K)$ is that we now consider actual curves in $\mathbb{P}^{n-1}$ (or the set of ramification points of $\mathscr{C} \to \mathbb{P}^1$ when $n = 2$), instead of defining equations.

Let $\mathcal{H}_n^+(\mathbb{C})$ be the space of positive definite Hermitian $n \times n$ matrices, and $\mathcal{H}_n^+(\mathbb{R})$ the space of positive definite symmetric real $n \times n$ matrices. We can identify these spaces with the spaces of positive definite Hermitian and real quadratic forms in n variables, respectively. There are natural and compatible (left) actions of $\mathrm{SL}_n(K)$ on $\mathcal{Y}_n(K)$ and $\mathcal{H}_n^+(K)$ given by the canonical map

$$\mathrm{SL}_n(K) \to \mathrm{PGL}_n(K) = \mathrm{Aut}(\mathbb{P}_K^{n-1})$$

on the one hand and by

$$g \cdot M = \bar{g}^{-t} M g^{-1}$$

on the other, where γ^{-t} denotes the inverse transpose of the matrix γ. If we identify the matrix $M \in \mathcal{H}_n^+(K)$ with the quadratic or Hermitian form $Q(x) = \bar{x}^t M x$, then the compatibility of the actions means that $(g \cdot Q)(gx) = Q(x)$.

Theorem 6.1. *For each $n \geq 2$ there is a unique $\mathrm{SL}_n(\mathbb{C})$-covariant map*

$$\varphi_{\mathbb{C}} : \mathcal{Y}_n(\mathbb{C}) \to \mathcal{H}_n^+(\mathbb{C})/\mathbb{R}_{>0}^{\times}.$$

This map is compatible with complex conjugation, and so restricts to an $\mathrm{SL}_n(\mathbb{R})$-covariant map

$$\varphi_{\mathbb{R}} : \mathcal{Y}_n(\mathbb{R}) \to \mathcal{H}_n^+(\mathbb{R})/\mathbb{R}_{>0}^{\times}.$$

Proof. Let $\mathscr{C} \to \mathbb{P}^{n-1}$ be a genus-one normal curve defined over $\mathbb{C}$, with Jacobian E. The action of $E[n]$ on $\mathscr{C}$ extends to $\mathbb{P}^{n-1}$ and hence defines a group

homomorphism $\chi : E[n](\mathbb{C}) \to \mathrm{PGL}_n(\mathbb{C})$. Lifting to $\mathrm{SL}_n(\mathbb{C})$ we obtain a diagram

$$\begin{tikzcd} 0 \arrow[r] & \mu_n \arrow[r] \arrow[d, equal] & H_n \arrow[r] \arrow[d] & E[n](\mathbb{C}) \arrow[r] \arrow[d, "\chi"] & 0 \\ 0 \arrow[r] & \mu_n \arrow[r] & \mathrm{SL}_n(\mathbb{C}) \arrow[r] & \mathrm{PGL}_n(\mathbb{C}) \arrow[r] & 0. \end{tikzcd}$$

The Heisenberg group H_n is a nonabelian group of order n^3. It comes with a natural n-dimensional representation, called the Schrödinger representation, which is known to be irreducible (since it is equivalent to the standard representation mentioned above). Now by the Weyl unitary trick, every irreducible complex representation of a finite group has a unique invariant inner product. (Recall that existence is proved by averaging over the group, and uniqueness (up to $\mathbb{R}^\times_{>0}$) using Schur's lemma.)

We define $\varphi_{\mathbb{C}}(\mathscr{C})$ to be the (matrix of the) Heisenberg invariant inner product, i.e., $\varphi_{\mathbb{C}}(\mathscr{C})$ is uniquely determined up to positive real scalars by the property that

$$\bar{h}^{-t}\varphi_{\mathbb{C}}(\mathscr{C})h^{-1} = \varphi_{\mathbb{C}}(\mathscr{C})$$

for all $h \in H_n$. If $g \in \mathrm{SL}_n(\mathbb{C})$, the Heisenberg groups H_n and H_n' of $\mathscr{C}$ and $g \cdot \mathscr{C}$ are related by $H_n' = gH_ng^{-1}$. Then $g \cdot \varphi_{\mathbb{C}}(\mathscr{C}) = \bar{g}^{-t}\varphi_{\mathbb{C}}(\mathscr{C})g^{-1}$ is an H_n'-invariant inner product, and so must be equal to $\varphi_{\mathbb{C}}(g \cdot \mathscr{C})$. Hence $\varphi_{\mathbb{C}}$ is $\mathrm{SL}_n(\mathbb{C})$-covariant. Moreover, since $H_n \subset \mathrm{SL}_n(\mathbb{C})$, this choice of covariant is forced on us. The compatibility with complex conjugation is seen in the same way. □

Remark 6.2. In general $\varphi_{\mathbb{R}}$ is not the only $\mathrm{SL}_n(\mathbb{R})$-covariant. However, it is if the points of $E[n]$ are defined over $\mathbb{R}$, as happens in the case $n = 2$ and $\Delta > 0$ [Stoll and Cremona 2003, Lemma 3.2].

In practical terms, we have the following corollary.

Corollary 6.3. *Let $M_T \in \mathrm{GL}_n(\mathbb{C})$ describe the action of $T \in E[n](\mathbb{C})$ on $\mathscr{C} \to \mathbb{P}^{n-1}$. Then the reduction covariant $\varphi_{\mathbb{C}}(\mathscr{C})$ is*

$$\sum_{T \in E[n](\mathbb{C})} \frac{1}{|\det M_T|^{2/n}} \overline{M}_T^t M_T.$$

Proof. To get an invariant inner product, we can take any inner product and average over its orbit under the action of H_n. Applying this to the standard inner product, we find that we can take, up to scaling,

$$\varphi_{\mathbb{C}}(\mathscr{C}) = \sum_{h \in H_n} \bar{h}^{-t}h^{-1} = \sum_{h \in H_n} \bar{h}^t h. \tag{6-1}$$

In the statement of the corollary, $M_T \in \mathrm{GL}_n(\mathbb{C})$ is any lift of the element τ_T in $\mathrm{PGL}_n(\mathbb{C})$ describing the action of T on $\mathbb{P}^{n-1}(\mathbb{C})$. The various preimages of τ_T in

H_n are given by $h = \alpha^{-1} M_T$ where $\alpha \in \mathbb{C}$ with $\alpha^n = \det M_T$. We then have

$$\bar{h}^t h = \bar{\alpha}^{-1} \alpha^{-1} \overline{M}_T^t M_T = \frac{1}{|\det M_T|^{2/n}} \overline{M}_T^t M_T.$$

Since this only depends on T, it is sufficient to take the sum in (6-1) just over $T \in E[n](\mathbb{C})$, instead of over $h \in H_n$. □

We can now define what we mean by a reduced genus-one normal curve.

Definition 6.4. A genus-one normal curve $\mathscr{C} \to \mathbb{P}^{n-1}$ defined over $\mathbb{R}$ is *Minkowski reduced* if $\varphi_{\mathbb{R}}(\mathscr{C})$ is the Gram matrix of a Minkowski reduced lattice basis, and likewise with "LLL" instead of "Minkowski".

Note that a lattice basis is (Minkowski or LLL) reduced if it is close to the standard basis of the standard lattice in the sense that the basis vectors are (short and) nearly orthogonal. The notion of a Minkowski reduced model has nice theoretical properties (it is optimal and essentially unique), whereas for practical purposes, it is important to be able to compute a reduced lattice basis efficiently; this is possible when using LLL reduced models.

If we start with some given (minimal) model $\mathscr{C} \to \mathbb{P}^{n-1}$, then in order to reduce it, we first compute its reduction covariant $\varphi_{\mathbb{R}}(\mathscr{C})$. We apply the LLL algorithm [Lenstra et al. 1982] to this Gram matrix, resulting in a unimodular transformation U and an LLL reduced Gram matrix M, such that $M = U^t \varphi_{\mathbb{R}}(\mathscr{C}) U = U^{-1} \cdot \varphi_{\mathbb{R}}(\mathscr{C})$. We then apply the transformation U^{-1} to our model $\mathscr{C}$. Since $\varphi_{\mathbb{R}}(\mathscr{C})$ is a covariant, we will have that $\varphi_{\mathbb{R}}(U^{-1} \cdot \mathscr{C}) = M$ is LLL reduced. Therefore $U^{-1} \cdot \mathscr{C}$ is the (minimal and) reduced model we are looking for.

Remark 6.5. Let $H : X_n \to X_n$ be the Hessian, as defined in [Fisher 2006a]. Then every nonsingular model in the pencil spanned by Φ and $H(\Phi)$ defines a curve with the same reduction covariant — since they have the same Heisenberg group. The case $n = 2$ was previously given as [Cremona 1999, Propositions 10 and 13].

In the following subsections we discuss how to compute $\varphi_{\mathbb{R}}$. There are two basic approaches. One is to find the hyperosculating points of $\mathscr{C}(\mathbb{C})$ numerically and to compute the covariant from them. If $n = 2$, we are looking for the ramification points of the covering $\mathscr{C} \to \mathbb{P}^1$; if $n = 3$, for the flex points of the plane cubic curve $\mathscr{C} \subset \mathbb{P}^2$. The other approach is to use the n-torsion points in $E(\mathbb{C})$ instead and compute their action on $\mathbb{P}^{n-1}$. Generally speaking, the first approach leads to simpler formulas, whereas the second approach tends to be numerically more stable.

6B. *Reduction of 2-coverings.* We identify $\mathscr{H}_2^+(\mathbb{R})$ with the space of real positive definite binary quadratic forms, and $\mathscr{H}_2^+(\mathbb{R})/\mathbb{R}_{>0}^\times$ with the upper half-plane. This identification maps a real positive definite binary quadratic form to its unique root in the upper half-plane.

Using the ramification points. Let $F(x,z) \in \mathbb{R}[x,z]$ be homogeneous of degree 4. We assume that $f(X) = F(X,1)$ has degree 4 as well. (If the leading coefficient is zero, make a change of coordinates first.) Let $\theta_1, \dots, \theta_4 \in \mathbb{C}$ be the roots of f. It is shown in [Stoll and Cremona 2003] that $\varphi_{\mathbb{R}}$ is given by

$$\varphi_{\mathbb{R}}(F)(x,z) = \sum_{i=1}^{4} \frac{1}{|f'(\theta_i)|}(x-\theta_i z)(x-\overline{\theta}_i z).$$

This goes back to [Julia 1917], where three different formulas are given according to the number of real roots of f; see also [Cremona 1999].

The formula is still valid for $\varphi_{\mathbb{C}}$, in the form

$$\varphi_{\mathbb{C}}(F)(x,z) = \sum_{i=1}^{4} \frac{1}{|f'(\theta_i)|}|x-\theta_i z|^2.$$

In practice one should first numerically compute the roots of the resolvent cubic (which is not changed by reduction) and then compute the roots of f from these.

Using the 2*-torsion of* E. The binary quartic

$$F(x,z) = ax^4 + bx^3z + cx^2z^2 + dxz^3 + ez^4$$

has invariants I and J (see Section 2) and resolvent cubic $r(X) = X^3 - 3IX + J$. For φ a root of r we set

$$\begin{aligned}
\alpha_1(\varphi) &= 4a\varphi - 8ac + 3b^2,\\
\alpha_2(\varphi) &= b\varphi - 6ad + bc,\\
\alpha_3(\varphi) &= \frac{-2\varphi^2 + 2c\varphi - 9bd + 4c^2}{3},
\end{aligned}$$

and

$$W = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \qquad A_\varphi = \begin{pmatrix} \alpha_1(\varphi) & \alpha_2(\varphi) \\ \alpha_2(\varphi) & \alpha_3(\varphi) \end{pmatrix}.$$

Lemma 6.6. *If* $\alpha_1(\varphi) \neq 0$, *then the action of the corresponding point* $T \in E[2]$ *on* $\mathbb{P}^1$ *is given by* $M_T = WA_\varphi$.

Proof. Let $H(x,z)$ be the Hessian of F. The pencil spanned by F and H contains three degenerate quartics: for each root φ' of the resolvent cubic, we have

$$\alpha_1(\varphi')\big(4\varphi' F(x,z) - \tfrac{1}{3}H(x,z)\big) = \big(\alpha_1(\varphi')x^2 + 2\alpha_2(\varphi')xz + \alpha_3(\varphi')z^2\big)^2.$$

Since the action of T leaves both F and H invariant, M_T must induce an involution on $\mathbb{P}^1$ that either fixes or swaps the roots of the quadratic on the right hand side; there is exactly one root φ' such that the roots of the corresponding quadratic are fixed. Therefore $\varphi' = \varphi$, and the lemma follows by checking that WA_φ does indeed fix the roots of the relevant quadratic. □

Lemma 6.7. *If $M_T \in \mathrm{GL}_2$ describes the action of $T \in E[2]$ on $\mathscr{C} \to \mathbb{P}^1$ then*

$$\sum_{T \in E[2]} \frac{1}{\det M_T} M_T^t M_T = 0. \tag{6-2}$$

Proof. We can verify this generically using the formula of Lemma 6.6. □

Proposition 6.8. *Let $F \in \mathbb{R}[x, z]$ be a nonsingular binary quartic, with resolvent cubic $r(X) = X^3 - 3IX + J$.*

(i) *If $\Delta(F) > 0$ then the reduction covariant is $\pm A_\varphi$ where φ is the unique root of r with $\det(A_\varphi) > 0$ and the sign is that of $\alpha_1(\varphi)$.*

(ii) *If $\Delta(F) < 0$ then the reduction covariant is*

$$\mathrm{Re}\Big(\frac{1}{|\det A_\varphi|} \bar{A}_\varphi A_\varphi - \frac{1}{\det A_\varphi} A_\varphi^2\Big)$$

where φ is a complex root of r.

Proof. If $\Delta(F) > 0$, then r has three real roots. Since $\det(A_\varphi) = -\alpha_1(\varphi) r'(\varphi)/3$, the analysis in [Cremona 1999] shows that there is a unique root φ of r with $\det(A_\varphi) > 0$ (in particular, $\alpha_1(\varphi) \neq 0$). By Lemmas 6.6 and 6.7 the reduction covariant simplifies (up to a factor of 2) to

$$\sum_{T \in E[2],\, \det M_T > 0} \frac{1}{\det M_T} M_T^t M_T = I_2 + \frac{1}{\det A_\varphi} A_\varphi^2 = \frac{\mathrm{tr} A_\varphi}{\det A_\varphi} A_\varphi,$$

by the Cayley–Hamilton theorem. So $\pm A_\varphi$ is the positive definite symmetric matrix we are looking for, with the sign that makes the top left entry positive.

If $\Delta(F) < 0$, then r has a pair of complex conjugate roots, say φ and $\bar{\varphi}$. If $E[2] = \{0, S, T, \bar{T}\}$, then we can take $M_S = M_T \overline{M}_T$, so

$$\det(M_S) = |\det(M_T)|^2 > 0.$$

By Lemmas 6.6 and 6.7 again, the reduction covariant simplifies to

$$\mathrm{Re}\Big(\frac{1}{|\det M_T|} \overline{M}_T^t M_T - \frac{1}{\det M_T} M_T^t M_T\Big) = \mathrm{Re}\Big(\frac{1}{|\det A_\varphi|} \bar{A}_\varphi A_\varphi - \frac{1}{\det A_\varphi} A_\varphi^2\Big).$$

Notice that we cannot have $\alpha_1(\varphi) = \alpha_1(\bar{\varphi}) = 0$, since then the resolvent cubic would have a repeated root, contradicting the fact that F is nonsingular. □

The cross terms. So far, we have shown how to find a unimodular transformation of the coordinates on $\mathbb{P}^1$ that reduces the 2-covering. (If we start with a generalised binary quartic (P, Q) then we work with $F = P^2 + 4Q$.) There is still an ambiguity coming from the possibility of making a y-substitution in the general form of a 2-covering. The most reasonable convention seems to be to arrange that the cross term coefficients l, m and n are 0 or 1.

6C. ***Reduction of* 3*-coverings.***

Using the flex points. Let $F(x, y, z) \in \mathbb{R}[x, y, z]$ be a nonsingular ternary cubic. In order to find its reduction covariant (as a positive definite quadratic form $Q(x, y, z)$), we proceed as follows. Let $H(x, y, z)$ be the Hessian of F as defined in Section 2. Then the intersection of $F = 0$ and $H = 0$ consists of nine distinct points, the flex points of F. Three of them are real, the others come in three complex conjugate pairs.

There are twelve lines each containing three of the flex points, coming in four triples of lines that do not meet in a flex point. (These triples are the "syzygetic triangles" mentioned below.) One of these triples has all three lines real, call them L_{11}, L_{12} and L_{13}. Another one has one line real, call it L_{21}, and two complex conjugate lines, call them L_{22} and L_{23}. Then Q spans the one-dimensional intersection of the spaces spanned by L_{11}^2, L_{12}^2 and L_{13}^2, and by L_{21}^2 and $L_{22}L_{23}$, respectively.

In order to see why this recipe works, first observe that it clearly defines an $\mathrm{SL}_3(\mathbb{R})$-covariant map. We can always make an $\mathrm{SL}_3(\mathbb{R})$-transformation to bring F into the standard Hesse form

$$F(x, y, z) = a(x^3 + y^3 + z^3) + bxyz.$$

Then L_{11}, L_{22} and L_{33} are x, y and z, and L_{21}, L_{22} and L_{23} are $x + y + z$, $x + \zeta_3 y + \zeta_3^2 z$ and $x + \zeta_3^2 y + \zeta_3 z$ (where ζ_3 is a primitive cube root of unity). One then looks at the intersection

$$\langle x^2, y^2, z^2\rangle \cap \langle (x + y + z)^2, x^2 + y^2 + z^2 - xy - yz - zx\rangle$$

and finds it is one-dimensional, spanned by $x^2 + y^2 + z^2$, which is the reduction covariant of any F in Hesse form.

The only way we know to implement this method in practice is by numerically solving for the flex points. If the given model is far from reduced, then usually several of the flex points are very close to one another, which makes the computation of the lines difficult. Another practical problem is that the two spaces of quadrics we compute are only approximate and therefore will usually not have a nontrivial intersection.

Using the 3*-torsion on* E. This is the method described in [Fisher 2006b, §9.5]. Let $F(x, y, z)$ be a ternary cubic with invariants c_4 and c_6 and Hessian H as defined in Section 2. Let $T = (x_T, y_T)$ be a 3-torsion point on the Jacobian

$$E: \quad y^2 = x^3 - 27c_4x - 54c_6.$$

Then the cubic $\mathcal{T}(x, y, z) = 2x_T F - 3H$ is the product of 3 linear forms. (In [Hilbert 1993, II.7] it is called a "syzygetic triangle".) Making a change of coordinates (if

necessary) we may suppose $\mathcal{T}(1,0,0) \neq 0$. We label the coefficients

$$\mathcal{T}(x,y,z) = rx^3 + s_1x^2y + s_2xy^2 + s_3y^3 + t_1x^2z + t_2xz^2 + t_3z^3 + uxyz + vy^2z + wyz^2.$$

The proof of [Fisher 2006b, Theorem 7.1] describes how to compute a formula for M_T, where the entries are polynomials in r, s_1, $s_2, \dots, w$ and y_T. Up to a scaling, this works out as $M_T = rA + 2y_T B$ where

$$\begin{aligned}
A_{11} &= -12rs_2w - 36rs_3t_2 + 12ruv + 4s_1^2w + 4s_1s_2t_2 - 8s_1t_1v - s_1u^2 + 12s_3t_1^2,\\
A_{12} &= -54rs_3w + 18rv^2 + 6s_1s_2w - 3s_1uv - 6s_2t_1v + 9s_3t_1u,\\
A_{13} &= -81rs_3t_3 + 9rvw + 9s_1s_2t_3 - 3s_1t_2v - 3s_2t_1w + 9s_3t_1t_2,\\
A_{21} &= 36rs_2t_2 - 9ru^2 - 12s_1^2t_2 + 12s_1t_1u - 12s_2t_1^2,\\
A_{22} &= 24rs_2w + 18rs_3t_2 - 15ruv - 8s_1^2w - 2s_1s_2t_2 + 10s_1t_1v + 2s_1u^2 - 3s_2t_1u - 6s_3t_1^2,\\
A_{23} &= 54rs_2t_3 - 9ruw - 18s_1^2t_3 + 6s_1t_1w + 3s_1t_2u - 6s_2t_1t_2,\\
A_{31} &= 0,\\
A_{32} &= -18rs_2v + 27rs_3u + 6s_1^2v - 3s_1s_2u - 18s_1s_3t_1 + 6s_2^2t_1,\\
A_{33} &= -12rs_2w + 18rs_3t_2 + 3ruv + 4s_1^2w - 2s_1s_2t_2 - 2s_1t_1v - s_1u^2 + 3s_2t_1u - 6s_3t_1^2
\end{aligned}$$

and $B = rB_1 + (s_1^2t_2 - s_1t_1u + s_2t_1^2)E_{13}$ with

$$B_1 = \begin{pmatrix} s_1u - 2s_2t_1 & s_1v - 3s_3t_1 & s_1w - 4s_2t_2 - t_1v + u^2 \\ -3ru + 2s_1t_1 & -3rv + s_2t_1 & -3rw + s_1t_2 \\ 6rs_2 - 2s_1^2 & 9rs_3 - s_1s_2 & 3rv - s_1u + s_2t_1 \end{pmatrix}.$$

(Notes: E_{ij} is the 3 by 3 matrix with (i,j) entry 1 and all other entries 0. Our matrices A and B would be called $r^3(\det P)A$ and r^3B in the notation of [Fisher 2006b].) This formula comes with the caveat (see [Fisher 2006b, Remark 7.2]) that it may give zero. However, as this will not happen for both T and $-T$, we may get around the problem by computing M_T as $(M_{-T})^{-1}$ if necessary.

Once we have computed M_T for all $T \in E[3]$ the reduction covariant is computed using Corollary 6.3.

6D. *Reduction of* 4-*coverings*. We could again try to find the reduction covariant starting from the 16 hyperosculating points on $\mathscr{C}$ and the quadruples of planes containing four of them, which are the analogue of the syzygetic triangles. However, this approach does not seem to be very promising.

Instead, we use the fact that below the given 4-covering $\mathscr{C}$, there is a 2-covering $\mathscr{C}_2$; let $\pi : \mathscr{C} \to \mathscr{C}_2$ be the covering map. If A and B are the symmetric matrices corresponding to the quadrics defining $\mathscr{C} \subset \mathbb{P}^3$, then $\mathscr{C}_2$ has equation $y^2 = F(x,z)$ where

$$F(x,z) := \det(Ax + Bz).$$

Applying reduction to the quartic on the right hand side, we find a good basis of the pencil of quadrics. It remains to find the reduction covariant of $\mathscr{C}$.

Let $\theta_j \in \mathbb{C}$ ($j = 1, 2, 3, 4$) be the ramification points of $\mathscr{C}_2 \to \mathbb{P}^1$, i.e., the roots of $f(X) = F(X, 1)$. Let G_j ($j = 1, 2, 3, 4$) be a linear form (unique up to scaling) describing the preimage of θ_j on $\mathscr{C} \subset \mathbb{P}^3$. Then (fixing the polynomials giving the covering map $\pi : \mathscr{C} \to \mathscr{C}_2$) there are $\alpha_j \in \mathbb{C}^\times$ such that

$$(x - \theta_j z) \circ \pi = \alpha_j G_j^2.$$

Now the action of $T \in E[4]$ on $\mathscr{C}$ induces the action of $2T \in E[2]$ on $\mathscr{C}_2$. Therefore the action of $T \in E[2]$ on $\mathscr{C}$ will be trivial on $\mathscr{C}_2$, hence the corresponding matrix $M_T \in \mathrm{SL}_4$ will fix the G_j up to sign. In fact, it can be checked that the action of $E[2]$ on $\mathbb{P}^3$ lifts to a representation on $\mathbb{C}^4$, which is isomorphic to the regular representation, and the G_j span the four eigenspaces. So any Hermitian form that is invariant under H_4 must be invariant under $E[2]$ and thus be of the form

$$\sum_{j=1}^{4} \lambda_j |G_j|^2.$$

It remains to determine the coefficients λ_j.

Proposition 6.9. *Keep the notation introduced so far, and let $f(X) = F(X, 1)$. Then the reduction covariant of $\mathscr{C}$ is the positive definite Hermitian form*

$$\varphi_{\mathbb{C}}(\mathscr{C}) = \sum_{j=1}^{4} \frac{|\alpha_j|}{|f'(\theta_j)|^{1/2}} |G_j|^2.$$

If $\mathscr{C}$ is defined over $\mathbb{R}$, then the restriction of this Hermitian form to $\mathbb{R}^4$ will be the positive definite quadratic form $\varphi_{\mathbb{R}}(\mathscr{C})$.

Proof. We first check that the given form is invariant under $\mathrm{SL}_2(\mathbb{C})$ acting on $\mathbb{P}^1$ (that is, does not depend on the choice of basis of the pencil of quadrics). We know (see page 805) that $\sum_{j=1}^{4} |f'(\theta_j)|^{-1} |x - \theta_j z|^2$ is an $\mathrm{SL}_2(\mathbb{C})$-covariant; the same computation (which deals with each summand separately) shows that $\sum_{j=1}^{4} |f'(\theta_j)|^{-1/2} |x - \theta_j z|$ is a covariant as well. But $|x - \theta_j z| = |\alpha_j G_j^2|$, and the coordinates in G_j are not affected by the $\mathrm{SL}_2(\mathbb{C})$-action, so the expression given in the statement is invariant.

Now we check that the given form is covariant with respect to the action of $\mathrm{SL}_4(\mathbb{C})$. But this is clear since every $\alpha_j G_j^2$ is covariant.

Since we can move any $\mathscr{C}$ into standard form by the action of $\mathrm{SL}_2(\mathbb{C}) \times \mathrm{SL}_4(\mathbb{C})$, it now suffices to verify that our formula gives the correct result when $\mathscr{C}$ is in standard form

$$a(x_0^2 + x_2^2) + 2bx_1x_3 = a(x_1^2 + x_3^2) + 2bx_0x_2 = 0.$$

In this case, the 2-covering $\mathscr{C}_2$ is given by

$$y^2 = (a^4 + b^4)x^2z^2 - a^2b^2(x^4 + z^4)$$

and the map π (see Lemma 4.6 for formulae), followed by the map $\mathscr{C}_2 \to \mathbb{P}^1$, is given by

$$(x : z) = \big(b^3(x_1^2 + x_3^2) + 2a^3x_0x_2 : -b^3(x_0^2 + x_2^2) - 2a^3x_1x_3\big).$$

The roots θ_j of $f(X) = -a^2b^2X^4 + (a^4 + b^4)X^2 - a^2b^2$ are a/b, $-a/b$, b/a and $-b/a$, and up to a common factor $b^4 - a^4$, we can take $\alpha_j = 1/b$, $1/b$, $1/a$, $-1/a$ and $G_j = x_1 - x_3$, $x_1 + x_3$, $x_0 - x_2$, $x_0 + x_2$. Also, $|f'(\theta_j)| = c|\theta_j|$ for some constant c. Since $|\alpha_j|/|\theta_j|^{1/2}$ has the same value $|ab|^{-1/2}$ for all j, our expression gives, up to a constant factor again,

$$|x_1 - x_3|^2 + |x_1 + x_3|^2 + |x_0 - x_2|^2 + |x_0 + x_2|^2 = 2(|x_0|^2 + |x_1|^2 + |x_2|^2 + |x_3|),$$

which is the correct result for a 4-covering in standard form. □

In order to find the α_j and G_j, we can make use of a result from [Fisher 2008b], where it is observed that $\alpha_j G_j^2$ is the quadratic form corresponding to the matrix

$$e\theta_j^{-1}A + M_1 + \theta_j M_2 + a\theta_j^2 B;$$

here $F(x, z) = \det(Ax + Bz) = ax^4 + bx^3z + cx^2z^2 + dxz^3 + ez^4$ and M_1 and M_2 are obtained from relation (4-3).

7. Examples

In this section we illustrate minimisation and reduction for two explicit examples over $\mathbb{Q}$ (one a 3-covering and the other a 4-covering). We then give references to further examples.

7A. ***Minimisation and reduction of a 3-covering.*** We consider the elliptic curve 105630d1 in [Cremona n.d.] with Weierstrass equation

$$E: \qquad y^2 + xy = x^3 + x^2 - 114848533x - 472424007827.$$

Computing the 3-Selmer group [Schaefer and Stoll 2004] we find $\mathrm{Sel}^{(3)}(\mathbb{Q}, E) \cong \mathbb{Z}/3\mathbb{Z}$. In [Cremona et al. 2008; 2009; n.d.] we show how to write down elements of the 3-Selmer group explicitly as 3-coverings of E. In this case our MAGMA programs find (before minimisation and reduction) that a generator is represented by the 3-covering $\mathscr{C} \subset \mathbb{P}^2$ with equation

$$\begin{aligned} F_1(x, y, z) = {} & 27089x^3 + 2142y^3 + 291938z^3 + 10008x^2y - 127341x^2z \\ & + 92937xy^2 + 104736y^2z + 21093xz^2 - 71172yz^2 - 2655xyz. \end{aligned}$$

(Random choices in the programs mean it need not return the same cubic every time. However, the answer will always be $\mathbb{Q}$-equivalent to F_1, and this can be checked using the algorithm in [Fisher 2006b].) The discriminant of this ternary cubic is $\Delta(F_1) = 3^{12} \cdot 503^{12} \cdot \Delta_E$ where $\Delta_E = 2^{39} \cdot 3 \cdot 5^9 \cdot 7^3 \cdot 503$ is the minimal discriminant of E. So F_1 has level 1 at the primes 3 and 503. Reducing mod 3 we find $F_1(x, y, z) = 2(x+z)^3 \pmod 3$. The level is decreased by the first iteration of our algorithm (see Theorem 4.3). Explicitly, we put

$$F_2(x, y, z) = \frac{1}{3^2} F_1(3x - y, z, y).$$

Likewise we find $F_2(x, y, z) \equiv 284(x+329y+33z)^3 \pmod{503}$ and our algorithm puts

$$\begin{aligned} F_3(x, y, z) &= \frac{1}{503^2} F_2(503x - 33y + z, z, y - 10z) \\ &= 40877301x^3 - 11504y^3 + 12z^3 - 8035425x^2y - 64887x^2z \\ &\qquad + 526580xy^2 - 200y^2z + 5803xz^2 - 383yz^2 + 7307xyz. \end{aligned}$$

The 3-torsion of $y^2 = x^3 - 27c_4x - 54c_6$ over $\mathbb{C}$ is generated by

$$\begin{aligned} S &= (667989.968057, 420236746.168), \\ T &= (-264330.994609, 34120617.5970i). \end{aligned}$$

The formulae in Section 6C show that S and T act on $\{F_3 = 0\}$ via

$$M_S = \begin{pmatrix} 285.46 & -19.022 & 3.4264 \\ 4352.6 & -290.04 & 52.341 \\ 509.05 & -33.785 & 4.5806 \end{pmatrix}$$

(entries being given to five-digit precision) and

$$M_T = \begin{pmatrix} -50.656 + 47.060i & 3.2758 - 3.3464i & 0.11909 + 2.2683i \\ -786.55 + 717.15i & 50.871 - 51.000i & 1.8675 \ \ + 34.587i \\ -119.84 + 93.073i & 7.8268 - 6.5354i & -0.21547 + 3.9405i \end{pmatrix}.$$

We have scaled these matrices to have determinant 1. By Corollary 6.3 the matrix of the reduction covariant (given to twelve-digit precision) is

$$A = \begin{pmatrix} 176413988.185 & -11560848.1174 & 3471.84429193 \\ -11560848.1174 & 757736.524016 & -1499.92503970 \\ 3471.84429193 & -1499.92503970 & 13237.5156939 \end{pmatrix}.$$

Running the LLL algorithm on the lattice with Gram matrix A results in the unimodular transformation

$$U = \begin{pmatrix} 0 & 0 & 1 \\ 4 & 61 & 6 \\ -3 & -46 & -4 \end{pmatrix}.$$

Accordingly we put $F_4(x, y, z) = F_3(4y - 3z, 61y - 46z, x + 6y - 4z)$ and find

$$F_4(x, y, z) = \\ 12x^3 + 12y^3 + 171z^3 + 65x^2y + 65x^2z - 94y^2z + 87xz^2 + 101yz^2 + 7xyz.$$

This ternary cubic has solution

$$(x : y : z) = (345420 : -1638959 : -373029),$$

which by the formulae in [An et al. 2001] maps down to a point

$$x = \frac{-748726207736084226230587579149810652 17}{109435039457696221^2},$$

$$y = \frac{5104304702532038917609849430784779872295822806191640758 7}{109435039457696221^3}$$

on $E(\mathbb{Q})$ of canonical height $86.5313\ldots$. Since the torsion subgroup of $E(\mathbb{Q})$ is trivial, it follows that rank $E(\mathbb{Q}) = 1$. It is equally convenient to find this generator using Heegner points.

Note that the MAGMA implementation of 3-descent does the minimisation and reduction automatically. To extract the intermediate model $F_1(x, y, z) = 0$, one should first specify that 3-descent prints out some of its working, using the command `SetVerbose("ThreeDescent",1);`

7B. ***Minimisation and reduction of a 4-covering.*** In [Skorobogatov 2001, §8.1], an example is given of a 4-covering $\mathscr{C}$ of the elliptic curve $E : y^2 = x^3 - 1221$ that represents an element of exact order 4 in the Shafarevich–Tate group of E. The symmetric matrices corresponding to the two quadrics defining $\mathscr{C} \subset \mathbb{P}^3$ are given as (to keep with our convention, we multiply by 2 so that entries are the second partial derivatives)

$$A = 2\begin{pmatrix} -1 & 11 & -66 & 396 \\ 11 & -66 & 396 & -2520 \\ -66 & 396 & -2520 & 16335 \\ 396 & -2520 & 16335 & -105786 \end{pmatrix}, \quad B = 2\begin{pmatrix} -1 & -3 & 33 & -198 \\ -3 & 33 & -198 & 1188 \\ 33 & -198 & 1188 & -7560 \\ -198 & 1188 & -7560 & 49005 \end{pmatrix}.$$

We will use $x_1, \ldots, x_4$ as the coordinates on $\mathbb{P}^3$. We find that

$$\det(Ax + Bz) = 2^4 \cdot 3^8(-9x^4 + 13x^3z - 18x^2z^2 + 3z^4),$$

which makes it clear that the model is non-minimal at $p=2$ and $p=3$. We compute that the discriminant of our quadric intersection is $(2\cdot 3^4)^{12}$ times the (minimal) discriminant $-2^4\,3^5\,11^2\,37^2$ of E, which shows that the level at 2 is 1 and the level at 3 is 4; the model is already minimal at all other primes.

We first minimise at $p=3$. According to our algorithm (see Section 4C), we have to look at the reductions of A and B mod 3, which are

$$\bar{A}=\begin{pmatrix}1&1&0&0\\1&0&0&0\\0&0&0&0\\0&0&0&0\end{pmatrix}\quad\text{and}\quad \bar{B}=\begin{pmatrix}1&0&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{pmatrix}$$

The common nullity is $s=2$, and the reduced quadratic forms already involve only the first two variables. They represent zero simultaneously over $\mathbb{F}_3$; the plane $x_1=0$ is contained in the reduction of the curve. So we apply the transformation $[\frac{1}{3}I_2, \mathrm{Diag}(3,1,1,1)]$, resulting in the new pair of matrices (which we will again denote by A and B)

$$A=\begin{pmatrix}-6&22&-132&792\\22&-44&264&-1680\\-132&264&-1680&10890\\792&-1680&10890&-70524\end{pmatrix},\quad B=\begin{pmatrix}-6&-6&66&-396\\-6&22&-132&792\\66&-132&792&-5040\\-396&792&-5040&32670\end{pmatrix}$$

The level at $p=3$ of the new model is 3. Reducing mod 3, we have now

$$\bar{A}=\begin{pmatrix}0&1&0&0\\1&1&0&0\\0&0&0&0\\0&0&0&0\end{pmatrix},\quad \bar{B}=\begin{pmatrix}0&0&0&0\\0&1&0&0\\0&0&0&0\\0&0&0&0\end{pmatrix}$$

The common nullity is again $s=2$, and there is a plane contained in the reduction. This time, the plane is $x_2=0$, so we swap x_1 and x_2 before we apply $[\frac{1}{3}I_2, \mathrm{Diag}(3,1,1,1)]$. The result is a model of level 2:

$$A=\begin{pmatrix}-132&22&264&-1680\\22&-2&-44&264\\264&-44&-560&3630\\-1680&264&3630&-23508\end{pmatrix},\quad B=\begin{pmatrix}66&-6&-132&792\\-6&-2&22&-132\\-132&22&264&-1680\\792&-132&-1680&10890\end{pmatrix}$$

Now we get a different situation mod 3:

$$\bar{A}=\begin{pmatrix}0&1&0&0\\1&1&1&0\\0&1&1&0\\0&0&0&0\end{pmatrix},\quad \bar{B}=\begin{pmatrix}0&0&0&0\\0&1&1&0\\0&1&0&0\\0&0&0&0\end{pmatrix}$$

The common nullity is $s = 1$. We swap x_1 and x_4 so that the reduced forms only involve the last three variables. Then we see that we are in 'Situation 2', so we apply the transformation $[I_2, \mathrm{Diag}(\frac{1}{3}, 1, 1, 1)]$. This results in a model of level 1, given by

$$A = \begin{pmatrix} -2612 & 88 & 1210 & -560 \\ 88 & -2 & -44 & 22 \\ 1210 & -44 & -560 & 264 \\ -560 & 22 & 264 & -132 \end{pmatrix}, \quad B = \begin{pmatrix} 1210 & -44 & -560 & 264 \\ -44 & -2 & 22 & -6 \\ -560 & 22 & 264 & -132 \\ 264 & -6 & -132 & 66 \end{pmatrix}$$

In the last minimisation step at $p = 3$, the reductions are now

$$\bar{A} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}, \quad \bar{B} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

The common nullity is once more $s = 1$, while the common kernel is spanned by $(1, -1, 0, 0)$. We move it to $(1, 0, 0, 0)$ and are in Situation 2 again. After applying $[I_2, \mathrm{Diag}(\frac{1}{3}, 1, 1, 1)]$, we obtain a model that is now minimal at $p = 3$.

$$A = \begin{pmatrix} -310 & 30 & 418 & -194 \\ 30 & -2 & -44 & 22 \\ 418 & -44 & -560 & 264 \\ -194 & 22 & 264 & -132 \end{pmatrix}, \quad B = \begin{pmatrix} 144 & -14 & -194 & 90 \\ -14 & -2 & 22 & -6 \\ -194 & 22 & 264 & -132 \\ 90 & -6 & -132 & 66 \end{pmatrix}$$

We still have to minimise at $p = 2$, using the algorithm described in Section 4D. We first find the "double" of our model:

$$\begin{aligned} \mathfrak{d}'(A, B) = (P, Q) = \big(&2^2(6413x^2 - 5665xz + 1248z^2), \\ &2^2(41126578x^4 - 72659303x^3z \\ &\quad + 48099091x^2z^2 - 14139840xz^3 + 1557501z^4)\big) \end{aligned}$$

We see that we already have $v_2(P) \geq 1$ and $v_2(Q) \geq 2$. The common kernel of the reductions mod 2 of the two quadratic forms is spanned by $(1, 1, 0, 1)$ and $(0, 0, 1, 0)$, so the common nullity is $s = 2$. We change coordinates so that the common kernel is given by $x_1 = x_2 = 0$. Then the reductions of the quadrics are x_1^2 and x_2^2, so they do not simultaneously represent zero. We apply the 'flip-flop' transformation $[\frac{1}{2}I_2, \mathrm{Diag}(2, 2, 1, 1)]$, after which the reductions are x_3x_4 and x_4^2, so now there is the plane $x_4 = 0$ contained in the reduction of the curve. We swap x_1 and x_4 and then apply $[\frac{1}{2}I_2, \mathrm{Diag}(2, 1, 1, 1)]$ to obtain a pair of matrices

representing a globally minimal model:

$$A = \begin{pmatrix} -728 & -424 & 319 & -474 \\ -424 & -252 & 187 & -280 \\ 319 & 187 & -140 & 209 \\ -474 & -280 & 209 & -310 \end{pmatrix}, \quad B = \begin{pmatrix} 348 & 198 & -152 & 220 \\ 198 & 114 & -86 & 130 \\ -152 & -86 & 66 & -97 \\ 220 & 130 & -97 & 144 \end{pmatrix}$$

We now apply reduction to this model as described in Section 6D. We have

$$\det(Ax + Bz) = 4(-9x^4 + 13x^3z - 18x^2z^2 + 3z^4)\,.$$

Following [An et al. 2001] and [Fisher 2008b], we compute the quadratic forms T_1, T_2 whose symmetric matrices M_1, M_2 are given by

$$\operatorname{adj}\big(\operatorname{adj}(A)x + \operatorname{adj}(B)z)\big) = 4^2 \cdot 81Ax^3 - 4 \cdot 9M_1x^2z + 4 \cdot 3M_2xz^2 + 4^2 \cdot 9Bz^3\,.$$

Then, writing Q_1 and Q_2 for the quadratic forms corresponding to A and B,

$$\alpha G^2 = 12\theta^{-1}Q_1 + T_1 + \theta T_2 - 36\theta^2 Q_2$$

for θ a root of $f(X) = \det(XA + B)$. We can for example take

$$G = (-18\theta^3 - 28\theta^2 + 6\theta + 2)x_1 + (18\theta^3 - 26\theta^2 + 2)x_2 + (18\theta^2 + \theta - 3)x_3 - 2x_4$$

and $\alpha = -1395\theta^3 + 1367\theta^2 - 2155\theta - 1001$. Also, $f'(\theta) = 12(-12\theta^3 + 13\theta^2 - 12\theta)$. The matrix corresponding to $\sqrt{12}\sum_\theta |\alpha||G|^2/|f'(\theta)|^{1/2}$ is (to five decimal places)

$$\begin{pmatrix} 8857.72019 & 5117.00780 & -3885.97776 & 5665.67630 \\ 5117.00780 & 3080.24124 & -2279.16858 & 3348.18401 \\ -3885.97776 & -2279.16858 & 1716.07038 & -2498.36286 \\ 5665.67630 & 3348.18401 & -2498.36286 & 3706.96839 \end{pmatrix}.$$

We apply LLL to this Gram matrix and obtain the reducing transformation matrix

$$U = \begin{pmatrix} -5 & -2 & -6 & 0 \\ -6 & -3 & -7 & -1 \\ -15 & -7 & -17 & 0 \\ 3 & 1 & 4 & 1 \end{pmatrix},$$

which finally brings the two matrices defining $\mathscr{C}$ into the form

$$U^tAU = \begin{pmatrix} -2 & 0 & -1 & -2 \\ 0 & -2 & -1 & 0 \\ -1 & -1 & -2 & 2 \\ -2 & 0 & 2 & -2 \end{pmatrix}, \quad U^tBU = \begin{pmatrix} 0 & 0 & -1 & 1 \\ 0 & 2 & -1 & -1 \\ -1 & -1 & 0 & -1 \\ 1 & -1 & -1 & -2 \end{pmatrix}.$$

These correspond, after a sign change, to the quadratic forms

$$Q_1 = x_1^2 + x_1x_3 + 2x_1x_4 + x_2^2 + x_2x_3 + x_3^2 - 2x_3x_4 + x_4^2,$$
$$Q_2 = x_1x_3 - x_1x_4 - x_2^2 + x_2x_3 + x_2x_4 + x_3x_4 + x_4^2\,.$$

7C. ***Further examples and applications.*** One useful application of the methods described in this paper is to help find large generators in the Mordell–Weil group of an elliptic curve E. This has already be demonstrated in Section 7A. Each rational point $P \in E(\mathbb{Q})$ lifts to one of the n-coverings of E. If we have a nice and small (*i.e.*, minimised and reduced) model $\mathscr{C}$ of this n-covering, then the logarithmic height with respect to $\mathscr{C} \to \mathbb{P}^{n-1}$ of the preimage Q of P in $\mathscr{C}(\mathbb{Q})$ will be smaller by a factor of about $\frac{1}{2n}$ than the logarithmic x-coordinate height of P — standard properties of heights imply that

$$h(Q) = \frac{1}{2n} h_x(P) + O(1)$$

where the implied constant depends on the equations defining $\mathscr{C} \to \mathbb{P}^{n-1}$. If the equations have small coefficients, this constant should be small as well. Therefore we can hope to find P much more easily by searching for Q on $\mathscr{C}$. In fact, this application was the motivation for the first tentative steps towards reduction of 4-coverings. The story begins with [Gebel et al. 1998], where the authors determined Mordell–Weil generators for all Mordell curves $y^2 = x^3 + D$, with D a nonzero integer of absolute value at most 10^4 (in order to determine all the integral points on these curves), with one exception, $D = 7823$. The analytic rank of this curve is 1, so we know that the Mordell–Weil rank must be also 1; however the Birch–Swinnerton-Dyer Conjecture predicts a generator of fairly large height. One of us (Stoll) used minimisation and reduction of 4-coverings in a fairly *ad hoc* fashion to find a good model of the one relevant 4-covering of $E : y^2 = x^3 + 7823$, so that a point search on this 4-covering curve was successful, thus resolving this last open case. The result was reported in a posting [Stoll 2002] to the `NMBRTHRY` mailing list. We give a short summary of the steps and the result. By a standard 2-descent, one obtains a 2-covering curve

$$C : y^2 = -18x^4 + 116x^3 + 48x^2 - 12x + 30\,.$$

A second 2-descent on C following [Merriman et al. 1996] produces a 4-covering of E, whose initial model was given by quadrics with coefficients of up to 15 decimal digits. Using the methods described here, one finds a model $D \subset \mathbb{P}^3$ given by

$$2x_1x_2 + x_1x_3 + x_1x_4 + x_2x_4 + x_3^2 - 2x_4^2 = 0,$$
$$x_1^2 + x_1x_3 - x_1x_4 + 2x_2^2 - x_2x_3 + 2x_2x_4 - x_3^2 - x_3x_4 + x_4^2 = 0.$$

It is not very difficult to find the point $P = (116 : 207 : 474 : -332)$ on D. This point then gives rise to the point

$$Q = \left(\frac{53463613}{32109353}, \frac{23963346820191122}{32109353^2}\right)$$

on C, which in turn finally produces the Mordell–Weil generator on E, with coordinates

$$x = \frac{226358214332142150210020923351777 7}{11981673410095561^2},$$

$$y = \frac{1863981525846233056248375514855967700281447766557 56}{11981673410095561^3}.$$

Note that in the version given in the mailing list posting, the model was not minimal at 2 (in fact, it had level 2 at 2).

4-descent including minimisation and reduction was also used to find some of the elliptic curves of high rank and prescribed torsion listed in [Dujella n.d.], for example the curve with $E(\mathbb{Q}) \cong \mathbb{Z}/12\mathbb{Z} \times \mathbb{Z}^4$.

Minimised and reduced models of 2-, 3-, and 4-coverings provide the starting point for the computation of 6- and 12-coverings as described in [Fisher 2008c]. These then allow us to find even larger generators (of logarithmic canonical height > 600). For example, this method was used to find the last missing generators for curves of prime conductor and rank at least 2 in the Stein–Watkins database [2002].

A table giving representatives of all elements of order 3 in the Shafarevich–Tate groups of all elliptic curves of conductor $< 130\,000$ can be found at [Fisher n.d.]. (It is only known that the table is complete if one assumes the conjecture of Birch and Swinnerton-Dyer.) The final form of these ternary cubics was obtained by applying the methods described in this paper to the original models produced by the algorithms described in [Schaefer and Stoll 2004] and [Cremona et al. 2008; 2009; n.d.].

References

[An et al. 2001] S. Y. An, S. Y. Kim, D. C. Marshall, S. H. Marshall, W. G. McCallum, and A. R. Perlis, "Jacobians of genus one curves", *J. Number Theory* **90**:2 (2001), 304–315. MR 2002g:14040 Zbl 1066.14035

[Artin et al. 2005] M. Artin, F. Rodriguez-Villegas, and J. Tate, "On the Jacobians of plane cubics", *Adv. Math.* **198**:1 (2005), 366–382. MR 2006h:14043 Zbl 1092.14054

[Birch and Swinnerton-Dyer 1963] B. J. Birch and H. P. F. Swinnerton-Dyer, "Notes on elliptic curves. I", *J. Reine Angew. Math.* **212** (1963), 7–25. MR 26 #3669 Zbl 0118.27601

[Bosch et al. 1990] S. Bosch, W. Lütkebohmert, and M. Raynaud, *Néron models*, Ergebnisse der Math. (3) **21**, Springer, Berlin, 1990. MR 91i:14034

[Bosma et al. 1997] W. Bosma, J. Cannon, and C. Playoust, "The Magma algebra system, I: The user language", *J. Symbolic Comput.* **24**:3-4 (1997), 235–265. MR 1484478

[Connell 1996] I. Connell, "Elliptic curve handbook", on-line notes, McGill University, 1996, available at http://www.math.mcgill.ca/connell/public/ECH1/.

[Cremona 1997] J. E. Cremona, *Algorithms for modular elliptic curves*, 2nd ed., Cambridge University Press, Cambridge, 1997. MR 99e:11068 Zbl 0872.14041

[Cremona 1999] J. E. Cremona, "Reduction of binary cubic and quartic forms", *LMS J. Comput. Math.* **2** (1999), 64–94. MR 2000f:11040

[Cremona n.d.] J. E. Cremona, Elliptic curve data, available at http://www.warwick.ac.uk/staff/J.E.Cremona/ftp/data/INDEX.html.

[Cremona et al. 2008] J. E. Cremona, T. A. Fisher, C. O'Neil, D. Simon, and M. Stoll, "Explicit n-descent on elliptic curves, I: Algebra", *J. Reine Angew. Math.* **615** (2008), 121–155. MR 2009g:11067

[Cremona et al. 2009] J. E. Cremona, T. A. Fisher, C. O'Neil, D. Simon, and M. Stoll, "Explicit n-descent on elliptic curves, II: Geometry", *J. Reine Angew. Math.* **632** (2009), 63–84. MR 2544143

[Cremona et al. n.d.] J. E. Cremona, T. A. Fisher, C. O'Neil, D. Simon, and M. Stoll, "Explicit n-descent on elliptic curves, III: Algorithms", in preparation.

[Deligne 1975] P. Deligne, "Courbes elliptiques: formulaire d'après J. Tate", pp. 53–73 in *Modular functions of one variable, IV* (Antwerp, 1972), edited by B. J. Birch and W. Kuyk, Lecture Notes in Math. **476**, Springer, Berlin, 1975. MR 52 #8135

[Djabri and Smart 1998] Z. Djabri and N. P. Smart, "A comparison of direct and indirect methods for computing Selmer groups of an elliptic curve", pp. 502–513 in *Algorithmic number theory* (Portland, OR, 1998), edited by J. Buhler, Lecture Notes in Comput. Sci. **1423**, Springer, Berlin, 1998. MR 2001f:11086 Zbl 0915.11034

[Dolgachev 2003] I. Dolgachev, *Lectures on invariant theory*, London Mathematical Society Lecture Note Series **296**, Cambridge University Press, Cambridge, 2003. MR 2004g:14051 Zbl 1023.13006

[Dujella n.d.] A. Dujella, "High rank elliptic curves with prescribed torsion", online table, available at http://web.math.hr/~duje/tors/tors.html.

[Fisher 2006a] T. Fisher, "The Hessian of a genus one curve", preprint, 2006.

[Fisher 2006b] T. Fisher, "Testing equivalence of ternary cubics", pp. 333–345 in *Algorithmic number theory*, edited by F. Hess et al., Lecture Notes in Comput. Sci. **4076**, Springer, Berlin, 2006. MR 2007j:11074 Zbl 1143.11325

[Fisher 2007] T. Fisher, "A new approach to minimising binary quartics and ternary cubics", *Math. Res. Lett.* **14**:4 (2007), 597–613. MR 2008k:11058 Zbl 1142.11038

[Fisher 2008a] T. Fisher, "The invariants of a genus one curve", *Proc. Lond. Math. Soc.* (3) **97**:3 (2008), 753–782. MR 2009j:11087 Zbl 05365466

[Fisher 2008b] T. Fisher, "Some improvements to 4-descent on an elliptic curve", pp. 125–138 in *Algorithmic number theory*, edited by A. van der Poorten and A. Stein, Lecture Notes in Comput. Sci. **5011**, Springer, Berlin, 2008. MR 2009m:11078 Zbl 05279282

[Fisher 2008c] T. A. Fisher, "Finding rational points on elliptic curves using 6-descent and 12-descent", *J. Algebra* **320**:2 (2008), 853–884. MR 2009g:11068 Zbl 1149.14025

[Fisher n.d.] T. A. Fisher, "Elements of order 3 in the Tate-Shafarevich group", online table, available at http://www.dpmms.cam.ac.uk/~taf1000/g1data/order3.html.

[Gebel et al. 1998] J. Gebel, A. Pethő, and H. G. Zimmer, "On Mordell's equation", *Compositio Math.* **110**:3 (1998), 335–367. MR 98m:11049 Zbl 0899.11013

[Hilbert 1993] D. Hilbert, *Theory of algebraic invariants*, Cambridge University Press, Cambridge, 1993. Translated by Reinhard C. Laubenbacher from handwritten course notes, taken by Sophus Marxsen. MR 97j:01049 Zbl 0801.13001

[Hodge and Pedoe 1952] W. V. D. Hodge and D. Pedoe, *Methods of algebraic geometry*, vol. II, Cambridge University Press, Cambridge, 1952. MR 95d:14002b Zbl 0048.14502

[Hulek 1986] K. Hulek, *Projective geometry of elliptic curves*, Astérisque **137**, Société Mathématique de France, Paris, 1986. MR 88c:14046 Zbl 0602.14024

[Jacobson 1985] N. Jacobson, *Basic algebra, I*, 2nd ed., W. H. Freeman and Company, New York, 1985. MR 86d:00001 Zbl 0557.16001

[Julia 1917] G. Julia, "Étude sur les formes binaires non quadratiques à indeterminées réelles ou complexes", *Mem. Acad. Sci. l'Inst. France* **55** (1917), 1–293.

[Kollár 1997] J. Kollár, "Polynomials with integral coefficients, equivalent to a given polynomial", *Electron. Res. Announc. Amer. Math. Soc.* **3** (1997), 17–27. MR 98g:11076 Zbl 0867.11047

[Kraus 1989] A. Kraus, "Quelques remarques à propos des invariants c_4, c_6 et Δ d'une courbe elliptique", *Acta Arith.* **54**:1 (1989), 75–80. MR 90j:11045 Zbl 0628.14024

[Laska 1982] M. Laska, "An algorithm for finding a minimal Weierstrass equation for an elliptic curve", *Math. Comp.* **38**:157 (1982), 257–260. MR 84e:14033 Zbl 0493.14016

[Lenstra et al. 1982] A. K. Lenstra, H. W. Lenstra, Jr., and L. Lovász, "Factoring polynomials with rational coefficients", *Math. Ann.* **261**:4 (1982), 515–534. MR 84a:12002 Zbl 0488.12001

[Liu 1996] Q. Liu, "Modèles entiers des courbes hyperelliptiques sur un corps de valuation discrète", *Trans. Amer. Math. Soc.* **348**:11 (1996), 4577–4610. MR 97h:11062 Zbl 0926.11043

[Merriman et al. 1996] J. R. Merriman, S. Siksek, and N. P. Smart, "Explicit 4-descents on an elliptic curve", *Acta Arith.* **77**:4 (1996), 385–404. MR 97j:11027 Zbl 0873.11036

[Milne 2008] J. S. Milne, "Lectures on étale cohomology", v. 2.10, 2008, available at http://www.jmilne.org/math/CourseNotes/LEC.pdf.

[Poonen 2001] B. Poonen, "An explicit algebraic family of genus-one curves violating the Hasse principle", *J. Théor. Nombres Bordeaux* **13**:1 (2001), 263–274. MR 2002e:14036 Zbl 1046.11038

[Raynaud 1970] M. Raynaud, *Anneaux locaux henséliens*, Lecture Notes in Math. **169**, Springer, Berlin, 1970. MR 43 #3252 Zbl 0203.05102

[Sadek 2009] M. Sadek, *Models of genus one curves*, Ph.D. thesis, University of Cambridge, 2009.

[Schaefer and Stoll 2004] E. F. Schaefer and M. Stoll, "How to do a p-descent on an elliptic curve", *Trans. Amer. Math. Soc.* **356**:3 (2004), 1209–1231. MR 2004g:11045 Zbl 1119.11029

[Serre 1979] J.-P. Serre, *Local fields*, Graduate Texts in Mathematics **67**, Springer, New York, 1979. MR 82e:12016 Zbl 0423.12016

[Siksek 1995] S. Siksek, *Descent on curves of genus one*, Ph.D. thesis, University of Exeter, 1995, available at http://www.warwick.ac.uk/staff/S.Siksek/papers/phdnew.pdf.

[Silverman 1986] J. H. Silverman, *The arithmetic of elliptic curves*, Graduate Texts in Mathematics **106**, Springer, New York, 1986. MR 87g:11070 Zbl 0585.14026

[Silverman 1994] J. H. Silverman, *Advanced topics in the arithmetic of elliptic curves*, Graduate Texts in Mathematics **151**, Springer, New York, 1994. MR 96b:11074 Zbl 0911.14015

[Skorobogatov 2001] A. Skorobogatov, *Torsors and rational points*, Cambridge Tracts in Mathematics **144**, Cambridge University Press, Cambridge, 2001. MR 2002d:14032 Zbl 0972.14015

[Stein and Watkins 2002] W. A. Stein and M. Watkins, "A database of elliptic curves: first report", pp. 267–275 in *Algorithmic number theory* (Sydney, 2002), Lecture Notes in Comput. Sci. **2369**, Springer, Berlin, 2002. MR 2005h:11113 Zbl 1058.11036

[Stoll 2002] M. Stoll, Posting to NMBRTHRY mailing list, 2002, available at http://tinyurl.com/2bgpxfd.

[Stoll and Cremona 2002] M. Stoll and J. E. Cremona, "Minimal models for 2-coverings of elliptic curves", *LMS J. Comput. Math.* **5** (2002), 220–243. MR 2003j:11062 Zbl 1067.11031

[Stoll and Cremona 2003] M. Stoll and J. E. Cremona, "On the reduction theory of binary forms", *J. Reine Angew. Math.* **565** (2003), 79–99. MR 2005e:11091 Zbl 1153.11317

[Tate 1975] J. Tate, "Algorithm for determining the type of a singular fiber in an elliptic pencil", pp. 33–52 in *Modular functions of one variable, IV* (Antwerp, 1972), edited by B. J. Birch and W. Kuyk, Lecture Notes in Math. **476**, Springer, Berlin, 1975. MR 52 #13850

[Weil 1954] A. Weil, "Remarques sur un mémoire d'Hermite", *Arch. Math.* (*Basel*) **5** (1954), 197–202. MR 15,896d Zbl 0056.03402

[Weil 1983] A. Weil, "Euler and the Jacobians of elliptic curves", pp. 353–359 in *Arithmetic and geometry, I*, edited by M. Artin and J. Tate, Progr. Math. **35**, Birkhäuser, Boston, MA, 1983. MR 85d:14060 Zbl 0554.01014

[Womack 2003] T. Womack, *Explicit descent on elliptic curves*, Ph.D. thesis, University of Nottingham, 2003, available at http://www.warwick.ac.uk/staff/J.E.Cremona/theses/.

Communicated by Karl Rubin
Received 2010-01-19 Accepted 2010-07-18

J.E.Cremona@warwick.ac.uk *Mathematics Institute, Zeeman Building, University of Warwick, Coventry, CV4 7AL, United Kingdom*
http://www.warwick.ac.uk/staff/J.E.Cremona

T.A.Fisher@dpmms.cam.ac.uk *DPMMS, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WB, United Kingdom*
http://www.dpmms.cam.ac.uk/~taf1000/

Michael.Stoll@uni-bayreuth.de *Universität Bayreuth, Mathematisches Institut, 95440 Bayreuth, Germany*
http://www.mathe2.uni-bayreuth.de/stoll/

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at the ANT website.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles in ANT are usually in English, but articles written in other languages are welcome.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and subject classifications for the article, and, for each author, postal address, affiliation (if appropriate), and email address.

Format. Authors are encouraged to use LaTeX but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages (Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc.) allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@mathscipub.org with details about how your graphics were generated.

White space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

Algebra & Number Theory

Volume 4 No. 6 2010

Generalized moonshine I: Genus-zero functions 649
SCOTT CARNAHAN

Integral trace forms associated to cubic extensions 681
GUILLERMO MANTILLA-SOLER

Parabolic induction and Hecke modules in characteristic p for p-adic GL_n 701
RACHEL OLLIVIER

Patching and admissibility over two-dimensional complete local domains 743
DANNY NEFTIN and ELAD PARAN

Minimisation and reduction of 2-, 3- and 4-coverings of elliptic curves 763
JOHN E. CREMONA, TOM A. FISHER and MICHAEL STOLL