CONTEMPORARY MATHEMATICS

133

p-Adic Methods in Number Theory and Algebraic Geometry

American Mathematical Society

p-Adic Methods in Number Theory and Algebraic Geometry

Recent Titles in This Series

133 **Alan Adolphson, Steven Sperber, and Marvin Tretkoff, Editors,** p-Adic methods in number theory and algebraic geometry, 1992

132 **Mark Gotay, Jerrold Marsden, and Vincent Moncrief,** Mathematical aspects of classical field theory, 1992

131 **L. A. Bokut', Yu. L. Ershov, and A. I. Kostrikin, Editors,** Proceedings of the International Conference on Algebra Dedicated to the Memory of A. I. Mal'cev, Part 1, 2, and 3, 1992

130 **L. Fuchs, K. R. Goodearl, J. T. Stafford, and C. Vinsonhaler, Editors,** Abelian groups and noncommutative rings, 1992

129 **John R. Graef and Jack K. Hale,** Oscillation and dynamics in delay equations, 1992

128 **Ridgley Lange and Shengwang Wang,** New approaches in spectral decomposition, 1992

127 **Vladimir Oliker and Andrejs Treibergs, Editors,** Geometry and nonlinear partial differential equations, 1992

126 **R. Keith Dennis, Claudio Pedrini, and Michael R. Stein, Editors,** Algebraic K-theory, commutative algebra, and algebraic geometry, 1992

125 **F. Thomas Bruss, Thomas S. Ferguson, and Stephen M. Samuels, Editors,** Strategies for sequential search and selection in real time, 1992

124 **Darrell Haile and James Osterburg, Editors,** Azumaya algebras, actions, and modules, 1992

123 **Steven L. Kleiman and Anders Thorup, Editors,** Enumerative algebraic geometry, 1991

122 **D. H. Sattinger, C. A. Tracy, and S. Venakides, Editors,** Inverse scattering and applications, 1991

121 **Alex J. Feingold, Igor B. Frenkel, and John F. X. Ries,** Spinor construction of vertex operator algebras, triality, and $E_8^{(1)}$, 1991

120 **Robert S. Doran, Editor,** Selfadjoint and nonselfadjoint operator algebras and operator theory, 1991

119 **Robert A. Melter, Azriel Rosenfeld, and Prabir Bhattacharya, Editors,** Vision geometry, 1991

118 **Yan Shi-Jian, Wang Jiagang, and Yang Chung-chun, Editors,** Probability theory and its applications in China, 1991

117 **Morton Brown, Editor,** Continuum theory and dynamical systems, 1991

116 **Brian Harbourne and Robert Speiser, Editors,** Algebraic geometry: Sundance 1988, 1991

115 **Nancy Flournoy and Robert K. Tsutakawa, Editors,** Statistical multiple integration, 1991

114 **Jeffrey C. Lagarias and Michael J. Todd, Editors,** Mathematical developments arising from linear programming, 1990

113 **Eric Grinberg and Eric Todd Quinto, Editors,** Integral geometry and tomography, 1990

112 **Philip J. Brown and Wayne A. Fuller, Editors,** Statistical analysis of measurement error models and applications, 1990

111 **Earl S. Kramer and Spyros S. Magliveras, Editors,** Finite geometries and combinatorial designs, 1990

110 **Georgia Benkart and J. Marshall Osborn, Editors,** Lie algebras and related topics, 1990

109 **Benjamin Fine, Anthony Gaglione, and Francis C. Y. Tang, Editors,** Combinatorial group theory, 1990

108 **Melvyn S. Berger, Editor,** Mathematics of nonlinear science, 1990

107 **Mario Milman and Tomas Schonbek, Editors,** Harmonic analysis and partial differential equations, 1990

106 **Wilfried Sieg, Editor,** Logic and computation, 1990

105 **Jerome Kaminker, Editor,** Geometric and topological invariants of elliptic operators, 1990

(*Continued in the back of this publication*)

CONTEMPORARY MATHEMATICS

133

p-Adic Methods in Number Theory and Algebraic Geometry

Alan Adolphson
Steven Sperber
Marvin Tretkoff
Editors

American Mathematical Society
Providence, Rhode Island

1991 *Mathematics Subject Classification*. Primary 11-06, 14-06.

Library of Congress Cataloging-in-Publication Data

p-adic methods in number theory and algebraic geometry/Alan Adolphson, Steven Sperber, Marvin Tretkoff, editors.

p. cm.—(Contemporary mathematics, ISSN 0271-4132; v. 133)

Includes bibliographical references.

ISBN 0-8218-5145-4 (alk. paper)

1. Number theory. 2. Geometry Algebraic. 3. p-adic analysis. I. Adolphson, Alan, 1951-. II. Sperber, Steven, 1945-. III. Tretkoff, Marvin, 1943-. IV. Series: Contemporary mathematics (American Mathematical Society); v. 133.

QA241.P24 1992 — 92-20147

512′.7—dc20 — CIP

This volume was printed directly from author-prepared copy.
Portions of the volume were typeset by the authors using AMS-TEX, the American Mathematical Society's TEX macro system.

10 9 8 7 6 5 4 3 2 1 97 96 95 94 93 92

Contents

Preface

In the fall of 1989, there were meetings of the American Mathematical Society held at Stevens Institute of Technology and Ball State University. Among the Special Sessions held at these meetings, two focussed on the role of p-adic methods in number theory and algebraic geometry. These Special Sessions are the starting point for this volume. The articles presented here are drawn from a wide area of mathematics; despite that, some aspects of current research in p-adic arithmetic do not appear here at all. There is no attempt here at comprehensiveness. Our goal is to give a sampling of the wide range of concerns and applications of the p-adic approach.

Alan Adolphson
Steven Sperber
Marvin Tretkoff

This page intentionally left blank

Contemporary Mathematics
Volume 133, 1992

On Christol's theorem. A generalization to systems of PDE's with logarithmic singularities depending upon parameters

Francesco Baldassarri and Bruno Chiarellotto

§0 Introduction

This article is dedicated to a better understanding of the beautiful theorem of Christol proved in [CH1] (cf. the generalization in [A2, Résultat 1]) and to its extension (cf. theorem 4.9 below) to first-order systems of relative PDE's defined on a tubular neighborhood of an affinoid space, with logarithmic singularities at the center. This extension is necessary in view of our program of presenting a proof of a general comparison theorem between the algebraic and the rigid cohomologies of a smooth scheme of characteristic $p > 0$, with suitable coefficients. These coefficients should be algebraically defined overconvergent isocrystals, with nice logarithmic singularities. For more details, we refer the reader to [BE] and [BA-CT3].

Christol's theorem deals with first-order systems of ordinary differential equations of the type

$$(\delta - G)y = 0 \tag{0.1}$$

where $\delta = x\frac{d}{dx}$ and G is an $n \times n$ matrix of functions analytic in the p-adic open unit disk $D(0, 1^-)$. He, as well as subsequent authors (André, Christol-Dwork), writes a formal matrix solution of (0.1) at the origin as $Y = Y_G x^{C(G)}$ where Y_G is an $n \times n$ matrix with formally meromorphic entries and $C(G)$ is a constant matrix whose eigenvalues may be assumed to be among those of $G(0)$. He then gives conditions for the actual p-adic meromorphy of Y_G in the disk $D(0, 1^-)$. The matrix Y_G is well-known to be itself a formally meromorphic solution of a system of the type (0.1) (of order n^2), so Christol's theorem really gives conditions for a formally meromorphic solution of such a system to be p-adically meromorphic in the disk $D(0, 1^-)$. This viewpoint turns out to be more easily generalized to the case of several variables (also with dependence upon parameters) and in fact to give a slightly stronger result even in the ordinary case. This is why we need to review Christol's proof. The proof we present here seems in fact to be

---

This paper is in final form and no version of it will be submitted for publication elsewhere.

1991 Mathematics Subject Classification. Primary 12H25.

0271-4132/92 $1.00 + $.25 per page

simpler than the original one, even in the ordinary case, partly because of the obviously simpler statement we use and partly because we take advantage of new arguments of Christol and Dwork [CH-DW]. In that paper, those authors prove, among other results, only a special case of [CH1], but assert that their arguments could be used to simplify the proof of the full result of [CH1], as we actually do.

In the part of the present article that deals with ordinary differential equations, we follow step by step the proofs in [CH-DW] indicating the necessary modifications. Our theorem 1.6 provides a small improvement on André's Résultat 1 [A2].

A formulation of a theorem for systems of PDE's precisely in the form given by Christol [CH1] in the one dimensional case, naturally requires a detailed study of the formal theory for such systems. This theory, developed by Gérard and Levelt in [G-L] in the absolute case, was extended to systems depending upon parameters in [BA2], following a suggestion of N. Katz. We feel however that that theory deserves a more extensive treatment than the one it previously received. We refer to the forthcoming [BA-CT2] for such a treatment, that will include as an application the expected formulation of the generalized Christol theorem.

We replaced the notion of "*p-adic non-Liouville*" numbers (see (1.2) below) by the weaker notion of number "*of positive type* ρ", for $\rho \in (0,1]$ (loc. cit.). This also represents an improvement on previous notions introduced by André [A2]. We are grateful to Professor Dwork for suggesting this definition and for the subsequent gain in generality and clarity of exposition.

We hope that our reconsideration of this important work of Christol, Dwork and André will be useful to the reader.

The methods of the present article may also be applied to free André's transfer theorem ([A1, V.4.2], [A2, Résultat 2] and [CH-DW, sect.6,7,8]) from some restrictions on the exponents. For clarity of exposition, we decided to postpone the exposition of the generalized André's theorem, to a subsequent article [BA-CT1]. Our most immediate geometric needs ([BA-CT2],[BA-CT3]) are filled by the present discussion, but André's theorem has important number theoretic and geometric applications. In particular, it should in the end allow an extension of our theorem of comparison of cohomology [BA-CT3] to a more general type of coefficients. Our interest in the case of irrational exponents is instead mainly motivated by variation of p-adic cohomology and the so-called *Boyarski Principle* [DW2].

§1 The one-dimensional result.

1.1 We recall the notation of [CH-DW].
p= a prime number;

k= an algebraically closed field of characteristic zero, complete under a non-archimedean absolute value

$$| \ | : k \longrightarrow \mathbf{R}_{\geq 0}$$

extending the p-adic absolute value of $\mathbf{Q}_p$ and normalized by $|p| = p^{-1}$;

$\mathcal{V}$= ring of integers of k;

$D(0, \rho^-)=\{x \in k \mid |x| < \rho\}$, the *open* disk of radius $\rho \geq 0$ centered at 0;

π= a fixed $(p-1)$-st root of $-p$ in k, $\pi^{p-1} = -p$;

E= the completion of $k(x)$ under the Gauss norm which we will still indicate by $| - |$;

$E_\circ$= the ring of analytic elements on $D(0, 1^-)$. The elements of $E_\circ$ are uniform limits on $D(0, 1^-)$ of sequences of rational functions without poles on $D(0, 1^-)$; therefore $E_\circ \subseteq E$ and the sup-norm on $D(0, 1^-)$ coincides, for elements of $E_\circ$, with the Gauss norm $| - |$.

$E'_\circ$= the quotient field of $E_\circ$.

Finite dimensional spaces over E with a natural basis are provided with a norm derived from $| - |$ in the standard way: we still denote such a norm by $| - |$. This applies to n-tuples and to matrices. In particular an element of $GL(n, E)$ is said to be *unimodular* if it and its inverse are bounded by 1.

1.2 For an element $\lambda \in k$, we define the *type* of λ as the radius of convergence $\rho(\lambda)$ of the series

$$g_\lambda(x) = \sum_{\substack{s=0 \\ \lambda - s \neq 0}}^{\infty} \frac{x^s}{\lambda - s}. \tag{1.2.1}$$

We therefore have:

$$\rho(\lambda) = \liminf_s |\lambda - s|^{\frac{1}{s}} \tag{1.2.2}$$

and $\rho(\lambda) \in [0, 1]$. The type of a number is clearly invariant under translations by integers.

Let now C be a square matrix with entries in k, $C \in \mathcal{M}_n(k)$, and $\prod_\lambda (x-\lambda)^{e_\lambda}$ be its characteristic polynomial $det(x - C)$. We define the *type* $\rho(C)$ of C as the product

$$\rho(C) = \prod_\lambda \rho(\lambda)^{e_\lambda}. \tag{1.2.3}$$

The type of a matrix therefore only depends upon the types of its eigenvalues; in particular two matrices having the same number (counting multiplicities) of eigenvalues in each class of $k/\mathbf{Z}$, have the same type. For any $G \in \mathcal{M}_n(k[\![x]\!])$ we

will also say that $\rho(G(0))$ is the *type* of the system (0.1) at 0 and, more abusively, of G.

We recall (cf. [CH1], [A2]) that an element $\lambda \in k$ is said to be *p-adically Liouville (in the weak sense)* if there exists an $r \in \mathbf{R}_{>0}$ such that for an infinite set of $n \in \mathbf{Z}$ one has

$$|\lambda - n| \leq \inf(r^n, r^{-n}). \tag{1.2.4}$$

Therefore a number λ is p-adically non-Liouville iff both λ and $-\lambda$ are of type 1. It is clear that p-adically Liouville numbers are in $\mathbf{Z}_p$. It is known by the p-adic Roth theorem [LA, VI.1, Theorem1] that if λ is an algebraic number the series $g_\lambda(x)$ has logarithmic growth δ in $D(0,1^-)$ [CH2, 2.3] for any $\delta > 2$. In particular, algebraic numbers are p-adically non-Liouville. In [A2], André considers numbers $\lambda \in k$ which " are not p-adically Liouville in the strong sense" and associates to such a λ a positive real number r. To us this means that λ and $-\lambda$ have positive type greater or equal to r. We now have:

Lemma 1.3. *If $\lambda \in \mathbf{Z}_p$ is a number of type ρ then an element $\lambda' \in \mathbf{Z}_p$ such that $p\lambda' - \lambda = n \in \mathbf{Z}$, is of type ρ^p.*

Proof. We can assume that $n = 0$ and that $\lambda \notin \mathbf{N}$. The series

$$\sum_{s \in \mathbf{N}} \frac{x^s}{\lambda - s}$$

converges in $D(0, \rho^-)$. But:

$$\sum_{s \in \mathbf{N}} \frac{x^s}{\lambda - s} = \sum_{s \in \mathbf{N}} \frac{x^s}{p\lambda' - s} = \sum_{\substack{s \not\equiv 0 \\ mod.p}} \frac{x^s}{p\lambda' - s} + \sum_{\substack{s \equiv 0 \\ mod.p}} \frac{x^s}{p\lambda' - s}.$$

But, by hypothesis, $\lambda' \in \mathbf{Z}_p$, so the series

$$\sum_{\substack{s \not\equiv 0 \\ mod.p}} \frac{x^s}{p\lambda' - s}$$

converges in $D(0,1^-)$. The second series (in the last term of the previous equation) may be re-written as:

$$\sum_{r \in \mathbf{N}} \frac{x^{pr}}{p(\lambda' - r)}$$

and it converges for $x \in D(0, \rho^-)$. It follows that λ' is of type ρ^p. But the argument could be reversed. **Q.E.D.**

1.4 For any differential ring extension (R,δ) of $(k[x],\delta)$, for $G \in \mathcal{M}_n(R)$ and $H \in GL(n,R)$ we define

$$G_{[H]} = (\delta H)H^{-1} + HGH^{-1}. \tag{1.4.1}$$

Therefore the substitution $z = Hy$ transforms the system $\delta y = Gy$ into $\delta z = G_{[H]}z$. Under these circumstances we will say that the two previous systems are *equivalent over* R. For $u \in E$ we define by u^ϕ to be the composition $\dot{x} \to u(x^p)$.

Let $(\Omega, |-|)$ be an algebraically closed and complete valued extension field of $(k, |-|)$, containing an element t of absolute value 1 whose residue class is trascendental over the residue field of k. We call t a *generic point* (for k). For $\rho \geq 0$ we will denote by $D(t,\rho^-)$ (resp. $D(t,\rho)$) the *open* (resp. *closed*) disk of radius ρ centered at t, in Ω:

$$D(t,\rho^-) = \{x \in \Omega \mid |x-t| < \rho\}$$

(resp.

$$D(t,\rho) = \{x \in \Omega \mid |x-t| \leq \rho\}).$$

Let $\mathcal{S}$ be the set of all $G \in \mathcal{M}_n(E)$ satisfying the following two conditions:

$\mathcal{S}1$. $G \in \mathcal{M}_n(E_o)$.

$\mathcal{S}2$. The system $(\delta - G)y = 0$ has a solution matrix $\mathcal{U}_{G,t}$ at t (normalized by $\mathcal{U}_{G,t}(t) = I_n$) which converges in $D(t,1^-) = \{x \in \Omega \mid |x-t| < 1\}$.

We define, for any differential extension ring (R,δ) of $(k[x],\delta)$ and $G \in \mathcal{M}_n(R)$, $G_0 = I_n$, $G_1 = G$ and for $m \in \mathbf{N}$:

$$G_{m+1} = \delta G_m + G_m G - mG_m. \tag{1.4.2}$$

Condition $\mathcal{S}2$ is then equivalent to the following

$\mathcal{S}'2$. For any $\eta \in (0,1)$:

$$\lim_{m \to +\infty} |\frac{G_m}{m!}|\eta^m = 0. \tag{1.4.3}$$

This is clear since

$$\mathcal{U}_{G,t}(x) = \sum_{m=0}^{\infty} \frac{G_m(t)}{m!}(\frac{x-t}{t})^m. \tag{1.4.4}$$

On the other hand condition $\mathcal{S}2$ is known to imply [CH2, Proposition 6.2.9] that the eigenvalues of $G(0)$ are in $\mathbf{Z}_p$.

We consider also the conditions:

$\mathcal{S}3$. $|G| \leq 1$.

$\mathcal{S}4$. If an eigenvalue of $G(0)$ is an integer then it is zero.

If condition $\mathcal{S}1$ is satisfied and the difference between two eigenvalues of $G(0)$ is never a non-zero integer, the equation $(\delta - G)y = 0$ has a formal solution matrix at 0 which may be written in the form $Y_G x^{G(0)}$, where

$$Y_G \in GL(n, k[[x]]) \ , \ Y_G(0) = I_n. \tag{1.4.5}$$

This is best understood within the framework of Gérard-Levelt's theory [G-L], according to which, under the previous assumptions, the systems $(\delta - G)y = 0$ and $(\delta - G(0))y = 0$ are equivalent over $k[[x]]$. In the general case, under the assumption $\mathcal{S}1$ we can only assume the existence of a formal solution matrix at 0 in the form: $Y_G x^{C(G)}$, where $Y_G \in GL(n, k((x)))$ and $C(G) \in \mathcal{M}_n(k)$. One could also assume that the eigenvalues of $C(G)$ lie among those of $G(0)$ and that two of them never differ by a non-zero integer ([G-L]).

Remark 1.5 We already recalled that the matrix Y_G viewed as a column vector of length n^2 satisfies a system of differential equations:

$$(\delta - H)Y_G = 0, \tag{1.5.1}$$

where $H \in \mathcal{M}_{n^2}(E_o)$ if G belongs to $\mathcal{M}_n(E_o)$. The eigenvalues of $H(0)$ are differences of two eigenvalues of $G(0)$: this explains the conditions (namely, the differences of exponents should be non-Liouville numbers, hence numbers of type 1) imposed by Christol [CH1] in order to estimate the radius of convergence of Y_G. Using the formulation $\mathcal{S}2$, one can show that the system $(\delta - C(G))y = 0$ satisfies $\mathcal{S}2$ if $(\delta - G)y = 0$ did, since $C(G) = G_{[N]}(0)$ for a matrix $N \in \mathcal{M}(k[x, \frac{1}{x}])$. We conclude that if the system $(\delta - G)y = 0$ has property $\mathcal{S}2$ then $(\delta - H)y = 0$ also has it : in fact the two systems, $(\delta - C(G))y = 0$ and $(\delta - G)y = 0$, are equivalent to the constant system $\delta y = 0$ over the ring of functions analytic in the generic disk of radius 1.

We will prove:

Theorem 1.6. *Let $G \in \mathcal{S}$ be of type $\rho(G)$. Let $y \in k((x))^n$ be a column solution of the system*

$$(\delta - G)Y = 0. \tag{1.6.1}$$

Then the entries of y are meromorphic in $D(0, \rho(G)^-)$ (with possible poles only at $x = 0$).

Corollary 1.6.2. *Let $G \in \mathcal{S}$ be such that the exponents of $G(0)$ are p-adically non-Liouville. Let $y \in k((x))^n$ be as in the theorem. Then the entries of y are meromorphic in $D(0, 1^-)$ (with possible poles only at $x = 0$) .*

Notice that our theorem 1.6 and, in particular, corollary 1.6.2 are stronger than the main result of [CH1], while their proof will be easier.

Another immediate corollary is the following result. Let $\mathcal{A}_r$ (resp. $\mathcal{B}_r$), for $r \in (0,1]$, denote the ring of analytic (resp. bounded analytic) functions on $D(0,r^-)$:

$$\mathcal{A}_r = \{\sum_{i=0}^{\infty} a_i x^i \in k[\![x]\!] \mid \lim_{i\to\infty} |a_i|\rho^i = 0, \forall \rho \in (0,r)\}.$$

$$\mathcal{B}_r = \{\sum_{i=0}^{\infty} a_i x^i \in k[\![x]\!] \mid \sup_{i\in\mathbf{N}} |a_i| r^i < \infty\}.$$

The sup-norm on $D(0,r^-)$ for an element $f(x) = \sum_{i=0}^{\infty} a_i x^i$ of $\mathcal{B}_r$, will be denoted by

$$|f(x)|(r) = \sup_{i\in\mathbf{N}} |a_i| r^i.$$

We will also denote by $|-|(r)$ the norm (resp.semi-norm) derived in the standard way on matrices with entries in $\mathcal{B}_r$ (resp. $\mathcal{A}_s$, for any $s \in (r,1]$). Let $\mathcal{S}_{\mathcal{B}_r}$ (resp.$\mathcal{S}_{\mathcal{A}_r}$) denote the set of matrices $G \in \mathcal{M}_n(\mathcal{B}_r)$ (resp. $G \in \mathcal{M}_n(\mathcal{A}_r)$) such that:

$\mathcal{S}_{\mathcal{B}_r}$ For every $\eta \in (0,1)$

$$\lim_{m\to+\infty} |\frac{G_m}{m!}|(r)\eta^m = 0.$$

(resp.

$\mathcal{S}_{\mathcal{A}_r}$ For every $\eta \in (0,1)$ and $\sigma \in (0,r)$

$$\lim_{m\to+\infty} |\frac{G_m}{m!}|(\sigma)\eta^m = 0. \quad)$$

Corollary 1.7. *Let $G \in \mathcal{S}_{\mathcal{B}_r}$ (resp. $G \in \mathcal{S}_{\mathcal{A}_r}$) be of type ρ and $y \in k((x))^n$ be a column solution of the system (1.6.1). Then the entries of y are meromorphic in $D(0,(r\rho)^-)$ (with possible pole only at $x = 0$).*

Proof (of the corollary, granting the theorem). Let $\alpha \in k^*$ be such that $|\alpha| < r$ and let $H(x) = G(\alpha x)$. Then $H_m(x) = G_m(\alpha x)$, so that $H(x) \in \mathcal{S}$. So, if $y(x) \in k((x))^n$ is a vector solution of (1.6.1), it follows that $y(\alpha x)$ is a vector solution of $(\delta - H)y = 0$ and, by theorem 1.6, it is meromorphic in $D(0,\rho^-)$. Therefore $y(x)$ is meromorphic in $D(0,(r\rho)^-)$. **Q.E.D.**

Remark 1.7.1. The previous condition appeared in [CH1] as the condition of convergence of the solution matrix in the open generic disk of radius r.

We will denote by $\overline{\mathcal{S}} \subset \mathcal{S}$ the set of $G \in \mathcal{S}$ satisfying also conditions $\mathcal{S}3$ and $\mathcal{S}4$. We have:

Lemma 1.8. *Let F be an element of $\mathcal{S}$. There exists a matrix $Q \in GL(n, E_o[\frac{1}{x}])$ such that $F_{[Q]} \in \overline{\mathcal{S}}$ and $\rho(F_{[Q]}) = \rho(F)$.*

Proof. We follow the procedure of section 4 of [CH-DW], up to line -6 of that section (condition $\mathcal{R}i$ of [loc.cit.] coincides with our condition $\mathcal{S}i$ for $i = 1, 2$, while $\mathcal{S}3 = \mathcal{R}4$). Thus, in the notation of [loc.cit.], $F_{[H_1 H]}$ also satisfies $\mathcal{S}1$, $\mathcal{S}2$, $\mathcal{S}3$. We recall that $H \in GL(n, E_o')$ and H_1 is a unimodular matrix in $GL(n, k(x))$. By Lemma 1.10 below, the eigenvalues of $F_{[H_1 H]}(0)$ differ by integers from those of $F(0)$. It follows that $F_{[H_1 H]}$ is also of type $\rho(F)$. By (2.3) of [loc.cit.] there exists $H_2 \in GL(n, k[x, x^{-1}])$ unimodular such that $(F_{[H_1 H]})_{[H_2]} = F_{[H_2 H_1 H]}$ satisfies $\mathcal{S}4$. Conditions $\mathcal{S}1$, $\mathcal{S}2$, $\mathcal{S}3$ are automatically valid for $F_{[H_2 H_1 H]}$, also of type $\rho(F)$. We then put $Q = H_2 H_1 H \in GL(n, E_o')$ and conclude that $G = F_{[Q]} \in \overline{\mathcal{S}}$. We want to show that $Q \in GL(n, E_o[\frac{1}{x}])$. This follows from the fact that the systems $(\delta - G)Y = 0$ and $(\delta - F)Y = 0$ have no singularities in $D(0, 1^-) \setminus \{0\}$. **Q.E.D.**

Our theorem 1.6 is therefore equivalent to the statement:

Theorem 1.9. *Let $G \in \overline{\mathcal{S}}$ be of type ρ and let $y \in k[[x]]^n$ be a column solution of the system (1.6.1). Then the entries of y are analytic in $D(0, \rho^-)$.*

We recall here the elementary :

Lemma 1.10. *Let $G, H \in \mathcal{M}_n(k[[x]])$ and $P \in GL(n, k((x)))$ such that $H = G_{[P]}$. Then for any coset $\alpha + \mathbf{Z} \in k/\mathbf{Z}$, the matrices $G(0)$ and $H(0)$ have the same number, counting multiplicities, of eigenvalues in $\alpha + \mathbf{Z}$.*

Proof. We write $P = A \Delta B$, with $A, B \in GL(n, k[[x]])$ and Δ a diagonal matrix with elements of the form x^r, $r \in \mathbf{Z}$. We then have:

$$H = G_{[A\Delta B]} = ((G_{[B]})_{[\Delta]})_{[A]}$$

and we are reduced to the case $P = \Delta$ and $H = G_{[\Delta]}$. We can subdivide our matrices H, G, Δ into rectangular blocks $H = (H_{i,j})$, $G = (G_{i,j})$, $\Delta = (\Delta_{i,j})$ for $i, j = 1, \ldots, s$, such that $\Delta_{i,j} = 0$ if $i \neq j$, $\Delta_{i,i} = r_i I_{e_i}$ and $r_i < r_j$ for $i < j$. We then have:

$$H = (\delta\Delta)\Delta^{-1} + \Delta G \Delta^{-1}$$

$$H_{i,j}(x) = r_i \delta_{ij} + x^{r_i - r_j} G_{ij}(x) \quad i, j = 1, \ldots, s.$$

By evaluating at $x = 0$ we deduce that if $i > j$

$$H_{ij}(0) = 0$$

and if $i < j$

$$G_{ij}(0) = 0.$$

On the other hand, for $i = 1, \ldots, s$:

$$H_{ii}(0) = r_i + G_{ii}(0)$$

Let $P_i(\lambda) \in k[\lambda]$ be the characteristic polynomial of $G_{ii}(0)$. We conclude that the characteristic polynomials of $H(0)$ and $G(0)$ are, respectively, $\prod_{i=1}^{s} P_i(\lambda)$ and $\prod_{i=1}^{s} P_i(\lambda - r_i)$. **Q.E.D.**

§2 The trivial estimate.

For a matrix $C \in \mathcal{M}_n(k)$ and $s \in \mathbf{N}$, we define $(C)_0 = I_n$ and $(C)_s = C(C+I_n)(C+2I_n)\dots(C+(s-1)I_n)$. The following proposition replaces lemma 3.1 of [CH1].

Proposition 2.1. *Let $C \in \mathcal{M}_n(k)$ be a matrix none of whose eigenvalues is a positive integer. Then the series*

$$h_C(x) = \sum_{i=0}^{\infty} \frac{x^i}{det(I_n - C)_i} = \sum_{i=0}^{\infty} \frac{(-1)^{ni} x^i}{det(C - I_n)(C - 2I_n)\dots(C - iI_n)}$$

converges in $D(0, (|\pi|^n \rho(C))^-)$.

Proof. We first notice that the series $h_C(x)$ is the Hadamard product of the series $h_\lambda(x)$ each taken with the multiplicity of λ as eigenvalue of C. We may then assume $n = 1$. By the formal Kummer identity [WA, 4.42]

$${}_1F_1(1; 1-\lambda; x) = e^x \, {}_1F_1(-\lambda; 1-\lambda; -x)$$

one obtains for $\lambda \neq 0, 1, 2\dots$:

(2.1.1)

$$\sum_{i=0}^{\infty} \frac{x^i}{(1-\lambda)_i} = e^x \sum_{i=0}^{\infty} \frac{(-x)^i (-\lambda)_i}{(1-\lambda)_i i!} = e^x \sum_{i=0}^{\infty} \frac{(-x)^i (-\lambda)}{(-\lambda+i) i!} = e^x \lambda \sum_{i=0}^{\infty} \frac{(-x)^i}{(\lambda - i) i!},$$

while for $\lambda = 0$, $h_0(x) = e^x$. The last series in (2.1.1) is the Hadamard product of $e^{(-x)}$ and $g_\lambda(-x)$: its radius of convergence is therefore the product of the radii of convergence of those two series, namely $|\pi|$ and $\rho(\lambda)$. Since e^x converges in $D(0, |\pi|^-)$, we conclude that $\sum_{i=0}^{\infty} \frac{x^i}{(1-\lambda)_i}$ converges in $D(0, (|\pi|\rho(\lambda))^-)$. **Q.E.D.**

We have:

Proposition 2.2. *Let $G \in \mathcal{M}_n(E_o)$ be of type ρ and satisfy $\mathcal{S}4$. Let $y \in k[[x]]^n$ be a column solution of (1.6.1). Then the entries of y are analytic in $D(0, (\max(1, |G|)^{-n} |\pi|^n \rho)^-)$. In particular, let $G \in \overline{\mathcal{S}}$ be of type ρ and $y \in k[[x]]^n$ be a column solution of (1.6.1). Then the entries of y are analytic in $D(0, (|\pi|^n \rho)^-)$.*

Proof. We write

$$G = \sum_{j=0}^{+\infty} G^{(j)} x^j,$$

$G^{(j)} \in \mathcal{M}_n(k)$ and $y(x) = \sum_{i=0}^{\infty} y_i x^i$, with $y_i \in k^n$. From $(\delta - G)y(x) = 0$ we obtain

$$G(0)y_0 = 0 \tag{2.2.1}$$

and

$$(G(0) - i)y_i = -\sum_{j=0}^{i-1} G^{(i-j)}y_j, \quad for\ i \geq 1. \tag{2.2.2}$$

We assume with no restriction that $|y_0| = 1$, and we deduce :

$$|y_i| \leq \max(1, |G|)^i |[(G(0) - I_n)(G(0) - 2I_n) \dots (G(0) - iI_n)]^{-1}| \leq$$

(2.2.3)

$$\leq \frac{\max(1, |G|)^i (\max(1, |G(0)|))^{(n-1)i}}{|det(I_n - G(0))_i|},$$

hence

$$|y_i| \leq \frac{1}{|det(I_n - G(0))_i|} \max(1, |G|)^{ni} \tag{2.2.4}$$

(by assumption $\mathcal{S}4$, $G(0) - j$ is invertible whenever $j \in \mathbf{N}$, $j \neq 0$). We conclude that y is dominated by the Hadamard product of the series $h_{G(0)}(x)$ and a series of radius of convergence $\max(1, |G|)^{-n}$. **Q.E.D.**

Corollary 2.3. *Let $G \in \mathcal{M}_n(E_o)$ be of type ρ. Let $y \in k((x))^n$ be a column solution of (1.6.1). Then the entries of y are meromorphic in*

$$D(0, (\max(1, |G|)^{-n} |\pi|^n \rho)^-)$$

with poles at most at 0.

Proof. We need only to apply some shearing transformations (cf. [CH-DW, prop. 2.3]) to reduce to the situation of the previous statement ($G \in \mathcal{S}4$).**Q.E.D.**

§3 Frobenius action.

We need some lemmas:

Lemma 3.0. *Let $G \in \overline{\mathcal{S}}$ be of type ρ. There exists $P \in GL(n, k[x, x^{-1}])$ unimodular, such that $G_{[P]} \in \overline{\mathcal{S}}$ is of type ρ and the eigenvalues of $G_{[P]}(0)$ are in $p\mathbf{Z}_p$.*

Proof. We apply the "shearing transformation" of (2.3) of [CH-DW]. **Q.E.D.**

Lemma 3.1. *Let $r \geq 0$.*

(i) The map $x \to x^p$ sends the closed disk $D(t,r)$ into $D(t^p, max(r^p, p^{-1}r))$. In particular, if $r \geq |\pi|$, $x \to x^p$ sends $D(t,r)$ into $D(t^p, r^p)$.

(ii) The image of $D(t,r)$ under the map $x \to x^p$ contains $D(t^p, r^p)$ (without restriction on r).

(iii) For $r \geq |\pi|$, the map $x \to x^p$ induces a surjection of $D(t,r)$ onto $D(t^p, r^p)$ each fiber of which has exactly p elements.

(We omit the easy proof).

The next lemma fills a gap in the proof of the theorem in section 3 (resp. of Lemma 8.2) of [CH-DW], where it is asserted that pF lies in $\mathcal{M}_n(E_\circ)$ (resp. in $\mathcal{M}_n(E'_\circ)$).

Lemma 3.2. *Let $r \geq |\pi|$ and f be meromorphic in the closed disk $D(t^p, r^p)$. (So, by the previous Lemma, $f(x^p)$ is meromorphic on $D(t,r)$). Suppose $f(x^p)$ coincides on $D(t,r)$ with $g \in E_\circ$ (resp.$E'_\circ$) . Then f extends to an element of $E_\circ$ (resp.$E'_\circ$).*

Proof. We may restrict ourselves to the case $g \in E_\circ$. Let $h(x) = \frac{1}{p}\sum_{z^p = x} g(z)$. If $g \in E_\circ \cap k(x)$, one sees as in [DW1, 3.5] that also $h \in E_\circ \cap k(x)$; it then follows from [DW1, 5.2] that for any $g \in E_\circ$, $h \in E_\circ$. We assert that $h = f$ on $D(t^p, r^p)$. For $x \in D(t,r)$, $h(x^p) = \frac{1}{p}\sum_{z^p = x^p} g(z) = \frac{1}{p}\sum_{\zeta^p = 1} g(x\zeta)$. Now $x\zeta \in D(t,r)$, hence $g(x\zeta) = f((x\zeta)^p) = f(x^p)$, i.e. $h(x^p) = f(x^p)$ for all $x \in D(t,r)$. Hence we have $h = f$ on the image of $D(t,r)$ under the p^{th} power map and this image is $D(t^p, r^p)$ by the previous lemma. **Q.E.D.**

Theorem 3.3. *Let $G \in \overline{\mathcal{S}}$ be such that the eigenvalues of $G(0)$ are in $p\mathbf{Z}_p$. Then there exist $H \in GL(n, E_\circ)$, H unimodular, and $F \in \mathcal{M}_n(E_\circ)$ satisfying $\mathcal{S}1$,$\mathcal{S}2$ such that*

$$pF^\phi = G_{[H]}. \tag{3.3.1}$$

We have $\rho(F) = \rho(G)^p$.

Proof. We follow the proof of the theorem in section 3 of [CH-DW] up to formula (3.2.1.1) included. We have:

$$pH(0) = \sum_{s=0}^{\infty} p\binom{G(0)}{s} z_s s!\,,$$

where z_s denotes a p-adic integer. Therefore:

$$pH(0) = pI_n + p\sum_{s=1}^{\infty} z_s G(0)(G(0)-1)\dots(G(0)-s+1).$$

We may now assume that $G(0)$ is a triangular matrix whose diagonal entries are in $p\mathbf{Z}_p$, to conclude that:

$$|detH(0)| = 1. \tag{3.3.2}$$

From this we deduce, as in [loc.cit.], that $H \in GL(n, E_\circ)$ and that it is unimodular. We still follow the proof in [loc.cit.], and, in particular, adopt definition (3.3) of $V(x)$ and (3.3.1). We conclude, as in [loc.cit.] but with the explanations given in the previous lemma, that F satisfies $\mathcal{S}1$ (=$\mathcal{R}1$ in the notations of [loc.cit]). We are not interested in equation (3.0.2) of [loc.cit.], now replaced by the obvious:

$$pF(0) = H(0)G(0)H(0)^{-1}. \tag{3.3.3}$$

This relation shows, with the help of Lemma 1.3, that $\rho(F) = \rho(G)^p$. Property $\mathcal{S}2$ is seen to hold for F as in [loc.cit.] We are not interested in the remaining part of the proof of [loc.cit.], which aims at the proof of relation (3.0.3) of that paper. **Q.E.D.**

Corollary 3.4. *Let $G \in \overline{\mathcal{S}}$. There exist $G_1 \in \overline{\mathcal{S}}$ and U in $GL(n, E_\circ[\frac{1}{x}])$, such that*

$$(pG_1^{\phi})_{[U]} = G \tag{3.4.1}$$

and such $\rho(G_1) = \rho(G)^p$.

Proof. By lemma 3.0, we find $P \in GL(n, E_\circ[\frac{1}{x}])$, unimodular, such that $G_{[P]}$ satisfies the conditions of applicability of the theorem (for G replaced by $G_{[P]}$ and $\rho(G_{[P]}) = \rho(G)$). If $F \in \mathcal{M}_n(E_\circ)$ satisfies $\mathcal{S}1$, $\mathcal{S}2$, is of type $\rho(F) = \rho(G)^p$ and

$$pF^{\phi} = (G_{[P]})_{[H]} = G_{[HP]}$$

for some $H \in GL(n, E_\circ)$, H unimodular, as in the theorem, we put $L = HP$. By lemma 1.8, there exists $L_1 \in GL(n, E_\circ[\frac{1}{x}])$ such that $F = (G_1)_{[L_1]}$ with $G_1 \in \overline{\mathcal{S}}$ with $\rho(G_1) = \rho(F)$. We then have

$$p(G_{1[L_1]})^{\phi} = G_{[L]} \tag{3.4.2}$$

and therefore

$$(pG_1^{\phi})_{[L_1^{\phi}]} = G_{[L]}. \tag{3.4.3}$$

We then put $U = L^{-1}L_1^{\phi}$. **Q.E.D.**

3.5 For $H \in \mathcal{M}_n(E_\circ)$ let $Sol(H, k[[x]])$ be the k-vector space of column solutions $y \in k[[x]]^n$ of $(\delta - H)y = 0$. We also define $\mathcal{R}(H) \in [0,1]$ to be the supremum of the set:

$$\{r \in [0,1] \mid \text{every } y(x) \in Sol(H, k[[x]]) \text{ is analytic in } D(0, r^-)\}. \tag{3.5.1}$$

We have seen in section 2 that if $H \in \overline{\mathcal{S}}$ is of type $\rho(H)$,

$$\mathcal{R}(H) \geq |\pi|^n \rho(H). \tag{3.5.2}$$

Lemma 3.6. *Let G, G_1, U be as in corollary 3.4 . The map*

$$y \longrightarrow Uy^{\phi} \tag{3.6.0}$$

induces an isomorphism of k-vector spaces

$$Sol(G_1, k[[x]]) \longrightarrow Sol(G, k[[x]]). \tag{3.6.1}$$

Proof. The map (3.6.0) is clearly an injection and is stable on $k((x))^n$. The fact that zero is the only integral exponent for both $\delta - G$ and $\delta - G_1$ shows that (3.6.0) restricts to a map (3.6.1), still an injection. On the other hand the matrices $G_1(0)$ and $pG_1^{\phi}(0) = pG_1(0)$ have the same nullity, which coincides with the dimension of the k-vector space $Sol(G_1, k[[x]])$ and also with the dimension of $Sol(pG_1^{\phi}, k[[x]]) = Sol(pG_1^{\phi}, k((x)))$ since $pG_1^{\phi}(0)$ has no non-zero integral eigenvalues (the eigenvalues of $G_1(0)$ are in $\mathbf{Z}_p$). Multiplication by U clearly gives an isomorphism between $Sol(pG_1^{\phi}, k((x)))$ and $Sol(G, k((x))) = Sol(G, k[[x]])$. **Q.E.D.**

Corollary 3.7. *In the notation of the lemma, we have*

$$\mathcal{R}(G) = \mathcal{R}(G_1)^{\frac{1}{p}}. \tag{3.7.1}$$

3.8 We can now prove theorem 1.9. We know from section 2 that $\mathcal{R}(G) \geq |\pi|^n \rho(G)$. Choose $G_1 \in \overline{\mathcal{S}}$ as in theorem 3.3 for G. We can repeat the process in order to obtain a sequences of matrices in $\overline{\mathcal{S}}$: $G_1, G_2, \ldots, G_l, \ldots$ (the matrix G_l is related to G_{l-1} as G_1 is to G in theorem 3.3). We have:

$$\mathcal{R}(G) = \mathcal{R}(G_l)^{\frac{1}{p^l}}.$$

and

$$\rho(G_l) = \rho(G)^{p^l}.$$

But from our weak estimate (§2) we obtain

$$\mathcal{R}(G_l) \geq |\pi|^n \rho(G_l) \geq |\pi|^n \rho(G)^{p^l},$$

we conclude that:

$$\mathcal{R}(G) \geq |\pi|^{\frac{n}{p^l}} \rho(G).$$

For $l \to \infty$, we conclude.**Q.E.D.**

§4 The case of several variables.

The present section deals with a generalization of the previous results to the case of several variables over an affinoid k-algebra A.

We need to introduce some more notation.

4.1 Let A be an integral affinoid k-algebra [BGR]. We denote by $|-|_A$ its supremum norm, which is a Banach k-algebra norm on A (it is actually equivalent to every complete k-algebra norm on A [BGR, 6.2.4 Theorem 1]).

By the Noether Normalization Lemma there exists a finite monomorphism of k-algebras [BGR, 6.1.2 Corollary 2] :

$$T_d \longrightarrow A$$

where T_d is the free Tate k-algebra in d indeterminates. It's well-known that T_d is integrally closed. The supremum norm on A, $|-|_A$, is induced by the spectral norm on the quotient field of A deduced from the Gauss norm (actually, an absolute value) on T_d [BGR, 3.8.1 Proposition 7]. Hence there exist a finite number of inequivalent absolute values on A [BGR 3.3.3 Proposition 1]:

$$|-|_1, \ldots, |-|_l \tag{4.1.1}$$

which extend the Gauss norm on T_d and such that for every $a \in A$:

$$|a|_A = \max_{j=1,\ldots,l} |a|_j. \tag{4.1.2}$$

We will denote by $K = \mathcal{Q}(A)$, the quotient field of A. For each $j = 1, \ldots, l$, let K_j denote an algebraically closed extension of K which is complete with respect to an absolute value extending the absolute value $|-|_j$: we will retain the same notation, $|-|_j$, for the extended absolute value on K_j.

4.2 Let $x_1, \ldots, x_s$ be indeterminates over A. For $\alpha \in \mathbf{N}^s$, $\alpha \in (\alpha_1, \ldots, \alpha_s)$ we define $|\alpha|_{\mathbf{R}}$ to be the standard euclidean norm of $\alpha \in \mathbf{R}^s$ and $x^\alpha = x_1^{\alpha_1} \ldots x_s^{\alpha_s}$. We will refer to the s-tuple $(0, \ldots, 1, \ldots 0)$ (1 in the position i) as δ_i. Let $r \in (0, 1]$. We indicate by $A\{\{x_1 \ldots, x_s\}\}_r = A\{\{x\}\}_r$ the ring of power series

$$\{ \sum_{\alpha \in \mathbf{N}^s} a_\alpha x^\alpha,\ a_\alpha \in A \mid \lim_{|\alpha|_{\mathbf{R}} \to +\infty} |a_\alpha|_A \eta^{|\alpha|_{\mathbf{R}}} = 0\ \forall \eta \in [0, r) \}. \tag{4.2.1}$$

We define a subring $A\{\{x\}\}_r^b$ of $A\{\{x\}\}_r$ by :

$$A\{\{x\}\}_r^b = \{ \sum_{\alpha \in \mathbf{N}^s} a_\alpha x^\alpha,\ a_\alpha \in A \mid \sup_{\alpha \in \mathbf{N}^s} |a_\alpha|_A r^{|\alpha|_\mathbf{R}} < +\infty \}. \tag{4.2.2}$$

This is a Banach k-algebra for the norm $|-|(r)$ [BGR 2.1.5], defined for $\sum_{\alpha \in \mathbf{N}^s} a_\alpha x^\alpha \in A\{\{x\}\}_r^b$, as

$$|\sum_{\alpha \in \mathbf{N}^s} a_\alpha x^\alpha|(r) = \sup_{\alpha \in \mathbf{N}^s} |a_\alpha|_A r^{|\alpha|_\mathbf{R}}.$$

The same symbol will denote the obvious extension of $|-|(r)$ to a norm on matrices with coefficients in $A\{\{x\}\}_r^b$ (resp. in $A\{\{x\}\}_c$, for any $c \in (r,1]$). We will also write:

$$A[[x]][\frac{1}{x}] = A[[x_1, \ldots, x_s]][\frac{1}{x_1 \ldots x_s}] \tag{4.2.3}$$

and

$$A\{\{x\}\}_r[\frac{1}{x}] = A\{\{x_1, \ldots, x_s\}\}_r[\frac{1}{x_1 \ldots x_s}]. \tag{4.2.4}$$

We note, in view of future application, that $A\{\{x\}\}_r$ may also be viewed as the ring of the analytic functions on the trivial bundle in open polydisks of radius r and dimension s over the affinoid base space SpA [BE], i.e. on :

$$SpA \times D(0, r^-)^s.$$

It follows that $A\{\{x\}\}_r^b$ represents the ring of bounded analytic functions on the previous fiber space and, finally, that $A\{\{x\}\}_r[\frac{1}{x}]$ is the ring of meromorphic functions on $SpA \times D(0,r^-)^s$ with poles only along the divisor $x_1 \ldots x_s = 0$.

4.3 Given a finite dimensional vector space V over a field L and m commuting L-linear endomorphisms of V, $(\varphi_1, \ldots, \varphi_m)$, we define a *multieigenvalue* of $(\varphi_1, \ldots, \varphi_m)$ as an m-tuple of elements in the algebraic closure $\overline{L}$ of L

$$(\lambda_1, \ldots, \lambda_m) \in \overline{L}^m$$

such that if $\overline{V} = V \otimes_L \overline{L}$ and $\overline{\varphi}_i$, $i = 1, \ldots, m$, denotes the obvious $\overline{L}$-linear extension of φ_i to $\overline{V}$, one has:

$$\bigcap_{i=1}^m Ker_{\overline{V}}(\overline{\varphi}_i - \lambda_i) \neq 0. \tag{4.3.1}$$

The dimension of the $\overline{L}$-vector space

$$\bigcap_{i=1}^m Ker_{\overline{V}}(\overline{\varphi}_i - \lambda_i)^{dim\, V} \tag{4.3.2}$$

will be called the *multiplicity* of the multieigenvalue $(\lambda_1, \ldots, \lambda_m)$.

In case for any multieigenvalue $(\lambda_1, \ldots, \lambda_m)$ of $(\varphi_1, \ldots, \varphi_m)$ the sum $\lambda_1 + \ldots + \lambda_m$ is in k, we will define the *type* of $(\varphi_1, \ldots, \varphi_m)$ to be radius of convergence of the series

$$g_{(\varphi_1, \ldots, \varphi_m)}(x), \tag{4.3.3}$$

Hadamard product of the series

$$g_{\lambda_1 + \ldots + \lambda_m}(x)$$

each taken with the multiplicity of $(\lambda_1, \ldots, \lambda_m)$ as multieigenvalue of $(\varphi_1, \ldots, \varphi_m)$.

4.4 Let $r \in (0, 1]$. A *standard system of PDE's on* $A\{\{x\}\}_r$ *relative to* A *with logarithmic singularities at* $x_1 \ldots x_s = 0$ is an integrable system of partial differential equations of the type:

$$x_i \frac{\partial}{\partial x_i} y = H_{\delta_i}(x) y \qquad i = 1, \ldots, s \tag{4.4.1}$$

where $H_{\delta_i}(x) \in \mathcal{M}_n(A\{\{x\}\}_r)$. The integrability condition means:

(4.4.2)

$$x_j \frac{\partial}{\partial x_j} H_{\delta_i}(x) + H_{\delta_i}(x) H_{\delta_j}(x) = x_i \frac{\partial}{\partial x_i} H_{\delta_j}(x) + H_{\delta_j}(x) H_{\delta_i}(x) \quad \text{for } i, j = 1, \ldots, s.$$

By evaluating the previous formulas at $x_1 = x_2 = \ldots = x_s = 0$, one gets:

$$H_{\delta_i}(0) H_{\delta_j}(0) = H_{\delta_j}(0) H_{\delta_i}(0), \quad \textit{for } i, j = 1, \ldots, s$$

where $H_{\delta_i}(0) \in \mathcal{M}_n(A)$. It follows that the system

$$x_i \frac{\partial}{\partial x_i} y = H_{\delta_i}(0) y \qquad i = 1, \ldots, s \tag{4.4.3}$$

with $H_{\delta_i}(0) \in \mathcal{M}(A)$, $i = 1, \ldots, s$, is still of the same form as (4.4.1). If Y is a solution matrix of (4.4.1) and, for $\alpha \in \mathbf{N}^s$ we put

$$\frac{\partial^\alpha}{\partial x^\alpha} = \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \cdot \ldots \cdot \frac{\partial^{\alpha_s}}{\partial x_s^{\alpha_s}},$$

we have

$$x^\alpha \frac{\partial^\alpha}{\partial x^\alpha} Y = H_\alpha(x) Y, \tag{4.4.4}$$

where $H_\alpha \in \mathcal{M}_n(A\{\{x\}\})$. The matrices $H_\alpha(x)$ may be constructed by recursive formulas. In fact for $\alpha \in \mathbf{N}^s$, $H_\alpha(x)$ as in (4.4.4), we have

$$H_{\alpha + \delta_i}(x) = x_i \frac{\partial}{\partial x_i}(H_\alpha(x)) + H_\alpha(x)(H_{\delta_i}(x) - \alpha_i). \tag{4.4.5}$$

One should note that the analogous recursion formulas for the derived system (4.4.3) give precisely the matrices of (4.4.5) evaluated at $x_1 = x_2 = \ldots = x_s = 0$: i.e., if Y is a solution of (4.4.3), then for $\alpha \in \mathbf{N}^s$, we have

$$x^\alpha \frac{\partial^\alpha}{\partial x^\alpha} Y = H_\alpha(0)Y.$$

We may express the previous remark symbolically by the equation:

$$H_\alpha(0) = H(0)_\alpha \tag{4.4.6}$$

for $\alpha \in \mathbf{N}^s$.

One deduces also from (4.4.5) that if $H_{\delta_i}(x) \in \mathcal{M}_n(A\{\{x\}\}_r^b)$ for $i = 1, \ldots, s$, then the matrices $H_\alpha \in \mathcal{M}_n(A\{\{x\}\}_r^b)$ for every $\alpha \in \mathbf{N}^s$.

4.5 We will indicate by ${}_{A,s}\mathcal{S}_{\mathcal{B}_r}$ (resp. ${}_{A,s}\mathcal{S}_{\mathcal{A}_r}$) the set of standard systems (as in (4.4.1))

$$x_i \frac{\partial}{\partial x_i} y = G_{\delta_i}(x)y \tag{4.5.1}$$

with $G_{\delta_i}(x) \in \mathcal{M}_n(A\{\{x\}\}_r^b)$ (resp.$G_{\delta_i}(x) \in \mathcal{M}_n(A\{\{x\}\}_r)$ for $i = 1, \ldots, s$, which satisfy the following condition :

${}_{A,s}\mathcal{S}_{\mathcal{B}_r}$ For each $\eta \in \mathbf{R}$, $0 < \eta < 1$, on setting $\alpha! = \alpha_1! \ldots \alpha_s!$, in the previous notation we get :

$$\lim_{|\alpha|_{\mathbf{R}} \to +\infty} |\frac{G_\alpha(x)}{\alpha!}|(r)\eta^{|\alpha|_{\mathbf{R}}} = 0. \tag{4.5.2}$$

(resp.

${}_{A,s}\mathcal{S}_{\mathcal{A}_r}$ For each $\eta, \sigma \in \mathbf{R}$, $0 < \eta < 1, 0 < \sigma < r$, in the previous notation we have :

$$\lim_{|\alpha|_{\mathbf{R}} \to +\infty} |\frac{G_\alpha(x)}{\alpha!}|(\sigma)\eta^{|\alpha|_{\mathbf{R}}} = 0. \quad) \tag{4.5.2)'}$$

Notice that

$${}_{A,s}\mathcal{S}_{\mathcal{A}_r} = \bigcap_{\sigma \in (0,r)} {}_{A,s}\mathcal{S}_{\mathcal{B}_\sigma} \tag{4.5.3}$$

Remark 4.5.4 The multieigenvalues $(\lambda_1, \ldots, \lambda_s)$ of $(G_{\delta_1}(0), \ldots, G_{\delta_s}(0))$, are usually referred to as the *multiexponents* of system (4.5.1) (*exponents* if $s = 1$) and their multiplicity as the *multiplicity* as multiexponents. (Cf. [BA2, 2.24], [BA-CT2]).

Remark 4.5.5 We will see later (5.2) that condition ${}_{A,s}\mathcal{S}_{\mathcal{B}_r}$ (resp. ${}_{A,s}\mathcal{S}_{\mathcal{A}_r}$) implies that any eigenvalue λ_i of $G_{\delta_i}(0)$ is in $\mathbf{Z}_p$ for $i = 1, \ldots, s$.

Definition 4.5.6 For a standard system in ${}_{A,s}\mathcal{S}_{\mathcal{A}_r}$, as (4.5.1), we define its *type* to be the type of $(G_{\delta_1}(0), \ldots, G_{\delta_s}(0))$ (cf. the previous remark)

4.6 For $A = k$, $s = 1$ we have the equalities

$$k\{\{x\}\}_r = \mathcal{A}_r$$

$$k\{\{x\}\}_r^b = \mathcal{B}_r$$

(cf. §1). Moreover, condition ${}_{k,1}\mathcal{S}_{\mathcal{B}_r}$ coincides with $\mathcal{S}_{\mathcal{B}_r}$, and condition ${}_{k,1}\mathcal{S}_{\mathcal{A}_r}$ coincides with $\mathcal{S}_{\mathcal{A}_r}$.

4.7 It should be noted that if a system (cf. (4.5.1))

$$x_i \frac{\partial}{\partial x_i} y = G_{\delta_i}(x) y \qquad i = 1, \ldots, s \tag{4.7.0}$$

is in ${}_{A,s}\mathcal{S}_{\mathcal{B}_r}$ (resp. ${}_{A,s}\mathcal{S}_{\mathcal{A}_r}$), then

$$x_i \frac{\partial}{\partial x_i} y = G_{\delta_i}(0) y \qquad i = 1, \ldots, s \tag{4.7.1}$$

is also in ${}_{A,s}\mathcal{S}_{\mathcal{B}_r}$ (resp. ${}_{A,s}\mathcal{S}_{\mathcal{A}_r}$).

4.8 Finally, by use of the previous notation, we can state our generalization of Christol's theorem:

Theorem 4.9. *Let*

$$y(x) \in (A[[x]][\frac{1}{x}])^n$$

be a formal column solution of a standard system (4.7.0) in ${}_{A,s}\mathcal{S}_{\mathcal{A}_r}$. *Let* ρ *be the type of (4.7.0). Then* $y(x) \in (A\{\{x\}\}_{r\rho}[\frac{1}{x}])^n$.

Corollary 4.9.1. *Let*

$$y(x) \in (A[[x]][\frac{1}{x}])^n$$

be a formal column solution of a standard system (4.7.0), in ${}_{A,s}\mathcal{S}_{\mathcal{A}_1}$. *Suppose for each multiexponent* $(\lambda_1, \ldots, \lambda_s)$ *of (4.7.0) the number* $\lambda_1 + \ldots + \lambda_s \in \mathbf{Z}_p$ *is p-adically non-Liouville. Then* $y(x) \in (A\{\{x\}\}_1[\frac{1}{x}])^n$.

§5 Proof of Theorem 4.9

5.1 We will assume, with no restriction (cf. (4.5.3)), that our system (4.7.0) is in fact in ${}_{A,s}\mathcal{S}_{\mathcal{B}_r}$. By (4.1.2) we observe that $g = \sum_{\alpha \in \mathbf{N}^s} a_\alpha x^\alpha \in A[[x]]$ belongs to $A\{\{x\}\}_r$ if and only if for every absolute value in (4.4.1) $|-|_j$, $j = 1, \ldots, l$, and for every $\eta \in \mathbf{R}$, $0 < \eta < r$:

$$\lim_{|\alpha|_{\mathbf{R}} \to +\infty} |a_\alpha|_j \eta^{|\alpha|_{\mathbf{R}}} = 0.$$

On the other hand if $g \in A\{\{x\}\}_r^b$ we have:

$$|g|(r) = \sup_{\alpha \in \mathbf{N}^s} |a_\alpha|_A r^{|\alpha|_{\mathbf{R}}} = \max_j \left(\sup_{\alpha \in \mathbf{N}^s} |a_\alpha|_j r^{|\alpha|_{\mathbf{R}}} \right).$$

It follows that a standard system

$$x_i \frac{\partial}{\partial x_i} y = G_{\delta_i}(x) y \qquad i = 1, \ldots, s \tag{5.1.1}$$

with $G_{\delta_i}(x) \in \mathcal{M}(A\{\{x\}\}_r^b)$, is in ${}_{A,s}\mathcal{S}_{\mathcal{B}_r}$ if and only if the same system (5.1.1) belongs to ${}_{K_j,s}\mathcal{S}_{\mathcal{B}_r}$ for each $j = 1, \ldots, s$ when regarding each $G_{\delta_i}(x)$, $i = 1, \ldots, s$, as an element $\mathcal{M}_n(K_j\{\{x\}\}_r^b)$.

5.2 In particular for each absolute value $|-|_j$ in (4.1.1), the ordinary (in the variable x_i) system associated to (5.1.1)

$$x_i \frac{d}{dx_i} y = G_{\delta_i}(0) y \tag{5.2.1}$$

with $G_{\delta_i}(0) \in \mathcal{M}(K_j)$, may be regarded as an element of ${}_{K_j,1}\mathcal{S}_{\mathcal{B}_r}$. This is true because the recursion matrices $G_{\delta_i}(0)_m$, $m \in \mathbf{N}$, for (5.2.1), in the sense of (1.1.3), coincide with the recursion matrices $G_{m\delta_i}(0) = G(0)_{m\delta_i}$, $m \in \mathbf{N}$, for the system (4.7.1) (cf. the symbolism of (4.4.6)).

It follows that for any $i \in \{1, \ldots, s\}$ the eigenvalues of $G_{\delta_i}(0)$ are in $\mathbf{Z}_p$ (cfr. §1) and this purely as a consequence of condition ${}_{K_j,1}\mathcal{S}_{\mathcal{B}_r}$ for one j in $\{1, \ldots, l\}$.

5.3 Since each field $(K_j, |-|_j)$ is an algebraic closed field of characteristic zero, complete under the absolute value $|-|_j$, in order to prove theorem 4.9 we need, by 5.1, only to prove its following special case:

Theorem 5.4. *Suppose*

$$y(x) \in (k[[x]][\frac{1}{x}])^n$$

is a formal column solution of a system (5.1.1) in ${}_{k,s}\mathcal{S}_{\mathcal{B}_r}$ of type ρ. Then $y(x)$ is an element of $(k\{\{x\}\}_{r\rho}[\frac{1}{x}])^n$.

Proof. The idea is to reduce theorem 5.4 to corollary 1.7 by the method of the "generic line" ([S-S],[BA2, §3]). We consider a system of indeterminates $\gamma_1, \ldots, \gamma_s$ over k. We denote by k_γ the free Tate algebra [BGR]:

$$k_\gamma = k < \gamma_1, \ldots, \gamma_s >$$

endowed with the Gauss norm, which is actually an absolute value. Hence we can extend such an absolute value to the quotient field and consider an algebraically closed extension field of k_γ complete with respect to an absolute value extending the absolute value in k_γ. We will denote this field by $\mathcal{K}_\gamma$ and its absolute value still by $|-|$.

We consider the substitution $x_i = \gamma_i z$, $i = 1, \ldots, s$. Notice that for an element $g(x) \in k\{\{x\}\}_r^b$ we have $g(\gamma_1 z, \ldots, \gamma_s z) \in \mathcal{K}_\gamma\{\{z\}\}_r^b$ and $|g(x)|(r) = |g(\gamma_1 z, \ldots, \gamma_s z)|(r)$.

Suppose $y(x)$ satisfies a system of the form (5.1.1) in ${}_{k,s}\mathcal{S}_{\mathcal{B}_r}$. We obtain a solution

$$y(z) = y(\gamma_1 z, \ldots, \gamma_s z) \in (\mathcal{K}_\gamma[[z]][\frac{1}{z}])^n,$$

of the system:

$$(z\frac{d}{dz} - G(z))y(z) = 0 \tag{5.4.1}$$

where:

$$G(z) = \sum_{i=1}^{s} G_{\delta_i}(\gamma_1 z, \ldots, \gamma_s z) \in \mathcal{M}_n(\mathcal{B}_r)$$

($\mathcal{B}_r = \mathcal{K}_\gamma\{\{z\}\}_r^b$ is defined as in §0, with k replaced by $\mathcal{K}_\gamma$). By [BA2, Lemma 3.3.2], for any $\epsilon > 0$, $y(z)$ converges for $|z| < \epsilon$ if and only if $y(x)$ converges in the open polydisk $\{x \in \mathcal{K}_\gamma^s \mid |x_i| < \epsilon \text{ for } i = 1, \ldots, s\}$. We refer to (5.4.1) as *the system induced on the generic line by* (5.1.1).

We shall now prove that the system (5.4.1) is an element of $\mathcal{S}_{\mathcal{B}_r}$ (cf. §1) and that its type is also ρ. If we denote by $G_m(z)$, $m \in \mathbf{N}$, the recursion matrices for the ordinary system (5.4.1) (according to (1.1.3)), and retain formulas (4.4.4) and (4.4.5) for the system (5.1.1), we have:

Lemma 5.4.2. *For each* $m \in \mathbf{N}$:

$$\frac{G_m(z)}{m!} = \sum_{|\alpha|_{\mathbf{R}}=m} \frac{G_\alpha(\gamma_1 z, \ldots, \gamma_s z)}{\alpha!}. \tag{5.4.3}$$

Proof. We prove formula (5.4.3) by induction on m. We suppose (5.4.3) true for m. By applying (1.1.3) to (5.4.1) we obtain:

$$(m+1)\frac{G_{m+1}(z)}{(m+1)!} = z\frac{d}{dz}(\frac{G_m(z)}{m!}) + \frac{G_m(z)}{m!}(G(z) - m). \tag{5.4.4}$$

By induction, the matrix on the right hand side comes from the matrix

$$\sum_{i=1}^{s} x_i \frac{\partial}{\partial x_i}(\sum_{|\alpha|_{\mathbf{R}}=m} \frac{G_\alpha(x)}{\alpha!}) + (\sum_{|\alpha|_{\mathbf{R}}=m} \frac{G_\alpha(x)}{\alpha!})(\sum_{i=1}^{s} G_{\delta_i}(x) - m) \tag{5.4.5}$$

simply by the substitution $x_i = \gamma_i z$, $i = 1, \ldots, s$. By rearranging the terms, (5.4.5) is seen to be equal to

$$\sum_{i=1}^{s} \sum_{|\alpha|_R = m} [x_i \frac{\partial}{\partial x_i}(\frac{G_\alpha(x)}{\alpha!}) + \frac{G_\alpha(x)}{\alpha!}(G_{\delta_i}(x) - \alpha_i)]. \tag{5.4.6}$$

By (4.4.5) we have:

$$(\alpha_i + 1)\frac{G_{\alpha+\delta_i}(x)}{(\alpha+\delta_i)!} = x_i \frac{\partial}{\partial x_i}(\frac{G_\alpha(x)}{\alpha!}) + \frac{G_\alpha(x)}{\alpha!}(G_{\delta_i}(x) - \alpha_i). \tag{5.4.7}$$

Therefore (5.4.6) is equal to

$$\sum_{i=1}^{s} \sum_{|\alpha|_R = m} (\alpha_i + 1)\frac{G_{\alpha+\delta_i}(x)}{(\alpha+\delta_i)!} = \sum_{i=1}^{s} \sum_{|\beta|_R = m+1} \beta_i \frac{G_\beta(x)}{\beta!} = (m+1) \sum_{|\alpha|_R = m+1} \frac{G_\alpha(x)}{\alpha!}.$$

We conclude that (5.4.3) holds with m replaced by $m+1$. **Q.E.D.**

Condition $\mathcal{S}_{\mathcal{B}_r}$ may be expressed for the system (5.4.1) with $G(z) \in \mathcal{M}_n(\mathcal{B}_r)$, by

$$\lim_{m \to \infty} |\frac{G_m(z)}{m!}|(r)\eta^m = 0$$

for every $0 < \eta < 1$. This holds as a consequence of lemma 5.4.2 , because:

$$|\frac{G_m(z)}{m!}|(r) = |\sum_{|\alpha|_R = m} (\frac{G_\alpha(\gamma_1 z, \ldots, \gamma_s z)}{\alpha!})|(r) =$$

$$= \sup_{|\alpha|_R = m} |\frac{G_\alpha(\gamma_1 z, \ldots, \gamma_s z)}{\alpha!}|(r) = \sup_{|\alpha|_R = m} |\frac{G_\alpha(x)}{\alpha!}|(r). \tag{5.4.8}$$

As for the type of (5.4.1), we simply notice that its exponents are precisely the numbers $\mu = \lambda_1 + \ldots + \lambda_s$ each taken with multiplicity

$$e(\mu) = \sum_{\sum \lambda_i = \mu} e(\lambda_1, \ldots, \lambda_s)$$

where $e(\lambda_1, \ldots, \lambda_s)$ is the multiplicity of $(\lambda_1, \ldots, \lambda_s)$ as multiexponent of (4.7.0). From the previous discussion it follows that $y(z)$ is a formal solution of the system (5.4.1), which is in the set $\mathcal{S}_{\mathcal{B}_r}$ and is of type ρ . We can then apply corollary 1.7 to conclude that $y(z)$ is meromorphic in the disk of radius $r\rho$ with possible pole only at $z = 0$.The formal solution $y(x)$ therefore converges in the open polydisk

of radius $r\rho$ with possible poles at $x_1 \dots x_s = 0$. This concludes the proof of theorem 5.4 and also of theorem 4.9. **Q.E.D.**

Remark 5.5 Consider a system (cf. (4.7.0))

$$x_i \frac{\partial}{\partial x_i} y = G_{\delta_i}(x) y \qquad i = 1, \dots, s \tag{5.5.1}$$

with $G_{\delta_i}(x) \in \mathcal{M}_n(A\{\{x\}\}_r^b)$ (resp. $G_{\delta_i}(x) \in \mathcal{M}_n(A\{\{x\}\}_r)$) in ${}_{A,s}\mathcal{S}_{\mathcal{B}_r}$ (resp. in ${}_{A,s}\mathcal{S}_{\mathcal{A}_r}$). Let us assume there exists a system (cf. [BA-CT2]):

$$x_i \frac{\partial}{\partial x_i} y = C_{\delta_i}(G) y \qquad i = 1, \dots, s \tag{5.5.2}$$

with $C_{\delta_i}(G) \in \mathcal{M}_n(A)$ is in ${}_{A,s}\mathcal{S}_{\mathcal{B}_r}$ (resp. in ${}_{A,s}\mathcal{S}_{\mathcal{A}_r}$) formally equivalent to (5.5.1). This means that there exists a matrix $Y(x) \in GL_n(A[[x]][\frac{1}{x}])$ satisfying:

$$x_i \frac{\partial}{x_i \partial} Y(x) = G_{\delta_i}(x) Y(x) - Y(x) C_{\delta_i}(G), \quad i = 1, \dots, s. \tag{5.5.3}$$

Then $Y(x)$ is a solution of a standard system of rank n^2 (where $Y(x)$ is written as a column vector in some definite order):

$$x_i \frac{\partial}{\partial x_i} Y(x) = \tilde{G}_{\delta_i}(x) Y(x) \qquad i = 1, \dots, s \tag{5.5.4}$$

with $\tilde{G}_{\delta_i}(x) \in \mathcal{M}_{n^2}(A\{\{x\}\}_r^b)$ (resp.$\tilde{G}_{\delta_i}(x) \in \mathcal{M}_{n^2}(A\{\{x\}\}_r)$). Under the previous assumptions the system (5.5.4) satisfies property ${}_{A,s}\mathcal{S}_{\mathcal{B}_r}$ (resp. ${}_{A,s}\mathcal{S}_{\mathcal{A}_r}$). To show this, we may restrict ourselves to the case of systems in ${}_{A,s}\mathcal{S}_{\mathcal{B}_r}$. By use of the arguments explained in 5.1, we need only check the assertion in the case when the ring A is replaced by a valued field complete and algebraically closed. For any of the systems (5.5.j), $j = 1, 2, 3, 4$, where we assume $A = k$, let (5.5.j)$_\gamma$ denote the system induced by (5.5.j) on the generic line $x_i = \gamma_i z$, $i = 1, \dots, s$. Lemma 5.4.2 shows that (5.5.4) satisfies ${}_{k,s}\mathcal{S}_{\mathcal{B}_r}$ iff (5.5.4)$_\gamma$ satisfies ${}_{\mathcal{K}_\gamma,1}\mathcal{S}_{\mathcal{B}_r}$. On the other hand (5.5.4)$_\gamma$ is the system obtained from (5.5.1)$_\gamma$ and (5.5.2)$_\gamma$ as (1.5.1) was obtained from $(\delta - G)y = 0$ and $(\delta - C(G))y = 0$. Remark 1.5 then shows that (5.5.4)$_\gamma$ satisfies ${}_{\mathcal{K}_\gamma,1}\mathcal{S}_{\mathcal{B}_r}$. The theory developed in the present paper, shows then that, if (5.5.4) is of type $\rho > 0$, the systems (5.5.1) and (5.5.2) are equivalent over $A\{\{x\}\}_{r\rho}[\frac{1}{x}]$. Actually, under the previous assumptions, the type of (5.5.4) may be foreseen from the types of the differences of the multiexponents of (5.5.1), exactly as in the 1-variable case.

Our next step will be to show that, under certain circumstances, there exists a system of the type of (5.5.2), equivalent to (5.5.1) over $A[[x]][\frac{1}{x}]$. We will show that this is the case when $A = k$, but also when A is a regular Tate algebra and the system (5.5.1) can be completed to an absolute (i.e. relative to k) integrable system by the action of all continuous derivations of A/k.

This will complete the analogy with Christol's work [CH1] and will be the subject of a forthcoming paper ([BA-CT2]).

Bibliography

[A1] André Y. : "*G-functions and Geometry*", Vieweg-Verlag, Bonn 1989.

[A2] André Y. : " Spécialisation dans les disques singuliers réguliers", in Groupe d'Etude d'Analyse Ultramétrique 1985/86; Y.Amice, G.Christol, P.Robba, Eds. ; Publ. Math. Université Paris VII.

[BA1] Baldassarri F. : " Comparaison entre la cohomologie algébrique et la cohomologie p-adique rigide à coefficients dans un module différentiel I" , Inv. Math. 87 (1987), 83-99 .

[BA2] Baldassarri F. : " Comparaison entre la cohomologie algébrique et la cohomologie p-adique rigide à coefficients dans un module différentiel II", Math. Ann.280 (1988), 417-439.

[BA-CT1] Baldassarri F.,Chiarellotto B. :" On André's transfer theorem", in this volume.

[BA-CT2] Baldassarri F.,Chiarellotto B. :" Formal and p-adic theory of differential systems with logarithmic singularities depending upon parameters", Article in preparation.

[BA-CT3] Baldassarri F.,Chiarellotto B. :" Comparison between algebraic and rigid cohomology with logarithmic coefficients", Article in preparation.

[BE] Berthelot P. : "Cohomologie rigide et cohomologie rigide à support propre", to appear in Astérisque.

[BGR] Bosch S., Güntzer U., Remmert R.: "*Non Archimedean Analysis*", Grundlehren der Math. Wissenschaften 261, Springer–Verlag, Berlin-Heidelberg- New York-Tokyo 1984.

[CH1] Christol G.: "Un théorème de transfert pour les disques singuliers réguliers", Astérisque 119-120 (1984),151-168.

[CH2] Christol G. : "*Modules Différentiels et Equations Différentielles p-adiques*", Queen's Papers in Pure and Applied Math. 66, Queen's University, Kingston 1983.

[CH-DW] Christol G., Dwork B. : " Effective p-adic bounds at regular singular points", Duke Math. J. Vol. 62, No.2 (1991).

[CL] Clark D. : "A note on the p-adic convergence of solutions of linear differential equations", Proc.Am. Math.Soc. 17 (1966),262-269.

[DW1] Dwork B. : "*Lectures on p-adic Differential Equations*", Grundleheren der mathematischen Wissenschaften 253, Springer-Verlag, New York-Heidelberg-Berlin 1982.

Bibliography

[DW2] Dwork B. : "*Generalized Hypergeometric Functions*", Oxford Mathematical Monographs, Clarendon Press, Oxford 1990.

[G-L] Gérard R., Levelt A.H.M. : " Sur les connexions à singularités régulières dans le cas de plusieurs variables", Funkcialaj Ekvacioj 19 (1976), 149-173.

[LA] Lang S. : " *Diophantine Geometry* ", Interscience Tracts in Pure and Applied Math.11,J. Wiley and Sons, New York-London 1962.

[S-S] Sibuya Y., Sperber S. : "Arithmetic properties of power series solutions of algebraic differential equations", Ann. Math. 113 (1981), 111-157.

[WA] Watson G.N. :"*A Treatise on the Theory of Bessel Functions*", Cambridge University Press, Cambridge 1966.

Authors' address:
Dipartimento di Matematica
Università di Padova
Via Belzoni 7
I-35131 PADOVA - ITALY
E-mail:
"BALDASSARRI@PDMAT1.UNIPD.IT"
"CHIARELLOTTO@PDMAT1.UNIPD.IT"
Fax: (39-49-)8758596

Contemporary Mathematics
Volume 133, 1992

On André's transfer theorem.

Francesco Baldassarri and Bruno Chiarellotto

§0 Introduction

In this article we consider a system of linear differential equations of the form

$$(\delta - G)y = 0 \tag{0.1}$$

where $\delta = x\frac{d}{dx}$ and G is an $n \times n$ matrix of functions holomorphic in the p-adic open unit disk $D(0, 1^-)$. We refer to system (0.1) as to "the system $\delta - G$".

We are interested in both understanding and improving André's results ([A1, Résultat 2], [A2, V.4.2], [CH-DW1, 8.4]), giving information on the radius of meromorphy of the uniform part of the solution matrix of (0.1) at 0, under some mild assumptions on convergence for the solution matrix at a generic point. As in our previous article [BA-CT0], we rather take the viewpoint of estimating the radius of p-adic meromorphy of formally meromorphic solutions of (0.1) at 0. We take advantage of arguments used by Christol and Dwork ([CH-DW1]) in the construction of a Frobenius transform of (0.1) when $G(0)$ is nilpotent. We are able to extend those arguments to the general case. This generalization and the strong form of the so-called (cf.[CH2, 4.8.1]) Dwork-Frobenius theorem proved in [CH-DW2, 4.2], permit some improvement upon previous results. The reader should compare part *(i)* of Corollary 3.5 with [A1, *loc.cit.*], [A2, *loc.cit.*], [CH-DW1, *loc. cit.*], [CH-DW2, 4.4], noticing that we are not imposing any restriction on the Jordan form of $G(0)$. Part *(ii)* of Corollary 3.5, however, gives sharper results in the semisimple case; the results of section 3 of [DW-VP] follow from that statement.

As in all the previous authors, the technique used here is that of iterating the Frobenius transformation in order to refine a first crude estimate of the radius of convergence of the solutions of (0.1) at 0.

We are indebted to Professor Dwork for his encouragement and for showing us remarkable simplifications on our original proofs. We acknowledge some useful discussions with G.Gerotto.

We hope our considerations will lead to improvements in estimates for the growth of solutions of (0.1) at 0 when approaching the boundary of their disk of convergence ([CH-DW1]).

This paper is in final form and no version of it will be submitted for publication elsewhere.

1991 Mathematics Subject Classification. Primary 12H25.

§1 Notation and basic results ([CH-DW1],[CH-DW2], [BA-CT0]).

1.1 We recall the notation of [BA-CT0].

p= a prime number;

k= an algebraically closed field of characteristic zero, complete under a non-archimedean absolute value

$$| \ | : k \longrightarrow \mathbf{R}_{\geq 0}$$

extending the p-adic absolute value of $\mathbf{Q}_p$ and normalized by $|p| = p^{-1}$; we also use the additive form:

$$\mathrm{ord}(\lambda) = -\frac{\log|\lambda|}{\log p}$$

for $\lambda \in k$ ($\mathrm{ord}(0) = \infty$);

$\mathcal{V}$= ring of integers of k.

Let $(\Omega, |-|)$ be an algebraically closed and complete valued extension field of $(k, |-|)$, containing an element t of absolute value 1 whose residue class is trascendental over the residue field of k. We call $(\Omega, |-|)$ a *universal domain for* k and t a *generic point* (for k). For $\rho \geq 0$ and $c \in \Omega$ we will denote by $D(c, \rho^-)$ (resp. $D(c, \rho)$) the *open* (resp. *closed*) disk of radius ρ centered at c, in Ω:

$$D(c, \rho^-) = \{x \in \Omega \mid |x - c| < \rho\}$$

(resp.

$$D(c, \rho) = \{x \in \Omega \mid |x - c| \leq \rho\}).$$

We also put:

π= a fixed $(p-1)$-st root of $-p$ in k, $\pi^{p-1} = -p$;

E= the completion of $k(x)$ under the Gauss norm which we will still indicate by $|-|$;

$E_\circ$= the ring of analytic elements (defined over k) on $D(0, 1^-)$. The elements of $E_\circ$ are uniform limits on $D(0, 1^-)$ of sequences of rational functions defined over k without poles on $D(0, 1^-)$; therefore $E_\circ \subseteq E$ and the sup-norm on $D(0, 1^-)$ coincides, for elements of $E_\circ$, with the Gauss norm $|-|$.

$E'_\circ$= the quotient field of $E_\circ$.

Finite dimensional spaces over E with a natural basis are provided with a norm derived from $|-|$ in the standard way: we still denote such a norm by $|-|$. This applies to n-tuples and to matrices. In particular an element of $GL(n, E)$ is said to be *unimodular* if it and its inverse are bounded by 1.

1.2 For an element $\lambda \in k$, we define the *type* of λ as the radius of convergence $\rho(\lambda)$ of the series

$$g_\lambda(x) = \sum_{\substack{s=0 \\ \lambda - s \neq 0}}^{\infty} \frac{x^s}{\lambda - s}. \tag{1.2.1}$$

The type of a number is invariant under translations by integers.

Let now C be a square matrix with entries in k, that is $C \in \mathcal{M}_n(k)$, and let $\prod_\lambda (x-\lambda)^{e_\lambda}$ be its characteristic polynomial $det(x-C)$. We define the *type* $\rho(C)$ of C as the product

$$\rho(C) = \prod_\lambda \rho(\lambda)^{e_\lambda}. \tag{1.2.2}$$

The type of a matrix therefore only depends upon the types of its eigenvalues; in particular two matrices having the same number (counting multiplicities) of eigenvalues in each class of $k/\mathbf{Z}$, have the same type. For a k-linear transformation $h : V \to V$ of a finite dimensional k-vector space V, we will refer to the type of any matrix representation of h as *the type of* h. For any $G \in \mathcal{M}_n(k[\![x]\!])$ we will also say that $\rho(G(0))$ is the *type of the system* (0.1) *at* 0 and, more abusively, *of* G: we then denote it by $\rho(G)$.

1.3 For any differential ring extension (R,δ) of $(k[x],\delta)$, for $G \in \mathcal{M}_n(R)$ and $H \in GL(n,R)$ we define

$$G_{[H]} = (\delta H)H^{-1} + HGH^{-1}. \tag{1.3.1}$$

Therefore the substitution $z = Hy$ transforms the system $\delta y = Gy$ into $\delta z = G_{[H]}z$. Under these circumstances we will say that the two previous systems are *equivalent over* R. We recall Lemma 1.10 of [BA-CT0]:

Lemma 1.4. *Let $G, H \in \mathcal{M}_n(k[[x]])$ and $P \in GL(n,k((x)))$ be such that $H = G_{[P]}$. Then for any coset $\alpha + \mathbf{Z} \in k/\mathbf{Z}$, the matrices $G(0)$ and $H(0)$ have the same number, counting multiplicities, of eigenvalues in $\alpha + \mathbf{Z}$.*

Corollary 1.5. *Let $G, H \in \mathcal{M}_n(k[[x]])$ be such that the systems $\delta - G$ and $\delta - H$ are equivalent over $k((x))$. Then $\rho(G) = \rho(H)$.*

We now have:

Lemma 1.6. *If $\lambda, \lambda' \in \mathcal{V}$ are such that $p^a\lambda' - \lambda \in \mathbf{Z}$, for some $a = 0, 1, \ldots$, then $\rho(\lambda') = \rho(\lambda)^{p^a}$.*

Proof. We can assume $a = 1$; the proof of [BA-CT0] then applies to the present case. **Q.E.D.**

1.7 We recall that a point $c \in \Omega$ is said to be an *apparent singularity* for the system $\delta - G$, where G is meromorphic in a neighborhood of c, if there exists a full solution matrix of $\delta - G$ meromorphic at c. Let $\mathcal{H}$ be the set of all matrices $G \in \mathcal{M}_n(E_0')$ satisfying the following conditions:

$\mathcal{H}1$. G is holomorphic at 0.

$\mathcal{H}2$. The system $(\delta - G)y = 0$ has at most apparent singularities in $D(0,1^-)\setminus\{0\}$.

If $G \in \mathcal{H}$, the entries of G are meromorphic in the region

$$\mathcal{D} = D(0,1^-) \cup D(t,1^-).$$

Let G be a matrix in $\mathcal{H}$ and $c \in \mathcal{D}$. Let

$$\mathcal{S}_c(G) = Sol(\delta - G, \Omega((x-c)))$$

denote the Ω-vector space of vector solutions $y \in \Omega((x-c))^n$ of $(\delta - G)y = 0$. For $G \in \mathcal{H}$ and $c \in \mathcal{D}$ we then define:

$R_c(G) = \inf(1, \text{maximum common radius of meromorphy of the elements of } \mathcal{S}_c(G))$

We note that $R_c(G) = R_c(G_{[H]})$ if $H \in GL(n, E_0')$. If G is holomorphic at c, we denote by $\mathcal{U}_{G,c}(x) \in \mathcal{M}_n(\Omega[[x-c]])$ the solution matrix of the system $\delta - G$ normalized by $\mathcal{U}_{G,c}(c) = I_n$. We define, for any differential extension ring (R, δ) of $(k[x], \delta)$ and $G \in \mathcal{M}_n(R)$, $G_0 = I_n$, $G_1 = G$ and for $m \in \mathbf{N}$:

$$G_{m+1} = \delta G_m + G_m G - m G_m. \tag{1.7.1}$$

We then have, for $G \in \mathcal{H}$:

$$\mathcal{U}_{G,t}(x) = \sum_{m=0}^{\infty} \frac{G_m(t)}{m!} \left(\frac{x-t}{t}\right)^m. \tag{1.7.2}$$

Notice that for $G \in \mathcal{H}$, we have:

$$G_m(0) = G(0)(G(0) - I_n) \ldots (G(0) - (m-1)I_n)\,. \tag{1.7.3}$$

Lemma 1.8. *Let $\sigma : k/\mathbf{Z} \longrightarrow k$ be a section of the canonical projection $k \longrightarrow k/\mathbf{Z}$. Let $G \in \mathcal{M}_n(k[\![x]\!])$ There exists $S \in GL(n, \mathcal{V}[x, x^{-1}])$ such that $G_{[S]} \in \mathcal{M}_n(k[\![x]\!])$ and $G_{[S]}(0)$ is an upper triangular matrix with diagonal entries in the image of σ.*

Proof. Cf.[CH-DW1, 2.2 and 2.3]. **Q.E.D.**

We will refer to the transformation $G \longmapsto G_{[S]}$ of the previous lemma as a *shearing transformation.*

Lemma 1.9 (Elimination of apparent singularities). *Let G be an element of $\mathcal{H}$. Then there exists $H \in Gl(n, k(x))$, H unimodular, H and H^{-1} analytic at 0, such that $G_{[H]} \in \mathcal{M}_n(E_0)$.*

Proof. Cf. [CH-DW1,2.4], [CH2, 5.4.3].**Q.E.D.**

Lemma 1.10 (Dwork-Frobenius). *Let G be an element of $\mathcal{H}$. Then there exists $H \in Gl(n. E_0')$, such that $G_{[H]} \in \mathcal{H}$ and*

$$|G_{[H]}| \leq \sup(1, \frac{|\pi|}{R_t(G)}).$$

Proof. Cf. [CH-DW2,4.2].**Q.E.D.**

Lemma 1.11 (The trivial estimate). *Let G be an element of $\mathcal{H}$. Then*

$$R_0(G) \geq |\pi|^n \rho(G) \max(1, |G|)^{-n}.$$

Proof. It follows from [BA-CT0,2.3], after elimination of apparent singularities. **Q.E.D.**

Corollary 1.12 (The basic estimate). *Let G be an element of $\mathcal{H}$. Then*

$$R_0(G) \geq |\pi|^n \rho(G) \sup(1, \frac{|\pi|}{R_t(G)})^{-n}.$$

Proof. We reduce to the previous statement via Dwork-Frobenius.**Q.E.D.**

§2 Analytic calculations.

For a matrix $\alpha \in \mathcal{M}_n(k)$, we define its *spectral norm* $|\alpha|_{sp}$ as:

$$|\alpha|_{sp} = \max\{|\lambda| \mid \det(\lambda I_n - \alpha) = 0\}.$$

Let $\alpha \in \mathcal{M}_n(k)$; we define a matrix of formal power series:

$$\theta(\alpha, x) = \sum_{s=0}^{\infty} \binom{\alpha}{s} x^s \in \mathcal{M}_n(k[\![x]\!]). \tag{2.1}$$

This applies in particular to the case $n = 1$, for $\alpha = \lambda \in k$, to give a formal power series $\theta(\lambda, x) \in k[\![x]\!]$. We recall that $d(\lambda, \mathbf{Z}_p)$ denotes the distance between λ and $\mathbf{Z}_p$.

Proposition 2.2. *Let $l \in \mathbf{N}$, $l = 1, 2, 3 \ldots$, and $\lambda \in k$. Suppose that the series $\theta(\lambda, x)$ converges in the closed disk $D(0, |\pi|^{\frac{1}{p^{l-1}}})$ i.e. that:*

$$\lim_{m \to +\infty} |\binom{\lambda}{m}| |\pi|^{\frac{m}{p^{l-1}}} = 0 \tag{2.2.1}$$

Then $d(\lambda, \mathbf{Z}_p) < p^{-l+1}$.

Proof. (Cf. [CH2, 6.2.9], [CL]). Our hyphothesis implies $|\lambda| \leq 1$. In fact otherwise:

$$|\binom{\lambda}{p^m}| > \frac{1}{|p^m!|} = p^{\frac{p^m - 1}{p-1}}.$$

But

$$p^{\frac{p^m - 1}{p-1}} |\pi|^{\frac{p^m}{p^{l-1}}} = p^{\frac{p^m - p^{m-l+1} - 1}{p-1}},$$

which does not converge to 0 for $m \to +\infty$.

Suppose on the other hand that $\lambda \notin \mathbf{Z}_p$. Let

$$d(\lambda, \mathbf{Z}_p) = p^{-r}$$

where r is a positive real number and $r = \sigma + \epsilon$, where $\sigma = [r]$ is the integral part of r and $\epsilon \in [0, 1)$.

Consider $m \in \mathbf{N}$ such that

$$|\lambda + m| = p^{-r}.$$

For $s \in \mathbf{N}$ and $s > m$ we have:

$$(\lambda)_s = (\lambda)_m(\lambda + m)_{s-m}.$$

It follows that for each $s \in \mathbf{N}$ with $s > m$:

$$|(\lambda)_s| = c|(\lambda + m)_{s-m}|,$$

where $c \in \mathbf{R}_{>0}$ is a constant independent of s. Let $\alpha = \lambda + m$, so that $\mathrm{ord}(\alpha) = \sigma + \epsilon \geq \mathrm{ord}(\alpha + i)$, $\forall i = 0, 1, 2, \ldots$. An elementary calculation shows that :

$$\mathrm{ord}(\alpha)_s = \mathrm{ord}(\alpha) + [\frac{s-1}{p}] + [\frac{s-1}{p^2}] \ldots + [\frac{s-1}{p^\sigma}] + \epsilon[\frac{s-1}{p^{\sigma+1}}]$$

Therefore

$$\mathrm{ord}(\lambda)_s = s(\frac{1}{p} + \frac{1}{p^2} + \ldots + \frac{1}{p^\sigma}) + \frac{s\epsilon}{p^{\sigma+1}} + O(1) = s(\frac{1}{p-1} - \frac{1}{p^\sigma}\frac{1}{p-1} + \frac{\epsilon}{p^{\sigma+1}}) + O(1),$$

where $O(1)$ denotes a constant depending upon $d(\lambda, \mathbf{Z}_p)$ but not upon λ,s. For fixed λ and $s \to +\infty$ we then get the asymptotic estimate:

$$\mathrm{ord}\binom{\lambda}{s} = s(-\frac{1}{p^\sigma}\frac{1}{p-1} + \frac{\epsilon}{p^{\sigma+1}}) + O(log(s)) = \frac{s}{p^\sigma}(\frac{\epsilon}{p} - \frac{1}{p-1}) + O(log(s)).$$

So, the series $\theta(\lambda, x)$ converges exactly for:

$$|x| < p^{\frac{1}{p^\sigma}(\frac{\epsilon}{p} - \frac{1}{p-1})}.$$

Our hypotheses then say that :

$$-\frac{1}{p^{l-1}(p-1)} < \frac{1}{p^\sigma}(\frac{\epsilon}{p} - \frac{1}{p-1}).$$

Then, if $\epsilon = 0$ we have:

$$\frac{1}{p^{l-1}(p-1)} > \frac{1}{p^{\sigma}(p-1)}$$

which implies $l - 1 < \sigma$. If on the other hand $\epsilon \neq 0$, we have:

$$\frac{1}{p^{l-1}(p-1)} > \frac{1}{p^{\sigma}}\left(\frac{1}{p-1} - \frac{1}{p}\right)$$

i.e.

$$\frac{1}{p^{l-1}(p-1)} > \frac{1}{p^{\sigma+1}(p-1)},$$

so $l - 1 < \sigma + 1$. But $\sigma, l \in \mathbf{N}$, so that

$$l - 1 < \sigma + \epsilon.$$

Q.E.D.

We may generalize the previous result:

Proposition 2.3. *Suppose that for a matrix $\alpha \in \mathcal{M}_n(k)$ the series $\theta(\alpha, x)$ converges in the closed disk*

$$D(0, |\pi|^{\frac{1}{p^{l-1}}})$$

for some positive integer l. Then for each eigenvalue λ of α,

$$d(\lambda, \mathbf{Z}_p) < p^{1-l}.$$

Proof. This is a consequence of 2.2. **Q.E.D.**

2.4 For $\alpha \in \mathcal{M}_n(k)$, we have the formal identity (cf. [CH-DW1, 2.8]):

$$\sum_{s=0}^{+\infty} \binom{\alpha}{s} x^s = \sum_{j=0}^{+\infty} \frac{\alpha^j}{j!} L(x)^j \tag{2.4.1}$$

where

$$L(x) = \log(1+x) = \sum_{i=1}^{+\infty} (-1)^{i-1} \frac{x^i}{i}.$$

Lemma 2.5. *Let l be a positive integer and assume $\lambda \in k$ is such that $|\lambda| < p^{1-l}$. Then the series of functions*

$$\sum_{j=0}^{+\infty} \frac{\lambda^j}{j!} L(x)^j$$

converges uniformly in $D(0, |\pi|^{\frac{1}{p^{l-1}}})$.

Proof. Consideration of the Newton polygon of

$$L(x) = \sum_{i=1}^{+\infty} (-1)^{i-1} \frac{x^i}{i}$$

shows that if $|x| \leq |\pi|^{\frac{1}{p^{l-1}}} = p^{-\frac{1}{p^{l-1}(p-1)}}$ then

$$|L(x)| \leq p^{l-1-\frac{1}{p-1}} = |\pi| p^{l-1}.$$

The statement now follows from our hyphotheses about λ. **Q.E.D.**

Remark 2.5.1. One may avoid the use of Newton polygons by the following argument. For $j = 1, 2, \ldots$, let

$$h(j) = \frac{j}{(p-1)p^l} - \mathrm{ord}(j) .$$

It is enough to show that

$$\inf_{j \geq 1} h(j) \geq -l + \frac{1}{p-1} .$$

Since $p \not| m$ implies $h(mj) \geq h(j)$, we may restrict our attention to $j = 1, p, p^2, \ldots$. Let $h_1(s) = h(p^s)$. A trivial calculation shows that $h_1(s) \leq h_1(s+1)$ if and only if $s \geq l$. Thus the maximum of h is assumed at $j = p^l$, which completes the proof.

As a consequence of the previous lemma we can state:

Corollary 2.6. *Let l be a positive integer and let $\lambda \in k$ be such that $|\lambda| < p^{1-l}$. Then:*

(i) $\theta(\lambda, x)$ converges in the closed disk $D(0, |\pi|^{\frac{1}{p^{l-1}}})$

(ii) If $\zeta \in k$ satisfies $\zeta^{p^l} = 1$, we have:

$$\theta(\lambda, \zeta - 1) = 1.$$

Proof. For the (i) part we simply notice that formula (2.4.1) presents $\theta(\lambda, x)$ as the Taylor expansion at 0 of the analytic function

$$\sum_{j=0}^{+\infty} \frac{\lambda^j}{j!} L(x)^j;$$

we then apply lemma 2.5.

For (ii), we remark that if $\zeta \in k$ satisfies $\zeta^{p^l} = 1$ then $|\zeta - 1| = |\pi|^{\frac{1}{p^{l-1}}}$. **Q.E.D.**

Finally we can state:

Theorem 2.7. *Let l a positive integer and $\alpha \in \mathcal{M}_n(k)$ be such that*

$$|\alpha|_{sp} < p^{1-l}.$$

Then if $\zeta \in k$ satisfies $\zeta^{p^l} = 1$,

$$\theta(\alpha, \zeta - 1) = I_n.$$

Proof. We decompose $\alpha \in \mathcal{M}_n(k)$ as

$$\alpha = \alpha_{ss} + \alpha_{nilp}$$

where $\alpha_{ss}, \alpha_{nilp} \in \mathcal{M}_n(k)$, α_{ss} is semisimple, α_{nilp} is nilpotent and

$$[\alpha_{nilp}, \alpha_{ss}] = 0.$$

We then obtain the formal identity:

$$\theta(\alpha, x) = \theta(\alpha_{nilp}, x)\theta(\alpha_{ss}, x). \tag{2.7.2}$$

In [CH-DW1, 2.8] it is proved that:

a) $\theta(\alpha_{nilp}, x)$ converges in $D(0, 1^-)$,

b) $\theta(\alpha_{nilp}, \zeta - 1) = I_n$ for any $\zeta \in k$ such that $\zeta^{p^N} = 1$, for some $N = 1, 2, \ldots$.

We may then suppose α semisimple and finally $\alpha = \lambda \in k$ (so, λ is an eigenvalue of the original α) .

The hypotheses of the theorem imply that $|\lambda| < p^{1-l}$: we then apply part (ii) of corollary 2.6. **Q.E.D.**

§3 Frobenius action and main theorem.

For a positive integer l, $q = p^l$, and $u \in \Omega((x))$ we define by u^{ϕ_q} to be the composition $x \to u(x^q)$. We sometimes informally refer to the map $x \to x^q$ as the *Frobenius transformation of order l*. We consider, for $l = 1, 2, 3 \ldots$ and $G \in \mathcal{H}$, the condition :

$\mathcal{H}_l$: *The formal matrix solution $\mathcal{U}_{G,t}$ converges in the closed disk $D(t, |\pi|^{\frac{1}{p^{l-1}}})$.*

We will also denote by $\mathcal{H}_l$ the set of matrices in $\mathcal{H}$ satisfying condition $\mathcal{H}_l$.

Proposition 3.1. *If $G \in \mathcal{H}_l$ there exists a unimodular $H \in GL(n, k(x))$ such that the following two conditions hold:*

(i) $G_{[H]} \in \mathcal{H}_l \cap \mathcal{M}_n(E_\circ)$,

(ii) for each eigenvalue λ of $G_{[H]}(0)$

$$d(\lambda, \mathbf{Z}_p) < p^{1-l}.$$

Proof. We use the matrix H of lemma 1.9 to eliminate apparent singularities. The fact that H is holomorphic in the open disk $D(t, 1^-)$, while $\det H$ does not vanish on that disk, shows that $G_{[H]} \in \mathcal{H}_l$. According to formulas (1.7.2) and (1.7.3), our hypotheses imply that

$$\theta(G_{[H]}(0), x)$$

converges in $D(0, |\pi|^{\frac{1}{p^{l-1}}})$. We then apply proposition 2.3. **Q.E.D.**

Corollary 3.2. *Let $G \in \mathcal{H}_l$. There exists a unimodular $U \in GL(n, k(x))$ such that $G_{[U]} \in \mathcal{H}_l \cap \mathcal{M}_n(E_\circ)$ and*

$$|G_{[U]}(0)|_{sp} < p^{1-l}.$$

Proof. We write, using the matrix H of the previous lemma 3.1,

$$U = TH$$

for a suitable " shearing transformation" (cf.(1.8)) $T \in GL(n, \mathcal{V}[x, x^{-1}])$.**Q.E.D.**

In the sequel we will refer to the matrix $G_{[U]}$ of corollary 3.2 as a *normal form* of $G \in \mathcal{H}_l$. A matrix $G \in \mathcal{H}_l$ will be said to be *in normal form* , if $G \in \mathcal{H}_l \cap \mathcal{M}_n(E_\circ)$ and $|G(0)|_{sp} < p^{1-l}$.

We can finally state :

Theorem 3.3. *Let $G \in \mathcal{H}_l$ be in normal form. Then there exists $H \in \mathcal{M}_n(E_\circ) \cap GL(n, E'_\circ)$, $H(0) = I_n$ and $F \in \mathcal{H}$ such that for $q = p^l$*

$$qF^{\phi_q} = G_{[H]}. \tag{3.3.1}$$

We have

$$\rho(F(0)) = \rho(G(0))^q \tag{3.3.2).}$$

For each $c \in \mathcal{D}$ the map:

$$\mathcal{S}_{c^q}(F) \longrightarrow \mathcal{S}_c(G)$$

$$y \longrightarrow H^{-1} y^{\phi_q}$$

is an isomorphism of Ω-vector spaces. In particular, for $c = 0$:

$$R_0(F) = R_0(G)^q \,, \tag{3.3.3}$$

and, for $c = t$:

$$R_t(F) = R_{t^q}(F) = R_t(G)^q \,. \tag{3.3.4}$$

Proof. We follow the proof of Theorem 8.2 of [CH-DW1].

For $G \in \mathcal{H}_l$ we define a series of functions of $(x, z) \in (\mathcal{D} \setminus \{0\}) \times (\mathcal{D} \setminus \{0\})$ (cf. (1.7.1))

$$\mathcal{Y}(x, z) = \sum_{s=0}^{+\infty} \frac{G_s(x)}{s!} \left(\frac{z - x}{x}\right)^s .$$

By our hypotheses this is a convergent series in the region

$$(x, z) \in (\mathcal{D} \setminus \{0\}) \times (\mathcal{D} \setminus \{0\}) \quad , \quad \left|\frac{z - x}{x}\right| \leq |\pi|^{\frac{1}{p^{l-1}}} .$$

For $c, x, z \in \mathcal{D} \setminus \{0\}$ and for

$$\left|\frac{z - x}{x}\right| \leq |\pi|^{\frac{1}{p^{l-1}}} \quad , \quad \left|\frac{x - c}{c}\right| \leq |\pi|^{\frac{1}{p^{l-1}}}$$

we have:

$$\mathcal{U}_{G,c}(z) = \mathcal{Y}(x, z)\mathcal{U}_{G,c}(x).$$

For $q = p^l$, we define:

$$H = \frac{1}{q} \sum_{\zeta^q = 1} \mathcal{Y}(x, \zeta x) = \sum_{m=0}^{+\infty} \left(\sum_{\zeta^q = 1} \frac{(\zeta - 1)^m}{q} \right) \frac{G_m(x)}{m!} \,.$$

Since for $\zeta^q = 1$

$$|\zeta - 1| \leq |\pi|^{\frac{1}{p^{l-1}}},$$

$H(x) \in \mathcal{M}_n(E_0)$. We now calculate $H(0)$. We have:

$$\sum_{m=0}^{+\infty} \frac{G_m(0)}{m!} (\zeta - 1)^m = \theta(G(0), \zeta - 1) = I_n$$

by theorem 2.7 and by our hypothesis that G be in normal form; therefore

$$H(0) = \frac{1}{q} \sum_{\zeta^q = 1} \theta(G(0), \zeta - 1) = I_n.$$

We conclude that $H(x) \in Gl(n, E_0')$. For the remaining part of the proof we refer to [CH-DW1, 8.2]. In particular we observe that (8.2.2) of *loc. cit.* implies that $F \in \mathcal{M}_n(E_0')$ by Lemma 3.2 of [BA-CT0]. Notice that, while (3.3.3) is obvious, (3.3.4) follows from [CH-DW1, Lemma 6.3] and the hypothesis $\mathcal{H}_l$. **Q.E.D.**

Finally our main theorem:

Theorem 3.4. *Let $G \in \mathcal{H}$. Then:*

$$R_0(G) \geq \rho(G(0))R_t(G)^n.$$

Proof. If G does not belong to $\mathcal{H}_1$ we deduce the result from the basic estimate (1.12). Otherwise, let l be maximal such that $G \in \mathcal{H}_l$; we may assume G to be in normal form. In the notation of theorem 3.3, we have $R_0(G) = R_0(F)^{\frac{1}{q}}$, $\rho(F(0)) = \rho(G(0))^q$, $R_t(F) = R_{tq}(F) = R_t(G)^q$ and

$$R_0(F) \geq |\pi|^n \rho(F(0)) \sup(1, \frac{|\pi|}{R_t(F)})^{-n}$$

by the basic estimate (1.12). We then deduce

$$R_0(G) \geq |\pi|^{\frac{n}{q}} \rho(G(0)) \sup(1, \frac{|\pi|}{R_t(G)^q})^{-\frac{n}{q}}. \tag{3.4.1}$$

By the maximality of l

$$|\pi|^{\frac{1}{p^{l-1}}} \leq R_t(G) \leq |\pi|^{\frac{1}{p^l}},$$

so that, for $q = p^l$

$$|\pi|^p \leq R_t(G)^q \leq |\pi|,$$

now (3.4.1) gives

$$R_0(G) \geq |\pi|^{\frac{n}{q}} \rho(G(0))(\frac{R_t(G)^q}{|\pi|})^{\frac{n}{q}} \geq \rho(G(0))R_t(G)^n .$$

Q.E.D.

Corollary 3.5. *Let $G \in \mathcal{H}$ and let ρ be the type of the linear transformation:*

$$\mathcal{M}_n(k) \longrightarrow \mathcal{M}_n(k)$$

$$X \longmapsto G(0)X - XG(0).$$

(i) Let $Y_G(x)x^{G(0)}$ (cf. [CH-DW1, Introduction], [A2, III.1.1]) be a formal solution matrix of (0.1) at 0, with $Y_G(x) \in \mathcal{M}_n(k((x)))$. Then $Y_G(x)$ is meromorphic in the open disk

$$D(0, R_t(G)^{n^2}\rho^-) .$$

(ii) Let $\lambda \in k$ be an eigenvalue of $G(0)$, and let $y(x)x^\lambda$, with $y(x) \in k((x))^n$, be a formal column solution of (0.1) at 0. Then $y(x)$ is meromorphic in the open disk

$$D(0, R_t(G)^n \rho(G(0) - \lambda I_n)^-) .$$

Bibliography

[A1] André Y. : " Spécialisation dans les disques singuliers réguliers", in Groupe d'Etude d'Analyse Ultramétrique 1985/86; Y.Amice, G.Christol, P.Robba, Eds. ; Publ. Math. Université Paris VII.

[A2] André Y. : "*G-functions and Geometry*", Vieweg-Verlag, Bonn 1989.

[BA-CT0] Baldassarri F.,Chiarellotto B. :" On Christol's theorem. A generalization to systems of PDE's with logarithmic singularities depending upon parameters", in this volume.

[CH1] Christol G.: "Un théorème de transfert pour les disques singuliers réguliers", Astérisque 119-120 (1984),151-168.

[CH2] Christol G. : "*Modules Différentiels et Equations Différentielles p-adiques*", Queen's Papers in Pure and Applied Math. 66, Queen's University, Kingston 1983.

[CH-DW 1] Christol G., Dwork B. : " Effective p-adic bounds at regular singular points", Duke Math. J. Vol. 62, No.2 (1991).

[CH-DW 2] Christol G., Dwork B. : " Differential modules of bounded spectral norm ", in this volume.

[CL] Clark D. : "A note on the p-adic convergence of solution of linear differential equations", Proc. Am. Math. Soc. 17 (1966), 262-269.

[DW-VP] Dwork B., van der Poorten A. : "The Eisenstein constant", Duke Math. Journal 65 (1992), 23-43.

Authors' address:
Dipartimento di Matematica
Università di Padova
Via Belzoni 7
I-35131 PADOVA - ITALY
E-mail:
"BALDASSARRI@PDMAT1.UNIPD.IT"
"CHIARELLOTTO@PDMAT1.UNIPD.IT"
Fax: (39-49-)8758596

This page intentionally left blank

Contemporary Mathematics
Volume 133, 1992

Differential modules of bounded spectral norm

G. Christol and B. Dwork

INTRODUCTION

Let k be a field of characteristic zero which is complete under a non-archimedean rank 1 valuation with residue class field $\overline{k}$ of characteristic $p > 0$. Let E be the completion of $k(x)$ under the gauss norm.

Let M be a n-dimensional differential module over E. Thus we have a mapping $\delta: M \to M$ such that for $u \in E, m \in E$ we have $\delta(um) = u\delta M + (Du)m$ where D is the extension to E of d/dx as operator on $k(x)$. As operator on E, D has norm 1 and spectral norm $|\pi|$, where $\pi^{p-1} = -p$.

M has a unique structure as topological vector space over E but the norm itself is not unique. If $e_1, \ldots, e_n$ is a basis then each element of M is of the form $z = \sum_{i=1}^{n} Y_i e_i, (Y_i \in E, 1 \leq i \leq n)$ and we may construct the associated norm, $|z| = \operatorname{Sup}_i |Y_i|$. For this norm $\{e_1, \ldots, e_n\}$ is an orthonormal basis. If $\{e_1', \ldots, e_n'\}$ is another basis then the associated norm coincides with the previous one if and only if in the change of basis matrix lies in $G\ell(n, O_E)$, where O_E is the closed unit ball of E. We shall use the word "norm" only in this sense.

If v_1, v_2 are two norms on M then the identity map has two norms, one as map of M_{v_1} into M_{v_2} and the other as map of M_{v_2} into M_{v_1}. We define $\theta(v_2, v_1)$ to be the product of these two norms, i.e.

$$\theta(v_1, v_2) = \alpha_1 \alpha_2$$

where

$$\alpha_i = \operatorname*{Sup}_{|z|_{v_i}=1} |z|_{v_j} \quad , \quad i \neq j; \ \ i, j \in \{1, 2\}.$$

Equivalently $\theta(v_1, v_2) = |H|\,|H^{-1}|$ where H is the change of basis matrix for an orthonormal basis relative to v_1 to a corresponding basis for v_2.

We observe that the spectral norm, $|\delta|_{sp} = \lim \sup |\delta^s|_v^{\frac{1}{s}}$ is independent of the choice of norm v and is unchanged if δ is replaced by $u\delta$ where u is a unit

1991 Subject Classification, Primary 12H25 Secondary 11S80.

This paper is in final form and no version of it will be submitted for publication elsewhere.

0271-4132/92 $1.00 + $.25 per page

in E provided $|\delta|_{sp} \geq |\pi|$. Of course if u is a unit then $|u\delta|_v = |\delta|_v$ but does depend upon the choice of v.

A norm of M will be said to be of **cyclic vector type** if there exists an orthonormal basis of the form $\{z, \delta z, \cdots, \delta^{n-1}z\}$.

We may associate with M a system of differential equations, most simply by choosing $G \in M_n(E)$ such that

$$\delta \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} = G \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} . \tag{0.1}$$

for some basis $\{e_1, \ldots, e_n\}$ of M. The associated differential equation will be taken to be

$$Dy = Gy \tag{0.1'}$$

(Strictly speaking this is the equation for horizontality in the dual module).

It is well known that

(0.2) the spectral norm of δ is bounded by $|\pi|$, if and only if equation (0.1') is solvable in the generic unit disk $D(t, 1^-)$

(0.3) The differential equation (0.1') has a solution matrix converging on the generic open disk of radius $|\pi|$, $D(t, |\pi|^-)$, if and only if δ has spectral norm bounded by unity.

The case of spectral norm greater than unity will be considered in §4.

The following elementary result [Chr 1, Prop. 8.1], [Chr. 2, Prop. 4.8.1], [Chr-Dw, Prop. 7.1], [A. IV 2.1] plays an important role in our considerations .

(0.4) Under hypothesis (0.2), $|\delta|_v \leq 1$ for each metric v of cyclic vector type.

The standard proof is by induction on the order and uses the estimate $|D| \leq 1/r$ for D as operator on functions analytic on the set $|x - t| = 1/r$. We give a new proof based upon the spectral norm and will indeed show more:

(0.5) Under hypothesis (0.3), $|\delta|_v \leq 1$ for each norm v of cyclic vector type.

Our main object is the discussion of the following conjectures.

Conjecture A.

Subject to hypothesis (0.3), for each norm v of M there exists a norm v' such that

$$A.1 \qquad |\delta|_{v'} \leq 1$$

$$A.2 \qquad \theta(v, v') \leq \sup\,(1, |\delta|_v)^{n-1}$$

Conjecture B.

Subject to hypothesis (0.3), for each norm v of M there exists a norm, v', of cyclic vector type such that

$$B.1 \qquad \theta(v, v') \leq \sup (1, |\delta|_v)^{n-1} / |(n-1)!|.$$

(Thus aside from the factor $(n-1)!$, Conjecture B implies Conjecture A by use of (0.5).

We have proven [Chr-Dw, § 1].

(0.6) Given $\epsilon > 0$ then for each norm v there esists a norm v' of cyclic vector type such that

$$\theta(v, v') \quad \leq (1+\epsilon) \quad \operatorname{Sup} (1, |\delta|_v)^{2(n-1)} \Big/ |(n-1)!|$$

(Here hypotheses 0.2, 0.3 play no role.)

(0.7) Given $\epsilon > 0$ then subject to hypothesis (0.2) there exists a norm v' of cyclic vector type such that

$$\theta(v, v') \quad \leq (1+\epsilon) \quad \operatorname{Sup} (1, |\delta|_v)^{n-1} / |(n-1)! \prod_{j=1}^{n-1} \binom{n}{j}|$$

Our proof of (0.7) used hypothesis (0.2) only to be able to use the conclusion of (0.4) and hence it will follow from the present article that

(0.7′) The conclusion of (0.7) holds subject to hypotheses (0.3). Thus neglecting the factor $(1+\epsilon)$ we may conclude that Conjecture A holds for $p \geq n+1$.

We have [Chr-Dw, A.5] demonstrated conjecture A for $n = 2$ subject to hypothesis (0.2). Once again we may by (0.5) replace (0.2) by hypothesis (0.3).

In the present article we demonstrate Conjecture A for $n = 3$.

We cannot believe that condition (0.3) has not been previously used in this type of investigation.

Conjecture A has application in the determination of effective p-adic bounds at regular singular points. In this application we may strengthen the hypothesis to (0.2). In this application Conjecture B is not needed. We do not know if B has been verified even for $n = 2$.

§1 General Theory

Let v be a norm of the differential module M. Let M_v(resp: $\mathfrak{P}_v$) be the closed unit ball (resp: the open unit ball) of M in this norm. Let $\overline{M}_v = M_v / \mathfrak{P}_v$, a vector space over $\overline{k}(x)$. Since δ need not be stable on M_v there need not be a corresponding differential module structure of $\overline{M}_v$.

Lemma 1.1.

Let $|\delta|_v > 1$. Let $A \in k, |A| = |\delta|_v$ then $\tau = A^{-1}\delta$ is stable on M_v and by passage to quotients we obtain $\overline{\tau}$, an endomorphism of $\overline{M}_v$ as $\overline{k}(x)$ vector space. Subject to hypothesis (0.3), $\overline{\tau}$ is nilpotent.

Proof

Since $|A^{-1}D|_E = |A^{-1}| < 1$, it is clear that $\overline{\tau}$ is $\bar{k}(x)$-linear. By hypothesis $|\delta^s|_v / b^s \to 0$ for all $b > 1$ and hence letting $b = |A|$ there exists $s \in \mathbb{N}$ such that $|\tau^s|_v = |A^{-1}|^s |\delta^s|_v < 1$. Thus $\overline{\tau}^s = 0$.

Corollary 1.2.

Assertion (0.5) is valid.

Proof

Let z_o be a cyclic element of M. Then

$$\delta^n z_0 = (C_1\delta^{n-1} + \cdots + C_n)z_0 \qquad , C_j \in E(1 \le j \le n). \tag{1.2.1}$$

We assert (subject to (0.3)) that $|C_j| \le 1$ for $1 \le j \le n$).

Suppose otherwise. Replacing k (if necessary) by a finite extension field, there exists $\alpha \in k$ such that

$$\sup_j |C_j \,/\, \alpha^j| = 1 \tag{1.2.2}$$

Thus relative to the basis $\{z_0, z_1, \ldots, z_{n-1}\}, z_j = \left(\frac{1}{\alpha}\delta\right)^j z_0$, the operator $\frac{1}{\alpha}\delta$ is represented by the matrix G

$$\tfrac{1}{\alpha}\,\delta \begin{pmatrix} z_0 \\ z_1 \\ \vdots \\ z_{n-1} \end{pmatrix} = G \begin{pmatrix} z_0 \\ \vdots \\ z_{n-1} \end{pmatrix} \tag{1.2.3}$$

$$G = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 1 \\ \frac{C_n}{\alpha^n} & \frac{C_{n-1}}{\alpha^{n-1}} & & , \ldots , & \frac{C_1}{\alpha} \end{pmatrix} \tag{1.2.4}$$

which shows that if v is the norm corresponding to this basis then $|\frac{1}{\alpha}\delta|_v = 1$ and so $|\delta|_v = |\alpha| > 1$. Using this norm the map $\overline{\tau}$ obtained from $\tau = \frac{1}{\alpha}\delta$ must be nilpotent. But the images of $\{z_0, \ldots, z_{n-1}\}$ in $\overline{M}_v$ are linearly independent over $\overline{k}(x)$ and the corresponding matrix, $\overline{G}$, of $\overline{\tau}$ is given by the image of G under the residue class map. The images β_j of C_j / α^j are not all zero while the characteristic polynomial of $\overline{G}$ is

$$-\lambda^n + \beta_1\lambda^{n-1} + \ldots + \beta_n, \tag{1.2.5}$$

which shows that $\overline{\tau}$ has a non-zero eigenvalue, contradicting the lemma.

Corollary 1.3.

(a) *Under the hypothesis of the lemma there exists a v orthonormal basis of*

$M, \{e_1, \dots, e_n\}$ *such that* $\tau e_n \in \mathfrak{P}_v$ *and for* $j \leq n-1$ *we have one of the following three possibilities*

(i) $\delta\, e_j = Ae_{j+1}$

(ii) $\frac{1}{A}\,\delta e_j \in \mathfrak{P}_v$

$\delta\, e_{j+1} = Ae_{j+2}$

(iii) $\frac{1}{A}\,\delta e_{j'} \in \mathfrak{P}_v$ *for all* $j' \in [j, n]$.

Furthermore we may insist that $j = 1$ *falls into case (i)*

(b) *In particular if* $j \neq n$ *is minimal satisfying (iii) then* $Ae_j = \delta e_{j-1}$ *and* $e_{j+1}, \dots, e_n$ *may be chosen such that for* $s \geq j+1$

$$\delta e_s = \sum_{i=1}^{n} B_{s,i} e_i \ ,$$

1.3.1 $B_{s,i} = 0$ *for each* i *such that* $\delta e_{i-1} = Ae_i$

1.3.2 $|B_{s,s}| \geq 1$

Proof

By hypothesis, $\overline{\tau}$ is nilpotent but not trivial. Hence the Jordan normal form of $\overline{\tau}$ is rational over $\overline{k}(x)$ and there must be a block of dimension $r \geq 2$. Thus there exists $\overline{e} \in \overline{M}_v$ such that $\{\overline{e}, \overline{\tau}\,\overline{e}, \dots, \overline{\tau}^{r-1}\,\overline{e}\}$ are linearly independent. Let e_1 be a lifting of $\overline{e}$ to M_v then $\{e_1, \tau e_1, \dots, \tau^{r-1} e_1\}$ is a lifting of the indicated basis of the Jordan block and hence must be an orthonormal set in the v metric of M. Putting $e_{j+1} = \tau e_j (1 \leq j \leq r-1)$ we obtain $\{e_1, \dots, e_r\}$ and certainly $\tau e_r \in \mathfrak{P}_v$. The proof of (a) is completed by applying the same argument to the full Jordan normal form of $\overline{\tau}$.

For part (b) let $T = \{i | \delta e_{i-1} = Ae_i\}$. For $s \geq j+1$ let $e'_s = e_s - \sum_{i \in T} \lambda_i\, e_{i-1}$

Now for $s \geq j+1$

$$\delta e_s = \sum_{q=1}^{n} B_{s,q}\, e_q \quad , |B_{s,q}| < A$$

Hence

$$\delta e'_s = \sum_{q=1}^{n} B_{s,q} e_q - \sum_{i \in T} \left((D\lambda_i) e_{i-1} + A\lambda_i e_i \right)$$

We choose, $\lambda_i \in E$ such that for $i \in T$

$$B_{s,i} - D\lambda_{i+1} = A\lambda_i$$

Of course λ_{i+1} is understood to be zero if $i+1 \notin T$. Starting with i maximal in T this fixes each λ_i and $|\lambda_i| < 1$. The new basis obtained by replacing

$\{e_{j+1}, \ldots, e_n\}$ by $\{e'_{j+1}, \ldots, e'_n\}$ is clearly v-orthonormal. To verify 1.3.2, for $j+1 \leq s \leq n$ let $e''_s = e'_s$ if $|B_{s,s}| \geq 1$ while otherwise let $e''_s = xe'_s$. This has the effect of replacing $B_{s,s}$ by $B_{s,s} + x^{-1}$.

Corollary 1.4.

Let v be a metric of M and assume hypothesis (0.3). The conclusion of Conjecture A holds trivially if $|\delta|_v \leq 1$. It also holds if $|\delta|_v > 1$ and if $\overline{\tau}$ is nilpotent of maximal order, i.e. $\overline{\tau}^{n-1} \neq 0$

Proof

Suppose $|\delta|_v > 1$ and $\overline{\tau}$ is nilpotent of maximal order. By Corollary 1.3, there exists $z \in M_v$ such that $z, \tau z, \ldots, \tau^{n-1}z$ is an orthonormal basis for the v–norm. Let v' be the norm corresponding to the basis $\{z, \delta z, \ldots, \delta^{n-1}z\}$. The change in basis matrix is

$$H = \begin{pmatrix} 1 & & & 0 \\ & A & & \\ & & \ddots & \\ 0 & & & A^{n-1} \end{pmatrix}$$

where $A \in k$, $|A| = |\delta|_v > 1$. Thus $\theta(v, v') = |A|^{n-1}$ and condition A.1 is satisfied by Corollary 1.2.

Corollary 1.5.

Conjecture A holds for $n \leq 2$.

Proof

For $n = 1$ there is only one metric and the assertion is quite trivial. In any case it follows from 1.2.

For $n = 2$ either $|\delta|_v \leq 1$ or $|\delta|_v > 1$. In the second case $\overline{\tau}$ must be nilpotent but not trivial and hence is of maximal order.

§2 Estimates

Let G be the matrix representing the action of δ on a basis of M. We have shown that if $|G| > 1$ then subject to (0.3) the matrix G has no eigenvalue λ (in an algebraic extension of E) such that $|\lambda| \geq |G|$. Thus the characteristic polynomial of G is of the form $\lambda^n + \alpha_1\lambda^{n-1} + \cdots + \alpha_n$ with

$$|\alpha_j| < |G|^j$$

if $|G| > 1$ and hypothesis (0.3) holds. Corollary 1.2 suggests that perhaps $|\alpha_j| \leq 1$, $(1 \leq j \leq n)$. This is not the case. For example let $G = \begin{pmatrix} 0 & A \\ 0 & 0 \end{pmatrix}$, $A \in k, |A| > 1$.

Then $D\begin{pmatrix} 1 & Ax \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & A \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & Ax \\ 0 & 1 \end{pmatrix}$ which shows that (0.3) is satisfied.

Putting $H = \begin{pmatrix} 1 & x \\ x & 1+x^2 \end{pmatrix}$,

$$G_{[H]} = DH \cdot H^{-1} + HGH^{-1},$$

we obtain $G_{[H]} = \begin{pmatrix} -(A+1)x & A+1 \\ 1-(A+1)x^2 & (A+1)x \end{pmatrix}$ and so the characteristic polynomial of $G_{[H]}$ is $\lambda^2 - (A+1)$. This and other examples suggest that subject to hypothesis (0.3) we have

$$|\alpha_j| \leq \quad \mathrm{Sup}(1, |G|)^{j-1}.$$

Proposition.

Subject to (0.3) we have $|TrG| \leq 1$.

Proof

If M satisfies (0.3) and if $\{e_1, \ldots, e_n\}$ is a basis then the one dimensional module spanned by $e_1 \wedge \cdots \wedge e_n$ also satisfies (0.3) and so the assertion follows from Corollary 1.2.

§3 Conjecture $A, n = 3$

We may by Corollaries 1.3, 1.4 assume that M has a v–orthonormal basis $\{e_1, e_2, e_3\}$ such that

$$\delta \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix} = \begin{pmatrix} 0 & A & 0 \\ B_1 & B_2 & B_3 \\ C_1 & 0 & C_3 \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix} \tag{3.1}$$

where $A \in k, |A| > 1$. We may also assume

$$|A| > \mathrm{Sup}\ |B_i| \tag{3.2}$$

$$|A| > \sup_i |C_i|, \tag{3.2'}$$

$$\begin{aligned} |B_2 + C_3| &\leq 1 \\ |C_3| &\geq 1. \end{aligned} \tag{3.3}$$

Hence

$$\mathrm{Sup}\,(1, |B_2|) = |C_3| \tag{3.3'}$$

Proposition 3.4.

If $B_3 \neq 0$ then

$$|B_2C_3 - AB_1| \leq |C_3| \tag{3.4.1}$$

$$|C_3B_1 - C_1B_3| \leq \frac{1}{|A|}\operatorname{Sup}(1, |AB_1|) \tag{3.4.2}$$

Proof

$$e_2 = \frac{1}{A}\delta e_1 \tag{3.4.3.}$$

$$e_3 = \frac{1}{B_3}\big((\delta - B_2)e_2 - B_1e_1\big)$$

$$= \frac{1}{B_3}\Big[(\delta - B_2)\frac{1}{A}\delta - B_1\Big]e_1.$$

$$e_3 = \frac{1}{AB_3}\big[\delta^2 - B_2\delta - AB_1\big]e_1 \tag{3.4.4}$$

and so e_1 is annihilated by the cubic monic operator

$$Q = AB_3\Big\{(\delta - C_3)\frac{1}{AB_3}\big(\delta^2 - B_2\delta - AB_1\big) - C_1\Big\}$$

$$= -AC_1B_3 + B_3(\delta - C_3) \circ \frac{1}{B_3}(\delta^2 - B_2\delta - AB_1)$$

$$= \delta^3 + Q_1\delta^2 + Q_2\delta + Q_3$$

where

$$Q_1 = -(B_2 + C_3) + B_3D\left(\frac{1}{B_3}\right)$$

$$Q_2 = C_3B_2 - AB_1 - B_3\,D\left(\frac{B_2}{B_3}\right)$$

$$Q_3 = A(B_1C_3 - C_1B_3) - AB_3\,D\left(\frac{B_1}{B_3}\right).$$

By (1.2) $|Q_i| \leq 1$ for $1 = 1, 2, 3$. The proposition is now clear.

Proposition 3.5.

If $C_1 \neq 0$ then

$$|B_2C_3 - AB_1| \leq |C_3| \tag{3.5.1}$$

$$|B_1C_3 - C_1B_3| \leq \frac{1}{|A|}|C_3|^2 \tag{3.5.2}$$

Proof

$$e_1 = \frac{1}{C_1}(\delta - C_3)e_3 \tag{3.5.3}$$

$$e_2 = \frac{1}{A}\delta\, e_1 = \left(\frac{1}{A}\delta \circ \frac{1}{C_1}(\delta - C_3)\right)e_3 \tag{3.5.4}$$

and so

$$(\delta - B_2)e_2 - B_1e_1 - B_3e_3 = 0$$

implies that e_3 is annihilated by the monic cubic operator

$$\begin{aligned} F &= AC_1\left[(\delta - B_2) \circ \frac{1}{A} \circ \delta \circ \frac{1}{C_1}(\delta - C_3) - \frac{B_1}{C_1}(\delta - C_3) - B_3\right] \\ &= C_1(\delta - B_2) \circ \delta \circ \frac{1}{C_1}(\delta - C_3) - AB_1(\delta - C_3) - AC_1B_3 \end{aligned}$$

Thus

$$F = \delta^3 + F_1\delta^2 + F_2\delta + F_3$$

where

$$F_1 = -(C_3 + B_2) + 2C_1D\left(\frac{1}{C_1}\right)$$

$$F_2 = B_2C_3 - AB_1 + \left[-B_2C_1D\left(\frac{1}{C_1}\right) + C_1D^2\left(\frac{1}{C_1}\right) - 2C_1D\left(\frac{C_3}{C_1}\right)\right]$$

$$F_3 = A(B_1C_3 - C_1B_3) + \left[B_2C_1D\left(\frac{C_3}{C_1}\right) - C_1D^2\left(\frac{C_3}{C_1}\right)\right]$$

Again by (1.2) we know that $|F_i| \leq |, 1 = 1, 2, 3$ and the assertion is again clear.

3.6

If both B_3 and C_1 vanish then M is a direct sum and the conclusion of conjecture A follows from 1.5 (indeed with $\theta(v, v') \leq \mathrm{Sup}(1, |\delta|_v)$). Thus we may assume the B_3, C_1 not both zero.

Propostion.

If B_3, C_1 not both zero then

$$|B_2C_3 - AB_1| \leq |C_3| \tag{3.6.1}$$

$$|B_1C_3 - C_1B_3| \leq \frac{1}{|A|}|C_3|^2 \tag{3.6.2}$$

$$\text{If } |AB_1| > 1 \text{ then } |AB_1| = |B_2|^2 = |C_3|^2 \tag{3.6.3}$$

Proof

Equation (3.6.1) follows from (3.4.1), (3.5.1). If $|AB_1| > 1$, then $|B_2| = |C_3| > 1$ as otherwise $|B_2C_3|$ and $|C_3|$ are bounded by 1. Thus $|B_2C_3| = |C_3|^2 > |C_3|$ and so (3.6.3) follows from (3.6.1). So now (3.6.2) is trivial if $C_1 \neq 0$ while if $B_3 \neq 0$ the assertion follows from (3.4.2) and (3.6.3).

Proposition 3.7.

The conclusion of Conjecture A holds if either of the following two statements hold,

$$|B_3| \geq \frac{1}{|A|^2} \operatorname{Sup}(1, |B_2|)^3 \tag{3.7.1}$$

$$|C_1| \geq \frac{1}{|A|^2} \operatorname{Sup}(1, |B_2|)^3 \tag{3.7.2}$$

Proof

To verify (3.7.1) we may assume $B_3 \neq 0$ and from equations (3.4.3), (3.4.4)

$$\begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix} = H \begin{pmatrix} e_1 \\ \delta e_1 \\ \delta^2 e_1 \end{pmatrix}$$

$$H = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{A} & 0 \\ -\frac{AB_1}{AB_3} & -\frac{B_2}{AB_3} & \frac{1}{AB_3} \end{pmatrix}$$

$$H^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & A & 0 \\ AB_1 & AB_2 & AB_3 \end{pmatrix}$$

Thus using (3.6.3)

$$|H^{-1}| \leq |A| \operatorname{Sup}(1, |B_1|, |B_2|, |B_3|) \tag{3.7.3}$$

$$|H| \leq \operatorname{Sup}\left(1, \frac{1}{|B_3|\,|A|} \operatorname{Sup}(|C_3|, |AB_1|)\right) \tag{3.7.4}$$

$$\leq \operatorname{Sup}\left(1, \frac{1}{|B_3A|}|C_3|^2\right).$$

The condition $|H|\,|H^{-1}| \leq |A|^2$ certainly holds if the product of the right sides of (3.7.3) and (3.7.4) is bounded by $|A|^2$. The term 1 on the right side of (3.7.4) may be discarded since the right side of (3.7.3) is bounded by $|A|^2$. Likewise the term $|B_3|$ on the right side of (3.7.3) may be discarded since $|C_3| \leq |A|$. Thus it is enough if

$$|B_3|\,|A|^2 \geq |C_3|^2 \operatorname{Sup}\left(|C_3|, |B_1|\right).$$

Now by (3.6.3),

$$\operatorname{Sup}\left(1, |B_1|\right) \leq \operatorname{Sup}\left(1, |C_3|^2/|A|\right) \leq |C_3|$$

The assertion is now clear.

For the proof of (3.7.2) we may assume $C_1 \neq 0$ and from equations (3.5.3), (3.5.4).

$$\begin{pmatrix} e_3 \\ e_1 \\ e_2 \end{pmatrix} = H \begin{pmatrix} e_3 \\ \delta e_3 \\ \delta^2 e_3 \end{pmatrix}$$

where

$$H = \begin{pmatrix} 1 & 0 & 0 \\ -\frac{C_3}{C_1} & \frac{1}{C_1} & 0 \\ \frac{q_2}{AC_1} & \frac{q_1}{AC_1} & \frac{1}{AC_1} \end{pmatrix}$$

$$q_1 = -C_3 + C_1 D\left(\frac{1}{C_1}\right)$$

$$q_2 = -C_1\, D\left(\frac{C_3}{C_1}\right)$$

and so

$$H^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ C_3 & C_1 & 0 \\ -q_1 C_3 - q_2 & -q_1 C_1 & AC_1 \end{pmatrix}$$

Thus

$$|H| \leq \operatorname{Sup}\left(1, \left|\frac{C_3}{C_1}\right|\right) \tag{3.7.5}$$

$$|H^{-1}| \leq \operatorname{Sup}\left(\ |AC_1|, |C_3|^2\right) \tag{3.7.6}$$

since

$$|q_1 C_1| \leq |C_1 C_3| \leq |C_1 A|$$
$$-q_1 C_3 - q_2 = {C_3}^2 + D C_3$$

The condition that $|A|^2 \geq |H|\,|H^{-1}|$ is satisfied if $|A|^2$ dominates the product of the right sides of (3.7.5) and (3.7.6). The term involving $|C_1|$ on the right side of (3.7.6) may be discarded since $|C_3| \leq |A|$. We may also discard the term 1 on the right side of (3.7.5). The condition (3.7.2) is now seen to be sufficient.

Corollary 3.8.

If Conjecture A does not hold for (3.1) then

$$|B_3| < \frac{1}{|A|^2}\,|C_3|^3$$

$$|C_1| < \frac{1}{|A|^2}\,|C_3|^3 .$$

Proposition 3.9 (Strong estimates).

If Conjecture A does not hold for (3.1) then

$$|B_1 A| \leq 1 \tag{3.9.1}$$

$$\operatorname{Sup}\ \left(|B_2|, |C_3|\right) = 1 = |C_3| \tag{3.9.2}$$

$$\operatorname{Sup}\ \left(|C_1|, |B_3|\right) < 1/\ |A|^2 \tag{3.9.3}$$

Proof

If $|B_1A| \leq 1$ then by (3.6.1), Sup $(|B_2|, |C_3|) = 1$ since otherwise $|B_2C_3| = |C_3|^2 > |C_3|$ which contradicts the hypothesis. Thus (3.9.2) would hold and (3.9.3) would follow from 3.8.

If $|AB_1| > 1$ then by (3.6.3),

$$|AB_1| = |C_3|^2 \quad > \quad |C_3|$$

and hence

$$|C_3B_1| > |C_3^2|/|A| = |B_2^2|/|A|$$

and hence by (3.6.2)

$$|B_3C_1| = |C_3B_1| = |A|^{\frac{1}{2}}|B_1|^{\frac{3}{2}} \tag{3.9.4}$$

while by (3.8)

$$|C_1| < \frac{1}{|A|^2}|AB_1|^{\frac{3}{2}} = |B_1|^{\frac{3}{2}}/|A|^{\frac{1}{2}}.$$

Substituting in (3.9.4) shows

$$|B_3| = \frac{1}{C_1}|A|^{\frac{1}{2}}|B_1|^{\frac{3}{2}} > |A|$$

a contradiction. This completes the proof.

3.10 Reduction to $p = 2$.

We assume Conjecture A does not hold for (3.1). We construct a cyclic element $z = e_1 + ye_3$ where $y \in O_E$, the closed unit ball in E. Let v be the associated norm of cyclic type. By (1.2), $A1$ holds. Hence $A2$ does not hold for v. We shall use $\mathfrak{p}$ to refer to the open unit ball in E. Thus

$$\begin{pmatrix} z \\ \delta z \\ \delta^2 z \end{pmatrix} = H \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix}$$

where

$$H = \begin{pmatrix} 1 & 0 & y \\ yC_1 & A & y' + C_3 y \\ q_1 & q_2 & q_3 \end{pmatrix}$$

$$q_1 = 2C_1 y' + (C_1' + C_1 C_3) y + AB_1$$
$$q_2 = AB_2 + AC_1 y$$
$$q_3 = y'' + 2C_3 y' + (C_3' + C_3^2) y + AB_3 \ .$$

By 3.9,

(3.10.1) $$|q_1| \leq 1, \ |q_2| \leq |A|, \ |q_3| \leq 1.$$

We conclude that

(3.10.2) $$|H| = |A| = |\text{adj } H|$$

and since $|H|\,|H^{-1}| > |A|^2$,

(3.10.3) $$|\det H| < 1$$

for all $y \in O_E$. To compute $\det H \bmod \mathfrak{p}$ we may first reduce $H \bmod \frac{1}{A}\mathfrak{p}$ since all coefficients of H with two exceptions are bounded by 1, the two exceptions are bounded by $|A|$ and the product of the two exceptions does not appear in the calculation of the determinant. The terms involving C_1, B_3 may thus be dropped and we find modulo $\frac{1}{A}\mathfrak{p}$.

$$H \equiv H_1 = \begin{pmatrix} 1 & 0 & y \\ 0 & A & y' + C_3 y \\ AB_1 & AB_2 & \overline{q}_3 \end{pmatrix}$$

where

$$\overline{q}_3 = y'' + 2C_3 y' + (C_3' + C_3^2) y.$$

Thus

$$\det H \equiv \det H_1 = A\, y'' + q_4\, y' + q_5 y \qquad \bmod \mathfrak{p}$$

where

$$q_4 = \begin{vmatrix} 1 & 0 & 0 \\ 0 & A & 1 \\ AB_1 & AB_2 & 2C_3 \end{vmatrix}$$

$$q_5 = \begin{vmatrix} 1 & 0 & 1 \\ 0 & A & C_3 \\ AB_1 & AB_2 & C_3' + C_3^2 \end{vmatrix}.$$

Thus for all $y \in O_E$,

(3.10.4) $$Ay'' + q_4 y' + q_5 y \in \mathfrak{p} .$$

Thus putting $y = 1, x, x^2$

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & x \\ 2 & 2x & x^2 \end{pmatrix} \begin{pmatrix} A \\ q_4 \\ q_5 \end{pmatrix} \in \mathfrak{p}$$

and so

(3.10.5) $$q_5, q_4, 2A \in \mathfrak{p} .$$

Thus

(3.10.6) $$p = 2, \quad 1 < |A| < 2 .$$

Since $q_4 = 2AC_3 - AB_2$ and $2AC_3 \in \mathfrak{p}$ we conclude

(3.10.7) $$AB_2 \in \mathfrak{p}.$$

Thus

$$q_5 \equiv \begin{vmatrix} 1 & 0 & 1 \\ 0 & A & C_3 \\ AB_1 & 0 & C_3' + C_3^2 \end{vmatrix} \mod \mathfrak{p}$$

which by (3.10.5) gives

$$A^2B_1 \equiv A(C_3' + C_3^2) \mod \mathfrak{p} . \tag{3.10.8}$$

We now show that (3.1) is not a counter example. Recall that by 3.9.2, C_3 is a unit in E. We put $y = A/C_3$ and transform G by

$$H = \begin{pmatrix} 1 & y & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Trivially $|H|\,|H^{-1}| = |A|^2$. We assert that

$$|G_{[H]}| \leq 1.$$

The explicit calculation gives

$$G_{[H]} = \begin{pmatrix} yB_1 & A + yB_2 + y' - B_1y^2 & B_3y \\ B_1 & -B_1y + B_2 & B_3 \\ C_1 & C_1y & C_3 \end{pmatrix} .$$

By (3.9) each coefficient is bounded by unity except possibly for $A + yB_2 + y' - B_1y^2$. By (3.10.7) we may discard $y\,B_2$. We now compute using (3.10.5),

$$\begin{aligned} C_3^2(y' + A - B_1y^2) &\equiv (C_3^2y)' + AC_3^2 - B_1(C_3y)^2 \mod \mathfrak{p} \\ &= (AC_3)' + A\,C_3^2 - B_1A^2 \\ &= A(C_3' + C_3^2) - B_1A^2 \end{aligned}$$

and by 3.10.8 this term is bounded by unity which completes the proof.

§4 Spectral norms not bounded by unity.

The radius of convergence has a well known relation with the spectral norm of δ.

4.0 Let $\beta \geq |\pi|$. The differential equation (0.1′) has a solution matrix at t converging in $D(t, (|\pi|/\beta)^-)$ if and only if δ has spectral norm bounded by β.

Proposition 4.1.

Let z be a cyclic vector of M. Then subject to (4.0) we have

$$\delta^n z = C_1\delta^{n-1}z + \cdots + C_nz, \quad C_j \in E$$

and

$$|C_j| \leq \mathrm{Sup}(1, \beta)^j \qquad (1 \leq j \leq n).$$

Proof

By Corollary 1.2 we may assume $\beta > 1$. Suppose the assertion false. Then there exists $\beta' > \beta$ such that $\sup |C_j|/\beta'^j = 1$. Replacing k by a finite extension field there exists $\alpha \in k$ such that $|\alpha| = \beta'$. Following the proof of Corollary 1.2, let v be the norm of M given by the basis $\left\{\left(\frac{1}{\alpha}\delta\right)^j z\right\}_{0 \leq j \leq n-1}$ and let the matrix G of that proof represent $\frac{1}{\alpha}\delta$. Thus $|\delta|_v = |\alpha| > 1$. Since β' strictly exceeds the spectral norm of δ, we know that $|(\alpha^{-1}\delta)^s|_v = (\beta')^{-s}|\delta^s|_v \to 0$. But $\tau = \frac{1}{\alpha}\delta$ is stable on M_v and hence $\overline{\tau}$ is an endomorphism of $\overline{M}_v$ which we now see must be nilpotent. As in the proof of Corollary 1.2 this is contradicted by the characteristic polynomial of $\overline{G}$.

This result may be used to simplify and slightly improve the results of our previous article, [Chr-Dw].

We recall some definitions and notations of that article, k is complete and algebraically closed,

$$E_0 = \{\xi \in E | \xi \text{ is analytic on } D(0, 1^-)$$
$$E_0' \text{ is the quotient field of } E_0$$
$$\delta = x\,\frac{d}{dx}$$
$$G_{[H]} = \delta H \cdot H^{-1} + HGH^{-1}$$

We consider operators, $\delta - G, G \in \mathcal{M}_n(E)$. $U_{G,t}$ denotes the solution matrix at the generic point, $r(U_{G,t})$ is the radius of convergence of $U_{G,t}$.

We consider certain conditions:

$R1$: $G \in M_n(E_0)$

$R3$: $G(0)$ nilpotent.

Under hypothesis $R3$, the equation $\delta - G$ has a solution matrix $Y_G x^{G(0)}$, at $x = 0$, where $Y_G \in G\ell(n, k[[x]]), Y_G(0) = I_n$. We use $r(Y_G)$ to represent the radius of convergence of Y_G.

Proposition 4.2 [Chr-Dw, Prop. 7.1].

Let $G \in \mathcal{M}_n(E_0'), r(U_{G,t}) = R \in (0, 1)$ Then there exists $H \in G\ell(n, E_0')$ such that

$$\left|G_{[H]}\right| \leq \mathrm{Sup}\left(1, \frac{|\pi|}{R}\right)$$

Also if G is analytic at the origin then the same holds for $G_{[H]}$.

Proof

We choose H such that $G_{[H]}$ has the form of a cyclic element, i.e., by Proposition 4.1

$$G_{[H]} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 1 \\ C_n & C_{n_1} & & \cdots & C_1 \end{pmatrix}$$

with $|C_j| \leq \mathrm{Sup}\left(1, \frac{|\pi|}{R}\right)^j$. We may assume $\mathrm{Sup}\,|C_j| > 1$.

We choose $\alpha \in k$ such that $\sup |C_j/\alpha^j| = 1$. Thus $|\alpha| \leq \mathrm{Sup}\left(1, \frac{|\pi|}{R}\right)$. Let H_1^{-1} be the diagonal matrix

$$H_1^{-1} = \begin{pmatrix} 1 & & & \\ & \alpha & & 0 \\ & & \ddots & \\ 0 & & & \alpha^{n-1} \end{pmatrix}.$$

We find

$$G_{[H_1 H]} = \begin{pmatrix} 0 & \alpha & 0 & \cdots & 0 \\ 0 & 0 & \alpha & \cdots & 0 \\ 0 & 0 & 0 & \cdots & \alpha \\ \frac{C_n}{\alpha^{n-1}} & & & & C_1 \end{pmatrix}$$

and so $\left|G_{[H_1 H]}\right| = |\alpha| \leq \mathrm{Sup}\left(1, \frac{|\pi|}{R}\right)$.

Proposition 4.3 ([Chr-Dw Prop. 7.2]).

Let G satisfy $R1$ and $R3$. Let $r(U_{G,t}) \geq R \in (0,1)$. Then

$$r(Y_G) \geq |\pi|^{n^2} \Big/ \mathrm{Sup}\left(1, \frac{|\pi|}{R}\right)^n.$$

Proof

We follow the proof of the reference. Thus equation (7.2.1) of the reference may by Proposition 4.2 be replaced by

$$G_{[H]} \leq \mathrm{Sup}\left(1, \frac{|\pi|}{R}\right)$$

Let $\epsilon > 0$. We choose $\beta \in k$ such that

$$(1+\epsilon)\,\mathrm{Sup}\left(1, \frac{|\pi|}{R}\right) > |\beta| > \mathrm{Sup}\left(1, \frac{|\pi|}{R}\right).$$

Let Δ be the diagonal matrix of the reference. The T as defined there satisfies

$$|T(0)| < 1$$

$$|T| \leq (1+\epsilon)^n \operatorname{Sup}\left(1, \frac{|\pi|}{R}\right)^n.$$

Each coefficient, T_i in the expansion at zero of T satisfies the same inequality and hence by induction using equation 7.2.7 of the reference,

$$|Y_s| \leq \left(\operatorname{Sup}\left(1, \frac{|\pi|}{R}\right)\right)^{ns} (1+\epsilon)^{ns} \Big/ |s!|^{n^2}.$$

We conclude that Y_T converges for

$$|x| < |\pi|^{n^2} \Big/ \left((1+\epsilon)\operatorname{Sup}\left(1, \frac{|\pi|}{R}\right)\right)^n.$$

The assertion follows by letting $\epsilon \to 0$.

Proposition 4.4 (cf [Chr-Dw, Theorem 8.3]).

Let G satisfy R1 and R3 and let $U_{G,t}$ converge in the closed disk $D(t,R)$ where $1 > R \geq |\pi|$. We choose $q = p^{\ell+1}$ such that $\ell \geq 0, |\pi|^{1/q} > R \geq |\pi|^{p/q}$. Then

$$r(Y_G) \geq |\pi|^{(n^2-n)/q} R^n.$$

Proof

We follow the proof in the reference but on the basis of our Proposition 4.3 we may replace (8.3.10) of reference by

$$r(Y_T) \geq |\pi|^{n^2} \Big/ \operatorname{Sup}\left(1, \frac{|\pi|}{R^q}\right)^n$$

The assertion now follows from equation 8.3.8 of the reference.

Remark

This proposition is a slight improvement of a theorem of Y.André [A'].

REFERENCES

[A] André, Y. *G-functions and Geometry*, Vieweg (1989).

[*A'*] _____ *Spécilisation dans les disques singuliers réguliers*, GEAU Publ. Math de l' Université Paris VII **29** (1985/86), 1-14.

[Chr1] Christol, G., *Systèms différentiels, linéaires p-adiques: structure de Frobenius faible*, Bull. Soc. Math France **109** (1981), 83-122.

[Chr2] _____ *Modules différentiels p-adiques et équations différentiels p-adiques*, Queen's Papers in Pure and Applied Math **66** (1984), Queen's University, Kingston, Ontario.

[Chr-Dw] _____ *Effective p-adic bounds, at regular singular points*, Duke J. Math., **62** (1991), 689-720.

Contemporary Mathematics
Volume 133, 1992

The p-adic monodromy of a generic abelian scheme in characteristic p

RICHARD CREW

ABSTRACT. If A/S is a principally polarized abelian scheme with a principal level n structure, $n \geq 3$, then its relative rigid cohomology is an overconvergent isocrystal. We compute the monodromy group of this isocrystal, and of the corresponding convergent isocrystal if A/S is "generic" in a suitable sense. The "overconvergent" monodromy group is a symplectic group just as in the classical case; the "convergent" monodromy group is a symplectic group unless all fibers are ordinary, in which case the group is a maximal parabolic subgroup of the symplectic group.

Introduction

Let X/k be a smooth variety defined over a perfect field of characteristic $p > 0$. In [1], Berthelot introduced the categories of *convergent* and *overconvergent* isocrystals on X, in the hope of finding a suitable p-adic analogue of the category of ℓ-adic sheaves on X. In [8] we studied a kind of monodromy group attached to a convergent or overconvergent isocrystal, which one hopes will play the same role that the geometric monodromy group does for an ℓ-adic local system. Consider for example a proper smooth morphism $f : Y \to X$, where X/k is as above. Berthelot conjectures that the relative rigid cohomology $R^i f_{rig*}(Y/X)$ has the structure of an overconvergent isocrystal on X/K [1]. If so, then one can ask if the overconvergent monodromy group of $R^i f_{rig*}(Y/X)$ bears any resemblence to the geometric monodromy group of (any of) the ℓ-adic representations $R^i f_* \mathbf{Q}_\ell$ for any $\ell \neq p$. Of course one does not even know know if the geometry monodromy groups of the $R^i f_* \mathbf{Q}_\ell$ for variable ℓ are in any sense the same (although Larsen

1991 *Mathematics Subject Classification.* 14F30, 14G15.
Key words and phrases. abelian schemes, monodromy, p-adic cohomology.
Supported in part by the NSF and the NSA
This paper is in final form and no version of it will be submitted for publication elsewhere.

0271-4132/92 $1.00 + $.25 per page

and Pink [12] have some results in this direction), but if they are, we can ask if the overconvergent monodromy group is "the same" as the ℓ-adic ones. The current state of our understanding of relative rigid cohomology is such that the techniques of [12] are a long way from being applicable here, but one can at least calculate a few examples. In [8] we showed that for a generic family of elliptic curves, the overconvergent monodromy group is $SL(2)$, just as in the ℓ-adic case, while the convergent monodromy group is a Borel subgroup of $SL(2)$ if all fibers are ordinary, and $SL(2)$ otherwise. Isocrystals related to Kloosterman sums were studied in [9]; again the overconvergent group has the same size and shape as the ℓ-adic ones.

In this paper we will show that the relative one-dimensional rigid cohomology of a suitably generic principally polarized abelian scheme with a principal level structure has the structure of an overconvergent isocrystal, and we will compute its monodromy group both as an overconvergent and as a convergent isocrystal. In the overconvergent category the monodromy group is the full symplectic group, just as in the ℓ-adic case, while in the convergent category, the result depends on whether or not all fibers are ordinary. As in the elliptic curve case, a key point is to show that the monodromy representation on the étale quotient of the Tate module is as large as possible, which in the essential case was done by Chai and Faltings [5]. The higher-dimensional case involves some new complications, mainly because the analogue for F-isocrystals of Grothendieck's monodromy theorem is not known to be valid over a smooth base of dimension greater than one.

§1 Isocrystals, monodromy groups, rigid cohomology

1.1 Isocrystals. Let k be a perfect field of characteristic $p > 0$, R a complete discrete ring with residue field k and fraction field K of characteristic 0. If X/k is a separated scheme of finite type, then [1] and [2] give a construction of the categories $\mathrm{Isoc}(X/K)$, $\mathrm{Isoc}^\dagger(X/K)$ of *convergent* and *overconvergent isocrystals* on X/K. This is not the place to give even a summary of the construction, but if X/k is smooth and has a formally smooth lifting $\mathfrak{X}/R$, then a convergent isocrystal on X/K can be thought of as a locally free sheaf $\mathcal{M}$ on the rigid-analytic space $\mathfrak{X}^{an}$ endowed with a *convergent connection* ∇ in the sense of [1] 4.1. More generally, if X has an embedding $X \hookrightarrow \mathcal{P}$ into a formally smooth formal R-scheme $\mathcal{P}$, then a convergent isocrystal M on X/K can be identified with a locally free sheaf $\mathcal{M}_\mathcal{P}$ with a convergent connection on a suitable "tubular neighborhood" of X in $\mathcal{P}$; $\mathcal{M}_\mathcal{P}$ is called the *realization* of M on $\mathcal{P}$. An overconvergent isocrystal on X/K is a convergent isocrystal on X/K satisfying supplementary conditions "at infinity". For details the reader may consult [1].

The categories $\mathrm{Isoc}(X/K)$, $\mathrm{Isoc}^\dagger(X/K)$ are of local nature on X; i.e. isocrystals and morphisms of isocrystals (convergent or overconvergent) can be patched together from isocrystals on a Zariski-open cover of X. Furthermore, they are functorial in the following sense: if X'/k', X/k are separated and of finite type,

if $f : X' \to X$ is a morphism over $\mathrm{Spec}(k') \to \mathrm{Spec}(k)$, and if R', R have residue fields k', k and fraction fields K', K, then there are functors

$$\begin{aligned} f^* &: \mathrm{Isoc}(X'/K) \longrightarrow \mathrm{Isoc}(X/K') \\ f^* &: \mathrm{Isoc}^\dagger(X'/K) \longrightarrow \mathrm{Isoc}^\dagger(X/K'). \end{aligned} \tag{1.1.1}$$

If $X' = U$ is an open subscheme and $k' = k$, $K' = K$, we will write $M \mapsto M|U$ for the functor 1.1.1; if $X' = X$, then 1.1.1 becomes an "extension of scalars" which we write $M \mapsto M \otimes K'$. If $x \to X$ is a point with values in a perfect field, then we will denote 1.1.1 by $M \mapsto M_x$. Finally, if $F : X \to X$ is the absolute Frobenius morphism and σ is a lifting of the Frobenius on k to K, then F^* is a σ-linear functor from $\mathrm{Isoc}(X/K)$ or $\mathrm{Isoc}^\dagger(X/K)$ to itself. If M is a (convergent or overconvergent) isocrystal, then an isomorphism $\Phi : F^*M \simeq M$ is called a *Frobenius structure* on M, and an isocrystal endowed with a Frobenius structure is an *F-isocrystal*; the category of convergent (resp. overconvergent) isocrystals on X/K is denoted $F\text{-Isoc}(X/K)$ (resp. $F\text{-Isoc}^\dagger(X/K)$).

The natural forgetful functors

$$\begin{aligned} \mathrm{Isoc}^\dagger(X/K) &\longrightarrow \mathrm{Isoc}(X/K) \\ F\text{-Isoc}^\dagger(X/K) &\longrightarrow F\text{-Isoc}(X/K) \end{aligned} \tag{1.1.2}$$

are faithful, and are expected, but not known, to be fully faithful. We will nonetheless allow ourselves the abuse of saying that a convergent isocrystal M on X/K "is overconvergent" to mean that M is in the image of 1.1.1; it could however be the case that M is the image of several non-isomorphic objects of $\mathrm{Isoc}^\dagger(X/K)$. By the same token, if M is an overconvergent isocrystal, then we can not assert in our present state of knowledge that a Frobenius structure on $\hat{M}$ arises from a Frobenius structure on M. For M an object of $\mathrm{Isoc}^\dagger(X/K)$ or $F\text{-Isoc}^\dagger(X/K)$, we will write $M \mapsto \hat{M}$ for the functor 1.1.2

If $F\text{-Cris}(X/W)$ denotes the category of F-crystals (of $\mathcal{O}_{X/W}$-modules) on X, then there is a natural functor

$$\begin{aligned} F\text{-Cris}(X/W) &\to F\text{-Isoc}(X/K) \\ M &\mapsto M \otimes \mathbf{Q} \end{aligned} \tag{1.1.3}$$

(c.f. [2] 2.3.11) which is fully faithful up to isogeny; i.e. the functor induces an isomorphism

$$\mathrm{Hom}_{F-Cris}(M, N) \otimes K_0 \xrightarrow{\sim} \mathrm{Hom}_{F-Isoc}(M \otimes \mathbf{Q}, N \otimes \mathbf{Q}).$$

where K_0 denotes the fixed field of σ (*loc. cit.* 2.3.12). If X is smooth, 1.1.3 is essentially surjective up to Tate twists; furthermore 1.1.3 is fully faithful on the level of isocrystals, in the sense that

$$\mathrm{Hom}_{Cris}(M, N) \otimes K \xrightarrow{\sim} \mathrm{Hom}_{Isoc}(M \otimes \mathbf{Q}, N \otimes \mathbf{Q}). \tag{1.1.4}$$

Finally, we note that the F-isocrystals in the image of 1.1.3 are not necessarily overconvergent.

If $X = \mathrm{Spec}(k)$ is a point, then the categories $\mathrm{Isoc}(X/K)$ and $\mathrm{Isoc}^\dagger(X/K)$ are equivalent to the category Vec_K of K-vector spaces. Similarly, the categories $F\text{-Isoc}(X/K)$ and $F\text{-Isoc}^\dagger(X/K)$ are equivalent to the category of K-vector spaces with a σ-linear automorphism, i.e. the category of what are usually called F-isocrystals on K. If σ has been chosen so as to fix a uniformizer of K (and we will assume this to be the case from now on), then one can define in the usual way the Newton polygon of an F-isocrystal on K. If, once again, X/k is separated of finite type and $x \to X$ is a point with values in a perfect field, then the Newton polygon of an F-isocrystal (M, Φ) at x is the Newton polygon of the pullback (M_x, Φ_x).

A convergent or overconvergent F-isocrystal (M, Φ) is a *unit-root* F-isocrystal if for each closed point $x \to X$, the fiber (M_x, Φ_x) is unit-root in Manin's sense (c.f. [7] §1.9). If we denote by K_0 the fixed field of σ, and by $\mathrm{UR}(X/K)$ the category of convergent unit-root F-isocrystals on X/K, then there is a natural equivalence of categories ([7], 2.1)

$$G : \mathrm{Rep}_{K_0}^{contin}(\pi_1(X, x)) \xrightarrow{\sim} \mathrm{UR}(X/K) \tag{1.1.5}$$

where $\mathrm{Rep}_{K_0}^{contin}(\pi_1(X, x))$ is the category of continuous representations of π_1 on finite-dimensional K_0-vector spaces. One would like to be able to identify the pre-image under 1.1.5 of the overconvergent unit-root F-isocrystals, i.e. the objects of $\mathrm{UR}(X/K)$ that are in the essential image of 1.1.2. In general very little is known; if X/k is a smooth curve, then $G(\rho)$ is overconvergent for any ρ with finite local monodromy at infinity [7], 3.1, and any overconvergent unit-root F-isocrystal of rank one arises in this way [7], 4.12.

1.2 Monodromy groups. There is a natural $\otimes$-product for isocrystals which makes $\mathrm{Isoc}(X/K)$ and $\mathrm{Isoc}^\dagger(X/K)$ into rigid K-linear $\otimes$-categories in the sense of Saavedra [14]. If $x \to X$ is any point of X with values in a perfect field, and L plays the role of K' in 1.1.1, and if we identify $\mathrm{Isoc}(x/L)$, $\mathrm{Isoc}^\dagger(x/L)$ with Vec_L, then the pullback functors are

$$\begin{aligned} \omega_x &: \mathrm{Isoc}(X/K) \longrightarrow \mathrm{Vec}_L \\ \omega_x &: \mathrm{Isoc}^\dagger(X/K) \longrightarrow \mathrm{Vec}_L \\ M &\mapsto M_x. \end{aligned} \tag{1.2.1}$$

These are faithful, exact, and compatible with tensor products; i.e. $\otimes$-functors in the sense of [14]. In particular, if X has a k-rational point x, then we can take $L = K$; then the functors 1.2.1 are fiber functors in the sense of Saavedra's theory, and $\mathrm{Isoc}(X/K)$, $\mathrm{Isoc}^\dagger(X/K)$ are neutral Tannakian categories.

Assume that X is geometrically connected and has a k-rational point x, so that $\mathrm{Isoc}(X/K)$, $\mathrm{Isoc}^\dagger(X/K)$ are K-linear neutral Tannakian categories. From Saavedra's theory it follows that $\mathrm{Isoc}(X/K)$ and $\mathrm{Isoc}^\dagger(X/K)$ are equivalent as $\otimes$-categories to categories of representations of an affine K-group. For our purposes, it is convenient to work with slightly smaller categories, so we define

$\mathrm{Isoc}(X/K)_F$ (resp. $\mathrm{Isoc}^\dagger(X/K)_F$) to be the full $\otimes$-subcategory of $\mathrm{Isoc}(X/K)$ resp. $\mathrm{Isoc}^\dagger(X/K)$ generated by the image of the forgetful functor

$$F\text{-}\mathrm{Isoc}(X/K) \longrightarrow \mathrm{Isoc}(X/K)$$
$$\text{resp.} \qquad F\text{-}\mathrm{Isoc}^\dagger(X/K) \longrightarrow \mathrm{Isoc}^\dagger(X/K).$$

Thus an object of $\mathrm{Isoc}(X/K)$ is in $\mathrm{Isoc}(X/K)_F$ if it is a subquotient of a convergent isocrystal with a Frobenius structure. We then define

$$\begin{aligned} \pi_1^{CF}(X,x) &= \mathrm{Aut}\,\omega_x \,|\, \mathrm{Isoc}(X/K)_F \\ \pi_1^{OCF}(X,x) &= \mathrm{Aut}\,\omega_x \,|\, \mathrm{Isoc}^\dagger(X/K)_F \end{aligned} \tag{1.2.2}$$

and there are then equivalences of $\otimes$-categories

$$\begin{aligned} \mathrm{Isoc}(X/K)_F &\xrightarrow{\sim} \mathrm{Rep}_K(\pi_1^{CF}(X,x)) \\ \mathrm{Isoc}^\dagger(X/K)_F &\xrightarrow{\sim} \mathrm{Rep}_K(\pi_1^{OCF}(X,x)) \end{aligned} \tag{1.2.3}$$

If M is an object of $\mathrm{Isoc}(X/K)$ or $\mathrm{Isoc}^\dagger(X/K)$, then we let $[M]$ be the full $\otimes$-subcategory generated by M; note that if M is in $\mathrm{Isoc}(X/K)_F$, then $[M]$ is a full subcategory of $\mathrm{Isoc}(X/K)_F$, and similarly in the overconvergent case. For M in $\mathrm{Isoc}(X/K)$ or $\mathrm{Isoc}^\dagger(X/K)$ we define the *monodromy group* of M to be

$$DGal(M,x) = \mathrm{Aut}\,\omega_x \,|\, [M]$$

and then 1.2.3 restricts to

$$[M] \xrightarrow{\sim} \mathrm{Rep}_K(DGal(M,x))$$

We shall omit the reference to x if there is no chance of confusion; in any case the groups $DGal(M,x)$, $\pi_1^{CF}(X,x)$, $\pi_1^{OCF}(X,x)$ are independent of the choice of "basepoint" x up to non-canonical isomorphism. For M in $\mathrm{Isoc}(X/K)_F$, the inclusion functor $[M] \hookrightarrow \mathrm{Isoc}(X/K)_F$ induces a surjective homomorphism $\pi_1^{CF} \to DGal(M)$; similarly, for M in $\mathrm{Isoc}^\dagger(X/K)$ we get $\pi_1^{OCF} \to DGal(M)$. If M corresponds via 1.2.3 to a representation ρ of π_1^{CF}, then $DGal(M)$ is isomorphic to the image of ρ.

Note that for M in $\mathrm{Isoc}^\dagger(X/K)$ there are *two* monodromy groups that can be attached to M, for in addition to $DGal(M)$, there is also the monodromy group $DGal(\hat{M})$ of the image of M under the forgetful functor 1.1.2. We could call $DGal(M)$ and $DGal(\hat{M})$ the "convergent" and "overconvergent" monodromy groups of M. By [8], there is a natural inclusion

$$DGal(\hat{M}) \hookrightarrow DGal(M) \tag{1.2.4}$$

which is not, in general, an isomorphism.

1.3. The monodromy groups just defined obey a formalism similar to that of fundamental groups or differential galois groups; we shall recall some facts from [8] that we will need later. For the rest of this section all schemes are assumed to be geometrically connected.

Let X'/k', X/k be separated schemes of finite type, $f : X' \to X$ a morphism, and let K', K be as in the paragraph before 1.1.1. Suppose that x' is a k'-point of X' mapping under f to a k-point of X; then the functors f^* in 1.1.1 induce homomorphisms

$$\begin{aligned} \pi_1^{CF}(X', x') &\longrightarrow \pi_1^{CF}(X, x) \otimes K' \\ \pi_1^{OCF}(X', x') &\longrightarrow \pi_1^{OCF}(X, x) \otimes K' \end{aligned} \tag{1.3.1}$$

and if M is an object of $\mathrm{Isoc}(X/K)_F$ or $\mathrm{Isoc}^\dagger(X/K)_F$, the restriction of 1.3.1 to $[M]$ induces a closed immersion

$$DGal(f^*M) \hookrightarrow DGal(M) \otimes K'. \tag{1.3.2}$$

If L/K is a finite extension, then 1.3.2 becomes an isomorphism

$$DGal(M \otimes K') \xrightarrow{\sim} DGal(M) \otimes K' \tag{1.3.3}$$

If $k' = k$, $K' = K$, and $X' \to X$ is finite étale and galois with group G, then 1.3.2 can be refined as follows: there is an exact sequence

$$0 \longrightarrow \pi_1^{CF}(X') \longrightarrow \pi_1^{CF}(X) \longrightarrow G \longrightarrow 0 \tag{1.3.4}$$

and, if X is a smooth curve

$$0 \longrightarrow \pi_1^{OCF}(X') \longrightarrow \pi_1^{OCF}(X) \longrightarrow G \longrightarrow 0. \tag{1.3.5}$$

Furthermore, if X is smooth (resp. a smooth curve), the group of connected components of $\pi_1^{CF}(X)$ (resp. $\pi_1^{OCF}(X)$) is the usual fundamental group; i.e. there is a natural isomorphism

$$\pi_0(\pi_1^{CF}(X)) \xrightarrow{\sim} \pi_1(X) \qquad \text{resp.} \qquad \pi_0(\pi_1^{OCF}(X)) \xrightarrow{\sim} \pi_1(X). \tag{1.3.6}$$

One can deduce from 1.3.4–6 that if X is smooth and M is an object of $\mathrm{Isoc}(X/K)_F$ (resp. X is a smooth curve and M is an object of $\mathrm{Isoc}^\dagger(X/K)_F$), then there is a finite étale conver $f : X' \to X$ such that $DGal(f^*M)$ is the connected component of $DGal(M)$. The reason for using $\mathrm{Isoc}^\dagger(X/K)_F$ in place of $\mathrm{Isoc}^\dagger(X/K)$, incidentally, is to be able to prove statements like 1.3.5

If (M, Φ) is a convergent unit-root F-isocrystal which corresponds, via 1.1.5, to a representation ρ of $\pi_1(X)$, then one can ask if $DGal(M)$ can be computed in terms of ρ. At the moment this is possible if one is willing to pass to an extension field of K; if x is a *generic* point of X, and K' is as in 1.1.1 (i.e. an extension of K containing the fraction field of the Witt vectors of perfection of $k(X)$), then

$$DGal(M, x) \xrightarrow{\sim} \overline{\mathrm{Im}\ \rho|\pi_1(X \otimes k^{alg}, x)} \otimes K' \tag{1.3.7}$$

where the right hand side is the Zariski-closure of the image of the geometric fundamental group under ρ. One cannot conclude that 1.3.7 holds for any point x with values in a perfect field; for example, if x is a k-point of X, then the

most one can deduce is that $DGal(M,x)$ and $\overline{\text{Im } \rho|\pi_1(X \otimes k^{alg}, x)} \otimes K$ are inner twists of each other.

The main result of [8], which will be of key importance here, is the following version of Grothendieck's global monodromy theorem:

1.4 THEOREM. *Suppose that k is the perfection of an absolutely finitely generated field, and that X/k is a smooth curve. If M is an overconvergent isocrystal on X/K with a Frobenius structure, then the radical of the connected component of $DGal(M)$ is unipotent.*

Some kind of finiteness hypothesis on k is necessary. The most important case is when k is finite, but we will have to make use of larger fields in the proof of theorem 2.7. Although one would think that 1.4 should hold for any separated X/k of finite type, this seems unattainable at the moment, and the lack of a general result will complicate the proof of 2.7.

1.5 Rigid cohomology. One of the original motivations for constructing the categories of convergent and overconvergent isocrystals was to study the relative crystalline cohomology of a proper smooth morphism "up to isogeny." In [1] and [13] it is explained how to associate, to any smooth proper morphism $f : X \to S$, a set of convergent F-isocrystals $R^i f_{rig*}(X/S)$, which in the case $S = \text{Spec}(k)$ are just the crystalline cohomology groups tensored with K. More generally, if the relative crystalline cohomology $R^i f_{cris*}(\mathcal{O}_{X/W})$ is locally free (and therefore an F-crystal), then the F-isocrystal corresponding to it via 1.1.3 is the "relative rigid cohomology" $R^i f_{rig*}(X/S)$. Berthelot conjectures that if X/S is smooth and proper, then $R^i f_{rig*}(X/S)$ has the structure of an overconvergent F-isocrystal; at the moment this is only known if X/S is liftable in some sense. One case of this result will be useful to us: if X/S has a smooth lifting $\mathbb{X}/\mathbb{S}$ with $\mathbb{X}$, $\mathbb{S}$ flat R-schemes, then $R^i f_{rig*}(X/S)$ has the structure of an overconvergent F-isocrystal (c.f. the remark after Theorem 5 of [1]). In fact, if we denote by X^{an}/S^{an} the morphism of rigid analytic spaces corresponding to the extension of scalars to K of $\mathbb{X}/\mathbb{S}$, and by $\mathfrak{S}$ the formal completion of $\mathbb{S}$, then S^{an} is a "strict neighborhood" of the rigid-analytic space $\mathfrak{S}^{an}$ in the sense of [1] Remark 2.5 (b), and a realization on S^{an} of $R^i f_{rig*}(X/S)$ is furnished by the relative De Rham cohomology of X^{an}/S^{an} (which is just the analytification of the relative De Rham cohomology of $\mathbb{X}/\mathbb{S}$).

In the absolute case ($S = \text{Spec}(k)$) Berthelot constructs, for any X/k separated of finite type, a rigid cohomology theory $H^i_{rig}(X/K)$ generalizing the constructions of Washnitzer-Monsky, as well as generalizations of the latter with coefficients in an overconvergent isocrystal. Very little can be proven about these constructions at the moment, and we will not discuss them further.

§2 Abelian schemes

2.1 De Rham and crystalline cohomology of an abelian scheme. We will need to recall a few basic facts about the De Rham and crystalline cohomol-

ogy of an abelian scheme. If $f : A \to S$ is an abelian scheme, then the relative De Rham cohomology

$$R^{\cdot} f_{DR*}(A/S) = \mathbf{R}^{\cdot} f_*(\Omega^{\cdot}_{A/S})$$

is locally free on S and its formation commutes with arbitrary base change. If S is of characteristic $p > 0$, then the same is true of the relative crystalline cohomology $R^{\cdot} f_{cris*}(\mathcal{O}_{A/W})$. Finally the relative rigid cohomology $R^{\cdot} f_{rig*}(A/S)$ is the convergent F-isocrystal associated to F-crystal $R^{\cdot} f_{cris*}(\mathcal{O}_{A/W})$, and is therefore compatible with any base change $S' \to S$, for S'/k' of finite type.

It is well known (e.g. [3] 2.5.6 and 3.3.7) that $R^1 f_{cris*}(\mathcal{O}_{A/W})$ is canonically isomorphic to the Dieudonné crystal $\mathbb{D}(G)$ of the p-divisible group G of A. Furthermore if A/S is ordinary, then the unit-root sub-F-crystal

$$L \subset R^1 f_{cris*}(\mathcal{O}_{A/W})$$

is the Dieudonné module of the étale quotient G^{et} of G; thus the p-adic representation corresponding to the convergent F-isocrystal $L \otimes K$ by 1.1.5 is just the canonical representation of $\pi_1(S)$ on the étale quotient of G.

Next, we recall the construction of the pairing

$$\langle\ ,\ \rangle_A : R^1 f_{DR*}(A/S) \otimes R^1 \hat{f}_{DR*}(\hat{A}/S) \to \mathcal{O}_S \tag{2.1.1}$$

where $\hat{f} : \hat{A} \to S$ is the dual of the abelian scheme $f : A \to S$ (the only published reference for this construction seems to be [3], §5, whose treatment we will follow). Let $\mathcal{L}_A$ denote the Poincaré bundle on $\hat{A} \times A$, and consider the composite map

$$\begin{aligned} \mathrm{Pic}\,(\hat{A} \times A) &\xrightarrow{c_1} R^2(\hat{f} \times f)_{DR*}(\hat{A} \times A/S) \\ &\xrightarrow{proj} R^1 \hat{f}_{DR*}(\hat{A}/S) \otimes R^1 f_{DR*}(A/S) \end{aligned} \tag{2.1.2}$$

in which c_1 is the first De Rham Chern class, and the second map is the projection onto the $(1,1)$-Kunneth component. The image of $\mathcal{L}_A$ defines a section of $R^1 \hat{f}_{DR*}(\hat{A}/S) \otimes R^1 f_{DR*}(A/S)$, which can be viewed as a dual pair of morphisms

$$\begin{aligned} \Phi_A &: R^1 f_{DR*}(A/S)^{\vee} \longrightarrow R^1 f_{DR*}(\hat{A}/S) \\ \Psi_A &: R^1 \hat{f}_{DR*}(\hat{A}/S)^{\vee} \longrightarrow R^1 f_{DR*}(A/S). \end{aligned}$$

These are in fact isomorphisms [3], 5.1.6, so we may define 2.1.1 by

$$\begin{aligned} &R^1 f_{DR*}(A/S) \otimes R^1 \hat{f}_{DR*}(\hat{A}/S) \\ &\quad \xrightarrow{(id, \Phi_A^{-1})} R^1 f_{DR*}(A/S) \otimes R^1 f_{DR*}(A/S)^{\vee} \longrightarrow \mathcal{O}_S \end{aligned}$$

Since $\Psi_A = \hat{\Phi}_A$, this is the same as

$$\begin{aligned} &R^1 f_{DR*}(A/S) \otimes R^1 \hat{f}_{DR*}(\hat{A}/S) \\ &\quad \xrightarrow{(\Psi_A^{-1}, id)} R^1 \hat{f}_{DR*}(\hat{A}/S)^{\vee} \otimes R^1 \hat{f}_{DR*}(\hat{A}/S) \longrightarrow \mathcal{O}_S. \end{aligned}$$

It follows from the construction that 2.1.1 is compatible with arbitrary base change. Finally, if $p : A \xrightarrow{\sim} \hat{A}$ is a principal polarization, then we get a pairing

$$\langle\ ,\ \rangle = \langle\ ,\ \rangle_A \circ (id, (p^*)^{-1}) : R^1 f_{DR*}(A/S) \otimes R^1 f_{DR*}(A/S) \to \mathcal{O}_S \tag{2.1.3}$$

whose formation is compatible with arbitrary base change.

If $j : A \to \hat{\hat{A}}$ is the self-duality isomorphism and $s : A \times \hat{A} \to \hat{A} \times A$ is exchange of factors, then one has $\Phi_{\hat{A}} = -j^* \circ \Psi_A$ and $\Psi_{\hat{A}} = -j^* \circ \Phi_A$ by [**3, 5.1.5**], and therefore

$$-\langle\ ,\ \rangle_A = \langle\ ,\ \rangle_{\hat{A}} \circ ((1 \times j)s)^*. \tag{2.1.4}$$

It follows that for a principally polarized A/S, the pairing 2.1.3 is symplectic. Furthermore, the pairings 2.1.1, 2.1.3 are horizontal for the Gauss-Manin connection on the modules involved; in fact the $(1,1)$-Kunneth component of $c_1(\mathcal{L})$ is horizontal in $R^2(f \times \hat{f})_{DR*}(A \times \hat{A})$, and horizontality of the pairings follows immediately by construction.

There is a similar pairing defined on crystalline cohomology, which in the liftable case is compatible with the De Rham pairing. It is also compatible with the natural Frobenius structure on crystalline cohomology:

$$\langle F^*x, F^*y\rangle = p\langle x, y\rangle \tag{2.1.5}$$

We now take R, K, k to be as in §1: k is a perfect field of characteristc $p > 0$, R is a complete discrete valuation ring with residue field k and fraction field K of characteristic 0. Let $\mathbb{S}/R$ be an R-scheme, and let $f : \mathbb{A} \to \mathbb{S}$ be an abelian scheme. We denote by $f : A \to S$ the reduction of $\mathbb{A} \to \mathbb{S}$ modulo the maximal ideal of R, and by $f : A_0 \to S_0$ the extension of scalars to K. Finally, we denote by $f : \mathfrak{A} \to \mathfrak{S}$ the formal completion of $\mathbb{A} \to \mathbb{S}$, and by $f : \mathfrak{A}^{an} \to \mathfrak{S}^{an}$ the corresponding rigid analytic space. Then, as we remarked in 1.5, the relative rigid cohomology $R^i f_{rig*}(A/S)$ has the structure of an overconvergent F-isocrystal on S/K, whose realization on the strict neighborhood S_0^{an} of $\mathfrak{S}^{an}$ is $R^i f_{DR*}(A_0/S_0)^{an}$. The pairing 2.1.1 for A_0/S_0 induces a morphism

$$R^1 f_{rig*}(A/S) \otimes R^1 \hat{f}_{rig*}(\hat{A}/S) \to \mathcal{O}_{X/K} \tag{2.1.6}$$

in $\mathrm{Isoc}^\dagger(S/K)$; thus if $\mathbb{A}$ has a principal polarization, then there is a natural symplectic pairing on the overconvergent isocrystal $R^1 f_{rig*}(A/S)$.

2.2 PROPOSITION. *Let A/S be a principally polarized abelian scheme over a scheme S/k of characteristic $p > 0$, and suppose that A has a principal level n structure for some $n \geq 3$ that is prime to p. Then the relative rigid cohomology $R^1 f_{rig*}(A/S)$ has a natural structure of overconvergent F-isocrystal on S/K, and the natural symplectic pairing*

$$\wedge^2 R^1 f_{rig*}(A/S) \longrightarrow \mathcal{O}_{S/K} \tag{2.2.1}$$

is overconvergent.

PROOF. Since the relative rigid cohomology of an abelian scheme is compatible with base change, it is enough to check the universal case, i.e. A/S is the universal principally polarized abelian scheme of level n. In this case $f : A \to S$ has a lifting $f : \mathbb{A} \to \mathbb{S}$ as a principally polarized abelian scheme with level structure, and the proposition follows from the discussion immediately preceding the proposition. □

REMARK. Presumably this is true without the hypothesis on the existence of a level structure.

2.3 Isotrivial families. An abelian scheme X/S is *isotrivial* if there is a finite étale cover $S' \to S$ such that the pullback $X \times_S S'$ is constant. In characteristic p, this can be detected by looking at the relative crystalline cohomology.

2.4 LEMMA. *Let $f : X \to S$ be an abelian scheme over a formally smooth complete local ring of characteristic $p > 0$. If the relative crystalline cohomology $R^1 f_{cris*}(\mathcal{O}_{X/W})$ is a constant crystal, then X/S is constant.*

PROOF. Since $R^1 f_{cris*}(\mathcal{O}_{X/W})$ is isomorphic to the Dieudonné crystal $\mathbb{D}(G)$ of the p-divisible group G of X, $\mathbb{D}(G)$ is constant as a crystal, and hence as an F-crystal. Since S has a p-base, the Dieudonné functor is fully faithful by [4] 4.1.1, and G is constant. By the Serre-Tate deformation theory, X/S is constant. □

Suppose now that every fiber of X/S is ordinary. Then, as is well known, there is a filtration

$$0 \longrightarrow L \longrightarrow R^1 f_{cris*}(\mathcal{O}_{X/W}) \longrightarrow M \longrightarrow 0 \tag{2.4.1}$$

of $R^1 f_{cris*}(\mathcal{O}_{X/W})$ by F-crystals such that L is unit-root. If X is principally polarized, then 2.1.5 shows that L is isotropic for the symplectic pairing on $R^1 f_{cris*}(\mathcal{O}_{X/W})$, and thus M is dual to L: $M \simeq \check{L}$. From 2.4.1 one deduces a filtration

$$0 \longrightarrow L \otimes \mathbf{Q} \longrightarrow R^1 f_{rig*}(X/S) \longrightarrow M \otimes \mathbf{Q} \longrightarrow 0 \tag{2.4.2}$$

2.5 LEMMA. *Suppose that S/k is of finite type and geometrically integral, and X/S is principally polarized. If 2.4.2 splits in the category of convergent isocrystals, then X is isotrivial.*

PROOF. Choose an integer $n \geq 3$ prime to p; then on an étale cover $S' \to S$ X has a principal level-n structure, and is therefore a pullback from the universal principally polarized abelian scheme $X/A_{g,n}$ of dimension g and level n. Since S' is geometrically integral, it is enough to show that $X_{S'}/S'$ is constant; in other words, we can assume that X/S is pullback from $A/X_{g,n}$, and then show that X/S is constant.

If 2.4.2 splits, then by 1.1.4, 2.4.1 splits as an extension of crystals. Since S/k is of finite type and geometrically integral, the smooth locus of S is open in S.

Choose a point x in the smooth locus, and let $T \to S$ be the spectrum of the completion of the local ring at of S at x; finally let

$$0 \longrightarrow L_T \longrightarrow R^1 f_{Tcris*}(\mathcal{O}_{X_T}) \longrightarrow M_T \longrightarrow 0$$

denote the restriction of 2.4.1 to the crystalline site of T. Since L is unit-root, L_T is constant, and since X is polarized, $M_T = \check{L}_T$ is constant too. Thus if 2.4.1 splits, $R^1 f_{cris*}(\mathcal{O}_{X_T/W})$ is constant as a crystal, and since T is the spectrum of a formally smooth local ring, the restriction X_T is a constant abelian scheme by 2.4. If $\pi : S \to A_{g,n}$ is the universal map, then the image of $\pi|T$ is a closed point of $A_{g,n}$; since however the generic points of T and S have the same image under π, it follows that the universal map $\pi : S \to A_{g,n}$ has a single point as its image, and therefore X/S is constant. □

2.6 Generic abelian schemes. The proof of our main result will need a slightly more flexible notion of "generic abelian scheme" then the usual one. We begin with an auxiliary definition.

DEFINITION. Suppose that k'/k is an extension of fields with algebraic closures k'^{alg}, k^{alg}, and suppose that X'/k, X/k are geometrically connected. A morphism $X' \to X$ over $\mathrm{Spec}(k') \to \mathrm{Spec}(k)$ is π_1-*generic* if the induced map $\pi_1(X' \otimes k'^{alg}) \to \pi_1(X \otimes k^{alg})$ has finite cokernel.

One should be careful about the base fields in the definition; if X'/k', X/k are geometrically connected and $X' \to X$ is π_1-generic, it does not follow that the natural morphism $X' \to X \otimes k'$ is π_1-generic.

Suppose that X''/k'', X'/k', X/k are geometrically connected and that $f : X'' \to X'$, $g : X' \to X$ are morphisms respecting the base fields. Then we evidently have

(2.6.1) If f and g are π_1-generic, then so is gf.

(2.6.2) If gf is π_1-generic, then so is g.

Slightly less evident are

(2.6.3) If k'/k is an extension of fields and X/k is geometrically connected, then $X \otimes k' \to X$ is π_1-generic (c.f. the proof of [**SGA1**] X 1.8).

(2.6.4) If X/k is normal and geometrically connected and $U \hookrightarrow X$ is an open subscheme, then $U \hookrightarrow X$ is π_1-generic (c.f. the proof of [**SGA1**] X 3.3, and V 6.9).

Next, we have

(2.6.5) Suppose X'/k', X/k are normal and geometrically connected, and let $f : X' \to X$ be a morphism respecting base fields. If there is a dense open subscheme $U \hookrightarrow X$ such that $f^{-1}(U) \to U$ is π_1-generic, then $f : X' \to X$ is π_1-generic.

In fact if $f^{-1}(U) \to U$ and $U \hookrightarrow X$ are π_1-generic, then so is the composite $f^{-1}(U) \to X$ by 2.6.2; then since $f^{-1}(U) \hookrightarrow X'$ is π_1-generic by 3.6.4, so is $X' \to X$ by 2.6.2.

(2.6.6) An étale morphism of normal geometrically connected schemes is π_1-generic.

In fact if $f : X' \to X$ is étale, then since $X' \to X$ is of finite type, we can find an open subscheme $U \hookrightarrow X$ such that $f^{-1}(U) \to U$ is finite étale, and therefore π_1-generic. The assertion then follows from 2.6.5.

(2.6.7) A dominant morphism of finite type with geometrically connected fibers between normal geometrically connected schemes is π_1-generic.

If $f : X' \to X$ is dominant, then there is an open subscheme $U \hookrightarrow X$ such that $f^{-1}(U) \to U$ is faithfully flat and has geometrically connected fibers; thus by 2.6.4 we can assume that f is faithfully flat, in which case the induced map $\pi_1(X') \to \pi_1(X)$ is actually surjective by [**SGA 1**] IX 5.6.

Note that a π_1-generic morphism is not necessarily dominant; in fact it follows from the Bertini-Lefschetz theorem for π_1 that if Y is normal, connected, and projective over k, then there are *curves* C/k and k-morphisms $C \to Y$ that are π_1-generic (see the last remark of this paper). This will be quite useful, as we shall soon see.

Let g, $n \geq 3$ be positive integers such that n is prime to p. We denote by $f : X \to A_{g,n}$ the extension of scalars to $\overline{\mathbf{F}}_p$ of the universal principally polarized abelian scheme of dimension g with a principal level n structure. The base $A_{g,n}$ is smooth, quasiprojective, and (geometrically) connected.

DEFINITION. Let S/k be a geometrically connected k-scheme and A/S an abelian scheme of relative dimension g. Then A/S is *generic* if the generic fiber of A/S is ordinary, and if for some $n \geq 3$, A has a principal level n structure for which the classifying map $p : S \to A_{g,n}$ is π_1-generic.

The assumption that a level n structure exists is simply to avoid the use of stacks and of facts concerning their fundamental groups for which there are no references. On the other hand, the condition defining what we call generic neither implies, nor does it seem to be implied by, the classifying map $S \to A_{g,n}$ being dominant. For that matter, it is not at all clear that a π_1-generic map $S \to A_{g,n}$ maps the generic point of S into the ordinary locus, which is why we assume explicitly that the generic fiber is ordinary.

If X/S is polarized and has a principal level structure such that the classifying map $S \to A_{g,n}$ is étale, then A/S is generic in the above sense. In fact the ordinary locus of X is a dense open subset of $A_{g,n}$, and by 2.6.6 the classifying map is π_1-generic.

If $f : A \to S$ is generic, then the relative rigid cohomology $R^1_{rig*}(A/S)$ is an overconvergent F-isocrystal by 2.2. Furthermore, the ordinary locus $H \subseteq S$ of A is open and dense, and the convergent F-isocrystal $R^1 f_{rig*}(A/S)^\frown|H$ has a

filtration

$$0 \longrightarrow L \longrightarrow R^1 f_{rig*}(A/S)\hat{}|H \longrightarrow \check{L} \longrightarrow 0 \tag{2.6.8}$$

by F-isocrystals, where L is of rank g and isotropic with respect to the symplectic pairing 2.2.1; we have already remarked that it is the convergent isocrystal associated by 1.1.5 to the representation of $\pi_1(S)$ on the étale quotient of the p-divisible group of A.

Suppose now that H has a k-rational point x, and with the notation of 1.1.4 we set

$$V = \omega_x(R^1 f_{rig*}(A/S))$$
$$W = \omega_x(L).$$

The symplectic pairing on the rigid cohomology $R^1 f_{rig*}(A/S)$ gives rise to a symplectic form ψ on V, and $W \subset V$ is a maximal isotropic subspace. Since S is geometrically irreducible, the groups $DGal(R^1 f_{rig*}(A/S))$ and $DGal(R^1 f_{rig*}(A/S)\hat{})$ are defined, and we have inclusions

$$DGal(R^1 f_{rig*}(A/S)\hat{}) \hookrightarrow DGal(R^1 f_{rig*}(A/S)) \hookrightarrow Sp(V, \psi). \tag{2.6.9}$$

If A/S is ordinary, then $H = S$, and the subobject $L \subseteq R^1 f_{rig*}(A/S)$ forces

$$DGal(R^1 f_{rig*}(A/S)\hat{}) \hookrightarrow P \tag{2.6.10}$$

where $P \subset Sp(V, \psi)$ is the subgroup stabilizing $W \subset V$. Our main result is

2.7 THEOREM. *Suppose that k is the perfection of an absolutely finitely generated field, and that S/k is normal, separated of finite type, and geometrically integral. If A/S is a generic abelian scheme whose ordinary locus has a k-rational point, then*

$$DGal(R^1 f_{rig*}(A/S)) = Sp(V, \psi). \tag{2.7.1}$$

If A/S is ordinary, then

$$DGal(R^1 f_{rig*}(A/S)\hat{}) = P, \tag{2.7.2}$$

otherwise

$$DGal(R^1 f_{rig*}(A/S)\hat{}) = Sp(V, \psi). \tag{2.7.3}$$

The proof will occupy the rest of this paper. We see from 2.7 that the group $DGal(R^1 f_{rig*}(A/S))$ is unchanged if S is replaced by a dense open subset, while $DGal(R^1 f_{rig*}(A/S)\hat{})$ is not, i.e. $DGal(\hat{M})$ is not a birational invariant of $A_{g,n}$; it also follows from 2.7 that the sub-isocrystal $L \subset R^1 f_{rig*}(A/S)\hat{}$ is not overconvergent.

The first thing to observe is that P is a maximal proper parabolic subgroup of $Sp(V)$, and any Levi subgroup of P can be identified with $Gl(W)$ by means of the inclusion $W \subset V$. Our first task is to show that $DGal(R^1 f_{rig*}(A/S)\hat{})$ contains a Levi subgroup of P; this is accomplished by

2.8 LEMMA. *If A/S is ordinary and L is as in 2.6.1, we have $DGal(L,x) = GL(W)$.*

PROOF. We have already remarked that L with its Frobenius structure is the unit-root F-isocrystal associated to the representation of $\pi_1(S)$ on the étale quotient of the p-divisible group of A, and thus by 1.3.7 there is a an extension K'/K such that $DGal(L) \otimes K'$ is isomorphic to the extension of scalars to K' of the Zariski-closure in $Gl(g, \mathbf{Q}_p)$ of the image of the geometric fundamental group $\pi_1(S \otimes k^{alg})$. Since $DGal(L) \subseteq GL(W)$, it is therefore enough to show that this image is open in $Gl(g, \mathbf{Q}_p)$. But by [5], Proposition 7.1 we know that if S is the ordinary locus of $A_{g,n}$, then the image of $\pi_1(S \otimes \overline{\mathbf{F}}_p)$ is all of $Gl(g, \mathbf{Z}_p)$, and the desired assertion follows from this. □

Since P is a maximal parabolic subgroup of $Sp(V,\psi)$, and the parabolic subgroups of a reductive group G are the ones for G/H are complete, there are no subgroups $H \subset Sp(V,\psi)$ properly contained between P and $Sp(V,\psi)$. We will need the following related fact, which is of a more special nature:

2.9 LEMMA. *Let L be a Levi subgroup of $Sp(V,\psi)$ contained in P. If G is a connected subgroup of $Sp(V,\psi)$ such that $L \subseteq G \subseteq P$, then $G = L$ or $G = P$.*

PROOF. Since G is assumed to be connected, we can argue on the level of Lie algebras, so let $\mathfrak{g}$, $\mathfrak{p}$, and $\mathfrak{l}$ denote the Lie algebras of G, P and L. Choose a Borel subgroup B of $Sp(V,\psi)$ contained in P, and a maximal torus $T \subset B$ contained in L. Then the basis $\{\alpha_1, \ldots, \alpha_g\}$ of the root system of $sp(V,\psi)$ corresponding to the choice of $T \subset B$ is such that if α_g is the short root, then $\{\alpha_1, \ldots, \alpha_{g-1}\}$ is a basis of the root system of $\mathfrak{l}$, and $\mathfrak{p}$ is the parabolic corresponding to the subset $\{\alpha_1, \ldots, \alpha_{g-1}\}$ of the basis $\{\alpha_1, \ldots, \alpha_g\}$.

The root system of $\mathfrak{p}$ consists of all roots for which the coefficient of α_g is positive; in fact it follows from the description of the roots of $sp(V)$ in [**LIE**] VI planche 3 that the roots of $\mathfrak{p}$ all have the form $l + a\alpha_g$ where l is a root of $\mathfrak{l}$, and $a = 0$ or 1. Since $\mathfrak{g}$ contains a Cartan subalgebra of $\mathfrak{p}$, the roots of $\mathfrak{g}$ are roots of $\mathfrak{p}$; furthermore if α and β are roots of $\mathfrak{g}$ such that $\alpha + \beta$ is a root of $sp(V,\psi)$, then in fact $\alpha + \beta$ is a root of $\mathfrak{g}$ (c.f. [**LIE**] VIII §3.1 Prop. 1). Now if $\mathfrak{g}$ has no root of the form $l + \alpha_g$, then $\mathfrak{g} = \mathfrak{l}$; if on the other hand it does, then $-l \in \mathfrak{l}$ and so $\alpha_g \in \mathfrak{g}$, from which it follows thatthat $\mathfrak{g} = \mathfrak{p}$. □

REMARK. Lemma 2.9 does not generalize, as stated, to arbitrary semisimple groups, and in fact fails for certain maximal parabolic subgroups of simple groups of type B_n or G_2.

We can now begin the proofs of the assertions in 2.7. In what follows we set $M = R^1 f_{rig*}(A/S)$.

PROOF OF 2.7.2. Let G be the connected component of $DGal(\hat{M})$; by the paragraph after 1.3.6 we know that there is a finite étale cover $\pi : S' \to S$ such that

$$G = DGal(\pi^* \hat{M}) \hookrightarrow DGal(\hat{M}).$$

Since $DGal(\hat{M}) \subseteq P$, it is enough to prove the assertion for the base-change $A_{S'}/S'$; i.e. we can assume that $G = DGal(\hat{M})$ is connected. Now we have $L \subseteq G \subseteq P$ by 2.2 and 2.6.3; if $G = L$, then 2.4.2 splits, and 2.5 would imply that the pullback of A/S is isotrivial, which is absurd. By 2.9 we must have $G = P$. □

PROOF OF 2.7.3. We have $P \subseteq DGal(\hat{M}) \subseteq Sp(V)$, so by the paragraph before 2.9 it is enough to show that the connected component G of $DGal(\hat{M})$ is not equal to P. As before there is a finite étale cover $\pi : S' \to S$ such that $DGal(\pi^* \hat{M}) = G$, and if there is a nonordinary fiber of A/S, then the same will be true of the base-change $A_{S'}/S'$; thus we can assume that $S' = S$. Let T be the ordinary locus of A, so that $T \neq S$ is a dense open subscheme. If $DGal(M) = P$ then the isomorphism $DGal(M) \simeq DGal(M|T)$ consequent on 2.7.2 shows that the restriction functor $[M] \to [M|T]$ is an equivalence of categories. In particular, the subisocrystal $L \subset M|T$ extends to a subisocrystal L' of M, and the Frobenius structure on L extends to L'. It follows from the Grothendieck's specialization theorem [6] 2.1.3 that the Frobenius structure in L' is necessarily unit-root, and thus M has a sub-unit-root F of rank g. Therefore A/S is ordinary, a contradiction. □

PROOF OF 2.7.1. Of course if A/S has a nonordinary fiber, this follows from 2.7.3 and the inclusion $DGal(\hat{M}) \subseteq DGal(M)$. In any case we may, by passing to an étale cover, assume that S is smooth and affine, A/S is ordinary, and that $DGal(M)$ is connected. It follows from Bertini's theorem that there is an extension k'/k where k' is the perfection of an absolutely finitely generated field, a smooth curve C/k', and a π_1-generic morphism $C \to S$ (see the remark below). Since the pullback A_C/C is ordinary, it is generic by 2.6.1. If M_C denotes the pullback of M to C, then M_C can be identified with the relative rigid cohomology of the pullback of X to C. By 2.7.2 we have $DGal(\hat{M}_C) = P \otimes K'$, where K' is the fraction field of the ring of Witt vectors of k'. Since $DGal(\hat{M}_C) \hookrightarrow DGal(M_C)$, it follows from the paragraph before 2.9 that $DGal(M_C)$ is either $P \otimes K'$ or $Sp(V) \otimes K'$. However, the radical of $P \otimes K'$ has a nontrivial toric part, and since C is a smooth curve, 1.4 shows that $P \otimes K'$ cannot be the monodromy group of an overconvergent F-isocrystal on C/K'. Thus $DGal(M_C) = Sp(V) \otimes K'$ for some suitable K'/K, and since $DGal(M_C) \hookrightarrow DGal(M) \otimes K'$, we must have $DGal(M) = Sp(V)$. □

REMARK. The consequence of Bertini's theorem that is being used here is the following: let S/k be normal, separated of finite type and geometrically integral; then there is an extension F/k, where F is the perfection of an absolutely finitely generated field, and a smooth curve C/F such that $\pi_1(C \otimes F^{alg}) \to \pi_1(S \otimes k^{alg})$ is surjective. Since for any open $U \hookrightarrow S$ we have $\pi_1(U \otimes k^{alg}) \twoheadrightarrow \pi_1(S \otimes k^{alg})$, we can assume that S is affine and of dimension ≥ 2. Choose an embedding $S \hookrightarrow \mathbf{A}^n$ over k into an affine space; then a version of Bertini's theorem (c.f. [11] 6.3.18, applied to successive hyperplane sections) says that for H the generic hyperplane in $\mathbf{A}^n$ of dimension $n - \dim S + 1$, and any étale cover $S' \to S \otimes k^{alg}$, the fiber

product $S' \times_{\mathbf{A}^n} H$ is geometrically connected. H is defined over the function field F_0 of a certain Grassmanian over k, and if we let F be the perfection of F_0 and set $C = S \cap H$, then the surjectivity of $\pi_1(C \otimes F^{alg}) \to \pi_1(S \otimes k^{alg})$ follows by the usual criterion ([**SGA1**] V 6.9). Since k is the perfection of an absolutely finitely generated field and F_0 is finitely generated over k, it follows that F is also the perfection of an absolutely finitely generated field.

References

1. P. Berthelot, *Cohomologie rigide*, Introductions aux cohomologies p-adiques (D. Barsky and P. Robba, eds.), Memoire 23, Suppl. au Bull. Soc. Math. Fr. 114/2, Gauthier-Villars, 1986, pp. 7–32.
2. ———, *Cohomologie rigide et cohomologie rigide à support propre* (to appear).
3. P. Berthelot, L. Breen, and W. Messing, *Théorie de Dieudonné cristalline* II, Lecture notes in Mathematics 930, Springer-Verlag, 1982.
4. P. Berthelot and W. Messing, *Théorie de Dieudonné cristalline* III, The Grothendieck Festschrift, Progress in Mathematics 86, Birkhauser, 1991, pp. 173–247.
5. C.-L. Chai and G. Faltings, *Degeneration of abelian varieties*, Springer-Verlag, 1990.
6. R. Crew, *Specialization of crystalline cohomology*, Duke Math. J. **53** (1986), 749–757.
7. ———, *F-isocrystals and p-adic representations*, Algebraic Geometry – Bowdoin 1985, PSPM 46, pp. 111–138.
8. ———, *F-isocrystals and their monodromy groups* (to appear).
9. ———, *Kloosterman sums and the monodromy of a p-adic hypergeometric equation* (to appear).
10. P. Deligne, *La conjecture de Weil* II, Publ. Math. IHES **52** (1980), 137–252.
11. J.-P. Jouanolou, *Théoremes de Bertini et Applications*, Birkhauser, 1983.
12. M. Larsen and R. Pink.
13. A. Ogus, *F-isocrystals and De Rham Cohomology* II – *Convergent isocrystals*, Duke Math. J. **51** (1984), 765–850.
14. N. Saavedra R., *Categories Tannakiennes*, Lecture Notes in Mathematics 265, Springer-Verlag, 1972.

[SGA1] A. Grothendieck, *Revetements Etale et Groupe Fondamental*, Lecture Notes in Mathematics 224, Springer-Verlag, 1971.

[LIE] N. Bourbaki, *Groupes et algèbres de Lie*, Masson, Paris, 1981.

Department of Mathematics, University of Florida, Gainesville, FL 32611

E-mail address: crew@math.ufl.edu

Contemporary Mathematics
Volume 133, 1992

Factorization of Drinfeld Singular Moduli

David R. Dorman[1]

ABSTRACT. A formula is given enabling one to compute the prime factorization of the j-invariant associated to rank a 2 Drinfeld module defined over the function field in one variable having finite base field.

§0. Introduction

Let $\mathbb{F}_q$ be the finite field having $q = p^s$ elements. Further we need to have 2 invertible in so assume $p > 2$. Let $k = \mathbb{F}_q(T)$ (resp. $A = \mathbb{F}_q[T]$) be the field of rational functions (resp. the ring of polynomials) in the indeterminate T. Fix $T^{-1} = \infty$ as the infinite place of k and denote by k_∞ the ∞-adic completion of k and by $\bar{k}_\infty$ an algebraic closure of k_∞. Let C be the completion of $\bar{k}_\infty$. Finally, let $\mathfrak{H} = C - k_\infty$ denote the upper half plane.

$GL_2(A)$ acts on H via linear fractional transformations and there is an analytic parametrization of the rigid analytic space $GL_2(A)\backslash \mathfrak{H}$ to C (cf. Drin[3])

$$j: GL_2(A)\backslash \mathfrak{H} \xrightarrow{\sim} C.$$

The function j can be described through the Eisenstein series as follows. Let L be a rank 2 lattice and let

$$E_t(L) = \sum_{0 \neq \gamma \in L} \gamma^{-t}.$$

Then

$$j(L) = \frac{g(L)^{q+1}}{\Delta(L)}$$

1980 Mathematics Subject Classification (1985 Revision) 11G15, secondary 14K10, 14K22. The final detailed version of this paper has appeared in Comp. Math. **80** (1991) 235-256.

[1]Research partially supported by NSF grants RII-8610679 and DMS-8903463.

where

$$g(L) = (T - T^q)E_{q-1}(L)$$

and

$$\Delta(L) = (T^q - T)^q E_{q-1}(L)^{q+1} - (T^{q^2} - T)E_{q^2-1}(L).$$

The function j is analogous to the classical j function of Dedekind and has many interesting arithmetic properties. In particular let $\tau \in H$ be imaginary quadratic over k, that is t satisfies a quadratic polynomial $a\tau^2 + b\tau + c = 0$ with a, b, $c \in A$ and relatively prime, and ∞ is non-split in the extension $k(\tau)/k$. Then the value $j(\tau)$ is an algebraic integer over A. Such integers are called singular moduli since they correspond to isomorphism invariants of rank 2 Drinfeld A-modules having complex multiplication by an order in an imaginary quadratic extension of k. An extension K/k is called imaginary quadratic if it is degree 2 and ∞ is non-split in K. Theorem 1.4 gives a formula for computing the prime factorization of such invariants. The proof of this theorem as well as the details of the results will appear elsewhere [1].

The factorization of differences of singular moduli associated to elliptic curves defined over number fields was accomplished by Gross and Zagier [5] in the case of prime discriminants and extended by the author [2] to the case of relatively prime composite discriminants. These earlier works led naturally to the study of singular moduli associated to rank 2 Drinfeld A-modules over function fields.

§ 1. The main result

Since $\tau \in \mathfrak{H}$ is imaginary quadratic over k it satisfies a quadratic polynomial $a\tau^2 + b\tau + c = 0$ with a, b, $c \in A$ and relatively prime, and moreover, ∞ is non-split in the extension $k(\tau)/k$. The polynomial $b^2-4ac = d = \mathrm{disc}(\tau)$ is called the discriminant of τ. It is well defined up to the square of a unit in $\mathbb{F}_q$ and depends only on τ. We consider d fixed and require that d be fundamental, that is, d is a field discriminant, or equivalently, square free. Let $K = k(\tau) = k(\sqrt{d})$, thus $d = d_K$ where d_K is the discriminant of K. The class number of $\mathcal{O}_K$, the integers of K, is denoted by h(d).

The pertinent arithmetic facts regarding singular moduli of Drinfeld modules are [cf Gekeler [4] and Hayes [6]]:

1. $j(\tau)$ is an algebraic integer of degree h(d) over A.

2. The h(d) Galois conjugates of j(t) over $\mathcal{O}$ are the values $j(\tau')$ where τ' runs through the roots of all the distinct primitive quadratic polynomials of discriminant d.

3. The field $K(j(\tau))$ is the Hilbert class field of K which is split completely over K, and it is therefore abelian over K.

4. There exists a rank 2 Drinfeld A-module φ defined over $K(j(\tau))$, having complex multiplication by $\mathcal{O}_K$ with j invariant $j_\varphi = j(\tau)$.

Let d' be a second fundamental imaginary discriminant and consider the product

$$J(d,d') = \prod_{\substack{[\tau],[\tau'] \\ \mathrm{disc}(\tau)=d,\ \mathrm{disc}(\tau')=d'}} \left(j(\tau) - j(\tau')\right) \tag{1.1}$$

where [ ] denotes an equivalence class modulo $GL_2(A)$. The main focus of this study is to compute $\mathrm{ord}_{\mathfrak{p}} J(d)$ where $\mathfrak{p}$ is any non-zero prime ideal of A. One then tries to determine this number by counting the number of mod $\mathfrak{p}$ isomorphisms between the rank 2 Drinfeld A-modules that have the corresponding j values. This counting is done by determining units in the endomorphism rings of these Drinfeld modules.

Theorem 1.4 below treats the case where $d' = u$, where u is a non-square unit in $\mathbb{F}_q^*$ and d is an arbitrary discriminant relatively prime to u. In this setting $H = k(\sqrt{u}) = \mathbb{F}_{q^2}(T)$ is the constant field extension and there is a unique rank 2 Drinfeld A-module, r, with complex multiplication by $\mathcal{O} = \mathbb{F}_{q^2}[T]$ has $j = 0$. Thus (1.1) can be rewritten as

$$J(d) = \prod_{\substack{[\tau] \\ \mathrm{disc}(\tau)=d}} \left(j(\tau) - 0\right) = \prod_{\substack{[\tau] \\ \mathrm{disc}(\tau)=d}} j(\tau) \tag{1.2}$$

The problem is then to count the number of isomorphisms between the rank 2 Drinfeld A-modules having j invariants equal to the $j(\tau)$ with ρ. This is much simpler than in the general case since $\mathbb{F}_{q^2}[T]$ has class number 1 and so the arithmetic in this ring is straight forward.

We now state the result. Let A be an ideal of A having factorization

$$\mathfrak{a} = \prod_i \mathfrak{p}_i^{m_i} \prod_j \mathfrak{q}_j^{n_j}$$

with $\mathfrak{p}_i$ split and $\mathfrak{q}_j$ inert in $\mathbb{F}_{q^2}(T)$. Define $R(A)$ to be the number of ideals of $\mathbb{F}_{q^2}(T)$ having norm the ideal A. $R(A)$ can be calculated by

$$R(\mathfrak{a}) = \begin{cases} \prod (m_i + 1) & \text{if all the } n_j \text{ are even} \\ 0 & \text{otherwise.} \end{cases} \tag{1.3}$$

Theorem 1.4: *Let* $\mathfrak{p}$ *be a non-zero prime ideal of* A *and let* π *be a generator for* $\mathfrak{p}$. *Then*

$$\mathrm{ord}_{\mathfrak{p}} J(d) = \begin{cases} 0 & \text{if } \mathfrak{p} \text{ is split in K} \\ \dfrac{q+1}{2} \displaystyle\sum_{m \in A} \sum_{n \geq 1} R\left(\frac{d - um^2}{\pi^{2n-1}}\right) & \text{if } \mathfrak{p} \text{ is inert in K} \end{cases}$$

The quantity $(d-um^2)/\pi^{2n-1}$ is to be interpreted as the generator of an ideal A. If A is integral then $R(\mathfrak{a})$ is determined by (1.3) otherwise $R(\mathfrak{a})$ equals 0. A

necessary condition for the above quotient to generate an integral ideal A with R(A) $\neq$ 0 is that degree $\pi \leq$ degree d. In the proof of Theorem 1.4 it is shown that this sum is finite. A striking corollary is

Corollary 1.5: *If* $\pi | J(d)$ *then* degree $\pi \leq$ degree d.

Our proof relies on the work of Gekeler [4] which established the connections between the endomorphism ring of a supersingular Drinfeld A-module of rank 2 in characteristic $\mathfrak{p}$ with a maximal order in the definite quaternion algebra over A ramified only at P and ∞.

Remark: For comparison we state the analogue of theorem 1.4 in the number field setting. Now $k = \mathbb{Q}$, $A = \mathbb{Z}$ and for simplicity we consider only the case $K = \mathbb{Q}(\sqrt{-p})$ where p is a prime congruent to 3 modulo 4. Then the analogue of (1.2) is

$$J(-p) = \prod_{\substack{[\tau] \\ \mathrm{disc}(\tau)=p}} \left(j(\tau) - 0\right) = \prod_{\substack{[\tau] \\ \mathrm{disc}(\tau)=p}} j(\tau) \tag{1.6}$$

It is well known that J(-p) is a perfect cube. As before let R(m) be the number of integral primes of $\mathcal{O}_K$ having norm equal to m. Then we have

Theorem 1.7 *Let* ℓ *be a rational prime. Then*

$$\mathrm{ord}_{\ell} J(-p) = \begin{cases} 0 & \text{if } \ell \text{ is split in K.} \\ \dfrac{3}{2} \displaystyle\sum_{x \in \mathbb{Z}} \sum_{n \geq 1} R\left(\frac{3p - x^2}{4\ell^n}\right) & \text{if } \ell \text{ is non-split in K.} \end{cases}$$

A more general theorem for the factorization of J(d,d') where d and d' are relatively prime composite discriminants can be found in [2].

Bibliography

1. D. R. Dorman, *On singular moduli for rank 2 Drinfeld modules*, Comp. Math **80** (1991) 235-256.
2. D. R. Dorman, *Special values of the elliptic modular function and factorization formulae*, Jour. reine und ang. Math. **383** (1988), 207-220.

3. V. Drinfel'd, *Elliptic Modules*, (Russian) Math. Sbornik **94** (1974) 594-627, English Translation: Math. USSR-Sbornik **23** (1976) 561-592.
4. E.-U. Gekeler, *Zur Arithmetic von Drinfeld-Moduln*, Math. Ann. **262** (1983) 167-182.
5. B. H. Gross and D. B. Zagier, *On singular moduli*, Jour reine und ang. Math. **355** (1985), 191-220.
6. D. R. Hayes, *Explicit class field theory in global function fields*, in Studies in Algebra and Number Theory, Advances in Mathematics Supplementary Studies, ed. G. Rota, **16** (1979) 173-217, Academic Press.

Department of Mathematics
Middlebury College
Middlebury, VT 05753
email dorman@midd.cc.middlebury.edu

This page intentionally left blank

Contemporary Mathematics
Volume 133, 1992

Distinctness of Kloosterman Sums

BENJI FISHER

ABSTRACT. For any prime p, the $p-1$ Kloosterman sums

$$\mathrm{Kl}(p;a) \stackrel{def}{=} \sum_{\substack{xy=a \\ x,y\in\mathbb{F}_p{}^\times}} e^{2\pi i(x+y)/p} \qquad (a \in \mathbb{F}_p{}^\times)$$

(also denoted $S(1,a;p)$) are distinct, and the proof is elementary. This paper generalizes this result to Kloosterman sums in several variables, over arbitrary finite fields, twisted by multiplicative characters. The proof uses étale cohomology, including Deligne's "Weil II" theorem.

0. Introduction

0.1. Notation. Let p be a prime, $q = p^d$ a power of p, and $\mathbb{F}_q$ the field with q elements. Let $l \neq p$ be prime. Let $\psi\colon \mathbb{F}_q \to \overline{\mathbb{Q}}_l{}^\times$ be a non-trivial additive character; for simplicity, assume that ψ is defined over $\mathbb{F}_p$: *i.e.*, $\psi = \psi_0 \circ \mathrm{tr}_{\mathbb{F}_q/\mathbb{F}_p}$ with $\psi_0\colon \mathbb{F}_p \to \overline{\mathbb{Q}}_l{}^\times$, so that $\psi = \psi\circ\sigma$ for any $\sigma \in \mathrm{Gal}(\mathbb{F}_q/\mathbb{F}_p)$.

For $\chi = (\chi_1, \ldots, \chi_n)$ any n-tuple of multiplicative characters, $\chi_i\colon \mathbb{F}_q{}^\times \to \overline{\mathbb{Q}}_l{}^\times$, $E/\mathbb{F}_q$ any finite extension, and any element $a \in E^\times$, define the **Kloosterman sum**

$$\mathrm{Kl}_n(\psi,\chi,E,a) \stackrel{def}{=} \sum_{x\in(E^\times)^n,\,\Pi x=a} \psi\circ\mathrm{tr}_{E/\mathbb{F}_q}\Big(\sum_{i=1}^n x_i\Big) \prod_{i=1}^n \chi_i\circ \mathrm{N}_{E/\mathbb{F}_q}(x_i),$$

where $x = (x_1, \ldots, x_n)$, and Πx denotes $\prod_{i=1}^n x_i$.

We will use $\mathbf{1}$ to denote both the trivial multiplicative character and the n-tuple $(\mathbf{1}, \ldots, \mathbf{1})$.

1991 *Mathematics Subject Classification.* Primary 11L05; Secondary 11G25, 11L07, 14F20, 14G15.

The author was supported in part by the National Science Foundation and the Sloan Foundation.

This paper is in final form and no version of it will be submitted for publication elsewhere.

0271-4132/92 $1.00 + $.25 per page

0.2. Summary. Section 1 presents a proof that the $p-1$ Kloosterman sums $\mathrm{Kl}_2(\psi, \mathbf{1}, \mathbb{F}_p, a)$ $(1 \leq a \leq p-1)$ are distinct. It also contains a short discussion of Kloosterman sums over $\mathbb{F}_q$ with $q \neq p$, especially the cases $p = 2$ and $p = 3$. Finally, it shows how an "unexpected" equality of Kloosterman sums leads to a non-trivial identity of Gauss sums.

The basic properties of Kloosterman sheaves from [**Ka-1**] and [**Ka-2**] are reviewed in §2. In §3 we give two methods of constructing a sheaf on $\mathbb{G}_m \otimes \mathbb{F}_p$ whose trace at t is $\pm\mathrm{Kl}_n(\psi, \chi, \mathbb{F}_q, at^n)$ (given $a \in {\mathbb{F}_q}^\times$).

The main result, proved in §4, is a complement to the equidistribution proved in [**Ka-1**]. If χ is not Kummer-induced (*e.g.*, $\chi = \mathbf{1}$) and $p > (2n^{2d}+1)^2$ $(d = [\mathbb{F}_q : \mathbb{F}_p])$ then the $q-1$ Kloosterman sums $\mathrm{Kl}_n(\psi, \chi, \mathbb{F}_q, a)$ $(a \in {\mathbb{F}_q}^\times)$ are distinct, up to the action of $\mathrm{Gal}(\mathbb{F}_q/\mathbb{F}_p)$. (There is one more restriction if $n = 4$: basically, we exclude $\chi = (\mathbf{1}, \mathbf{1}, \mathbf{1}, \Lambda_2)$, where Λ_2 denotes the quadratic character on ${\mathbb{F}_q}^\times$.) For a discussion of the hypotheses, see Remarks 4.28.

The idea of the proof is that if $\mathrm{Kl}(a) = \mathrm{Kl}(b)$ then the sheaves constructed in §3 for a and b have the same traces at all points $t \in \mathbb{G}_m(\mathbb{F}_p)$; if $p \gg 0$ then, thanks to Weil II, this almost implies that the sheaves are isomorphic. It follows that $a = b$ (up to the action of the Galois group).

0.3. Acknowledgements. Besides teaching me all I know about this subject and suggesting this topic, my thesis advisor, Nicholas M. Katz, supplied several of the key ideas in this paper. Michael Larsen helped me learn the representation theory of semisimple Lie algebras. Bernard Dwork suggested that the connection with Gauss sums (*cf.* Remark 1.7) might be interesting.

My thesis tries to explain some of the motivation for the proof and gives more details on some of the calculations.

1. Elementary Results

In this section, we will prove the distinctness of the classical Kloosterman sums and discuss non-distinctness in various circumstances, including "small" values of p.

LEMMA 1.1. *For $p > 2$, let $\left(\frac{x}{p}\right)$ denote the Legendre symbol. (In particular, $\left(\frac{x}{p}\right) = 0$ if $x = 0 \in \mathbb{F}_p$.) Then the Kloosterman sum $\mathrm{Kl}(a) = \mathrm{Kl}_2(\psi, \mathbf{1}, \mathbb{F}_p, a)$ is given by*

$$\mathrm{Kl}(a) = \sum_{n=0}^{p-1} \left[1 + \left(\frac{n^2 - 4a}{p}\right)\right] \psi(n) = \sum_{n=0}^{p-1} \left(\frac{n^2 - 4a}{p}\right) \psi(n).$$

PROOF. The number of pairs $(x, y) \in \left({\mathbb{F}_p}^\times\right)^2$ satisfying $xy = a$, $x + y = n$ is the number of roots in $\mathbb{F}_p$ of $T^2 - nT + a$, or $1 + \left(\frac{n^2-4a}{p}\right)$. □

Remarks 1.2. (1) The same calculation, with ψ replaced by the constant 1, shows that $\sum_{n=0}^{p-1}\left(\frac{n^2-4a}{p}\right) = -1$.

(2) Lemma 1.1 is not new. Weil, in his classic note on Kloosterman sums [**Weil**] mentions it and refers to Davenport [**Dav**], who refers to Salié [**Sal**].

PROPOSITION 1.3. *If* $\mathrm{Kl}(a) = \mathrm{Kl}(b)$ $(a, b \in \mathbb{F}_p^{\times})$ *then* $a = b$.

PROOF. This is trivial if $p = 2$, so assume now that $p > 2$. If $\mathrm{Kl}(a) = \mathrm{Kl}(b)$ then Lemma 1.1 and Remark 1.2(1) imply that $\left(\frac{n^2-4a}{p}\right) = \left(\frac{n^2-4b}{p}\right)$ for all n in $\mathbb{F}_p$. (Indeed, the only $\mathbb{Q}$-linear relation among $1, \psi(1), \ldots, \psi(p-1)$ is $1 + \psi(1) + \cdots + \psi(p-1) = 0$.) Therefore there are at least $\frac{p+1}{2}$ values of $x \in \mathbb{F}_p$ (*viz.*, $x = n^2 - 4a$ for $n = 0, 1, \ldots, \frac{p-1}{2}$) for which $\left(\frac{x}{p}\right) = \left(\frac{x+4(a-b)}{p}\right)$. For such x, either $x = 0$ or $1 + 4(a-b)/x$ is a quadratic residue (mod p); since there are only $\frac{p-3}{2}$ residues other than 1, this forces $a = b$. □

Remark 1.4. If $q \neq p$ then the $q-1$ Kloosterman sums $\mathrm{Kl}(a) = \mathrm{Kl}_2(\psi, \mathbf{1}, \mathbb{F}_q, a)$ $(a \in \mathbb{F}_q^{\times})$ are not distinct. Indeed, for $\sigma \in \mathrm{Gal}(\mathbb{F}_q/\mathbb{F}_p)$ and $a \in \mathbb{F}_q^{\times}$,

$$\begin{aligned}
\mathrm{Kl}(a) &= \sum_{xy=a} \psi(x+y) = \sum_{\sigma(x)\sigma(y)=\sigma(a)} \psi(x+y) \\
&= \sum_{xy=\sigma(a)} \psi\circ\sigma^{-1}(x+y) = \sum_{xy=\sigma(a)} \psi(x+y) \\
&= \mathrm{Kl}\big(\sigma(a)\big).
\end{aligned}$$

(Recall that we assume that ψ is defined over $\mathbb{F}_p$.) Similarly, one finds that $\mathrm{Kl}_n(\psi, \chi, \mathbb{F}_q, a) = \mathrm{Kl}_n\big(\psi, \chi\circ\sigma^{-1}, \mathbb{F}_q, \sigma(a)\big)$, with $\chi\circ\sigma^{-1} = \big(\chi_1\circ\sigma^{-1}, \ldots, \chi_n\circ\sigma^{-1}\big)$.

Remarks 1.5. Let $\zeta_p = \psi(1)$ be a primitive p^{th} root of 1 and consider the Kloosterman sums $\mathrm{Kl}(a) = \mathrm{Kl}_2(\psi, \mathbf{1}, \mathbb{F}_q, a)$, with $q = p^d$. in $\mathbb{Z}[\zeta_p] \subseteq \overline{\mathbb{Q}}_l$,

$$\mathrm{Kl}(a) \equiv \sum_{x \in \mathbb{F}_q^{\times}} 1 = q - 1 \equiv -1 \pmod{1 - \zeta_p}.$$

Also, the Kloosterman sum is real. Thus $\mathrm{Kl}_\psi(a) \in \mathbb{Z}[\zeta_p] \cap \mathbb{R} = \mathbb{Z}[\zeta_p + \zeta_p^{-1}]$ and (if $p \neq 2$) the above congruence holds modulo $2 - (\zeta_p + \zeta_p^{-1})$. Less obviously, we have the Hasse-Weil estimate [**Weil**]

$$\big|\mathrm{Kl}(a)\big| \leq 2\sqrt{q}.$$

For $p = 2$ (respectively, $p = 3$) there are thus about $4\sqrt{q}/2 = 2\sqrt{q}$ (resp., $4\sqrt{q}/3$) possible values for $\mathrm{Kl}(a)$; there are about q/d non-conjugate values for $a \in \mathbb{F}_q^{\times}$. Thus for large d, the Kloosterman sums for $\mathbb{F}_{2^d}$ and $\mathbb{F}_{3^d}$ are not distinct. In fact, for $p = 2$, $d > 3$ suffices; and for $p = 3$, $d > 1$ suffices. (It helps to note that $\mathrm{Kl}(a) \equiv -1 \pmod 4$ if $p = 2$ and $d > 1$.) Furthermore, these bounds are sharp, as one sees by calculating the Kloosterman sums for $\mathbb{F}_4$ and $\mathbb{F}_8$.

Remark 1.6. For any multiplicative character Λ, let $\chi\Lambda = (\chi_1\Lambda, \ldots, \chi_n\Lambda)$; then $\mathrm{Kl}_n(\psi, \chi\Lambda, \mathbb{F}_q, a) = \Lambda(a)\ \mathrm{Kl}_n(\psi, \chi, \mathbb{F}_q, a)$. We say that χ is **Kummer-induced** if, for some non-trivial Λ, $\chi = \chi\Lambda$ as unordered n-tuples; *i.e.*, if there is a permutation $i \mapsto i'$ such that $\chi_i = \chi_{i'}\Lambda$. If this holds then $\mathrm{Kl}_n(\psi, \chi, \mathbb{F}_q, a) = 0$ whenever $\Lambda(a) \neq 1$; *i.e.*, whenever $a \notin (\mathbb{F}_q^\times)^k$, where k is the order of Λ. Note that $\chi = \chi\Lambda$ is equivalent to (renumbering if necessary) $\chi = (\chi_1, \ldots, \chi_n) = (\chi_1, \chi_1\Lambda, \ldots, \chi_1\Lambda^{k-1}, \ldots, \chi_{n/k}, \chi_{n/k}\Lambda, \ldots, \chi_{n/k}\Lambda^{k-1})$.

Remark 1.7. Any "unexpected" coincidence of Kloosterman sums (*i.e.*, one not predicted by Remarks 1.4 or 1.6) leads to a non-trivial identity of Gauss sums. Indeed, elementary Fourier inversion on the group $\mathbb{F}_q^\times$ leads to the formula

$$\mathrm{Kl}_n(\psi, \chi, \mathbb{F}_q, a) = \frac{1}{q-1} \sum_{\Lambda} \Lambda(a) \prod_{i=1}^{n} g\left(\psi, \Lambda^{-1}\chi_i\right),$$

where Λ runs over all multiplicative characters of $\mathbb{F}_q^\times$ and $g(\psi, \Lambda)$ is the Gauss sum $\sum_{x \in \mathbb{F}_q^\times} \psi(x)\Lambda(x)$. (*Cf.* [**Ka-1**, §4.0].)

Two examples follow. For more numerical data on when unexplained coincidences occur, see Remark 4.28(2).

Example 1.7.1. Let $\mathbb{F}_9 = \mathbb{F}_3[i]$ with $i^2 = -1$ and let $\mathrm{Kl}(a) = \mathrm{Kl}_2(\psi, \mathbf{1}, \mathbb{F}_9, a)$. One finds that $\mathrm{Kl}(2) = \mathrm{Kl}(2+i) = 2$. If we let $\zeta_8 \in \overline{\mathbb{Q}_l}^\times$ denote a primitive 8^{th} root of 1 and let Λ_k denote the multiplicative character defined by $\Lambda_k(2+i) = \zeta_8^k$ then we find (since $2 = (2+i)^4$)

$$\sum_{k=0}^{7} (-1)^k g(\psi, \Lambda_{-k})^2 = \sum_{k=0}^{7} \zeta_8^k g(\psi, \Lambda_{-k})^2.$$

Using the elementary formula $g(\psi, \Lambda^p) = g(\psi, \Lambda)$ (since ψ is defined over $\mathbb{F}_p$), letting $A \stackrel{def}{=} \zeta_8 + \zeta_8^3$, and noting that $\zeta_8^4 = -1$, this becomes

$$(2 - A)g(\psi, \Lambda_1)^2 + (2 + A)g(\psi, \Lambda_7)^2 = 2g(\psi, \Lambda_2)^2 + 2g(\psi, \Lambda_4)^2.$$

Noting that $g(\psi, \Lambda_2)^2 = g(\psi, \Lambda_4)^2 = 9$, $g(\psi, \Lambda_1)g(\psi, \Lambda_7) = -9$ (special cases of $g(\psi, \Lambda)g(\psi, \Lambda^{-1}) = \Lambda(-1)q$), and $A^2 = -2$, this gives a quadratic equation for $g(\psi, \Lambda_1)^2$ (with coefficients in $\mathbb{Z}[A]$).

Example 1.7.2. Similarly, let $\mathbb{F}_{25} = \mathbb{F}_5[\sqrt{2}]$ and $\mathrm{Kl}(a) = \mathrm{Kl}_3(\psi; \mathbf{1}, \mathbf{1}, \chi_2; \mathbb{F}_{25}, a)$ (where χ_2 denotes the quadratic character). One finds that $\mathrm{Kl}(\sqrt{2}) = \mathrm{Kl}(2\sqrt{2})$. Letting α be a cube root of $\sqrt{2}$ and Λ_i the character with $\Lambda_i(\alpha) = \zeta_{24}^i$ (so $\chi_2 = \Lambda_{12}$), this leads to

$$\sum_{i=0}^{23} (\zeta_8^i - \zeta_8^{3i}) g(\psi, \Lambda_{24-i})^2 g(\psi, \Lambda_{12-i}) = 0.$$

Remark 1.8. The referee points out that Remark 1.7, along with the Stickelberger relation, shows that the $p-1$ Kloosterman sums $\mathrm{Kl}_n(a) = \mathrm{Kl}_n(\psi, \mathbf{1}, \mathbb{F}_p, a)$ ($p \in \mathbb{F}_p{}^\times$) are distinct. Indeed, let $\omega\colon \mathbb{F}_p{}^\times \to \mathbb{Z}_p^\times$ denote the Teichmüller character, so that $\overline{\omega(a)} = a$ for $a \in \mathbb{F}_p{}^\times$, where the bar denotes the map $\mathbb{Z}_p \twoheadrightarrow \mathbb{Z}_p/p\mathbb{Z}_p \cong \mathbb{F}_p$. Then ω generates the multiplicative characters of $\mathbb{F}_p{}^\times$. The Stickelberger relation is $\mathrm{ord}_p\big(g(\psi,\omega^{-m})\big) = \frac{m}{p-1}$ for $0 \le m < p-1$. (*Cf.* [**Ir-Ro**, §14.4].) Therefore, in $\mathbb{Z}_p[\zeta_p]$,

$$\begin{aligned}(p-1)\mathrm{Kl}_n(a) &= \sum_{m=0}^{p-2} \omega^m(a) g(\psi,\omega^{-m})^n = (-1)^n + \sum_{m=1}^{p-2} \omega^m(a) g(\psi,\omega^{-m})^n \\ &\equiv (-1)^n + \omega(a) g(\psi,\omega^{-1})^n \pmod{\pi^{2n}},\end{aligned}$$

where $\pi = 1-\zeta_p$, $\mathrm{ord}_p(\pi) = \frac{1}{p-1}$. Thus if $\mathrm{Kl}_n(a) = \mathrm{Kl}_n(b)$ then $\omega(a) \equiv \omega(b) \pmod{\pi^n}$ and $a = \overline{\omega(a)} = \overline{\omega(b)} = b$.

This result is better than Theorem 4.22 in that it works for all p, but this author does not see how to extend it to $q > p$ and non-trivial characters χ.

2. Review of Kloosterman Sheaves

In this section, we will review some of Katz's results on Kloosterman sheaves from [**Ka-1**] and [**Ka-2**]. Let $k = \mathbb{F}_q$ be a finite field of characteristic p.

DEFINITION 2.1. For additive and multiplicative characters $\psi\colon \mathbb{F}_q \to \overline{\mathbb{Q}}_l{}^\times$, $\chi\colon \mathbb{F}_q{}^\times \to \overline{\mathbb{Q}}_l{}^\times$, let $\mathcal{L}_\psi$ and $\mathcal{L}_\chi$ denote the rank-one, lisse sheaves on $\mathbb{G}_m \otimes \mathbb{F}_q$ such that, for any finite extension $E/\mathbb{F}_q$ and any $a \in E^\times = \mathbb{G}_m(E)$,

$$\mathrm{tr}\big(F_a \mid \mathcal{L}_\psi\big) = \psi \circ \mathrm{tr}_{E/\mathbb{F}_q}(a); \qquad \mathrm{tr}\big(F_a \mid \mathcal{L}_\chi\big) = \chi \circ \mathrm{N}_{E/\mathbb{F}_q}(a),$$

where $F_a = F_{E,a} \in \pi_1(\mathbb{G}_m)$ denotes the geometric Frobenius element at a. The sheaves $\mathcal{L}_\psi$ and $\mathcal{L}_\chi$ are constructed from the Lang torsor (*cf.* [**Ka-1**, §4.3] or [**Ka-2**, §7.2]).

DEFINITION 2.2. For any n-tuple χ of multiplicative characters $\chi_i\colon k^\times \to \overline{\mathbb{Q}}_l{}^\times$, we define the **Kloosterman complex** $\mathrm{Kl}_n(\psi,\chi)$ to be the multiple convolution (in $D_c^b(\mathbb{G}_m \otimes k, \overline{\mathbb{Q}}_l)$, *cf.* [**Ka-2**, §8.1.8])

$$\mathrm{Kl}_n(\psi,\chi) = \overset{n}{\underset{i=1}{*_!}} \mathcal{L}_\psi \otimes \mathcal{L}_{\chi_i}[1],$$

so that for every finite extension E/k and all $a \in E^\times = \mathbb{G}_m(E)$,

$$\mathrm{tr}\big(F_a \mid \mathrm{Kl}_n(\psi,\chi)\big) = (-1)^n \mathrm{Kl}_n(\psi,\chi,E,a).$$

Cf. [**Ka-2**, §8.2 and Remark 8.4.3].

Remark 2.3. The basic properties of Kloosterman sheaves are given by the following Theorem.

THEOREM 2.4. *There is a lisse, l-adic sheaf on $\mathbb{G}_m \otimes k$, denoted $\mathcal{K}_n(\psi, \chi)$ (or simply $\mathcal{K}$) such that $\mathrm{Kl}_n(\psi, \chi) = \mathcal{K}[1]$ (i.e., the Kloosterman complex $\mathrm{Kl}_n(\psi, \chi)$ consists of this Kloosterman* sheaf *$\mathcal{K}$, placed in degree -1), so that for any finite extension E/k and any $a \in E^\times$,*

$$\mathrm{tr}(F_a \mid \mathcal{K}) = -\mathrm{tr}\big(F_a \mid \mathrm{Kl}_n(\psi, \chi)\big) = (-1)^{n-1}\mathrm{Kl}_n(\psi, \chi, E, a).$$

The sheaf $\mathcal{K}$ has rank n and is pure of weight $n-1$. $\mathcal{K}$ is tame at 0 and

$$\mathcal{K} \mid I_0 \cong \bigoplus_\chi \mathcal{L}_\chi \otimes \mathrm{Unip}(m_\chi),$$

where $m_\chi = \#\{i \mid \chi_i = \chi\}$ and $\mathrm{Unip}(n)$ is a unipotent Jordan block of size n; taking semisimplifications,

$$(\mathcal{K} \mid I_0)^{\mathrm{s/s}} \cong \bigoplus_{i=1}^{n} \mathcal{L}_{\chi_i}.$$

$\mathcal{K}$ is totally wild at ∞ with $\mathrm{Sw}_\infty(\mathcal{K}) = 1$, so all ∞-breaks of $\mathcal{K}$ are $1/n$. If $p \nmid n$ then

$$[n]^*\mathcal{K} \mid P_\infty \cong \bigoplus_{\zeta \in \mu_n(\overline{k})} \mathcal{L}_{\psi(\zeta x/n)} \mid P_\infty.$$

PROOF. [**Ka-2**, Theorem 8.4.2] or [**Ka-1**, Theorems 4.1.1, 7.4.1]; and [**Ka-1**, Remark 10.4.5]. □

Remark 2.5. Since a Kloosterman sheaf is totally wild with Swan conductor 1 at ∞, it is irreducible. We will need a stronger condition than irreducibility.

DEFINITION 2.6. Let X be a connected k-scheme. A lisse, l-adic sheaf $\mathcal{F}$, corresponding to a representation $\rho_\mathcal{F} : \pi_1(X) \to \mathrm{GL}(V)$, is called **Lie-irreducible** if and only if for every subgroup $\Gamma \subseteq \pi_1^{\mathrm{geom}}(X) \stackrel{def}{=} \pi_1\big(X \underset{k}{\otimes} \overline{k}\big)$ of finite index, V (or $\mathcal{F}$) is irreducible as a representation of Γ. *Cf.* [**Ka-2**, discussion preceding Theorem 7.2.6].

Remark 2.7. With notation as in the definition, let G be the geometric monodromy group of $\mathcal{F}$ (*i.e.*, the Zariski-closure of the image of $\rho_\mathcal{F}$). Then $\mathcal{F}$ is Lie-irreducible if and only if G^0, the identity component of G, acts irreducibly on $\mathcal{F}$ (or V). This holds if and only if $\mathcal{F}$ (or V) is irreducible as a representation of the Lie algebra of G.

DEFINITION 2.8. A lisse, l-adic sheaf $\mathcal{F}$ on $\mathbb{G}_m \otimes k$ or $\mathbb{G}_m \otimes \overline{k}$ is **Kummer-induced** if and only if $\mathcal{F} = [d]_*\mathcal{G}$ for some lisse, l-adic sheaf $\mathcal{G}$ on $\mathbb{G}_m \otimes k$ or $\mathbb{G}_m \otimes \overline{k}$, where $1 < d \in \mathbb{Z}$ and $[d] = (t \mapsto t^d) : \mathbb{G}_m \to \mathbb{G}_m$ denotes the Kummer map.

THEOREM 2.9. *Assume $p > 2n+1$. A Kloosterman sheaf $\mathcal{K} = \mathcal{K}_n(\psi,\chi)$ is Lie-irreducible unless it is geometrically Kummer-induced. Furthermore, if $\mathcal{K} = [d]_*\mathcal{G}$ as a representation of $\pi_1^{\text{geom}} = \pi_1(\mathbb{G}_m \otimes \overline{k})$ then $d|n$, $d|q-1$, and χ is Kummer-induced of order d; i.e., $\chi = \chi\Lambda$ (as unordered n-tuples), where $\Lambda\colon k^\times \to \overline{\mathbb{Q}}_l{}^\times$ is a character of order (exactly) d. (Cf. Remark* 1.6).

PROOF. By [**Ka-2**, Theorem 7.2.6 (4)], *any* irreducible sheaf of rank n on $\mathbb{G}_m \otimes \overline{k}$ is either Kummer-induced or Lie-irreducible. The form of a Kummer-induced Kloosterman sheaf follows from [**Ka-2**, Corollary 8.9.2]. □

PROPOSITION 2.10. *Let $\mathcal{K} = \mathcal{K}_n(\psi,\chi)$ and assume that $p > 2n+1$ and $\mathcal{K}$ is not geometrically Kummer-induced. Let $G = G_{\text{geom}}(\mathcal{K})$ be the geometric monodromy group of $\mathcal{K}$. If n is odd then $G^0 = \mathrm{SL}(n)$; if n is even then G^0 is one of $\mathrm{SL}(n)$, $\mathrm{SO}(n)$, or $\mathrm{Sp}(n)$; in all cases, the inclusion $G^0 \subseteq \mathrm{SL}(\mathcal{K}) \cong \mathrm{SL}(n)$ is the standard one. When n is even, G^0 is $\mathrm{SO}(n)$ or $\mathrm{Sp}(n)$ if and only if there is a multiplicative character η of k such that $\overline{\chi} = \chi\eta$ (as unordered n-tuples). (N.B.: Since $\mathcal{K}$ is not Kummer-induced, η is unique if it exists.) If this is the case then $\left(\prod_i \chi_i\right)^2 = \eta^{-n}$; if $\prod_i \chi_i = \eta^{-n/2}$ then $G^0 = \mathrm{Sp}(n)$ and if $\prod_i \chi_i = \Lambda_2\eta^{-n/2}$ (where Λ_2 is the quadratic character on $k^\times$) then $G^0 = \mathrm{SO}(n)$.*

PROOF. The initial classification is a special case of [**Ka-2**, Theorem 8.11.3]. The criteria for $G^0 = \mathrm{Sp}(n)$ and $G^0 = \mathrm{SO}(n)$ follow from [**Ka-2**, Proposition 8.11.5 and Theorems 8.8.1, 8.8.2]. □

Remarks 2.11. (1) In particular, if $n = 2$ then $G^0 = \mathrm{Sp}(2) \cong \mathrm{SL}(2)$ (and $\eta = (\chi_1\chi_2)^{-1}$).
(2) We will not actually use the details of the classification; we just wish to know that (the Lie algebra of) G^0 is simple, except in the case $G^0 = \mathrm{SO}(4)$. Still, it is comforting to know that if we are given χ then we can easily decide whether G^0 is $\mathrm{SL}(n)$, $\mathrm{SO}(n)$, or $\mathrm{Sp}(n)$.

3. Descent of Kloosterman Sheaves

Let k be a finite field of characteristic p and $\psi\colon k \to \overline{\mathbb{Q}}_l{}^\times$ a non-trivial additive character of k (not necessarily defined over $\mathbb{F}_p$). Fix $2 \le n \in \mathbb{Z}$ and fix an n-tuple χ of multiplicative characters $\chi_i\colon k^\times \to \overline{\mathbb{Q}}_l{}^\times$. The goal of this section is to construct a sheaf $\mathcal{F}_a(\chi)$ on $\mathbb{G}_m \otimes \mathbb{F}_p$ ($a \in k^\times$) such that for all t in $\mathbb{F}_p{}^\times$, $\mathrm{tr}\left(F_t \mid \mathcal{F}_a(\chi)\right) = \pm\mathrm{Kl}_n(\psi,\chi,k,at^n)$. This sheaf will be needed in §4.2 below. Actually, there is no need to work over the prime field in this section, so we fix a finite field $\mathbb{F}_q$ of characteristic p and let $k = \mathbb{F}_{q^d}$. We will give two independent constructions. Amusingly, they give different values to the $\pm$.

3.1. First Construction.

NOTATION 3.1. Let B denote a finite, étale $\mathbb{F}_q$-algebra of degree d. (In our application, $\mathbb{F}_q = \mathbb{F}_p$ and $B = k$ is a finite extension field.) Let $\psi\colon B \to \overline{\mathbb{Q}}_l{}^\times$ denote a *non-degenerate* additive character. (*I.e.*, assume that ψ makes B into

its own Pontrjagin dual. If B is a field then non-degeneracy is the same as non-triviality.)

DEFINITION 3.2. Let $\mathbb{B}^\times$ (respectively, $\mathbb{B}$) denote the connected, smooth, commutative $\mathbb{F}_q$-group-scheme such that, for any $\mathbb{F}_q$-algebra A, $\mathbb{B}^\times(A) = \left(A \underset{\mathbb{F}_q}{\otimes} B\right)^\times$ (resp., $\mathbb{B}(A) = A \underset{\mathbb{F}_q}{\otimes} B$). Let $j\colon \mathbb{B}^\times \hookrightarrow \mathbb{B}$ denote the open immersion corresponding to $\left(A \underset{\mathbb{F}_q}{\otimes} B\right)^\times \hookrightarrow A \underset{\mathbb{F}_q}{\otimes} B$.

DEFINITION 3.3. For $\chi\colon \mathbb{B}^\times(\mathbb{F}_q) = B^\times \to \overline{\mathbb{Q}}_l{}^\times$ any multiplicative character (respectively, $\tilde{\psi}\colon \mathbb{B}(\mathbb{F}_q) = B \to \overline{\mathbb{Q}}_l{}^\times$ any additive character) let L_χ (resp., $L_{\tilde{\psi}}$) denote the rank-one, lisse, l-adic sheaf on $\mathbb{B}^\times$ (resp., on $\mathbb{B}$) constructed from the Lang torsor (*cf.* Definition 2.1). Thus, for any finite extension $E/\mathbb{F}_q$ and any $t \in \mathbb{B}^\times(E)$ (resp., $t \in \mathbb{B}(E)$)

$$\begin{aligned} \operatorname{tr}\left(F_{E,t} \mid L_\chi\right) &= \chi\Big(\prod_{\sigma \in \operatorname{Gal}(E/\mathbb{F}_q)} \sigma(t)\Big); \\ \operatorname{tr}\left(F_{E,t} \mid L_{\tilde{\psi}}\right) &= \tilde{\psi}\Big(\sum_{\sigma \in \operatorname{Gal}(E/\mathbb{F}_q)} \sigma(t)\Big), \end{aligned}$$

where the product is in $\mathbb{B}^\times(E)$ and the sum is in $\mathbb{B}(E)$. We will often write $L_{\tilde{\psi}}$ instead of $j^* L_{\tilde{\psi}}$. Note that we distinguish notationally between L_χ and $L_{\tilde{\psi}}$, just defined, and the standard $\mathcal{L}_\chi$, $\mathcal{L}_\psi$ on $\mathbb{G}_m$ (Definition 2.1).

Remark 3.4. These objects are all defined in [**Ka-1**, §8.8]. However, there is no discussion there of convolution on $\mathbb{B}^\times$, as in the following definition.

DEFINITION 3.5. For an n-tuple χ of multiplicative characters $\chi_i\colon B^\times \to \overline{\mathbb{Q}}_l{}^\times$, define the **Kloosterman complex** $K = \mathrm{Kl}_{n,B}(\psi, \chi) \in D_c^b(\mathbb{B}^\times, \overline{\mathbb{Q}}_l)$ as the multiple convolution (*cf.* Definition 2.2)

$$\mathrm{Kl}_{n,B}(\psi, \chi) \overset{def}{=} \overset{n}{\underset{i=1}{*_!}} \left(L_{\chi_i} \otimes L_\psi[d]\right),$$

so that, for any finite extension $E/\mathbb{F}_q$ and any $t \in \mathbb{B}^\times(E)$,

$$\operatorname{tr}\left(F_{E,t} \mid K\right) = (-1)^{nd} \sum_{\substack{x_1 \cdots x_n = t \\ x_i \in \mathbb{B}^\times(E)}} \prod_{i=1}^{n} \psi\Big(\sum_\sigma \sigma(x_i)\Big) \chi_i\Big(\prod_\sigma \sigma(x_i)\Big),$$

where σ runs over $\operatorname{Gal}(E/\mathbb{F}_q)$.

NOTATION 3.6. For the rest of this section, let $B = k = \mathbb{F}_{q^d}$ and write G for $\mathbb{B}^\times$.

Remarks 3.7. (1) For any n-tuple χ of multiplicative characters $\chi_i\colon k^\times \to \overline{\mathbb{Q}}_l{}^\times$ and any $t \in k^\times = G(\mathbb{F}_q)$, we have

$$\operatorname{tr}\left(F_{\mathbb{F}_q,t} \mid \mathrm{Kl}_{n,k}(\psi, \chi)\right) = (-1)^{nd} \mathrm{Kl}_n(\psi, \chi, k, t).$$

(2) We would like to show that $K = \mathrm{Kl}_{n,k}(\psi, \chi)$ consists of a single lisse sheaf, placed in degree $-d$. It suffices to prove this for the stalk K_a at every geometric point $a \in G(\overline{\mathbb{F}}_q)$; alternately, it suffices to prove this for $K|G \otimes \overline{\mathbb{F}}_q$. Furthermore, we will later need additional information about $K|G \otimes \overline{\mathbb{F}}_q$. In fact, it suffices to pull back to $G \otimes k$.

LEMMA 3.8. $G \underset{\mathbb{F}_q}{\otimes} k \cong \prod_{\sigma \in \mathrm{Gal}(k/\mathbb{F}_q)} \mathbb{G}_m \otimes k$. *Under this isomorphism,* $K|G \otimes k$ *corresponds to the external tensor product* $\boxtimes_\sigma \mathrm{Kl}_n(\psi \circ \sigma^{-1}, \chi \circ \sigma^{-1})$.

PROOF. $G \otimes k$ represents the functor (on k-algebras) $A \mapsto \left(A \underset{\mathbb{F}_q}{\otimes} k\right)^\times$ and $\prod_\sigma \mathbb{G}_m \otimes k$ represents $A \mapsto \prod_\sigma A^\times$. Since

$$\left(A \underset{\mathbb{F}_q}{\otimes} k\right)^\times = \left(A \underset{k}{\otimes} k \underset{\mathbb{F}_q}{\otimes} k\right)^\times \xrightarrow{\sim} \left(A \underset{k}{\otimes} \prod_{\sigma \in \mathrm{Gal}(k/\mathbb{F}_q)} k\right)^\times = \prod_{\sigma \in \mathrm{Gal}(k/\mathbb{F}_q)} A^\times,$$

the first part follows.

Now let $\chi\colon k^\times \to \overline{\mathbb{Q}}_l{}^\times$ be a multiplicative character and consider $L_\chi \mid G \otimes k$. It is a lisse, rank-one sheaf, as is the external tensor product $\boxtimes_{\sigma \in \mathrm{Gal}(k/\mathbb{F}_q)} \mathcal{L}_{\chi \circ \sigma^{-1}}$. Checking traces (at $t \in G(E)$ for any finite extension E/k), one finds that L_χ on $G \otimes k$ corresponds to $\boxtimes_{\sigma \in \mathrm{Gal}(k/\mathbb{F}_q)} \mathcal{L}_{\chi \circ \sigma^{-1}}$ on $\prod_\sigma \mathbb{G}_m \otimes k$. Similarly, L_ψ on $G \otimes k$ corresponds to $\boxtimes_{\sigma \in \mathrm{Gal}(k/\mathbb{F}_q)} \mathcal{L}_{\psi \circ \sigma^{-1}}$ on $\prod_\sigma \mathbb{G}_m \otimes k$. Since $*_!$ commutes with base changes, it follows that

$$K \mid G \otimes k \overset{def}{=} \overset{n}{\underset{i=1}{*_!}} (L_{\chi_i} \otimes L_\psi[d]) \mid G \otimes k$$

corresponds to

$$\overset{n}{\underset{i=1}{*_!}} \boxtimes_\sigma (\mathcal{L}_{\chi \circ \sigma^{-1}} \otimes \mathcal{L}_{\psi \circ \sigma^{-1}}[1]) \cong \boxtimes_{\sigma \in \mathrm{Gal}(k/\mathbb{F}_q)} \mathrm{Kl}_n(\psi \circ \sigma^{-1}, \chi \circ \sigma^{-1}). \quad \square$$

COROLLARY 3.9. *The Kloosterman complex* K *consists of a single lisse sheaf placed in degree* $-d = -[k : \mathbb{F}_q]$; *say* $K = \mathcal{K}[d]$. *Thus (identifying* $k^\times$ *with* $G(\mathbb{F}_q)$*)*

$$(\forall t \in k^\times)\ \mathrm{tr}\left(F_{\mathbb{F}_q, t} \mid \mathcal{K}\right) = (-1)^{nd-d} \mathrm{Kl}_n(\psi, \chi, k, t).$$

PROOF. The second statement follows from the first and from Definition 3.5. As in Remark 3.7(2), the first statement follows from the corresponding fact about $K \mid G \otimes \overline{\mathbb{F}}_q$; by Lemma 3.8 and Theorem 2.4,

$$\begin{aligned} K \mid G \otimes \overline{\mathbb{F}}_q &= \boxtimes_{\sigma \in \mathrm{Gal}(k/\mathbb{F}_q)} \mathcal{K}_n\left(\psi \circ \sigma^{-1}, \chi \circ \sigma^{-1}\right)[1] \\ &= \left(\boxtimes_{\sigma \in \mathrm{Gal}(k/\mathbb{F}_q)} \mathcal{K}_n(\psi \circ \sigma^{-1} \chi \circ \sigma^{-1})\right)[d]. \quad \square \end{aligned}$$

DEFINITION 3.10. An additive character $\psi\colon k \to \overline{\mathbb{Q}}_l^{\times}$ is **defined over** $\mathbb{F}_q$ if and only if $\psi = \psi_0 \circ \mathrm{tr}_{k/\mathbb{F}_q}$ for some additive character $\psi_0\colon \mathbb{F}_q \to \overline{\mathbb{Q}}_l^{\times}$. A multiplicative character $\chi\colon k^{\times} \to \overline{\mathbb{Q}}_l^{\times}$ is **defined over** $\mathbb{F}_q$ if and only if $\chi = \chi_0 \circ \mathrm{N}_{k/\mathbb{F}_q}$ for some multiplicative character $\chi_0\colon \mathbb{F}_q^{\times} \to \overline{\mathbb{Q}}_l^{\times}$. An n-tuple χ of multiplicative characters $\chi_i\colon k^{\times} \to \overline{\mathbb{Q}}_l^{\times}$ is **defined over** $\mathbb{F}_q$ if and only if for all $\sigma \in \mathrm{Gal}(k/\mathbb{F}_q)$, there is a permutation s of $\{1, \dots, n\}$ such that $(\forall i)\ \chi_i \circ \sigma = \chi_{s(i)}$, *i.e.*, $\chi \circ \sigma = \chi$ as unordered n-tuples.

Example 3.10.1. The n-tuple $\mathbf{1} = (\mathbf{1}, \dots, \mathbf{1})$, which gives rise to the "standard" Kloosterman sheaf, is defined over $\mathbb{F}_q$.

COROLLARY 3.11. *With notation as in Lemma* 3.8 *and Corollary* 3.9, *assume that* ψ *and* χ *are defined over* $\mathbb{F}_q$. *Then*

$$\mathcal{K} \mid G \otimes k \xrightarrow{\sim} \boxtimes_{\sigma \in \mathrm{Gal}(k/\mathbb{F}_q)} \mathcal{K}_n(\psi, \chi). \quad \square$$

THEOREM 3.12. *For all* $a \in k^{\times}$ *and any* n*-tuple* χ *of multiplicative characters* $\chi_i\colon k^{\times} \to \overline{\mathbb{Q}}_l^{\times}$ *there is a lisse sheaf* $\mathcal{F}_a(\chi)$ *on* $\mathbb{G}_m \otimes \mathbb{F}_q$ *such that*

$$(\forall t \in \mathbb{F}_q^{\times})\ \mathrm{tr}\left(F_{\mathbb{F}_q, t} \mid \mathcal{F}_a(\chi)\right) = (-1)^{nd-d}\mathrm{Kl}_n(\psi, \chi, k, at^n);$$

furthermore,

$$\mathcal{F}_a(\chi) \mid \mathbb{G}_m \otimes k = \bigotimes_{\sigma \in \mathrm{Gal}(k/\mathbb{F}_q)} \left(t \mapsto \sigma(a)t^n\right)^* \mathcal{K}_n(\psi \circ \sigma^{-1}, \chi \circ \sigma^{-1}).$$

PROOF. There is a homomorphism $f_a(A)\colon A^{\times} \to \left(A \underset{\mathbb{F}_q}{\otimes} k\right)^{\times}$, where A is any $\mathbb{F}_q$-algebra, given by $t \mapsto t^n \otimes a$. (We apologize for the notation: $a \in k^{\times}$, $t \in A$.) Since these homomorphisms are compatible with maps $A \to A'$, they come from a map $f_a\colon \mathbb{G}_m \to G$. Let $\mathcal{F}_a \overset{def}{=} f_a^* \mathcal{K}$, with $\mathcal{K}$ as in Corollary 3.9; clearly $\mathcal{F}_a$ has the desired $\mathbb{F}_q$-traces.

For any k-algebra A, the composition

$$A^{\times} \xrightarrow{f_a(A)} \left(A \underset{\mathbb{F}_q}{\otimes} k\right)^{\times} \xrightarrow{\sim} \prod_{\sigma \in \mathrm{Gal}(k/\mathbb{F}_q)} A^{\times}$$

corresponding to $\mathbb{G}_m \otimes k \xrightarrow{f_a} G \otimes k \cong \prod_{\sigma} \mathbb{G}_m \otimes k$ is $t \mapsto \left(\sigma(a)t^n\right)_{\sigma}$, so the result follows from Lemma 3.8. $\square$

COROLLARY 3.13. *The sheaf* $\mathcal{F}_a(\chi)$ *of Theorem* 3.12 *has rank* n^d *and is pure of weight* $d(n-1)$. *It is tame at* 0 *and its* ∞*-breaks are at most* 1.

PROOF. The Kloosterman sheaves $\mathcal{K}_{n, \psi \circ \sigma^{-1}}\left(\chi_1 \circ \sigma^{-1}, \dots, \chi_n \circ \sigma^{-1}\right)$ have rank n; they are pure of weight $n-1$; they are tame at 0; and they have all ∞-breaks $\frac{1}{n}$ by Theorem 2.4. Therefore $\mathcal{F}_a \mid \mathbb{G}_m \otimes k$ has rank n^d; it is pure of weight $d(n-1)$; it is tame at 0; and all its ∞-breaks are ≤ 1. It follows that the same holds for $\mathcal{F}_a$ itself (*cf.* Remarks 3.7). $\square$

Remark 3.14. Similarly, the map $A^\times \to (A \underset{\mathbb{F}_q}{\otimes} k)^\times$ given by $t \mapsto t \otimes 1$ allows us to pull back the Lang-torsor sheaves L_χ, $L_{\tilde{\psi}}$ on G to lisse, rank-one sheaves $\mathcal{L}_\chi$, $\mathcal{L}_{\tilde{\psi}}$ (abuse of notation) on $\mathbb{G}_m \otimes \mathbb{F}_q$. We have, for $t \in \mathbb{F}_q{}^\times$, $\operatorname{tr}(F_t \mid \mathcal{L}_\chi) = \chi(t)$, $\operatorname{tr}(F_t \mid \mathcal{L}_{\tilde{\psi}}) = \tilde{\psi}(t)$.

3.2. Tensor-Induction. The definition of tensor-induction, and its basic properties, can be found in [**Ka-2**, §§10.3–10.5]. First we show how, in certain cases, to calculate the trace function of the tensor-induction of a lisse sheaf. Once this is done, it will be easy to give an alternate proof of Theorem 3.12.

PROPOSITION 3.15. *Let $G = F^{\mathbb{Z}}$ be an infinite cyclic group, $H = F^{n\mathbb{Z}} \subseteq G$, and let V be a representation of H. Then* $\operatorname{tr}(F \mid \otimes\text{-}\mathrm{Ind}_{H\subseteq G} V) = \operatorname{tr}(F^n \mid V)$.

PROOF. Let the characteristic polynomials of F on $\otimes\text{-}\mathrm{Ind}_{H\subseteq G} V$ and of F^n on V be

$$\begin{aligned} P_{\otimes n}(T) &\overset{def}{=} \det\left(T - F \mid \otimes\text{-}\mathrm{Ind}_{H\subseteq G} V\right); \\ P(T) &\overset{def}{=} \det\left(T - F^n \mid V\right) = \prod_{i=1}^{r} (T - \lambda_i). \end{aligned}$$

We have

$$\begin{aligned} P_{\otimes n}(T) &= T^{r^n} - \operatorname{tr}\left(F \mid \otimes\text{-}\mathrm{Ind}_{H\subseteq G} V\right) T^{r^n-1} + \cdots; \\ P(T) &= T^r - \operatorname{tr}\left(F^n \mid V\right) T^{r-1} + \cdots. \end{aligned}$$

By [**Ka-2**, 10.4.1 and 10.4.3],

$$P_{\otimes n}(T) = \prod_{Z} \left(T^{\#Z} - \lambda(Z)\right),$$

where Z runs over the orbits of $\{1, \ldots, r\}^n$ under cyclic permutation and if $(a_1, \ldots, a_n) = (a_1, \ldots, a_m, \ldots, a_1, \ldots, a_m) \in Z$ $(m = \#Z)$ then

$$\lambda(Z) \overset{def}{=} \prod_{i=1}^{\#Z} \lambda_{a_i}.$$

Separating out the singleton orbits $Z_i \overset{def}{=} \{(i, \ldots, i)\}$ (so that $\#Z_i = 1$),

$$\begin{aligned} P_{\otimes n}(T) &= \prod_{i=1}^{r} (T - \lambda_i) \cdot \prod_{\#Z>1} \left(T^{\#Z} - \lambda(Z)\right) \\ &= T^{r^n} - \sum_{i=1}^{r} \lambda_i \cdot T^{r^n-1} + \cdots, \end{aligned}$$

so $\operatorname{tr}(F \mid \otimes\text{-}\mathrm{Ind}_{H\subseteq G} V) = \sum_{i=1}^{r} \lambda_i = \operatorname{tr}(F^n \mid V)$. □

LEMMA 3.16. *Let E be a finite field, E'/E an extension of (finite) degree n, and fix an algebraic closure $\overline{E}$ of E, containing E'. Let $F_E \in \mathrm{Gal}(\overline{E}/E)$ be the (geometric) Frobenius element, $F_{E'} = (F_E)^n$, and let $\pi\colon \operatorname{Spec} E' \to \operatorname{Spec} E$ be the natural map. If $\mathcal{F}$ is a lisse sheaf on $\operatorname{Spec} E'$ (i.e., a $\mathrm{Gal}(\overline{E}/E')$-module) then $\operatorname{tr}(F_E \mid \pi_{\otimes *}\mathcal{F}) = \operatorname{tr}(F_{E'} \mid \mathcal{F})$.*

PROOF. We apply the "Mackey Subgroup Theorem for Tensor-Induction" ([**Ev**], quoted as [**Ka-2**, Theorem 10.3.3]) with $G = \mathrm{Gal}(\overline{E}/E) = {F_E}^{\hat{\mathbb{Z}}}$, $H = \mathrm{Gal}(\overline{E}/E') = {F_{E'}}^{\hat{\mathbb{Z}}} = {F_E}^{n\hat{\mathbb{Z}}}$, and $K = {F_E}^{\mathbb{Z}}$. Since $K \twoheadrightarrow G/H$, we have

$$\pi_{\otimes *}\mathcal{F} \mid K \overset{def}{=} \otimes\text{-}\mathrm{Ind}_{H \subseteq G}\, \mathcal{F} \mid K = \otimes\text{-}\mathrm{Ind}_{H \cap K \subseteq K}\, \mathcal{F}.$$

Now apply the previous proposition to $H \cap K \subseteq K$, *i.e.*, ${F_E}^{n\mathbb{Z}} \subseteq {F_E}^{\mathbb{Z}}$. □

PROPOSITION 3.17. *Let $f\colon Y \to X$ be a finite, étale map of degree n between connected schemes, E a finite field, $t \in X(E)$. Assume that the fiber $Y_t \overset{def}{=} Y \times_X \operatorname{Spec} E$ is connected. Then $Y_t = \operatorname{Spec} E'$, where E'/E is the extension of degree n; and for any lisse sheaf $\mathcal{F}$ on Y, $\operatorname{tr}(F_{E,t} \mid f_{\otimes *}\mathcal{F}) = \operatorname{tr}(F_{E',t_Y} \mid \mathcal{F})$.*

$$\begin{array}{ccccc} \operatorname{Spec} E' & = & Y_t & \xrightarrow{t_Y} & Y \\ & & {\scriptstyle f_E}\downarrow & & \downarrow{\scriptstyle f} \\ & & \operatorname{Spec} E & \xrightarrow[t]{} & X \end{array}$$

PROOF. Since $f_E\colon Y_t \to \operatorname{Spec} E$ is étale of degree n, the connectedness hypothesis implies that $Y_t = \operatorname{Spec} E'$. For any lisse sheaf $\mathcal{F}$ on Y,

$$\operatorname{tr}\left(F_{E,t} \mid f_{\otimes *}\mathcal{F}\right) = \operatorname{tr}\left(F_E \mid t^* f_{\otimes *}\mathcal{F}\right) = \operatorname{tr}\left(F_E \mid (f_E)_{\otimes *} t_Y{}^* \mathcal{F}\right)$$

(by the definition of $F_{E,t}$ and by [**Ka-2**, Proposition 10.5.3]); applying the lemma to $t_Y{}^*\mathcal{F}$, we obtain

$$\operatorname{tr}\left(F_E \mid (f_E)_{\otimes *} t_Y{}^* \mathcal{F}\right) = \operatorname{tr}\left(F_{E'} \mid t_Y{}^* \mathcal{F}\right) = \operatorname{tr}\left(F_{E',t_Y} \mid \mathcal{F}\right). \quad \square$$

ALTERNATE PROOF OF THEOREM 3.12. Let $\mathcal{K} \overset{def}{=} \mathcal{K}_n(\psi, \chi)$ be the standard Kloosterman sheaf on $\mathbb{G}_m \otimes k$. Let $\pi\colon \mathbb{G}_m \otimes k \to \mathbb{G}_m \otimes \mathbb{F}_q$ be the natural map and let

$$\mathcal{F}_a(\chi) \overset{def}{=} \pi_{\otimes *}\left((t \mapsto at^n)^* \mathcal{K}\right).$$

Up to an unimportant sign, $\mathcal{F}_a(\chi)$ has the desired $\mathbb{F}_q$-traces by the preceding proposition; to get rid of the sign, we could tensor it with $\alpha^{\deg}$, where $\alpha = (-1)^{(n-1)(d-1)}$. By [**Ka-2**, Proposition 10.5.2] $\mathcal{F}_a(\chi) \mid \mathbb{G}_m \otimes k$ has the desired tensor-decomposition. □

Remark 3.18. This method also gives a descent of $\mathcal{L}_\chi$ and $\mathcal{L}_\psi$ from $\mathbb{G}_m \otimes k$ to $\mathbb{G}_m \otimes \mathbb{F}_q$; *cf.* Remark 3.14.

4. Proof of Distinctness

Let $k = \mathbb{F}_q$ ($q = p^d$) be a finite field and let $\psi: k \to \overline{\mathbb{Q}}_l^{\times}$ ($l \neq p$) be a non-trivial additive character. Fix an integer $n \geq 2$. The main goal of this section is to show that if p is sufficiently large (with respect to n and d) then the $q-1$ Kloosterman sums

$$\mathrm{Kl}_n(a) = \mathrm{Kl}_n(\psi, \mathbf{1}, k, a) \quad (a \in k^{\times})$$

are distinct up to $\mathbb{F}_p$-conjugacy; more precisely, if ψ is defined over $\mathbb{F}_p$ and if $\mathrm{Kl}_n(a) = \mathrm{Kl}_n(b)$ then $b = \sigma(a)$ for some $\sigma \in \mathrm{Gal}(k/\mathbb{F}_p)$.

We will also consider Kloosterman sums with characters, $\mathrm{Kl}_n(\psi, \chi, k, a)$. There is an essential restriction: χ should not be Kummer-induced (*cf.* Remark 1.6). We also assume that (the identity component of) the geometric monodromy group of $\mathcal{K}_n(\psi, \chi)$ is not SO(4). (*Cf.* Remark 2.11(2).) This restriction, as well as the unsatisfactory bound $p > \left(2n^{2d}+1\right)^2$, may be an artifact of our method. *Cf.* Remarks 4.28.

After making some elementary observations, including the reduction to the case that ψ is defined over $\mathbb{F}_p$, we prove some technical lemmas. Then we show that if $\mathrm{Kl}_n(a) = \mathrm{Kl}_n(b)$ (and p is sufficiently large) then a and b are conjugate over $\mathbb{F}_p$; there are additional restrictions if we consider Kloosterman sums with characters. We also give some corollaries of the main theorem and consider how it might be strengthened.

4.1. Elementary remarks.

Remark 4.1. If ψ, $\psi': k \to \overline{\mathbb{Q}}_l^{\times}$ are non-trivial additive characters then, for some $\alpha \in k^{\times}$, $\psi'(x) = \psi(\alpha x)$.

PROPOSITION 4.2. *Let $\psi: k \to \overline{\mathbb{Q}}_l^{\times}$ be a non-trivial additive character defined over $\mathbb{F}_p$, $\alpha \in k^{\times}$, $\psi'(x) = \psi(\alpha x)$. Let $\chi = (\chi_1, \ldots, \chi_n)$ be an n-tuple of multiplicative characters. Then for $a, b \in k^{\times}$, $\mathrm{Kl}_n(\psi', \chi, k, a) = \mathrm{Kl}_n(\psi', \chi, k, b)$ if and only if $\mathrm{Kl}_n(\psi, \chi, k, a\alpha^n) = \mathrm{Kl}_n(\psi, \chi, k, b\alpha^n)$.*

PROOF. An easy calculation shows that, for any $t \in k^{\times}$,

$$\mathrm{Kl}_n(\psi', \chi, k, t) = \left(\prod_{i=1}^{n} \chi_i\right)\left(\frac{1}{\alpha}\right) \cdot \mathrm{Kl}_n(\psi, \chi, k, t\alpha^n). \quad \square$$

PROPOSITION 4.3. *Let ψ be a non-trivial additive character of k defined over $\mathbb{F}_p$. Let χ be an n-tuple of multiplicative characters $\chi_i: k^{\times} \to \overline{\mathbb{Q}}_l^{\times}$ and let $\sigma \in \mathrm{Gal}(k/\mathbb{F}_p)$, $a \in k^{\times}$. Then $\mathrm{Kl}_n(\psi, \chi, k, a) = \mathrm{Kl}_n(\psi, \chi \circ \sigma^{-1}, k, \sigma(a))$.*

PROOF. Since $\psi \circ \sigma^{-1} = \psi$, the result follows easily from the definition. $\square$

COROLLARY 4.4. *With the notations of Proposition 4.3, assume ψ and χ are defined over $\mathbb{F}_p$. Then $\mathrm{Kl}_n(\psi, \chi, k, a) = \mathrm{Kl}_n(\psi, \chi, k, \sigma(a))$.* $\square$

Remark 4.5. One consequence of Theorem 4.22 below is a converse of Corollary 4.4: if ψ and χ are defined over $\mathbb{F}_p$ and $\mathrm{Kl}_n(\psi, \chi, k, a) = \mathrm{Kl}_n(\psi, \chi, k, b)$ then a and b are conjugate over $\mathbb{F}_p$. Theorem 4.22 also provides a converse of Proposition 4.3.

4.2. Preliminary Lemmas.

Remarks 4.6. Fix a non-trivial additive character $\psi\colon k \to \overline{\mathbb{Q}}_l^{\times}$. The purpose of this subsection is to translate the equality of two Kloosterman sums into a non-trivial map of sheaves on $\mathbb{G}_m \otimes \overline{\mathbb{F}}_p$ (assuming p is sufficiently large). In the following lemma, we let $\zeta_p, \zeta_{q-1} \in \overline{\mathbb{Q}}_l^{\times}$ be primitive roots of unity ($q = \#k$), so that Kloosterman sums defined over k lie in $\mathbb{Q}[\zeta_p, \zeta_{q-1}] \subseteq \overline{\mathbb{Q}}_l$. Lemma 4.9 below is where we use the "big guns" of étale cohomology. We state it for sheaves over any finite field k even though we use it only for sheaves defined over $\mathbb{F}_p$.

Recall that if $\mathcal{F}$ is a lisse sheaf on a normal scheme of finite type over a finite field (such as $\mathbb{G}_m$ over k) and $\mathcal{F}$ is pure of some weight then $\mathcal{F}$ is geometrically semisimple [**Del**, Theorem 3.4.1]. This is important for the statement of Lemma 4.9 below.

LEMMA 4.7. *Let χ be an n-tuple of multiplicative characters $\chi_i\colon k^{\times} \to \overline{\mathbb{Q}}_l^{\times}$. Let $\mathrm{Kl}(b) = \mathrm{Kl}_n(\psi, \chi, k, b)$ $(b \in k^{\times})$. For $t \in \mathbb{F}_p^{\times}$, let $\sigma_t \in \mathrm{Gal}(\mathbb{Q}[\zeta_p, \zeta_{q-1}]/\mathbb{Q})$ denote the automorphism such that $\sigma_t(\zeta_p) = \zeta_p^t$, $\sigma_t(\zeta_{q-1}) = \zeta_{q-1}$. For all $a \in k^{\times}$ and $t \in \mathbb{F}_p^{\times}$,*

$$\sigma_t\big(\mathrm{Kl}(a)\big) = \prod_{i=1}^{n} \chi_i\left(\frac{1}{t}\right) \cdot \mathrm{Kl}(at^n).$$

PROOF. Just use the facts that, for all x in $k^{\times}$, $\sigma\big(\psi(x)\big) = \psi(x)^t = \psi(tx)$ and $\sigma\big(\chi_i(x)\big) = \chi_i(x)$. □

COROLLARY 4.8. *Let χ and ρ be n-tuples of multiplicative characters, say $\chi_i, \rho_j\colon k^{\times} \to \overline{\mathbb{Q}}_l^{\times}$, and let $a, b \in k^{\times}$. Let $\mathcal{F}_a(\chi)$, $\mathcal{F}_b(\rho)$ be the sheaves on $\mathbb{G}_m \otimes \mathbb{F}_p$ of Theorem* 3.12 *and let $\mathcal{L}_{\Pi\overline{\chi}}$, $\mathcal{L}_{\Pi\overline{\rho}}$ be as in Remark* 3.14 *or Remark* 3.18 *(where $\Pi\overline{\chi}$ denotes $\prod_{i=1}^n \overline{\chi}_i = \prod_{i=1}^n \chi_i{}^{-1}$; similarly for $\Pi\overline{\rho}$). If $\mathrm{Kl}(\psi, \chi, k, a) = \mathrm{Kl}_n(\psi, \rho, k, b)$ then for all $t \in \mathbb{F}_p^{\times}$,*

$$\mathrm{tr}\big(F_{\mathbb{F}_p, t} \mid \mathcal{F}_a(\chi) \otimes \mathcal{L}_{\Pi\overline{\chi}}\big) = \mathrm{tr}\big(F_{\mathbb{F}_p, t} \mid \mathcal{F}_b(\rho) \otimes \mathcal{L}_{\Pi\overline{\rho}}\big).$$

PROOF. If $\mathrm{Kl}(a) = \mathrm{Kl}(b)$ then, for $t \in \mathbb{F}_p^{\times}$, $\sigma_t\big(\mathrm{Kl}(a)\big) = \sigma_t\big(\mathrm{Kl}(b)\big)$. □

LEMMA 4.9. *Let $k = \mathbb{F}_q$ be a finite field of characteristic p. Let $\mathcal{F}$, $\mathcal{F}'$ be lisse sheaves on $\mathbb{G}_m \otimes k$ of the same rank r and pure of the same weight w. Assume $(\forall t \in k^{\times})$ $\mathrm{tr}\big(F_t \mid \mathcal{F}\big) = \mathrm{tr}\big(F_t \mid \mathcal{F}'\big)$. Let $\mathcal{G}$ be a geometrically irreducible sheaf of rank s on $\mathbb{G}_m \otimes k$, also pure of weight w, such that $\mathcal{G} \mid \mathbb{G}_m \otimes \overline{k}$ occurs* exactly *once in (the decomposition into irreducibles of) $\mathcal{F} \mid \mathbb{G}_m \otimes \overline{k}$. Then $\mathcal{G} \mid \mathbb{G}_m \otimes \overline{k}$ occurs* at least *once in $\mathcal{F}' \mid \mathbb{G}_m \otimes \overline{k}$, provided that $q > \big[2rs(M_0 + M_\infty) + 1\big]^2$,*

where M_0 is the largest 0-break of $\mathcal{F} \oplus \mathcal{F}'$ and M_∞ is the largest ∞-break of $\mathcal{F} \oplus \mathcal{F}'$.

PROOF. We will assume that $\mathcal{G} \mid \mathbb{G}_m \otimes \overline{k}$ does *not* occur in $\mathcal{F}' \mid \mathbb{G}_m \otimes \overline{k}$ and derive an upper bound for q. We reduce to the case $w = 0$ by replacing $\mathcal{F}$, $\mathcal{F}'$, and $\mathcal{G}$ by $\mathcal{F}(\frac{w}{2})$, $\mathcal{F}'(\frac{w}{2})$, and $\mathcal{G}(\frac{w}{2})$.

Applying the Lefschetz Trace Formula (*cf.* [**Ka-1**, §2.3]) to $\mathcal{G}^\vee \otimes \mathcal{F}$ and to $\mathcal{G}^\vee \otimes \mathcal{F}'$ and using (for t in $k^\times$)

$$\begin{aligned} \operatorname{tr}(F_t \mid \mathcal{G}^\vee \otimes \mathcal{F}) &= \operatorname{tr}(F_t \mid \mathcal{G}^\vee)\operatorname{tr}(F_t \mid \mathcal{F}) = \operatorname{tr}(F_t \mid \mathcal{G}^\vee)\operatorname{tr}(F_t \mid \mathcal{F}') \\ &= \operatorname{tr}(F_t \mid \mathcal{G}^\vee \otimes \mathcal{F}') \end{aligned}$$

we conclude that (letting $H^i_c(-) = H^i_c(\mathbb{G}_m \otimes \overline{k}, -)$)

$$\sum_{i=0}^{2}(-1)^i \operatorname{tr}\Big(F_k \mid H^i_c(\mathcal{G}^\vee \otimes \mathcal{F})\Big) = \sum_{i=0}^{2}(-1)^i \operatorname{tr}\Big(F_k \mid H^i_c(\mathcal{G}^\vee \otimes \mathcal{F}')\Big).$$

Since $\mathcal{G}^\vee \otimes \mathcal{F}$ and $\mathcal{G}^\vee \otimes \mathcal{F}'$ are lisse and $\mathbb{G}_m \otimes \overline{k}$ is affine, $H^0_c(\mathbb{G}_m \otimes \overline{k}, \mathcal{G}^\vee \otimes \mathcal{F}) = 0$ and $H^0_c(\mathbb{G}_m \otimes \overline{k}, \mathcal{G}^\vee \otimes \mathcal{F}') = 0$. Because we assume that $\mathcal{G} \mid \mathbb{G}_m \otimes \overline{k}$ occurs exactly once in $\mathcal{F} \mid \mathbb{G}_m \otimes \overline{k}$ and not at all in $\mathcal{F}' \mid \mathbb{G}_m \otimes \overline{k}$,

$$H^2_c(\mathbb{G}_m \otimes \overline{k}, \mathcal{G}^\vee \otimes \mathcal{F}) = Hom(\mathcal{G}, \mathcal{F})_{\pi_1(\mathbb{G}_m \otimes \overline{k})}(-1)$$

is one-dimensional and pure of weight 2, and

$$H^2_c(\mathbb{G}_m \otimes \overline{k}, \mathcal{G}^\vee \otimes \mathcal{F}') = Hom(\mathcal{G}, \mathcal{F}')_{\pi_1(\mathbb{G}_m \otimes \overline{k})}(-1) = 0.$$

By Weil II (*i.e.*, [**Del**, Théorème 3.3.1]), $H^1_c(\mathcal{G}^\vee \otimes \mathcal{F})$ and $H^1_c(\mathcal{G}^\vee \otimes \mathcal{F}')$ are mixed of weight ≤ 1. Therefore

$$\begin{aligned} q &= \left|\operatorname{tr}(F_k \mid H^2_c(\mathbb{G}_m \otimes \overline{k}, \mathcal{G}^\vee \otimes \mathcal{F}))\right| \\ &= \left|\operatorname{tr}(F_k \mid H^1_c(\mathbb{G}_m \otimes \overline{k}, \mathcal{G}^\vee \otimes \mathcal{F})) - \operatorname{tr}(F_k \mid H^1_c(\mathbb{G}_m \otimes \overline{k}, \mathcal{G}^\vee \otimes \mathcal{F}'))\right|; \\ q &\leq \sqrt{q}\left[h^1_c(\mathbb{G}_m \otimes \overline{k}, \mathcal{G}^\vee \otimes \mathcal{F}) + h^1_c(\mathbb{G}_m \otimes \overline{k}, \mathcal{G}^\vee \otimes \mathcal{F}')\right]. \end{aligned}$$

By the Euler-Poincaré Formula (*cf.* [**Ka-1**, §2.3]) (and the above values for H^0_c and H^2_c)

$$\begin{aligned} h^1_c(\mathbb{G}_m \otimes \overline{k}, \mathcal{G}^\vee \otimes \mathcal{F}) &= \operatorname{Sw}_0(\mathcal{G}^\vee \otimes \mathcal{F}) + \operatorname{Sw}_\infty(\mathcal{G}^\vee \otimes \mathcal{F}) + 1; \\ h^1_c(\mathbb{G}_m \otimes \overline{k}, \mathcal{G}^\vee \otimes \mathcal{F}') &= \operatorname{Sw}_0(\mathcal{G}^\vee \otimes \mathcal{F}') + \operatorname{Sw}_\infty(\mathcal{G}^\vee \otimes \mathcal{F}'). \end{aligned}$$

Since $\mathcal{G} \mid \mathbb{G}_m \otimes \overline{k} \hookrightarrow \mathcal{F} \mid \mathbb{G}_m \otimes \overline{k}$, the largest 0-break of $\mathcal{G}$ is at most M_0; thus the same is true of $\mathcal{G}^\vee$ and of $\mathcal{G}^\vee \otimes \mathcal{F}$ and $\mathcal{G}^\vee \otimes \mathcal{F}'$. Thus $\operatorname{Sw}_0(\mathcal{G}^\vee \otimes \mathcal{F})$, $\operatorname{Sw}_0(\mathcal{G}^\vee \otimes \mathcal{F}') \leq rsM_0$. Similarly, $\operatorname{Sw}_\infty(\mathcal{G}^\vee \otimes \mathcal{F})$, $\operatorname{Sw}_\infty(\mathcal{G}^\vee \otimes \mathcal{F}') \leq rsM_\infty$. Therefore

$$q \leq \left[2rs(M_0 + M_\infty) + 1\right]^2. \quad \square$$

Remarks 4.10. In Lemma 4.9, if $\mathcal{G} \subseteq \mathcal{F}$ (as sheaves on $\mathbb{G}_m \otimes k$) then we actually have $H_c^2(\mathbb{G}_m \otimes \overline{k}, \mathcal{G}^\vee \otimes \mathcal{F}) = \overline{\mathbb{Q}}_l(-1)$. For our purposes, it is enough to consider this case: the sheaf $\mathcal{G}$ we will construct is automatically a subsheaf of $\mathcal{F}$. Also, the special case $\mathcal{G} = \mathcal{F}$ can be useful: for example, when considering the distinctness of Kloosterman sums for $\mathbb{F}_p$, $\mathcal{F}_a(\chi) = (t \mapsto at^n)^*\mathcal{K}_n(\psi, \chi)$ is itself geometrically irreducible.

LEMMA 4.11. *Let $a, b \in \overline{\mathbb{F}}_q^\times$; let χ and ρ be n-tuples of multiplicative characters $\chi_i, \rho_j : k^\times \to \overline{\mathbb{Q}}_l^\times$ with $p \nmid n$; and let η be a tame multiplicative character of $\overline{\mathbb{F}}_q$. Let $\mathcal{H} \stackrel{def}{=} \mathcal{K}_n(\psi, \chi) \mid \mathbb{G}_m \otimes \overline{\mathbb{F}}_q$, $\mathcal{K} \stackrel{def}{=} \mathcal{K}_n(\psi, \rho) \mid \mathbb{G}_m \otimes \overline{\mathbb{F}}_q$. Then $\mathcal{L}_\eta \otimes (t \mapsto at)^*\mathcal{H} \cong (t \mapsto bt)^*\mathcal{K}$ if and only if $a = b$ and $\eta\chi = \rho$ as unordered n-tuples.*

Remarks 4.12. Any such η is clearly defined over k. If χ (or ρ) is not Kummer-induced then η, if it exists, is unique. In our application, $p > n$; in particular, $p \nmid n$.

PROOF. Let $\alpha, \beta \in \overline{\mathbb{F}}_q$ be any n^{th} roots of a, b. Use $[n]^*\circ(t \mapsto at)^*\mathcal{H} = (t \mapsto \alpha t)^*\circ[n]^*\mathcal{H}$ and $[n]^*\circ(t \mapsto bt)^*\mathcal{K} = (t \mapsto \beta t)^*\circ[n]^*\mathcal{K}$; use Theorem 2.4 to evaluate $[n]^*\mathcal{H}$ and $[n]^*\mathcal{K}$ on P_∞. We get

$$\bigoplus_{\zeta \in \mu_n(\overline{\mathbb{F}}_q)} \mathcal{L}_{\psi(\alpha\zeta x/n)} \mid P_\infty \cong \bigoplus_{\zeta \in \mu_n(\overline{\mathbb{F}}_q)} \mathcal{L}_{\psi(\beta\zeta x/n)} \mid P_\infty,$$

which implies $\{\zeta\alpha \mid \zeta \in \mu_n(\overline{\mathbb{F}}_q)\} = \{\zeta\beta \mid \zeta \in \mu_n(\overline{\mathbb{F}}_q)\}$. This implies $\alpha^n = \beta^n$, *i.e.*, $a = b$.

Restricting to I_0 and using Theorem 2.4, $\eta\chi = \rho$ as unordered n-tuples. This gives one direction; the converse follows from basic facts about convolution; *cf.* [**Ka-2**, 8.2.5]. □

4.3. The Main Theorem.

Remarks 4.13. Suppose that $\psi : k \to \overline{\mathbb{Q}}_l^\times$ is a non-trivial additive character defined over $\mathbb{F}_p$ and $\mathrm{Kl}_n(\psi, \chi, k, a) = \mathrm{Kl}_n(\psi, \rho, k, b)$. We would like to apply Lemma 4.9 with $\mathcal{F} = \mathcal{F}_a(\chi) \otimes \mathcal{L}_{\Pi\overline{\chi}}$, $\mathcal{F}' = \mathcal{F}_b(\rho) \otimes \mathcal{L}_{\Pi\overline{\rho}}$, but we need to produce a geometrically irreducible sheaf $\mathcal{G}$ as in the hypotheses of the Lemma. This is what we do in Lemma 4.17 and Proposition 4.18. We then show, in Theorem 4.22, that a and the χ_i are related to b and the ρ_j by some $\sigma \in \mathrm{Gal}(k/\mathbb{F}_p)$.

NOTATION 4.14. Let $k = \mathbb{F}_q$ ($q = p^d$) and fix a non-trivial additive character $\psi : k \to \overline{\mathbb{Q}}_l^\times$. For $\alpha \in \overline{\mathbb{F}}_q^\times$, let $T_\alpha = (t \mapsto \alpha t) : \mathbb{G}_m \otimes \overline{\mathbb{F}}_q \to \mathbb{G}_m \otimes \overline{\mathbb{F}}_q$. For any Kloosterman sheaf $\mathcal{K} = \mathcal{K}_n(\psi, \chi)$ and any $\sigma \in \mathrm{Gal}(k/\mathbb{F}_q)$, let $\mathcal{K}\circ\sigma$ denote $(\mathrm{Spec}\,\sigma^{-1})^*\mathcal{K} = \mathcal{K}_n(\psi\circ\sigma, \chi\circ\sigma)$. If $\mathcal{F}$ is a lisse sheaf on $\mathbb{G}_m \otimes \mathbb{F}_q$ or $\mathbb{G}_m \otimes \overline{\mathbb{F}}_q$, let $G_{\mathrm{geom}}(\mathcal{F})$ denote the geometric monodromy group of $\mathcal{F}$, *i.e.*, the Zariski-closure of the image of $\pi_1(\mathbb{G}_m \otimes \overline{\mathbb{F}}_q)$ in $\mathrm{GL}(\mathcal{F})$; and let $\mathfrak{g}(\mathcal{F})$ denote the Lie algebra of $G_{\mathrm{geom}}(\mathcal{F})^0$. If $\mathcal{F}$ is geometrically semisimple then, by [**Del**, Theorem 1.3.8], $G_{\mathrm{geom}}(\mathcal{F})^0$ is a connected, semisimple algebraic group over $\overline{\mathbb{Q}}_l$ and $\mathfrak{g}(\mathcal{F})$ is a

semisimple Lie algebra; and by [**Del**, Theorem 3.4.1], if $\mathcal{F}$ is mixed then it is geometrically semisimple. For any finite map $\phi\colon \mathbb{G}_m \to \mathbb{G}_m$ and any multiplicative character η of $k^\times$, $\mathfrak{g}(\mathcal{F}) = \mathfrak{g}(\phi^*\mathcal{F}) = \mathfrak{g}(\mathcal{F} \otimes \mathcal{L}_\eta)$.

DEFINITION 4.15. We say that a representation V (of a semisimple algebraic group or Lie algebra) has a **highest weight** λ (with respect to some choice of positive roots) if λ is a weight of V and if whenever μ is another such then $\lambda - \mu$ is a sum of positive roots (with non-negative, integral coefficients).

Remarks 4.16. (1) Changing the choice of positive roots corresponds to replacing λ with $w\lambda$, for some w in the Weyl group. In particular, V has a highest weight with respect to some choice of positive roots if and only if V has a highest weight with respect to any other such choice.
(2) Irreducible representations have highest weights. If V and W each have a highest weight then so does $V \otimes W$.

LEMMA 4.17. *Let $\mathcal{F}$ be a lisse, geometrically semisimple, l-adic sheaf on a normal, geometrically connected scheme X of finite type over a finite field E (e.g., $X = \mathbb{G}_m \otimes \mathbb{F}_p$). Let $G = G_{\mathrm{geom}}(\mathcal{F})$ be the geometric monodromy group of $\mathcal{F}$. Assume that $\mathcal{F}$ has a highest weight λ (as a representation of G^0) which occurs with multiplicity 1 and let $\mathcal{G} \subseteq \mathcal{F}$ be the irreducible sub-G^0-representation with highest weight λ. Then $\mathcal{G}$ is $\pi_1(X)$-stable, geometrically irreducible, and $\mathcal{G} \mid X \otimes \overline{E}$ occurs exactly once in $\mathcal{F} \mid X \otimes \overline{E}$. If $\mathcal{F}$ is pure of weight w then so is $\mathcal{G}$.*

PROOF. Let G_{arith} be the arithmetic monodromy group of $\mathcal{F}$ and choose any $A \in G_{\mathrm{arith}}$. Since G^0 is normal in G_{arith}, A permutes the G^0-irreducible constituents of $\mathcal{F}$. We wish to show that $A \cdot \mathcal{G} = \mathcal{G}$; since λ occurs in $\mathcal{F}$ with multiplicity 1, it suffices to show that $A \cdot \mathcal{G} \cong \mathcal{G}$ as G^0-representations. For $B \in G^0$, $v \in \mathcal{G}$, $B \cdot Av = A \cdot A^{-1}BAv$; thus the representation $G^0 \to \mathrm{GL}(A \cdot \mathcal{G})$ can be obtained from $G^0 \to \mathrm{GL}(\mathcal{G})$ by composing with the automorphism $\phi_A\colon B \mapsto A^{-1}BA$ of G^0 (and $A_*\colon \mathrm{GL}(\mathcal{G}) \to \mathrm{GL}(A \cdot \mathcal{G})$). Therefore ${\phi_A}^*(\lambda)$ is a weight of $A \cdot \mathcal{G}$.

Any automorphism of G^0 acts on the weights of G^0 as an element of the Weyl group composed with a permutation of the simple roots (an automorphism of the Dynkin diagram). Since any Weyl-conjugate of ${\phi_A}^*(\lambda)$ is a weight of $A \cdot \mathcal{G}$, there is a permutation s_A of the simple roots α such that, writing $\lambda = \sum_\alpha k_\alpha \alpha$ $(0 \le k_\alpha \in \mathbb{Q})$, $s_A \cdot \lambda \stackrel{def}{=} \sum_\alpha k_\alpha \cdot s_A(\alpha)$ is a weight of $A \cdot \mathcal{G}$. Since $A \cdot \mathcal{G} \subseteq \mathcal{F}$ and λ is the highest weight of $\mathcal{F}$, $\lambda - s_A \cdot \lambda$ is a sum of simple roots (with non-negative, integral coefficients). Therefore $\lambda = s_A \cdot \lambda$ and $A \cdot \mathcal{G} \cong \mathcal{G}$.

Since $\mathcal{G}$ is G_{arith}-stable, it is $\pi_1(X)$-stable. Since $\mathcal{G}$ is G^0-irreducible, it is geometrically irreducible. Similarly, $\mathcal{G} \mid X \otimes \overline{E}$ occurs exactly once in $\mathcal{F} \mid X \otimes \overline{E}$. The purity statement is clear. □

PROPOSITION 4.18. *Let $a \in k^\times$ and let χ be an n-tuple of multiplicative characters $\chi_i\colon k^\times \to \overline{\mathbb{Q}}_l{}^\times$. Assume that $p > 2n + 1$ and $\mathcal{H} \stackrel{def}{=} \mathcal{K}_n(\psi, \chi)|\mathbb{G}_m \otimes \overline{k}$*

is not Kummer-induced. Let $\mathcal{F}_a(\chi)$ be as in Theorem 3.12. *As a representation of $\mathfrak{g}(\mathcal{F}_a(\chi))$, $\mathcal{F}_a(\chi)$ has a highest weight $\lambda_a(\chi)$ (with multiplicity one) and $\mathcal{F}_a(\chi)$ has a subsheaf $\mathcal{G}_a(\chi)$ such that, as a representation of $\mathfrak{g}(\mathcal{F}_a(\chi))$, $\mathcal{G}_a(\chi)$ is* the *irreducible sub-representation with highest weight $\lambda_a(\chi)$. In particular, $\mathcal{G}_a(\chi)$ is geometrically irreducible and $\mathcal{G}_a(\chi)|\mathbb{G}_m \otimes \overline{\mathbb{F}}_p$ occurs with multiplicity* 1 *in $\mathcal{F}_a(\chi)|\mathbb{G}_m \otimes \overline{\mathbb{F}}_p$.*

PROOF. From Theorem 3.12,

$$\mathcal{F}_a(\chi)|\mathbb{G}_m \otimes \overline{\mathbb{F}}_p \cong \bigotimes_{\sigma \in \mathrm{Gal}(k/\mathbb{F}_p)} [n]^* T_{\sigma(a)}{}^* \left(\mathcal{H} \circ \sigma^{-1}\right).$$

Since $\mathcal{H}$ is not Kummer-induced, Theorem 2.9 shows that no $\mathcal{H} \circ \sigma^{-1}$ is Kummer-induced. (Alternately, $\mathcal{H} \circ \sigma^{-1} = (\operatorname{Spec} \sigma)^* \mathcal{H} \ldots .$) By the same Theorem, each $\mathcal{H} \circ \sigma^{-1}$ is Lie-irreducible, *i.e.*, irreducible as a representation of $\mathfrak{g}\left(\mathcal{H} \circ \sigma^{-1}\right)$. Since

$$\begin{aligned}
\mathfrak{g}(\mathcal{F}_a(\chi)) &\cong \mathfrak{g}\Big(\bigotimes_{\sigma \in \mathrm{Gal}(k/\mathbb{F}_p)} [n]^* T_{\sigma(a)}{}^* \left(\mathcal{H} \circ \sigma^{-1}\right) \Big) \\
&\cong \mathfrak{g}\Big(\bigoplus_{\sigma \in \mathrm{Gal}(k/\mathbb{F}_p)} [n]^* T_{\sigma(a)}{}^* \left(\mathcal{H} \circ \sigma^{-1}\right) \Big) \\
&\cong \mathfrak{g}\Big(\bigoplus_{\sigma \in \mathrm{Gal}(k/\mathbb{F}_p)} \mathcal{H} \circ \sigma^{-1} \Big),
\end{aligned}$$

it follows that each $\mathcal{H} \circ \sigma^{-1}$ is irreducible as a representation of $\mathfrak{g}(\mathcal{F}_a(\chi))$ and that $\mathcal{F}_a(\chi)$ has a highest weight, say $\lambda_a(\chi)$ (*cf.* Remark 4.16 (2)).

Let $\mathcal{G}_a(\chi) \subseteq \mathcal{F}_a(\chi)$ denote the irreducible sub-$\mathfrak{g}(\mathcal{F}_a(\chi))$-representation with highest weight $\lambda_a(\chi)$. By Lie theory, $\mathcal{G}_a(\chi)$ is an irreducible sub-$G_{\mathrm{geom}}(\mathcal{F}_a(\chi))^0$-representation with highest weight $\lambda_a(\chi)$. By Lemma 4.17, we can extend the action of $G_{\mathrm{geom}}(\mathcal{F}_a(\chi))^0$ to an action of $\pi_1(\mathbb{G}_m \otimes \mathbb{F}_p)$, thus making $\mathcal{G}_a(\chi)$ a sheaf on $\mathbb{G}_m \otimes \mathbb{F}_p$. □

Remark 4.19. If $\mathcal{F}'$ is some auxiliary sheaf then the statements and proofs of Lemma 4.17 and Proposition 4.18 remain true if we replace $G_{\mathrm{geom}}(\mathcal{F}(a))$ with $G_{\mathrm{geom}}(\mathcal{F}(a) \oplus \mathcal{F}')$ and $\mathfrak{g}(\mathcal{F}(a))$ with $\mathfrak{g}(\mathcal{F}(a) \oplus \mathcal{F}')$.

COROLLARY 4.20. *Let $a, b \in k^\times$ and let χ and ρ be n-tuples of multiplicative characters $\chi_i, \rho_j : k^\times \to \overline{\mathbb{Q}}_l{}^\times$. Let $\mathcal{F}_a(\chi)$ and $\mathcal{F}_b(\rho)$ be as in Theorem* 3.12 *and $\mathcal{G}_a(\chi) \subseteq \mathcal{F}_a(\chi)$ as in Proposition* 4.18. *Assume that $\mathcal{H} \stackrel{def}{=} \mathcal{K}_n(\psi, \chi)|\mathbb{G}_m \otimes \overline{k}$ is not Kummer-induced; $p > (2n^{2d} + 1)^2$; and $\mathrm{Kl}_n(\psi, \chi, k, a) = \mathrm{Kl}_n(\psi, \rho, k, b)$. Then $\mathcal{G}_a(\chi) \otimes \mathcal{L}_{\Pi\overline{\chi}} \mid \mathbb{G}_m \otimes \overline{\mathbb{F}}_p$ occurs at least once in $\mathcal{F}_b(\rho) \otimes \mathcal{L}_{\Pi\overline{\rho}} \mid \mathbb{G}_m \otimes \overline{\mathbb{F}}_p$.*

PROOF. We apply Lemma 4.9 with $\mathcal{F} = \mathcal{F}_a(\chi) \otimes \mathcal{L}_{\Pi\overline{\chi}}$, $\mathcal{F}' = \mathcal{F}_b(\rho) \otimes \mathcal{L}_{\Pi\overline{\rho}}$, $\mathcal{G} = \mathcal{G}_a(\chi) \otimes \mathcal{L}_{\Pi\overline{\chi}}$. By Corollary 3.13, $\mathcal{F}$ and $\mathcal{F}'$ have rank $r = n^d$; they are pure of weight $w = d(n-1)$; and (in the notation of Lemma 4.9) $M_0 = 0$, $M_\infty \leq 1$. For $t \in \mathbb{F}_p{}^\times$, we have $\mathrm{tr}(F_{\mathbb{F}_p,t} \mid \mathcal{F}) = \mathrm{tr}(F_{\mathbb{F}_p,t} \mid \mathcal{F}')$ by Corollary 4.8. Since $\mathcal{G}_a \subseteq \mathcal{F}_a$, $s \stackrel{def}{=} \operatorname{rank} \mathcal{G}_a \leq r$, so the bound on p in Lemma 4.9 is satisfied. □

Remark 4.21. One could improve the bound on p slightly by calculating $s \stackrel{def}{=} \operatorname{rank} \mathcal{G}_a(\chi)$ instead of using the trivial estimate $s \leq \operatorname{rank}(\mathcal{F}) = n^d$. The hypothesis on p becomes $p > (2n^d s + 1)^2$. For example, if $G_{\mathrm{geom}}(\mathcal{F})^0 \cong \mathrm{SL}(n)^{d/e}$ and $\mathcal{F} \cong \boxtimes_\tau (\mathrm{std}_{n,\tau})^{\otimes e}$ (where std_n denotes the canonical representation of $\mathrm{SL}(n)$) then $\mathcal{G}_a(\chi) \cong \boxtimes_\tau \operatorname{Sym}^e(\mathrm{std}_{n,\tau})$, so $s = \binom{n+e-1}{e}^{d/e}$.

THEOREM 4.22. *Let $a, b \in k^\times$ and let χ and ρ be n-tuples of multiplicative characters $\chi_i, \rho_j : k^\times \to \overline{\mathbb{Q}}_l^\times$. Assume that ψ is defined over $\mathbb{F}_p$. Letting $\mathcal{H} \stackrel{def}{=} \mathcal{K}_n(\psi, \chi) \mid \mathbb{G}_m \otimes \overline{k}$ and $\mathcal{K} \stackrel{def}{=} \mathcal{K}_n(\psi, \rho) \mid \mathbb{G}_m \otimes \overline{k}$, assume that $\mathcal{H}$ and $\mathcal{K}$ are not Kummer-induced and that $G(\mathcal{H})^0$, $G(\mathcal{K})^0 \not\cong \mathrm{SO}(4)$. If $p > (2n^{2d} + 1)^2$ and $\mathrm{Kl}_n(\psi, \chi, k, a) = \mathrm{Kl}_n(\psi, \rho, k, b)$ then for some $\sigma \in \mathrm{Gal}(k/\mathbb{F}_p)$ and some multiplicative character $\eta : k^\times \to \overline{\mathbb{Q}}_l^\times$, we have $b = \sigma(a)$, $\rho = \eta \cdot \chi \circ \sigma^{-1}$ as unordered n-tuples, and either both Kloosterman sums vanish or $\eta(b) = 1$.*

Remark 4.23. Since we assume that $\mathcal{H}$ and $\mathcal{K}$ are not Kummer-induced, there is at most one η such that $\rho = \eta \cdot \chi \circ \sigma^{-1}$ (as unordered n-tuples). In particular, if $\chi = \rho$ then $\eta = \mathbf{1}$.

PROOF. Let $\mathcal{H}_\sigma \stackrel{def}{=} T_{\sigma(a)}{}^* \mathcal{H} \circ \sigma^{-1}$, $\mathcal{K}_\tau \stackrel{def}{=} T_{\tau(a)}{}^* \mathcal{K} \circ \tau^{-1}$, and let

$$G \stackrel{def}{=} G\Big(\bigoplus_{\sigma \in \mathrm{Gal}(k/\mathbb{F}_p)} \mathcal{H}_\sigma \oplus \bigoplus_{\tau \in \mathrm{Gal}(k/\mathbb{F}_p)} \mathcal{K}_\tau \Big), \qquad \mathfrak{g} \stackrel{def}{=} \mathrm{Lie}(G^0),$$

so that all the sheaves that we are considering are representations of G and $\mathfrak{g}$. (We think of $\mathfrak{g}$ as "the Lie algebra of π_1.") Fix some choice of positive roots of $\mathfrak{g}$, so that we can talk unambiguously of highest weights and fundamental weights. Recall that $\mathcal{F}_a(\chi) | \mathbb{G}_m \otimes \overline{\mathbb{F}}_p \cong \bigotimes_\sigma [n]^* \mathcal{H}_\sigma$, $\mathcal{F}_b(\rho) | \mathbb{G}_m \otimes \overline{\mathbb{F}}_p \cong \bigotimes_\tau [n]^* \mathcal{K}_\tau$; and that $\mathcal{G}_a(\chi) \subseteq \mathcal{F}_a(\chi)$, $\mathcal{G}_b(\rho) \subseteq \mathcal{F}_b(\rho)$ are the irreducible subsheaves with the same highest weights, $\lambda_a(\chi)$, $\lambda_b(\rho)$, as $\mathcal{F}_a(\chi)$, $\mathcal{F}_b(\rho)$.

Since we assume $G(\mathcal{H})^0 \not\cong \mathrm{SO}(4)$, we also have (for all σ) $G(\mathcal{H}_\sigma)^0 \not\cong \mathrm{SO}(4)$, (*cf.* the proof of Proposition 4.18). Therefore Proposition 2.10 (along with Remark 2.11 (2)) implies that (for all σ) $\mathfrak{g}(\mathcal{H}_\sigma)$ is *simple* (and classical) and that $\mathcal{H}_\sigma$ is its "standard" representation. In particular, $\mathcal{H}_\sigma$ is a *fundamental* representation of $\mathfrak{g}(\mathcal{H}_\sigma)$; *i.e.*, with respect to any choice of positive roots of $\mathfrak{g}(\mathcal{H}_\sigma)$, $\mathcal{H}_\sigma$ is the irreducible representation corresponding to a fundamental weight of $\mathfrak{g}(\mathcal{H}_\sigma)$. Since $\mathfrak{g} \twoheadrightarrow \mathfrak{g}(\mathcal{H}_\sigma)$, $\mathcal{H}_\sigma$ is a fundamental representation of $\mathfrak{g}$. Let λ_σ denote the highest weight of $\mathcal{H}_\sigma$ (as a representation of $\mathfrak{g}$). Similarly, (for all τ) $\mathcal{K}_\tau$ is a fundamental representation of $\mathfrak{g}$, say with highest weight μ_τ. We have $\lambda_a(\chi) = \sum_\sigma \lambda_\sigma$, $\lambda_b(\rho) = \sum_\tau \mu_\tau$.

Applying Corollary 4.20 (along with Remark 4.19) twice, there are inclusions $\mathcal{G}_a(\chi) \hookrightarrow \mathcal{F}_b(\rho)$ and $\mathcal{G}_b(\rho) \hookrightarrow \mathcal{F}_a(\chi)$ (as representations of $\mathfrak{g}$). Since $\mathcal{G}_a(\chi)$, $\mathcal{F}_a(\chi)$ have highest weight $\lambda_a(\chi)$ and $\mathcal{G}_b(\rho)$, $\mathcal{F}_b(\rho)$ have highest weight $\lambda_b(\rho)$ (as representations of $\mathfrak{g}$), this shows that $\lambda_b(\rho) - \lambda_a(\chi)$ and $\lambda_a(\chi) - \lambda_b(\rho)$ are both sums of positive roots; thus $\lambda_a(\chi) = \lambda_b(\rho)$, *i.e.*, $\sum_\sigma \lambda_\sigma = \sum_\tau \mu_\tau$.

Since all λ_σ, μ_τ are fundamental and the fundamental weights are linearly independent, this implies that for all τ there is a σ such that $\lambda_\sigma = \mu_\tau$. This means $\mathcal{H}_\sigma \cong \mathcal{K}_\tau$ (as representations of $\mathfrak{g}$ or G^0). It follows that, as sheaves on $\mathbb{G}_m \otimes \overline{k}$, $\mathcal{L} \otimes \mathcal{H}_\sigma \cong \mathcal{K}_\tau$, where $\mathcal{L}$ is some rank-one, lisse sheaf; by [**Ka-2**, Lemma 8.11.7.1] $\mathcal{L} = \mathcal{L}_\eta$ for some tame multiplicative character $\eta\colon \overline{\mathbb{F}}_p^\times \to \overline{\mathbb{Q}}_l^\times$. Taking τ to be the identity, there is some σ such that $\mathcal{L}_\eta \otimes T_{\sigma(a)}{}^*(\mathcal{H} \circ \sigma^{-1}) \cong T_b{}^*\mathcal{K}$.

Since ψ is defined over $\mathbb{F}_p$, $\mathcal{H} \circ \sigma^{-1} = T_{\sigma(a)}{}^*\mathcal{K}_n(\psi, \chi \circ \sigma^{-1})|\mathbb{G}_m \otimes \overline{k}$ and we can apply Lemma 4.11 and conclude that $b = \sigma(a)$ and $\rho = \eta \cdot \chi \circ \sigma^{-1}$ (as unordered n-tuples). Finally, since $\psi \circ \sigma^{-1} = \psi$, direct calculation shows that

$$\mathrm{Kl}_n(\psi, \rho, k, b) = \eta(b)\,\mathrm{Kl}_n(\psi, \chi, k, a),$$

so the Kloosterman sums themselves are equal if and only if $\eta(b) = 1$ or if they vanish. □

4.4. Some Corollaries. Corollaries 4.24–4.26 give some consequences of Theorem 4.22. Proposition 4.27 deals with absolute values of Kloosterman sums and Remarks 4.28 discuss the hypotheses of Theorem 4.22.

COROLLARY 4.24. *Keeping the hypotheses of Theorem* 4.22, *take* $\rho = \chi$ *and assume that* χ *is defined over* $\mathbb{F}_p$. *Then*

$$\mathrm{Kl}_n(\psi, \chi, k, a) = \mathrm{Kl}_n(\psi, \chi, k, b)$$

if and only if, for some $\sigma \in \mathrm{Gal}(k/\mathbb{F}_p)$, $b = \sigma(a)$. *In particular,*

$$\mathrm{Kl}_n(\psi, \mathbf{1}, k, a) = \mathrm{Kl}_n(\psi, \mathbf{1}, k, b)$$

if and only if a *and* b *are* $\mathbb{F}_p$*-conjugate. In particular, if* $a, b \in \mathbb{F}_p$ *then* $a = b$. □

COROLLARY 4.25. *Let* ψ *be defined over* $\mathbb{F}_p$; *take* $\chi_i = \mathbf{1}$ *for* $1 \le i < n$ *and let* χ_n *be a primitive character. (I.e., if* g *is a generator of* $\mathbb{F}_q{}^\times$ *then* $\chi(g)$ *is a primitive* $(q-1)$*-th root of unity.) If* $p > (2n^{2d}+1)^2$ *then the* $q-1$ *Kloosterman sums*

$$\mathrm{Kl}_n(a) = \mathrm{Kl}_n(\psi, \mathbf{1}, \mathbb{F}_q, a)$$

$(a \in \mathbb{F}_q{}^\times)$ *are all distinct.*

PROOF. Let $\mathcal{H} = \mathcal{K}_n(\psi, \chi)|\mathbb{G}_m \otimes \overline{k}$. Then $\mathcal{H}$ is Kummer-induced (respectively, $G(\mathcal{H})^0 \cong \mathrm{SO}(4)$) if and only if $n = 2$ (resp., $n = 4$) and χ_n is the quadratic character. This is only possible if $q = 3$, which is ruled out by the bound on p. Unless $n = 2$, $\chi = \eta \cdot \chi \circ \sigma^{-1}$ implies $\eta = \mathbf{1}$, whence (letting g be a generator of $\mathbb{F}_q{}^\times$) $g = \sigma^{-1}(g)$ and $\sigma = \mathrm{id}$. Thus $\mathrm{Kl}_n(a) = \mathrm{Kl}_n(b)$ implies $a = b$. If $n = 2$ then there is the additional possibility $\eta = \chi_n$, $\eta \cdot \chi_n \circ \sigma^{-1} = \mathbf{1}$. this implies $\chi_n(g) \cdot \chi_n \circ \sigma^{-1}(g) = 1$ or $\sigma^{-1}(g) = g^{-1}$, whence (for all a in $\mathbb{F}_q{}^\times$) $\sigma^{-1}(a) = a^{-1}$. Since σ^{-1} is an automorphism, this is only possible if $q = 2$ or $q = 3$, so we do not have to worry about this case. □

COROLLARY 4.26. *Keeping the hypotheses of Theorem* 4.22, *take* $\rho = \chi$ *and assume that, for all* $\sigma \in \mathrm{Gal}(\mathbb{F}_q/\mathbb{F}_p)$, $\chi\circ\sigma = \overline{\chi} = \chi$ (*as unordered n-tuples*). (*For example, all* χ_i *are either the trivial character or the quadratic character.*) *Consider the Kloosterman sums*

$$\mathrm{Kl}_n(a) = \mathrm{Kl}_n(\psi, \chi, \mathbb{F}_q, a)$$

as lying in $\mathbb{Q}[\zeta_p, \zeta_{q-1}] \subseteq \mathbb{C}$. *Then* $\mathrm{Kl}_n(a)$ *is real if and only if* a *is conjugate to* $(-1)^n a$.

PROOF.

$$\overline{\mathrm{Kl}_n(a)} = \mathrm{Kl}_n(\overline{\psi}, \overline{\chi}, \mathbb{F}_q, a) = \mathrm{Kl}_n(\psi, \overline{\chi}, \mathbb{F}_q, (-1)^n a). \quad \square$$

PROPOSITION 4.27. *Let* χ, ρ, a, b, ψ, $\mathcal{H}$, *and* $\mathcal{K}$ *be as in Theorem* 4.22, *with the same hypotheses. If* $p > \left(2n^{4d}+1\right)^2$ *and* $\left|\mathrm{Kl}_n(\psi, \chi, k, a)\right| = \left|\mathrm{Kl}_n(\psi, \rho, k, b)\right|$ *then for some* $\sigma \in \mathrm{Gal}(k/\mathbb{F}_p)$ *and some multiplicative character* $\eta\colon k^\times \to \overline{\mathbb{Q}}_l{}^\times$, *either* $b = \sigma(a)$ *and* $\rho = \eta\cdot\chi\circ\sigma^{-1}$ *as unordered n-tuples or* $b = (-1)^n\sigma(a)$ *and* $\rho = \eta\cdot\overline{\chi}\circ\sigma^{-1}$ *as unordered n-tuples.*

PROOF. With notation as in Theorem 3.12, $(\forall t \in \mathbb{F}_p{}^\times)$

$$\begin{aligned} \mathrm{tr}\left(F_t \mid \mathcal{F}_a(\chi)\right) &= \mathrm{Kl}_n(\psi, \chi, k, a); \\ \mathrm{tr}\left(F_t \mid \mathcal{F}_{(-1)^n a}(\overline{\chi})\right) &= \overline{\mathrm{Kl}_n(\psi, \chi, k, a)}; \end{aligned}$$

and similarly for $\mathcal{F}_b(\rho)$, $\mathcal{F}_{(-1)^n b}(\overline{\rho})$. Arguing as in Lemma 4.7 and Corollary 4.8, one finds that $(\forall t \in \mathbb{F}_p{}^\times)$

$$\mathrm{tr}\left(F_t \mid \mathcal{F}_a(\chi)\otimes\mathcal{F}_{(-1)^n a}(\overline{\chi})\right) = \mathrm{tr}\left(F_t \mid \mathcal{F}_b(\rho)\otimes\mathcal{F}_{(-1)^n b}(\overline{\rho})\right).$$

Following the proof of Theorem 4.22, one finds that, for some $\sigma \in \mathrm{Gal}(k/\mathbb{F}_p)$ and some multiplicative character $\eta\colon k^\times \to \overline{\mathbb{Q}}_l{}^\times$, either

$$\mathcal{L}_\eta \otimes T_{\sigma(a)}{}^*\mathrm{Kl}_n(\psi, \chi\circ\sigma^{-1}) \cong T_b{}^*\mathrm{Kl}_n(\psi, \rho)$$

or

$$\mathcal{L}_\eta \otimes T_{\sigma(a)}{}^*\mathrm{Kl}_n(\psi, \chi\circ\sigma^{-1}) \cong T_{(-1)^n b}{}^*\mathrm{Kl}_n(\psi, \overline{\rho}).$$

Applying Lemma 4.11, either $b = \sigma(a)$ and $\rho = \eta\cdot\chi\circ\sigma^{-1}$ as unordered n-tuples or $b = (-1)^n\sigma(a)$ and $\rho = \eta\cdot\overline{\chi}\circ\sigma^{-1}$ as unordered n-tuples. It is easy to check that either of these is sufficient. $\square$

Remarks 4.28. (1) In the proof of Theorem 4.22, the assumption that ψ is defined over $\mathbb{F}_p$ is not used until the last paragraph. If $\psi'(x) = \psi(\mu x)$ $(\mu \in k^\times)$ then, by [**Ka-2**, Lemma 8.7.2],

$$\mathcal{K}_n(\psi', \chi) \mid \mathbb{G}_m \otimes \overline{k} \cong T_{\mu^n}{}^*\mathcal{K}_n(\psi, \chi) \mid \mathbb{G}_m \otimes \overline{k},$$

so one can easily derive a (messy) variant form of Lemma 4.11. However, the condition that $\eta(b) = 1$ (or the Kloosterman sums vanish) is replaced by a

more complicated one unless (for example) the n-tuple χ is defined over $\mathbb{F}_p$. In Proposition 4.27, no such problem arises.

(2) Direct (computer) calculation suggests that $p > (2n^{2d}+1)^2$ is too restrictive, at least when all the multiplicative characters are either trivial or quadratic. The referee conjectures that, at least when $\chi = \mathbf{1}$, $p \geq nd$ is sufficient. The data for small values of p and d suggest that this is true. In fact, for ordinary Kloosterman sums, coincidences occur *if and only if* $p \geq 2d$, with the exception of $p = 2$, $d = 2, 3$, as far as I have checked: $q = p$, all p; $q = p^2$, $p \leq 149$; $q = p^3$, $p \leq 47$; $q = p^4$, $p \leq 19$; $q = p^5$, $p \leq 11$.

Larger values of n do *not* seem to require larger values of p for distinctness to hold. There is some heuristic evidence that $n = 2$ is the hardest case: there are more possible values for $\mathrm{Kl}_n(a)$ since $|\mathrm{Kl}_n(a)| \leq n\sqrt{q}$ (*cf.* Remark 1.5); there is the referee's proof in the case $q = p$, which is independent of n (Remark 1.8); and there is the observation (which the author made embarrassingly recently) that the Kloosterman sums of order n for $\mathbb{F}_q$ can be computed in $O(q^2pn)$ time.

(3) When non-trivial characters χ are used, the bound $p > (2n^{2d}+1)^2$ still seems excessive, although $p \geq nd$ is *not* sufficient: when $n = 3$, $d = 2$ and $\chi = (\mathbf{1}, \mathbf{1}, \Lambda_2)$, the Kloosterman sums are not distinct for $p = 5$ and $p = 7$. (Here Λ_2 denotes the quadratic character.) The condition $p > 2n + 1$ from Theorem 2.9 may be necessary. Similarly, the SO(4) restriction seems unnecessary: when $n = 4$, $d = 1$ (respectively, $d = 2$) and $\chi = (\mathbf{1}, \mathbf{1}, \mathbf{1}, \Lambda_2)$, the only known coincidences are when $p = 5$ (resp., $p = 3$ or 5).

References

[Dav] Harold Davenport, On certain exponential sums, *J. Reine Angew. Math.* **169** (1933) 158–176.

[Del] Pierre Deligne, La conjecture de Weil. II, *Publ. Math. IHES* **52** (1981) 313–428.

[Ev] Leonard Evens, A generalization of the transfer map in the cohomology of groups, *Trans. AMS* **108** (1963) pp. 54–65.

[Ha] H. Hasse, Theorie der relativ-zyklischen algebraischen Functionenkörper, *J. Reine Angew. Math.* **172** (1934) pp. 37–54.

[Ir-Ro] Kenneth Ireland and Michael Rosen, *A Classical Introduction to Modern Number Theory* Springer-Verlag, New York, 1982.

[Jou] J.-P. Jouanolou, Cohomologie l-adique, in *Cohomologie ℓ-Adique et Fonctions L* (SGA 5) Lecture Notes in Mathematics 589, Springer-Verlag, Berlin, 1977 pp. 251–281.

[Ka-1] Nicholas M. Katz, *Gauss Sums, Kloosterman Sums, and Monodromy Groups* Princeton University Press, Princeton, 1988.

[Ka-2] Nicholas M. Katz, *Exponential Sums and Differential Equations* Princeton University Press, Princeton, 1990.

[Kl] H. Kloosterman, On the representation of numbers in the form $ax^2 + by^2 + cz^2 + dz^2$, *Acta Math.* **49** (1926) pp. 407–464.

[Sal] Hans Salié, Über die Kloostermanschen Summen $S(u, v; q)$, *Math. Zeit.* **34** (1931-32) pp. 91–109.

[Weil] André Weil, On some exponential sums, *Proc. Nat'l Acad. Sci.* **34** (1948) pp. 204–207.

Department of Mathematics, Columbia University, New York, New York 10027
E-mail address: benji@math.columbia.edu

Contemporary Mathematics
Volume 133, 1992

Intersection Formulas for Mumford Curves *

Richard M. Freije †

Abstract

In this paper we will give explicit formulas for the intersection pairing of sections on certain curves defined over the ring of integers of a local field. The curves that we consider are given by the Mumford uniformization over a local field K. That is, the structure of these curves as schemes over the ring of integers O_K is determined by a p-adic Schottky group $\Gamma \subset PGL_2(K)$. We give formulas below in the case Γ is generated by one element (This is the case of a curve of genus 1.) and in the case Γ is co-compact in $PGL_2(K)$.

1 Schottky Groups and Mumford Curves

1.1 Notation

K will always denote a field complete with respect to a discrete valuation $v : K^* \to \mathbf{Z}$ with the usual convention $v(0) = +\infty$. We denote the ring of integers by O_K or simply O if there is no chance of confusion. We choose a generator π of the maximal ideal of O so $v(\pi) = 1$ and $v(x) = ord_\pi(x) = ord(x)$. We will assume the residue field $k = O/\pi O$ to be of finite order q and take the usual normalized absolute value on K that is; for $x \in K$, $|x| = q^{-v(x)}$.

As usual the projective line over K will be denoted $\mathbf{P}^1_K$ and the K - rational points $\mathbf{P}^1(K) = K \cup \infty$ can be given by homogeneous coordinates. i.e.

$$\mathbf{P}^1(K) = \{[x,y] \mid x,y \in K\}/[x,y] \sim \lambda[x,y]; \lambda \in K^*$$
$$\text{(We identify } K \text{ with } \{[x,1] \mid x \in K\} \text{)}$$

Let $G = PGL_2(K) = GL_2(K)/K^*$ where K^* is identified in $GL_2(K)$ by

$$a \to \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}.$$

*1991 Mathematics Subject Classification. Primary 11G20.
†Partially supported by a grant from the NSF.
This paper is in final form and will not be submitted for publication elsewhere.

0271-4132/92 $1.00 + $.25 per page

G acts on $\mathbf{P}^1(K)$ by linear fractional transformations so if $g \in G$ is represented by a matrix

$$g \sim \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

and $x \in \mathbf{P}^1(K)$ then:

$$g(x) = \frac{ax+b}{cx+d}.$$

1.2 Schottky Groups and Mumford Curves

Definition 1.1 *An element g of G is said to be hyperbolic if g has eigenvalues over K with distinct absolute values.*

In terms of matrices g hyperbolic means g is conjugate in G to a matrix of the form $\begin{pmatrix} r & 0 \\ 0 & 1 \end{pmatrix}$ with $r \in \pi O$. In terms of the action on $\mathbf{P}^1$ a hyperbolic g will have two distinct fixed points in $\mathbf{P}^1(K)$.

Definition 1.2 *A subgroup $\Gamma \subset G$ is called a Schottky group if γ is finitely generated and every $\gamma \in \Gamma$, $\gamma \neq$ id is hyperbolic.*

It follows from a theorem of Ihara [7] that any Schottky group is free.

Definition 1.3 *For any Schottky group Γ we denote by Σ_Γ (or just Σ) the closure in the topology given by the p-adic metric of the set of points in $\mathbf{P}^1(K)$ fixed by any element of Γ. We will always assume that $0, \infty \in \Sigma$.*

We denote the complement of Σ in $\mathbf{P}^1(K)$ by $\Omega(K) = \mathbf{P}^1(K) - \Sigma$. The action of Γ on $\Omega(K)$ is properly discontinuous, that is, we can view the quotient $\Omega(K)/\Gamma$ as a K-analytic Riemann surface. The following theorem says that this space also has an algebraic structure.

Theorem 1.1 (Mumford) *The quotient $X_\Gamma(K) = \Omega(K)/\Gamma$ has the structure of the K-rational points of a projective non-singular algebraic curve X_Γ which is defined over K. The genus of $X_\Gamma = g = rank(\Gamma)$.*

Proof: See [4],[11] or [12].

The group Γ then determines a curve X_Γ by giving its functor of points. Curves that arise in this way are called Mumford curves.

1.3 Examples

1. The projective line $\mathbf{P}^1$ is a Mumford curve by taking $\Gamma = id$.

2. (Tate Elliptic Curves) If Γ is generated by one element γ then up to conjugation we have $\gamma = \begin{pmatrix} r & 0 \\ 0 & 1 \end{pmatrix}$ with $r \in \pi O$. In this case $\Sigma = \{0\} \cup \{\infty\}$

so $\Omega(K) = K^*$ and $X_\Gamma(K) = K^*/r^{\mathbf{Z}}$. These are the Tate elliptic curves and correspond to elliptic curves over K with non-integral j-invariant. (See [15])

3. If Γ is a Schottky group co-compact in G then it can be shown that $\Sigma = \mathbf{P}^1(K)$. In other words $\Omega(K) = \emptyset$ so the corresponding X_Γ is defined over K but has no K-rational points. In this case we will concern ourselves with points on X rational over a quadratic extension L/K and view the curve over L, that is, think of $\Gamma \subset PGL_2(L)$. We have then $\Omega(L) = \mathbf{P}^1(L) - \mathbf{P}^1(K) = L - K$ so

$$X_\Gamma(L) = (L - K)/\Gamma.$$

This case of, of course, looks formally very much like the classical uniformization of a Riemann surface by a Fuchsian group acting on the complex upper half plane. When considering intersection pairings below we will restrict ourselves to these three examples for it is the relatively simple structure of Σ in these cases which translates to an ability to find simple formulas.

1.4 Intersection multiplicities

We review the definition of intersection multiplicity in this context.

In general, if we have a non-singular (projective) curve X defined over K then we can find a regular model $\mathcal{X}$ of this curve over O. Thus $\mathcal{X}$ is a regular scheme over O with general fibre isomorphic to X. If we begin with a K-rational point $x \in X(K)$, we can extend x to a divisor $\underline{x}$ on $\mathcal{X}$ by taking the closure of x in $\mathcal{X}$. Thus, $\underline{x}$ is an O-rational point of $\mathcal{X}$ or, in other words, $\underline{x}$ is given by a morphism $\underline{x} : Spec(O) \to \mathcal{X}$.

Now we begin with two points $x, y \in X(K)$, $x \neq y$ and we want to compute the intersection of the corresponding divisors $\underline{x}$ and $\underline{y}$ in the special fibre of $\mathcal{X}$. Since O is local we can find an affine open $U = spec(R)$ in $\mathcal{X}$ that contains the O-points $\underline{x}$ and $\underline{y}$. Thus, the morphism $\underline{x}$ induces a homomorphism

$$\underline{x}^* : R \to O.$$

In fact, $\underline{x}$ will induce a homomorphism which we again call

$$\underline{x}^* : R_{\underline{x}(\pi)} \to O$$

where $R_{\underline{x}(\pi)}$ is the local ring of the closed point $\underline{x}(\pi)$, π the closed point of $spec(O)$. Now, let $f_y \in R_{\underline{x}(\pi)}$ be a local equation for $\underline{y}$ and we define:

Definition 1.4 . *The intersection multiplicity* $(\underline{x}, \underline{y}) = ord(\underline{x}^*(f_y))$.

It is clear that the number $(\underline{x}, \underline{y})$ calculates "how congruent" the points x and y are. In other words $(\underline{x}, \underline{y}) = n$ where $x \equiv y \pmod{\pi^n}$ but $x \not\equiv y \pmod{\pi^{n+1}}$.

Let us make this explicit for $\mathbf{P}^1_O = ProjO[u, v]$. Two O-points can be represented by homogeneous coordinates

$$\underline{x} \mapsto [a_1, b_1] \ , \underline{y} \mapsto [a_2, b_2] \ , a_i, b_i \in O.$$

Now, choose $c \in O$ such that $c - b_1$ and $c - b_2$ are units then $a_1/(c - b_1)$ and $a_2/(c - b_2)$ are O-points of $U = Spec(O[u/(c - v)])$. In $O[u/(c - v)]) \approx O[t]$, $\underline{y}$ is given by a principal ideal, so we can take $f_y = t - (a_2/(c - b_2))$. The homomorphism $\underline{x}^*$ is evaluation at $a_1/(c - b_1)$. So:

$$(\underline{x}, \underline{y}) = ord[(a_1/(c - b_1)) - (a_2/(c - b_2)].$$

From a more elementary point of view we can calculate the intersection on $\mathbf{P}^1_O$ as follows:

Notice that we can express $\mathbf{P}^1(K)$ as $A_1 \cup A_2$ where

$A_1 = \{x \in \mathbf{P}^1(K) \mid x$can be represented by a homogeneous coordinate of the form $[a, 1]$ $a \in O\}$

$A_2 = \{x \in \mathbf{P}^1(K) \mid x$can be represented by a homogeneous coordinate of the form $[1, a]$ $a \in O\}$

Definition 1.5 *If $x, y \in \mathbf{P}^1(K)$, we define*

$$\underline{ord}(x - y) = \begin{cases} ord(a - b) & \text{if } x, y \in A_i \text{ (Say } A_1.\text{) with} \\ & x \leftrightarrow [a, 1] \text{ and } y \leftrightarrow [b, 1]. \\ 0 & \text{Otherwise.} \end{cases}$$

Then it is easy to see that $(\underline{x}, \underline{y}) = \underline{ord}(x - y)$.

2 The Structure of X_Γ as a Scheme over O

2.1 The tree T_K

The object that allows us to make calculations concerning the O structure of a Mumford curve is the tree T_K associated to the field K.

We construct a graph T_K (or just T if there is no chance for confusion) as follows:

The vertices of T correspond to two dimensional O-lattices $M \subset K \oplus K$ up to homothety. That is,

$$Vert(T) = \{M \subset K \oplus K \mid M \text{ an } O \text{ order}\}/\sim$$

Where $M \sim N$ if there exists a $\lambda \in K^*$ with $M = \lambda N$.

Given a lattice M we denote its class by $[M]$ and the corresponding vertex by v_M. Two vertices v_M and v_N are connected by an edge if there are lattices $M' \in [M]$ and $N' \in [N]$ such that $M' \subset N'$ and $N'/M' \approx O/\pi = k$. We list the following facts about T_K, their proofs can be found in [12], [13], [16].

Proposition 2.1 *1) T is a homogeneous tree of degree $q+1$.*
2) The $q+1$ vertices adjacent to a given vertex can be identified with the points of $\mathbf{P}^1(k)$.
3) The 1/2-lines in T based at a fixed vertex v (That is, the collection of infinite paths in T based at v), correspond to points in $\mathbf{P}^1(K)$. For $x \in \mathbf{P}^1(K)$ we write $\overline{x}_v$ for the corresponding 1/2-line from the vertex v. If we choose as a basepoint the vertex $v_{O\oplus O} = v_0$ then we write $\overline{x}$ for $\overline{x}_{v_{O\oplus O}}$.
4) The action of G on $K \oplus K$ induces an action of G on $Vert(T)$ which is isometric where distance is measured by the number of edges between two vertices. Thus we can think of G as acting on the tree T.
5) The action of G on $Vert(T)$ is transitive and the stabilizer of $v_{O\oplus O} = v_0$ is $U = PGL_2(O)$, thus $Vert(T) \approx G/U$.
6) If Γ is a Schottky group then Γ acts freely on T, that is it leaves no edges or vertices fixed.
7) A Schottky group Γ will leave invariant a subtree T_Γ where $T_\Gamma = \bigcup_{x\in\Sigma} \overline{x}$ and T_Γ is the minimal subtree with this property. Further, the quotient T_Γ/Γ is a finite graph.

By 3 we can think of $\mathbf{P}^1(K)$ as being the "boundary" of this infinite graph. Choosing a base point for the 1/2-lines corresponds to picking coordinates for the projective line at infinity in fact if we let v_0 be the vertex of T corresponding to $O\oplus O$ then we can label the vertices distance n away from $v_0$by the following scheme. (Distance is measured without backtracking.)

$$\{\text{Vertices distance n from } v_0\ \} \leftrightarrow \{v_L\}$$

Where:

$$L \in \{[\sum a_i\pi^i, 1]O + [\pi^n, 0]O\}$$

or

$$L \in \{[0, \pi^n]O + [1, \sum a_i\pi^i]O\}$$

Where a_i represent cosets of O/π. If we notice that any point in $\mathbf{P}^1(K)$ can be represented uniquely by a homogeneous coordinate of the form $[a, 1]$ or $[1, \pi a]$ with $a \in O$ then the above remark says that we can label vertices distance n from v_0 by these coordinates "mod π^n".

In more sophisticated language this last remark says that the 1/2-lines in T_K from v_0 correspond to elements of

$$\lim_{\leftarrow} \mathbf{P}^1(O/\pi^n O)$$

or, what is the same thing, the 1/2-lines correspond to O-points (sections) of the formal O-scheme obtained from completing the O-scheme $\mathbf{P}^1$ along its special fibre. Notice that we have the following correspondences:

$$\Omega(K) \leftrightarrow \{\overline{x} \mid \overline{x} \text{ contains vertices not in } T_\Gamma\}$$
$$\leftrightarrow \{\ 1/2 \text{ - lines in } T \text{ containing exactly one vertex in } T_\Gamma\}$$

For $x \in \Omega(K)$ we write $\overline{x}_\Gamma$ for the 1/2 - line given by the second correspondence.

2.2 Trees and formal schemes

Mumford's theory allows us to get information about the formal completion of a Mumford curve along its special fibre by looking at the quotient of T_K by Γ.

First we must, following Mumford [12], construct a formal scheme upon which Γ acts whose O-sections correspond to the points in $\Omega(K)$.

We begin with the scheme $\mathbf{P}^1$ it has a general fibre $\mathbf{P}^1_K$ and a special fibre $(\mathbf{P}^1)_0 = \mathbf{P}^1_k$. There are $q+1$ rational points on the special fibre. We blow up those points on the special fibre that are specializations of points in $\Sigma \subset \mathbf{P}^1(K)$. Blowing up a rational point in the special fibre has the effect of separating those O-sections that are congruent mod π but not mod π^2. We call the new scheme P_1 it is a model for $\mathbf{P}^1$ as its general fibre is still $\mathbf{P}^1_K$ but its special fibre will have several components

We repeat this process, blowing up those points in the special fibre of P_1 which are specializations of points in Σ, call the new scheme P_2. We can continue this process to get $\mathrm{P}_3, \mathrm{P}_4, \ldots$. Each P_n has general fibre $\mathbf{P}^1_K$ but at each stage the special fibre has more componants.

The idea now is to take a direct limit of these schemes, to do this notice there is a natural embedding of $\mathrm{P}'_n \hookrightarrow \mathrm{P}'_{n+1}$ where

$$\mathrm{P}'_n = \mathrm{P}_n - \{x \in (\mathrm{P}_n)_0 \mid \text{a point in } \sigma \text{ specializes to } x\}$$

Definition 2.1 $\mathrm{P}'_\Gamma = \varinjlim \mathrm{P}'_n$

The scheme P'_Γ again has general fibre $\mathbf{P}^1_K$ and its special fibre now has infinitely many componants, in fact its special fibre has dual graph T_Γ. The action of Γ on the graph T_Γ translates to an action on the special fibre.

Definition 2.2 *$\mathcal{P}_\Gamma$ is the formal completion of P'_Γ along the special fibre $(\mathrm{P}'_\Gamma)_0$.*

The O rational points of $\mathcal{P}_\Gamma$, that is the O - sections correspond to elements of

$$\varprojlim \mathrm{P}'_\Gamma(O/\pi^n O).$$

In terms of the trees T and T_Γ these sections correspond to infinite paths in T having exactly one vertex in T_Γ . In other words

$$\mathcal{P}_\Gamma(O) \leftrightarrow \Omega(K)$$

by the correspondence above and we get an action of Γ on this formal object, which preserves its O - structure.

We can now state a main result of Mumford's paper [12] as

Theorem 2.1 (Mumford) *a)The quotient $\mathcal{P}_\Gamma(O)/\Gamma = \hat{\mathcal{X}}_\Gamma(O)$. Where $\mathcal{X}_\Gamma$ denotes the regular minimal model of X_Γ and $\hat{\mathcal{X}}_\Gamma(O)$ denotes its formal completion along its special fibre.*
b)The finite graph, T_Γ/Γ corresponds to the dual graph of the special fibre of $\mathcal{X}_\Gamma$.

This theorem allows us to make the following identifications:

$$X_\Gamma(K) \leftrightarrow \mathcal{X}_\Gamma(O) \leftrightarrow \hat{\mathcal{X}}_\Gamma(O) \leftrightarrow$$
Infinite paths in T_K/Γwhich have only one vertex in T_Γ/Γ.

We write:

$$a \leftrightarrow \underline{a} \leftrightarrow \underline{a} \leftrightarrow \overline{a}_\Gamma.$$

3 Intersection Theory

3.1 The general intersection formula for Mumford curves.

We have the following:

Theorem 3.1 *If $a, b \in \mathbf{P}^1(K)$ and $\underline{a}, \underline{b}$ are their closures in $\mathbf{P}^1_O$ then the intersection number $(\underline{a}, \underline{b})$ = # edges $\overline{a}$ and $\overline{b}$ travel together in T.*

Proof: From the construction of T and the comments in section above it is clear that the number of edges these two 1/2-lines travel together calculates how congruent these points are and so calculates the intersection number.

This formula is true since the 1/2-lines correspond in a natural way to elements of $\lim_{\leftarrow} \mathbf{P}^1(O/\pi^n O)$. We have, from the construction in the last section:

$$\{1/2\text{-lines } \overline{x}_\Gamma\} \subset T \leftrightarrow \lim_{\leftarrow} \mathrm{P}'_\Gamma(O/\pi^n O) \leftrightarrow \Omega(K)$$

and

$$\{1/2\text{-lines } \overline{a}_\Gamma\} \subset T/\Gamma \leftrightarrow \lim_{\leftarrow} \mathcal{X}_\Gamma(O/\pi^n O) \leftrightarrow X_\Gamma(K).$$

This allows us to express the intersection number of divisors on P'_Γ and $\mathcal{X}_\Gamma$ (which come from points in $\Omega(K)$ and $X_\Gamma(K)$ respectively) in terms of the combinatorics of these graphs and so we have the following theorems.

Theorem 3.2 *a) If we denote by $(\ ,\ )_\Gamma$ the intersection on P'_Γ then if $a, b \in \Omega(K)$, $a \neq b$ and $\underline{a}, \underline{b}$ are the corresponding divisors on P'_Γ then:*

$$(\underline{a}, \underline{b})_\Gamma = \#\text{edges } \overline{a}_\Gamma \text{ and } \overline{b}_\Gamma \text{ travel together in } T$$

b) If $x, y \in X_\Gamma(K)$ $x \neq y$ and $\underline{x}, \underline{y}$ are the corresponding divisors on $\mathcal{X}_\Gamma$ then:

$$(\underline{x}, \underline{y}) = \#\text{edges } \overline{x}_\Gamma \text{ and } \overline{y}_\Gamma \text{ travel together in} T/\Gamma.$$

We can now use some simple graph theory to get a formula for this intersection number.

Theorem 3.3 (Intersection Formula) *If $x, y \in X_\Gamma(K)$ $x \neq y$ and $\underline{x}, \underline{y}$ are the corresponding divisors on $\mathcal{X}_\Gamma$ then:*

$$(\underline{x}, \underline{y}) = \sum_{\gamma \in \Gamma} (\underline{a}, \underline{\gamma b})_\Gamma$$

Where $a, b \in \Omega(K)$ lie above x and y respectively.

The main step in the proof of this theorem is the following:

Proposition 3.1 *The sum*

$$\sum_{\gamma \in \Gamma} (\underline{a}, \underline{\gamma b})_\Gamma$$

has at most one non-zero term.

Proof: Notice that for $\underline{a}$ and $\underline{b}$ to intersect means that the corresponding 1/2-lines $\overline{a}_\Gamma$ and $\overline{b}_\Gamma$ must (at least) meet the subtree T_Γ at a common vertex. The fact that Γ acts freely on the tree T_Γ along with the observation that

$$\gamma \overline{c}_\Gamma = \overline{\gamma c}_\Gamma$$

for any $c \in \Omega(K)$ insures that if $\overline{a}_\Gamma$ and $\overline{b}_\Gamma$ meet T_Γ at a common vertex that $\overline{a}_\Gamma$ and $\overline{\gamma b}_\Gamma$ do not meet at all.

Proof of Theorem 3.3:
Suppose $(\underline{x}, \underline{y}) = n$ for $x, y \in X_\Gamma(K)$. This says that the paths $\overline{x}_\Gamma$ and $\overline{y}_\Gamma$ must have n edges in common. Thus we can lift these paths to $\overline{\alpha}_\Gamma$ and $\overline{\beta}_\Gamma$ with $\alpha, \beta \in \Omega(K)$ and $\overline{\alpha}_\Gamma$ and $\overline{\beta}_\Gamma$ having n edges in common. Since α, β lie above x and y there exists $\mu, \nu \in \Gamma$ such that $\alpha = \mu a$ and $\beta = \nu b$ but then;

$$(\underline{\mu a}, \underline{\nu b})_\Gamma = (\underline{a}, \underline{\mu^{-1} \nu b})_\Gamma = n$$

and by proposition 3.1 above;

$$\sum_{\gamma \in \Gamma} (\underline{a}, \underline{\gamma b})_\Gamma = n.$$

3.2 Formulas in Special Cases

The rest of this section we will be concerned with deriving formulas for the intersection pairing in some special cases. Our ability to find a "nice" formula depends on having a simple description of the tree T_Γ.

1. The Projective Line $\mathbf{P}^1(K)$

We think of the line $\mathbf{P}^1(K)$ as a Mumford curve X_Γ with $\Gamma = (id)$. In this case, of course $\Sigma = \emptyset$, $T_\Gamma = v_0$. For $a, b \in \mathbf{P}^1(K)$ we have:

$$(\underline{a}, \underline{b}) = \underline{ord}(a - b)$$

2. Tate's Elliptic Curves

Here Γ is generated by one element γ and we can normalize so that

$$\gamma = \begin{pmatrix} r & 0 \\ 0 & 1 \end{pmatrix} \quad r \in \pi O.$$

$T_\Gamma = \overline{0} \cup \overline{\infty}$, that is a line in T. The action of γ on this line is translation through $m = ord(r)$ edges so T_Γ/Γ is a polygon with m sides and

$$X_\Gamma(K) = K^*/r^{\mathbf{Z}}.$$

So given $a, b \in K^*$, $a \neq b$ we want to calculate $(\underline{a}, \underline{b})_\Gamma$.

Proposition 3.2

$$(a,b)_\Gamma = ord^+\left(\frac{a-b}{a}\right) = ord^+\left(\frac{a-b}{b}\right).$$

Where for $x \in K$:

$$ord^+(x) = \begin{cases} ord(x) & \text{If } ord(x) > 0. \\ 0 & \text{Otherwise.} \end{cases}$$

Proof: In terms of the graphs, $(\underline{a}, \underline{b})_\Gamma = \#$ of edges the 1/2-lines $\overline{a}$ and $\overline{b}$ (these are based at v_0) travel together outside T_Γ.

In other words:

$$(\underline{a}, \underline{b})_\Gamma = \begin{cases} \#\text{of edges } \overline{a} \text{ and } \overline{b} \text{ travel together in } T- \\ \#\text{of edges they travel together in } T_\Gamma. & \text{If this is} > 0. \\ 0 & \text{Otherwise.} \end{cases}$$

So:

$$(\underline{a}, \underline{b})_\Gamma = \begin{cases} \underline{ord}(a-b)- \\ \#\text{of edges they travel together in } T_\Gamma. & \text{If this is} > 0. \\ 0 & \text{Otherwise.} \end{cases}$$

Notice that # of edges $\overline{a}$ travels with T_Γ is just $\underline{ord}(a)$ since $T_\Gamma = \overline{0} \cup \overline{\infty}$.

Also note that $(\underline{a},\underline{b})_\Gamma > 0$ implies that $\overline{a}$ and $\overline{b}$ travel the same number of edges in T_Γ thus this implies that $\underline{ord}(a) = \underline{ord}(b)$ and:

$$(\underline{a},\underline{b})_\Gamma = \begin{cases} \underline{ord}(a-b) - \underline{ord}(a) = \underline{ord}(a-b) - \underline{ord}(b) & \\ & \text{If this is } > 0. \\ 0 & \text{Otherwise.} \end{cases}$$

In order to check that this is the same as $ord^+\left(\frac{a-b}{a}\right)$ we notice that for $\underline{a}$ and $\underline{b}$ to have positive intersection it is necessary that they both lie in the same affine piece A_i this is the case exactly when $a, b \in O$ or $1/a, 1/b \in O$.

It is easy to see that if this is not the case then $ord\left(\frac{a-b}{a}\right) \leq 0$. If $a, b \in O$ then;

$$\underline{ord}(a-b) - \underline{ord}(a) = ord(a-b) - ord(a) = ord\left(\frac{a-b}{a}\right)$$

If $1/a, 1/b \in O$ then;

$$\underline{ord}(a-b) - \underline{ord}(a) = ord(1/a - 1/b) - ord(1/a) = ord\left(\frac{a-b}{a}\right)$$

So we have shown that:

$$(a,b)_\Gamma = ord^+\left(\frac{a-b}{a}\right).$$

and by interchanging the roles of a and b:

$$(a,b)_\Gamma = ord^+\left(\frac{a-b}{b}\right).$$

Corollary 3.3 *If X_Γ is a Tate elliptic curve with $\Gamma = (\gamma)$ and $\gamma = \begin{pmatrix} r & 0 \\ 0 & 1 \end{pmatrix}$, Then for $x, y \in X_\Gamma(K)$ and $a, b \in K^*$ above x and y respectively we have:*

$$(\underline{x},\underline{y}) = \sum_{-\infty}^{\infty} ord^+\left(\frac{a - r^n b}{a}\right)$$

Proof: Theorem 3.3 and proposition 3.2.

3.3 The case of co-compact Γ

Recall that in this case our curve X_Γ has no K-rational points, rather $\Sigma = \mathbf{P}^1(K)$ but we consider the L-rational points of X_Γ where L/K is a separable quadratic extension of K. Thus the L-rational points $X_\Gamma(L) = (L-K)/\Gamma$. The

role of T_Γ is played by the image of T in T_L. Just how this image sits in T_L depends on whether the extension L/K is ramified or not.

There is an inclusion $Vert(T_K) \to Vert(T_L)$ given by the map $M \to M \otimes O_L$ where M is an O_K lattice.

In the case L/K is unramified then T_L is a homogeneous tree of degree q^2+1 and we can identify the image of T_K in T_L as a subtree, in fact we have

$$Im(T_K) = \bigcup_{x \in \mathbf{P}^1(K) \subset \mathbf{P}^1(K)} \underline{x}.$$

In the ramified case we have $T_K \approx T_L$ and the image of T_K is no longer a homogeneous tree but rather looks like T_K with each edge subdivided into two.

For $a \in L$ denote by $\tilde{a}$ the image of a under the non-trivial element of $Gal(L/K)$. If (π_L) is the maximal ideal of O_L then for the rest of this paper we will write $ord(x) = ord_{\pi_L}(x)$ for $x \in L$.

Proposition 3.4 *In this case for $a, b \in L - K$, $a \neq b$, $a \neq \tilde{b}$ we have:*

$$(a,b)_\Gamma = ord^+\left(\frac{a-b}{a-\tilde{b}}\right)$$

Proof: Again we assume $\bar{a}$ and $\bar{b}$ are 1/2-lines based at v_0 in T_L and we calculate:

$$(\underline{a},\underline{b})_\Gamma = \begin{cases} \#\text{of edges } \bar{a} \text{ and } \bar{b} \text{ travel together in } T_L - \\ \#\text{of edges they travel together in } Im(T_\Gamma). & \text{If this is} > 0. \\ \\ 0 & \text{Otherwise.} \end{cases}$$

If this intersection is greater than zero we have:

$$(\underline{a},\underline{b})_\Gamma = \underline{ord}(a-b) - \ \#\text{ of edges } \bar{a} \text{ and } \bar{b} \text{ travel together in } Im(T_K).$$

Notice also that $(\underline{a},\underline{b})_\Gamma > 0$ implies that $\bar{a}$ and $\bar{b}$ have the same intersection with $Im(T_K)$.

Now for an $a \in L - K$ we want to calculate the number of edges $\bar{a}$ travels in $Im(T_K)$. In order to do this we assume $a \in A_i \approx O_L$ ($\mathbf{P}^1(L) = A_1 \cup A_2$ as above.) and under this isomorphism we have: $a = a_0 + a_1\pi_L + a_2\pi_L^2 + \ldots$ with a_i representing cosets of $O_L/\pi_L O_L$.

The number of edges $\bar{a}$ travels in $Im(T_K)$ is easily seen to be that n such that:

$$\sum_{i=0}^{n-1} a_i\pi^i \in K \text{ and } \sum_{i=0}^{n} a_i\pi^i \in L - K.$$

In other words, # of edges $\bar{a}$ travels in $Im(T_K) = \underline{ord}(a - \tilde{a})$. So we have that if $(\underline{a},\underline{b})_\Gamma > 0$,

$$\underline{ord}(a-\tilde{a}) = \underline{ord}(b-\tilde{b}) = \underline{ord}(a-\tilde{b})$$

Thus,

$$(\underline{a},\underline{b})_\Gamma = \begin{cases} \underline{ord}(a-\tilde{a}) - \underline{ord}(a-\tilde{b}) & \text{If this is } > 0. \\ 0 & \text{Otherwise.} \end{cases}$$

We omit the calculation that shows that this integer is also calculated by,

$$(\underline{a},\underline{b})_\Gamma = ord^+\left(\frac{a-b}{a-\tilde{b}}\right)$$

it is tedious but straightforward in the manner of the corresponding calculation in the proof of proposition 3.2.

Corollary 3.5 . *If X_Γ is a Mumford curve with co-compact Γ and $x, y \in X_\Gamma(L)$, $x \neq y$ with $a, b \in L - K$ above x and y respectively then:*

$$(\underline{x},\underline{y}) = \sum_{\gamma\in\Gamma} ord^+\left(\frac{a-\gamma b}{a-\gamma\tilde{b}}\right).$$

Proof: Theorem 3.3 and proposition 3.4 (Also notice that $\gamma\tilde{b} = (\gamma\tilde{b})$ since $\gamma \in PGL_2(K)$.).

3.4 Application to local heights.

One of the applications of intersection theory on curves over local rings is the calculation of the Neron height pairing for these curves. Formulas for this pairing have been found by Gross [5] in some special cases of Mumford curves without appealing to intersection theory and we are now in a position to recover these formulae as consequences of section 3.3 above.

In order to define the height pairing we need the following notation:
Let X be any complete non-singular curve over K.

$$Div(X,K) = \text{Set of divisors of } X \text{ rational over } K.$$

For $\mathcal{A} \in Div(X,K)$, $\mathcal{A} = \sum m_x(x)$ define $deg(\mathcal{A}) = \sum m_x$, the support of $\mathcal{A} = |\mathcal{A}| = \{x \mid m_x \neq 0\}$. Two divisors are said to be relatively prime if they have disjoint support.

$$Div^0(X,K) = \{\mathcal{A} \in Div(X,K) \mid deg(\mathcal{A}) = 0\}$$

$Z^0(X,K) =$ The elements of degree zero in the free abelian group on $X(K)$.

(Notice that $Z^0(X,K) \subset Div^0(X,K)$)

Finally, if f is a function on X over K with divisor relatively prime to $\mathcal{A} = \sum m_x(x)$, define $f(\mathcal{A}) = \prod f(x)^{m_x}$.

Theorem 3.4 *(Neron) There exists a unique pairing* $< \mathcal{A}, \mathcal{B} >$ *on relatively prime divisors,* $\mathcal{A} \in Z^0(X,K)$, $\mathcal{B} \in Div^0(X,K)$ *with values in* $\mathbf{R}$ *which satisfies:*

1. $< \mathcal{A}, \mathcal{B} > + < \mathcal{A}, \mathcal{C} > = < \mathcal{A}, \mathcal{B} + \mathcal{C} >$
2. $< \mathcal{A}, \mathcal{B} > = < \mathcal{B}, \mathcal{A} >$, *whenever* $\mathcal{B} \in Z^0(X,K)$.
3. $< \mathcal{A}, div(f) > = \log |f(\mathcal{A})|$
4. *For a fixed* $\mathcal{B}$ *and a point* $x_0 \in X(K) - |\mathcal{B}|$. *then the function;*

$$X(K) - |\mathcal{B}| \to \mathbf{R},$$

defined by

$$x \to < (x) - (x_0), \mathcal{B} >$$

is continuous.

Proof: See [5].

This pairing is called the height pairing of the curve and can be related to the intersection pairing as follows. First notice that a divisor $\mathcal{A} \in Z(X,K)$ has a natural extension to a divisor $\underline{\mathcal{A}}$ on $\mathcal{X}$ a regular model of X over O by taking $\underline{\mathcal{A}} = \sum m_x(\underline{x})$. (Where $\underline{x}$ is defined as above.) Also we extend the intersection pairing from points to divisors by extending linearly in each factor.

The relation between the height pairing and the intersection pairing is given by the following:

Theorem 3.5 *If* $\mathcal{A}, \mathcal{B} \in Z^0(X,K)$ *and* $\underline{\mathcal{A}}$ *and* $\underline{\mathcal{B}}$ *each have total intersection* 0 *with each component of the special fibre of* $\mathcal{X}$ *then:*

$$< \mathcal{A}, \mathcal{B} > = -(\underline{\mathcal{A}}, \underline{\mathcal{B}}) \log(q).$$

Proof: See [5]

N.B. The relationship between the intersection pairing and the height pairing is actually much more general than we have stated here. For details and more general theorems see [5] or [9,p. 287].

Formulas in Special Cases

1. The Projective Line $\mathbf{P}^1(K)$

Here we take $a, b, c, d \in \mathbf{P}^1(K)$ and form the divisors $\mathcal{A} = a - b$, $\mathcal{B} = c - d$ calculate the height:

$$< \mathcal{A}, \mathcal{B} > = -(\underline{\mathcal{A}}, \underline{\mathcal{B}}) \log(q)$$

$$< \mathcal{A}, \mathcal{B} > = -(\underline{a - b}, \underline{c - d}) \log(q)$$

$$< \mathcal{A}, \mathcal{B} > = -\log(q)[(\underline{a}, \underline{b}) - (\underline{b}, \underline{c}) + (\underline{b}, \underline{c}) - (\underline{a}, \underline{d})]$$

Now by the formula found in section 3.3:

$$< \mathcal{A}, \mathcal{B} >= -\log(q)[\underline{ord}(a-b) - \underline{ord}(b-c) + \underline{ord}(b-d) - \underline{ord}(a-d)]$$

It is a straightforward but tedious calculation to show:

$$\underline{ord}(a-b) - \underline{ord}(b-c) + \underline{ord}(b-d) - \underline{ord}(a-d) = ord\left(\frac{(a-b)(b-d)}{(b-c)(a-d)}\right)$$

Thus:

$$< \mathcal{A}, \mathcal{B} >= (-\log(q))ord\left(\frac{(a-b)(b-d)}{(b-c)(a-d)}\right) = \log\left|\frac{(a-b)(b-d)}{(b-c)(a-d)}\right|$$

The last equality being true since $\log|x| = -ord(x)\log(q)$. This recovers the formula for the height pairing calculated in [5].

2. The case of co-compact Γ

We will calculate the height pairing for divisors divisors $\mathcal{A} = x - \tilde{x}$, $\mathcal{B} = y - \tilde{y}$ $x, y \in X_\Gamma(L)$. We have:

$$< \mathcal{A}, \mathcal{B} >= -(\underline{\mathcal{A}}, \underline{\mathcal{B}})_\Gamma \log(q).$$

Since x and $\tilde{x}$ reduce to the same point of the special fibre, $\mathcal{A}$ (similarly $\mathcal{B}$) have total intersection 0 with each component of the special fibre.

$$< \mathcal{A}, \mathcal{B} >= -\log(q)[(\underline{x}, \underline{y})_\Gamma - (\underline{x}, \underline{\tilde{y}})_\Gamma + (\underline{\tilde{x}}, \underline{\tilde{y}})_\Gamma - (\underline{\tilde{x}}, \underline{y})_\Gamma]$$

$$< \mathcal{A}, \mathcal{B} >= -2\log(q)[(\underline{x}, \underline{y})_\Gamma - (\underline{x}, \underline{\tilde{y}})_\Gamma]$$

(Since $(\underline{x}, \underline{y})_\Gamma = (\underline{\tilde{x}}, \underline{\tilde{y}})_\Gamma$ and $(\underline{x}, \underline{\tilde{y}})_\Gamma = (\underline{\tilde{x}}, \underline{y})_\Gamma$)

So:

$$< \mathcal{A}, \mathcal{B} >= -2\log(q)\left(\sum_{\gamma\in\Gamma} ord^+\left(\frac{a-\gamma b}{a-\gamma\tilde{b}}\right) - \sum_{\gamma\in\Gamma} ord^+\left(\frac{a-\gamma\tilde{b}}{a-\gamma b}\right)\right)$$

Where a and b lie over x and y respectively and:

$$< \mathcal{A}, \mathcal{B} >= -2\log(q)\left(\sum_{\gamma\in\Gamma} ord^+\left(\frac{a-\gamma b}{a-\gamma\tilde{b}}\right) - \sum_{\gamma\in\Gamma} ord^-\left(\frac{a-\gamma b}{a-\gamma\tilde{b}}\right)\right)$$

Where:

$$ord^-(c) = \begin{cases} ord(c) & \text{If this is negative.} \\ 0 & \text{Otherwise.} \end{cases}$$

The last equality is true since $ord^+(c/d) = ord^-(d/c)$. Finally since $ord(c) = ord^+(c) - ord^-(c)$ we have:

$$< \mathcal{A}, \mathcal{B} >= -2\log(q) \left(\sum_{\gamma \in \Gamma} ord \left(\frac{a - \gamma b}{a - \gamma \tilde{b}} \right) \right)$$

and

$$< \mathcal{A}, \mathcal{B} >= \sum_{\gamma \in \Gamma} \log \left(\frac{a - \gamma \tilde{b}}{a - \gamma b} \right)$$

Which again recovers the formula found in [5].

References

[1] I. Cerednik, *Uniformization of algebraic curves by discrete subgroups of $PGL_2(k_w)$ with compact quotient space.* ; Math. of USSR Sb. 100 (1976), pp. 59-88..

[2] V. Drinfeld and Y. Manin, *Periods of p-adic Schottky groups.*; J. reine angew. Math. 262/263 (1973), pp. 239-247.

[3] L. Gerritzen, *Zur nichtarchimedischen Uniformisierung von Kurven.* Math. Ann. 169 (1972), pp. 323-346.

[4] L. Gerritzen and M. Van der Put, *Schottky Groups and Mumford Curves.* Springer Lecture Notes 817 (1980).

[5] B. Gross, *Local heights on curves.* In *Arithmetic Geometry*, G. Cornell and J. Silverman, Eds. Springer-Verlag, New York (1986)

[6] R. Hartshorne, *Algebraic Geometry.* Springer-Verlag, New York (1977).

[7] Y. Ihara, *On discrete subgroups of the two by two projective linear group over p-adic fields.* J. Math. Soc. Japan, 18 (1966), pp. 219-235.

[8] A. Kurihara, *On some examples of equations defining Shimura curves and the Mumford uniformization.* J. Faculty Science Univ. Tokyo, Sec. 1A, 25, n0 3 (1979), pp. 277-300.

[9] S. Lang *Fundamentals of Diophantine Geometry.* Springer-Verlag, New York (1983).

[10] S. Lichtenbaum, *Curves over discrete valuation rings.* American J. of Math. 90 (1968), pp. 380-405.

[11] Y. Manin, *P-adic automorphic forms.* (English translation) Soviet J. of Math. (1976), pp. 279-331.

[12] D. Mumford, *An analytic construction of degenerating curves over complete local rings.* Compositio Math. Vol. 24, Fasc. 2 (1972), pp. 129-174.

[13] J. Myers, *P-Adic Schottky Groups.* Ph.D. Thesis, Harvard University (1974).

[14] M. Raynaud, *Construction analytique des courbes en geometrie non archimedienne.* Seminar Bourbaki 427 (1972/73).

[15] P. Roquette, *Analytic Theory of Elliptic Functions Over a Local Field.* Hamburger Mathematische Einzelschriften, Neue Folge, Heft 1 (1970).

[16] J. P. Serre, *Trees.* Springer-Verlag, New York (1980).

[17] I. R. Shafarevich, *Lectures on minimal models and birational transformations of two dimensional schemes.* Tata Institute, Bombay (1966).

[18] M. F. Vigneras, *Arithmetique des Agebres de Quaternions.* Springer Lecture Notes 800 (1980).

Department of Mathematics, Mount Holyoke College, South Hadley, MA 01075
Current address: 31 Blithewood Ave. #906, Worcester, MA 01604

Contemporary Mathematics
Volume 133, 1992

L-series of Grössencharakters of Type A_0 for Function Fields

DAVID GOSS

ABSTRACT. Let E be a Drinfeld module over a finite extension of the base field $\mathbf{k}$. By following classical theory, one can define its L-series in a straightforward fashion. If E has complex multiplication, then we can factor this L-series into the product of L-series of grössencharakters of type A_0. We shall show how such L-series have an analytic continuation and strong v-adic interpolations for finite primes v. We also discuss questions that now arise from the analogy to the L-series of abelian varieties with complex multiplication.

0. Introduction

One of the most important and interesting areas of modern number theory is the subject of zeta-functions of elliptic curves (abelian varieties, algebraic varieties, or, even, "motives") over number fields, [15], [2], [12]. The collection of interesting questions and beautiful results that have arisen in the theory is both wide and deep. Moreover, recent history has witnessed surprising connections to such classical problems as Fermat's Last Theorem or class numbers of quadratic fields.

The process of constructing such zeta-functions mimics that of the quintessential example of the Riemann zeta-function: Let E be an elliptic curve over a number field k, and let $\mathcal{P}$ be one of the primes of k where E has good reduction (one knows that almost all primes have this property). Let $E_{\mathcal{P}}$ denote this reduction. Let l be a prime integer with $(l, \mathcal{P}) = 1$ and let T_l be the l-adic Tate module. On T_l we have the action of the Frobenius morphism, $\mathrm{Frob}_{\mathcal{P}}$, and we let $f_{\mathcal{P}}(t)$ be its (inverse) characteristic polynomial. One knows that $f_{\mathcal{P}}(t)$ has integral coefficients which are invariant of l; it is the "Euler factor at $\mathcal{P}$". If $\mathcal{P}$

1991 *Mathematics Subject Classification.* Primary 11G09; Secondary 11R58.
Partially supported by NSF Grant DMS-890110 and NSA Grant MDA 904-89-H-2059.
This paper is in final form and no version of it will be submitted for publication elsewhere

has bad reduction one can also define an Euler factor, $f_{\mathcal{P}}(t)$, at $\mathcal{P}$ in a fairly similar fashion. This factor gives a measure of the singularities of the reduction of E at $\mathcal{P}$. One then formally defines

$$L(E,s) = \prod_{\text{all finite } \mathcal{P}} f_{\mathcal{P}}(N\mathcal{P}^{-s})^{-1};$$

where N denotes the norm on ideals. The Riemann Hypothesis for curves over finite fields, applied to the primes of good reduction, assures us that this Euler-product converges to an analytic function in some half-plane of the complex numbers. One then conjectures that $L(E,s)$ has an analytic continuation, functional equation, etc..

Suppose now that E has complex multiplication. It is then well-known that $L(E,s)$ factors into a finite product of L-series of grössencharakters of type A_0, (see [11], [17]). As Hecke has shown that such L-series have an analytic continuation and a functional equation, [18], one also deduces the same for $L(E,s)$. Thus the case of complex multiplication marks the first step in understanding general ("motivic") zeta-functions.

In the theory of function fields there is a similar construction based on Drinfeld modules (or, more generally, G. Anderson's "motives" [1]): Let E be a Drinfeld module defined over an algebraic extension L of the base field $\mathbf{k}$. Let $\mathcal{P}$ be one of the primes of L where E has good reduction (again, one knows that almost all primes L have this property) and let $E_{\mathcal{P}}$ denote the reduction. As above, there is a Frobenius morphism, $\text{Frob}_{\mathcal{P}}$, at $\mathcal{P}$. Let Λ be a prime of the base ring $\mathbf{A}$ *not* lying under $\mathcal{P}$. Then one forms the Λ-adic Tate module T_{Λ} of $E_{\mathcal{P}}$ in the standard fashion, and we again denote the inverse characteristic polynomial of $\text{Frob}_{\mathcal{P}}$ by $f_{\mathcal{P}}(t)$. This polynomial has coefficients in $\mathbf{A}$ and is invariant of Λ [13, Lemma 1]. For the finite number of primes of bad reduction, the classical construction also tells us how to define the Euler factor $f_{\mathcal{P}}(t)$. Now, in [7], we showed how to define the analytic element $\mathbf{B}^{-s}$ for $\mathbf{B}$ an ideal of A. Thus, following the above prescription, we again set

$$L(E,s) = \prod_{\text{all finite } \mathcal{P}} f_{\mathcal{P}}(N\mathcal{P}^{-s})^{-1}.$$

The "Riemann Hypothesis" for the primes of good reduction [13, Prop. 3] assures us of a "half-plane" of convergence for $L(E,s)$.

Now assume that E has (sufficiently many) complex multiplications. One then uses the classical proof to check that $L(E,s)$ factors, as above, into L-series associated to "grössencharakters of type A_0" for function fields. It is thus the aim of this paper to present a reasonably self-contained study of the first properties of such L-series. We shall show that, like their classical counterparts mentioned above, they have an analytic continuation; moreover, we will also establish that they possess a "strong v-adic interpolation" (v in Spec($\mathbf{A}$)) which mirrors that found in our earlier work (e.g., [7], [8]). This strong v-adic interpolation seems

to be a general phenomenon not obviously seen in classical theory. We present further remarks along these lines below.

Our first chapter will discuss the types of grössencharakters of interest to us. We cannot handle all grössencharakters but only those with "nicely" behaved infinite components. We call these "admissible". Fortunately, this class is till quite large and, in fact, contains those grössencharakters that are "motivic" i.e., of type A_0. Our writing of this chapter was highly influenced by the classic paper [17] of A. Weil. In this context, the reader should also consult [8], [13], [16].

In our second chapter we define the L-series of an admissible grössencharakter and show that it has a half-plane of convergence. The main result of the chapter then will be to show that such functions have an analytic continuation. The proof is accomplished by examining the actions of certain large additive groups provided by the Riemann-Roch Theorem together with some elementary estimates. Moreover, as happens in the number field case, the proof depends essentially on the structure of the group of idele classes.

Our third, and last, chapter focuses on the L-series for those grössencharakters of type A_0. Here we shall see that much more can be said. Let i be a non-negative integer, $i \gg 0$. We shall show that, upon making a small change of variable, the L-series now possesses a "polynomial special-value" at $-i$ with algebraic coefficients. We previously saw this phenomenon in our earlier studies (cited above and where i can be chosen to be ≥ 0) and it's occurrence seems to be a fundamental mystery. The set of points between where the Euler factor converges and where the "good" special-values appear is a sort of "critical strip" for the L-series. Finally we shall show that these rational-polynomials interpolate to *v-adic entire functions* (in a sense to be made precise in the chapter). The interpolation of these polynomials themselves and not just some special values, constitutes the "strong v-adic interpolation" mentioned above.

There is a very rich theory associated to the L-series of grössencharakters of type A_0 and their special values in classical theory. For instance, there is a relationship between these special values and the "periods of the grössencharakter" and so on. In this context we refer the reader to the books of N. Schappacher [12], S. Lang [11], and the paper [10] of G. Henniart. Moreover, in general (i.e., without complex multiplication), one can still say a great deal about the L-series of abelian varieties by Faltings [3]. These results lead naturally to a large body of questions for the type of L-series considered here. We now mention only a few:

(1) Does there exist a version of Damarell's theorem on special-values for the L-series of grössencharakters of type A_0?

(2) Does the Isogeny conjecture [3] hold for Drinfeld-modules or, more generally, Anderson's motives? (For rank 1 objects, it follows as a consequence of class-field theory.) If it is not true, how can one attach to the L-series a factor that does insure it to be true? (The reader can find a treatment of some of the questions arising out of [3] in [14].)

(3) Is there a version of Birch-Tate-Swinnerton-Dyer for Drinfeld-modules that reflects the structure of the rational points as $\mathbf{A}$-module in the L-series? (See [5] where such an interpretation is given for the local factors, $f_{\mathfrak{p}}(t)$, at the good primes.)

(4) Do the L-series of general Drinfeld modules, motives, etc. come equipped with analytic continuations and v-adic interpolations as above?

This last question may be made to seem more reasonable in light of the following general observation. This observation concerns a feature of the function field theory which seems to have no obvious analog in classical theory: Since the characteristic polynomials of Frobenius have $\mathbf{A}$-coefficients, one can also form an Euler-product *v-adically* (as opposed to our construction given above at ∞), which *automatically* converges on a half-plane (in the sense of Chapter 3). Thus, we now have *two distinct* types of convergent Euler products associated to the Drinfeld-module, and it seems reasonable that there should be some way (i.e., analytic continuation and polynomial special-values) to pass between them. In other words, it appears reasonable to expect that there will be strong v-adic interpolation for *all* Drinfeld modules (t-modules, etc..) and not just those with complex multiplication!

It is our pleasure to thank the Institute for Advanced Study for its hospitality and Greg Anderson for his steady encouragement.

1. Admissible Grössencharakters

1.1. Notation and Background.

Let $\mathbf{k}$ be a function field over a finite field and ∞ a *fixed* closed point of $\mathbf{k}$. Let $\mathbf{A}$ be the ring of those elements in $\mathbf{k}$ which are regular outside ∞. Thus $\mathbf{A}$ is a Dedekind domain with finite class group.

Let $\mathbf{F}_r$, $r = p^m$, p a rational prime, be the field of constants of $\mathbf{k}$; so $\mathbf{A}^* = \mathbf{F}_r^*$. Put $\mathbf{K} = \mathbf{k}_\infty =$ the completion of $\mathbf{k}$ at the prime ∞ of $\mathbf{k}$. Thus $\mathbf{A} \subseteq \mathbf{K}$ discretely and $\mathbf{K}/\mathbf{A}$ is compact. We let $\mathbf{F}_\infty \subseteq \mathbf{K}$ be its field of constants.

Let $\overline{\mathbf{K}}$ be a fixed algebraic closure of $\mathbf{K}$ equipped with the canonical extension of the normalized absolute value $|?|_\infty$ of $\mathbf{K}$. *All* finite extensions of $\mathbf{k}$ are to be considered as lying in $\overline{\mathbf{K}}$. Let $\overline{\mathbf{k}} \subseteq \overline{\mathbf{K}}$ be the algebraic closure of $\mathbf{k}$ and let $\mathbf{k}^s \subseteq \overline{\mathbf{k}}$ be the separable closure.

Let D be a divisor of $\mathbf{k}$ which is rational over $\mathbf{F}_r$. Let $\deg(D)$ be its degree, and set

$$d = \deg(\infty);$$

so $\mathbf{F}_\infty \simeq \mathbf{F}_{r^d}$. If $\varrho \subseteq \mathbf{A}$ is an ideal, we set $\deg(\varrho)$ to be the degree of the underlying divisor. To be absolutely clear about this, note that if $a \in \mathbf{A}$ has a pole of order ν at ∞, then

$$\deg((a)) = \nu d.$$

Let $L \subseteq \mathbf{k}^s$ be a *fixed* finite, separable extension. Let $\mathbf{F} \subseteq L$ be its field of constants. Set $t = [\mathbf{F} : \mathbf{F}_r]$; so $\mathbf{F} \simeq \mathbf{F}_{r^t}$. Let D be a divisor of L rational over

$\mathbf{F}$. We let $\deg_1(D)$ be the degree of D over $\mathbf{F}$, and we let $\deg(D)$ be the degree over $\mathbf{F}_r$. Thus

$$\deg(D) = t \deg_1(D) .$$

Let $\mathbf{O} = \mathbf{O}_L$ be the ring of $\mathbf{A}$-integers of L. Unless otherwise specified (as in Section 3), we use the symbol "w" to denote an arbitrary prime of L and "v" to denote a finite prime (i.e., in Spec($\mathbf{O}$)). We let $\nu_w(?) = \mathrm{ord}_w(?)$ be the additive valuation measuring divisibility by w.

Let $\infty_1, \ldots, \infty_\lambda$ be the primes of L above ∞. Let $d_1, \ldots, d_\lambda$ be their degrees over $\mathbf{F}$, and let $f_1, \ldots, f_\lambda$ be the residue class degrees. Further, let $\sigma_1, \ldots, \sigma_\lambda$ be *fixed* injections of L into $\overline{\mathbf{K}}$ corresponding to $\infty_1, \ldots, \infty_\lambda$.

Put $G = \mathrm{Gal}(\overline{\mathbf{K}}/\mathbf{K})$ and let $\sigma \in G$. Then $\sigma \circ \sigma_i$ is again an injection of L into $\overline{\mathbf{K}}$. We denote the number of distinct injections obtained in this way by r_i. By standard theory, [18], $[L : \mathbf{k}] = \sum r_i$.

For each i, let $\mathbf{F}_i$ be the field of constants of the completion of $\sigma_i(L) \subseteq \overline{\mathbf{K}}$. So

$$\mathbf{F}_r \subseteq \mathbf{F} \subseteq \mathbf{F}_i .$$

Thus,

$$\mathbf{F}_i \simeq \mathbf{F}_{r^{td_i}} .$$

On the other hand, we have

$$\mathbf{F}_r \subseteq \mathbf{F}_\infty \subseteq \mathbf{F}_i .$$

Thus

$$\mathbf{F}_i \simeq \mathbf{F}_{r^{df_i}} ,$$

and

$$td_i = df_i .$$

We view σ_i as giving a continuous injection of the completion, L_{∞_i}, of L into $\overline{\mathbf{K}}$. From now on, we consider L_{∞_i} as being a subfield of $\overline{\mathbf{K}}$ unless otherwise specified. We denote L_{∞_i} by "L_i".

Let $U(i) \subseteq L_i$ be the group of units, and $U_1(i) \subseteq U(i)$ be the subgroup of 1-units. It is a well-known fact that $U_1(i)$ is topologically an infinite product of $\mathbf{Z}_p$'s. Let $\pi_i \in L_i$ be a *fixed* uniformizing element.

1.2. Quasi-characters.

Let G be a topological group and F a topological field. An *F-valued quasi-character* is a continuous F^*-valued homomorphism on G. If the image of χ is a finite set, then we call χ a (finite) *character*.

Our next set of definitions will specify those $\overline{\mathbf{K}}$-valued quasi-characters on L_i^* of interest to us.

DEFINITION 1.1. A *basic triple* on L_i is an ordered triple (x, m, h) such that

(a) $x \in \overline{\mathbf{K}}^*$ (= algebraic closure, $\overline{L_i}$, of L_i),

(b) $m \in \mathbf{Z}$, and

(c) $h \in \mathbf{Z}_p$.

DEFINITION 1.2. Let $\alpha_i \in L_i^*$ be written (uniquely) as

$$\alpha_i = \pi_i^{n_i} \beta_i u_i ;$$

where π_i is as above, $n_i \in \mathbf{Z}$, $\beta_i \in \mathbf{F}_i^*$, and $U_i \in U_1(i)$. A $\overline{\mathbf{K}}$-valued quasi-character ψ on L_i^* is said to be *basic* if and only if there exists $\sigma_\psi \in \mathrm{Gal}(\overline{\mathbf{K}}/\mathbf{K})$ and a basic triple (x_ψ, m_ψ, h_ψ) on L_i, such that for $\alpha_i \in L_i^*$ written as above,

$$\psi(\alpha_i) = \sigma_\psi(x_\psi^{n_i} \beta_i^{m_\psi} u_i^{h_\psi}) .$$

(Recall that σ_ψ is continuous, as is the endomorphism, $u \mapsto u^{h_\psi}$, of $U_1(i)$). A finite product of basic $\overline{\mathbf{K}}$-valued quasi-characters is said to be *admissible*. Each element of this product is called a *factor*.

Note that m_ψ is determined only modulo $(r^{td_i} - 1)$.

The mapping $\omega_\psi(\alpha_i) = \sigma_\psi(\beta_i^{m_\psi})$ is called *the unit-character associated to* ψ. Unless otherwise specified, *all* quasi-characters are $\overline{\mathbf{K}}$-valued.

It is clear that there are *many* non-admissible quasi-characters arising from the large group of continuous endomorphisms of 1-units.

DEFINITION 1.3. Let v be a finite prime of L with L_v the associated completion. A $\overline{\mathbf{K}}^*$-valued quasi-character of L_v^* is called *admissible* if its kernel is open. It is said to be *unramified* if its kernel is the full unit-group in L_v^*.

We let J_L be the idele group of L equipped with the standard topology, [18]. If $a = (a_w) \in J_L$ is any idele, we denote by (a) the *associated fractional ideal of* a in the standard fashion:

$$(a) = \prod_{v \text{ finite}} v^{\mathrm{ord}_v(a_v)} .$$

From the definition of J_L it is clear that (a) is an $\mathbf{O}$-fractional ideal, and that the map $a \mapsto (a)$ is a homomorphism. From standard theory, [18], $L^* \subseteq J_L$ *discretely* and we equip J_L/L^* with the canonical quotient topology.

Let H be a topological group and let $\psi : J_L \to H$ be a continuous homomorphism. Since L_w^* is embedded in J_L, we can restrict ψ to obtain a continuous map $\psi_w : L_w^* \to H$. We call ψ_w the *local component* of ψ at w.

DEFINITION 1.4. Let $\psi : J_L/L^* \to \overline{\mathbf{K}}^*$ be a quasi-character. We say ψ is *admissible* if for all primes w of L, ψ_w is admissible.

We call such a ψ an *admissible grössencharakter*. For $i = 1, \ldots, \lambda$, we put $\psi_i = \psi_{\infty_i}$. The maps $\{\psi_i\}$ are the *infinite components of* ψ. The reader will note that one purpose of the above definitions is to differentiate the roles played by the finite primes and the infinite primes. Unlike classical theory this does *not* follow automatically (by topology) and must be specified.

From now on *all* ψ are assumed to admissible. From the way we have constructed ψ, it has a *conductor* f defined in the standard fashion: Let v be one of the finite set of finite primes where ψ_v is *ramified* (i.e., is *not* unramified).

Let $n_v > 0$ be the smallest integer so that if $\alpha_v \in L_v$ is $\equiv 1 \pmod{v^{n_v}}$, then $\psi_v(\alpha_v) = 1$. Then

$$f = \prod_{v \text{ ramified}} v^{n_v}.$$

If ψ is unramified at all finite primes of L, we agree to set $f = (1) = \mathbf{O}$.

In contrast to the number field theory, the numbers n_v just mentioned are *always* 1. Indeed, the group of 1-units in L_v is a topological p-group. Thus, one easily sees that if ψ_v is trivial on some open subgroup of the 1-units then it is trivial on the *full* group.

Let $\deg_1$ also denote the standard extension of the degree mapping to J_L, [**18**, Chapter 6]; so $\deg_1 : J_L \to \mathbf{Z}$. We let J_L^0 be those ideles of degree 0. One knows that J_L^0/L^* is *compact*; thus, as ψ is continuous, the group $\psi(J_L^0)$ must consist of *units*. We therefore obtain the important fact that $|\psi(\alpha)|_\infty$ *depends only on* $\deg_1(\alpha)$. This can be restated as follows: Let π be a uniformizer of $\mathbf{K}$. Then there exists a rational number g_ψ (independent of π) so that

$$(*) \qquad \psi_i(\alpha_i) = \pi^{g_\psi \deg_1(\alpha_i)} \cdot \{\text{unit in } \overline{\mathbf{K}}\}.$$

This identity will be crucial when we discuss L-series.

The proofs of the following results come from their counterparts in [**17**]. We present them as they are short and are very instructive in the use of admissible grössencharakters.

Let $I(f)$ be the group of $\mathbf{O}$ fractional ideals of L involving primes *not* dividing f. Let $I \in I(f)$. For each finite prime v, let $\operatorname{ord}_v(I)$ be the power of v dividing I. Let (α_w^I) be *any* idele which has $\alpha_w^I = 1$ if w is infinite, $\alpha_w^I \equiv 1 \pmod{f\mathbf{O}_w}$ if $w|f$, and $\operatorname{ord}_w(\alpha_w^I) = \operatorname{ord}_w(I)$ for all other w.

DEFINITION 1.5. We set $\tilde{\psi}(I) = \psi((\alpha_w^I))$.

It is easy to check that $\tilde{\psi}(I)$ depends only on I, and that $\tilde{\psi}$ is a homomorphism of $I(f)$.

LEMMA 1.1. (a) *Let $\alpha \in L^*$ with $\alpha \equiv 1 \pmod{f}$. Then*

$$\tilde{\psi}((\alpha)) = \prod_{i=1}^{\lambda} \psi_i^{-1}(\alpha).$$

(b) *Conversely, let χ be a $\overline{\mathbf{K}}^*$-valued homomorphism of $I(\varrho)$ for some ideal ϱ. Assume that for $\alpha \in L$, $\alpha \equiv 1 \pmod{\varrho}$, there exists admissible quasi-characters ψ_i at ∞_i, all i, so that ψ_i satisfies * and*

$$\chi((\alpha)) = \prod_{i=1}^{\lambda} \psi_i^{-1}(\alpha).$$

Then, there exists an admissible grössencharakter ψ on J_L/L^* so that

$$\tilde{\psi} = \chi.$$

Furthermore, the conductor of ψ divides ϱ and $\{\psi_i\}$ are the infinite components of ψ.

PROOF.

(a) Let $\alpha \in L$ with $\alpha \equiv 1 \pmod{f}$. By definition,

$$\psi((\alpha)) = \prod_{\text{all } w} \psi_w(\alpha) = 1\,,$$

as ψ factors through L^*. On the other hand, for all $v|f$, $\psi_v(\alpha) = 1$. Part a) now follows immediately.

(b) Let $a = (a_w) \in J_L$. By the approximation lemma, one can find $\beta \in L^*$ such that

$$b = (b_w) = (\beta a_w)$$

has the property that $b_v \equiv 1 \pmod{v^{t_v}}$, $t_v = \mathrm{ord}_v(\varrho)$, for all v dividing ϱ. If β_1 is another such element of L^*, then clearly $\beta/\beta_1 \equiv 1 \pmod{\varrho}$.

Set

$$\psi(a) = \chi((b)) \prod_{i=1}^{\lambda} \psi_i(b_{\infty_i})\,.$$

By definition, this does *not* depend on our choice of β. One now sees that ψ is an admissible grössencharakter and that $\tilde{\psi} = \chi$. □

DEFINITION 1.6. Let ψ be an admissible grössencharakter with conductor f. We say ψ be of *type A* if for all i, $1 \leq i \leq \lambda$, *every* factor of ψ_i is associated to a triple (x, m, h) with $h = a \in \mathbf{Q} \cap \mathbf{Z}_p$ and $x = \pi_i^a$ (specified root). Each ψ_i is said to be of type A also.

Let $\alpha \in L$ with $\alpha \equiv 1 \pmod{f}$. Using Definition 1.6 (i.e., that $h = a \in \mathbf{Q}$), the decomposition of α_i given in Definition 1.2, and Lemma 1.1.a, it is simple to see that $\tilde{\psi}((\alpha))$ is algebraic. (N.B: Because we are using the decomposition given in Definition 1.2, there is *no* need for any restriction on π_i.) As such ideals are of finite index in $I(f)$, we have

PROPOSITION 1.1. *Let ψ be a grössencharakter of type A. Then $\tilde{\psi}$ takes on algebraic values.* □

DEFINITION 1.7. An admissible quasi-character $\psi : L_i^* \to \overline{\mathbf{K}}^*$ is said to be of type A_0 if there exists

(a) $\theta \in \mathbf{Z}[\mathrm{Gal}(\overline{\mathbf{K}}^*/\mathbf{K})]$, and

(b) a continuous character χ of L_i^* of *finite order*,

so that if α_i in L_i^*, then

$$\psi(\alpha_i) = \chi(\alpha_i)\alpha_i^{\theta}\,;$$

where we use exponential notation for the action of $\mathbf{Z}[\mathrm{Gal}(\overline{\mathbf{K}}/\mathbf{K})]$.

DEFINITION 1.8. Let $\psi : J_L/L^*$ be an admissible grössencharakter of type A. Then ψ is said to be of type A_0 if and only if each infinite local component is type A_0.

Let ψ of type A_0 and let f be its conductor. Let $\tilde{\psi}$ be the associated homomorphism of $I(f)$. Let $\mathbf{k}(\psi) = \mathbf{k}\{\tilde{\psi}(I) | I \in I(f)\}$. By proposition 1.1, we know that $\mathbf{k}(\psi) \subseteq \overline{\mathbf{k}}$.

PROPOSITION 1.2. *$\mathbf{k}(\psi)$ is a finite extension of $\mathbf{k}$.*

PROOF. Let H be the finite extension by adjoining the values of the finite characters that appear in the local components at infinity. Let H_1 be the composition of H and the Galois closure of L. If $\alpha \in L^*$ is $\equiv 1 \pmod{f}$, $f =$ conductor of ψ, then

$$\tilde{\psi}((\alpha)) \in H_1$$

as is clear form the definition. As such (α) are of finite index in $I(f)$, the result follows. □

Let $\{\psi_1, \ldots, \psi_\lambda\}$ be the infinite components of ψ, and let $\theta_i \in \mathbf{Z}[\mathrm{Gal}(\overline{\mathbf{K}}/\mathbf{K})]$ be the element associated to ψ_i by definition.

Let

$$\omega = \sum_1^\lambda \theta_i \circ \sigma_i ;$$

thus ω is a formal $\mathbf{Z}$-linear combination of *all* injections $L \to \overline{\mathbf{k}} \subseteq \overline{\mathbf{K}}$ over $\mathbf{k}$.

DEFINITION 1.9. Let ψ be as above. Let $\hat{\theta}_i$ be the continuous quasi-character of L_i^* given by $\alpha_i \mapsto \alpha_i^{-\theta_i}$. Set

$$\hat{\psi} = \psi \cdot \hat{\theta_1} \cdots \hat{\theta_\lambda} .$$

It is clear that $\hat{\psi}$ is a homomorphism of the ideles but is *not* trivial on L^*. In fact if $\alpha \in L^*$, then

$$\hat{\psi}(\alpha) = \alpha^{-\omega} .$$

Moreover, $\hat{\psi}(J_L) \subseteq \overline{\mathbf{k}}^*$, and the infinite components of $\hat{\psi}$ are just characters of finite order. The map $\hat{\psi}$ is called the *Hecke character of type A_0* associated to ψ. Its *infinity type* is $-\omega$.

Let γ be a prime of $\mathbf{A}$ and let $\bar{\gamma}$ be a prime of $\overline{\mathbf{k}}$ above it. Via $\bar{\gamma}$ we view $\overline{\mathbf{k}}$ as a subfield of the algebraic closure, $\overline{\mathbf{k}}_\gamma$, of the completion $\mathbf{k}_\gamma$. Let $v_1, \ldots, v_s$ be the primes of L above γ and $\eta_1, \ldots, \eta_s$ the corresponding injections of L_{v_i} into $\overline{\mathbf{k}}_\gamma$. We view L_{v_i} as being a subfield of $\overline{\mathbf{k}}_\gamma$ via η_i.

Via $\bar{\gamma}$ we now view ω as being a $\mathbf{Z}$-linear combination of *all* injections $L \to \overline{\mathbf{k}} \subseteq \overline{\mathbf{k}}_\gamma$. It is then clear that we can decompose ω as

$$\omega = \delta_1 \circ \eta_1 + \ldots + \delta \circ \eta_s ,$$

where

$$\delta_i \in \mathbf{Z}[\mathrm{Gal}(\overline{\mathbf{k}}_\gamma/\mathbf{k}_\gamma)] .$$

DEFINITION 1.10. Let $\hat{\delta}_i$ be the $\overline{\mathbf{k}}_\gamma$-valued quasi-character of L_{v_i} given by $\alpha_i \mapsto \alpha_i^{\delta_i}$. Set

$$\psi_{\bar{\gamma}} = \hat{\psi}\hat{\delta}_1 \cdots \hat{\delta}_s .$$

The next result is now easily established.

PROPOSITION 1.3. (a) $\psi_{\bar{\gamma}}(\alpha) = 1$ *for* $\alpha \in L^*$.
(b) $\psi_{\bar{\gamma}} : J_L/L^* \to \overline{\mathbf{k}}_\gamma$ *is a continuous quasi-character.* □

We call $\psi_{\bar{\gamma}}$ the γ-*adic grössencharakter* associated to ψ (or $\hat{\psi}$).

EXAMPLES 1.1.

(a) Let χ be a character of J_L/L^* arising from a finite extension via a class field theory. Then one sees easily that χ is of type A_0.

(b) Let $\mathbf{k} = \mathbf{F}_r(T)$ and $\mathbf{A} = \mathbf{F}_r[T]$. Let $\pi = \frac{1}{T}$ and, as in Definition 1.2, let $x \in \mathbf{K}^*$ be written uniquely as

$$x = \pi^n \beta u\,,$$

where $\beta \in \mathbf{F}_r^*$ and u is a 1-unit. We set

$$\beta = \operatorname{sgn}(x)\,.$$

It is clear that sgn is a continuous character of $\mathbf{K}^*$, (see the beginning of the next chapter). Let χ be the character of the ideal group of $\mathbf{A}$ which sends every ideal to its unique *monic* generator. It is trivial to see that

$$\chi((\alpha)) = \alpha/\operatorname{sgn}(\alpha)\,.$$

Thus, by Lemma 1.1, χ gives rise to an admissible grössencharakter ψ whose conductor is (1).

More generally, for any finite extension $L/\mathbf{k}$, we can compose χ with the norm map to obtain admissible grössencharakters of J_L/L^*. It is clear that these grössencharakters are of type A_0.

(c) Let $\mathbf{k}$, etc., be arbitrary as before. Let deg be the degree map on J_L. Let χ be non-zero. We then have the admissible grössencharakter given by

$$a \mapsto x^{\deg(a)}\,.$$

2. The L-Series of an Admissible Grössencharakter

Let ψ be an admissible grössencharakter. In this section we define its L-series and show how such L-series have an *analytic continuation.*

DEFINITION 2.1. Let $\mathbf{F}_\infty (\simeq \mathbf{F}_{r^d})$ be the constant field of $\mathbf{K}$. A *sign function*, sgn, is a continuous morphism from $\mathbf{K}^*$ to $\mathbf{F}_\infty^*$ which is the identity on $\mathbf{F}_\infty^*$.

As U_1 = group of 1-units $\subseteq \mathbf{K}^*$ is a topological p-group, it is clear that $\operatorname{sgn}(U_1) = 1$. Those elements $x \in \mathbf{K}^*$ with $\operatorname{sgn}(x) = 1$ are said to be *monic* or *positive.*

We now *fix* π to be a positive uniformizing element of $\mathbf{K}$, and we let ν_∞ be the additive valuation on $\overline{\mathbf{K}}$ with $\nu_\infty(\pi) = 1$.

DEFINITION 2.2. If $x \in \mathbf{k}^*$, then we set $\deg(x)$ to be the degree of the *finite part* of the divisor of x.

Thus $\deg(x) = -d\nu_\infty(x)$.

DEFINITION 2.3. Let $a \in k^*$. We set

$$\langle a \rangle = \langle a \rangle_\infty = \pi^{-\nu_\infty(a)} \mathrm{sgn}(a)^{-1} a .$$

Thus $\langle a \rangle \in U_1$, and clearly $\langle ab \rangle = \langle a \rangle \langle b \rangle$ for $a \cdot b \neq 0$.

DEFINITION 2.4. [6]

(a) Let $\Im$ be the group of $\mathbf{A}$-fractional ideals of $\mathbf{k}$.

(b) Let $m \in \mathbf{N}$. We let $\mathcal{P}(\pi^m)$ be the group of principal ideals of k generated by positive $\alpha \in \mathbf{k}$ with $\langle \alpha \rangle \equiv 1 \pmod{\pi^m}$. We put $\hat{\mathcal{P}} = \mathcal{P}(\pi^0)$.

(c) We let

$$\hat{\Im} = \varprojlim_m \Im / \mathcal{P}(\pi^m) .$$

There is the obvious exact sequence

$$0 \to \varprojlim_m \hat{\mathcal{P}} / \mathcal{P}(\pi^m) \to \hat{\Im} \to \Im / \hat{\mathcal{P}} \to 0 .$$

It is easy to see that $\Im / \hat{\mathcal{P}}$ is an extension of $\mathrm{Pic}(\mathbf{A})$ by $\mathbf{F}_\infty^* / \mathbf{F}_r^*$. Moreover, via $\langle\ \rangle$,

$$\hat{\hat{\mathcal{P}}} = \varprojlim_m \hat{\mathcal{P}} / \mathcal{P}(\pi^m) \simeq U_1 .$$

We therefore think of $\hat{\Im}$ as an extension of the finite group $\Im / \hat{\mathcal{P}}$ by U_1.

Let $y \in \mathbf{Z}_p$. It is clear that $\alpha \to \alpha^y$ is a continuous endomorphism of U_1 to itself. Via $\langle\ \rangle$, we lift this to a map of $\hat{\hat{\mathcal{P}}}$. As this map is to a field of characteristic p, we can (perhaps non-canonically) extend $\alpha \mapsto \alpha^y$ to a $\overline{\mathbf{K}}^*$-valued continuous quasi-character of $\hat{\Im}$. We *fix* one such lifting and, by abuse of language, denote the extension by

$$\alpha \in \hat{\Im} \mapsto \alpha^y .$$

If ϱ is an $\mathbf{A}$-fractional ideal, we denote by

$$\langle \varrho \rangle^y$$

the composition of the map $\Im \mapsto \hat{\Im}$ and the map $\alpha \mapsto \alpha^y$.

DEFINITION 2.5. (a) Set $S_\infty = \overline{\mathbf{K}}^* \times \mathbf{Z}_p$.

(b) Let ϱ be an $\mathbf{A}$-fractional ideal and $s = (x, y) \in S_\infty$. Set

$$\varrho^s = x^{\deg(\varrho)} \langle \varrho \rangle^y .$$

One easily sees that $\varrho^{s+u} = \varrho^s \varrho^u$ and $(\varrho_0 \varrho_1)^s = \varrho_0^s \varrho_1^s$.

Let $\alpha \in \mathbf{k}^*$. If α is positive, then

$$(\alpha)^s = x^{\deg(\alpha)} \langle \alpha \rangle^y .$$

If α is not positive, then, at least $(\alpha)^s$ and $x^{\deg(\alpha)}\langle\alpha\rangle^y$ differ by an $\frac{r^d-1}{r-1}$-st root of unity. Indeed, set $z = \frac{r^d-1}{r-1}$. Then

$$\begin{aligned}(\langle(\alpha)\rangle^y/\langle\alpha\rangle^y)^z &= (\langle(\alpha)\rangle^z/\langle\alpha\rangle^z)^y \\ &= (\langle(\alpha^z)\rangle/\langle\alpha^z\rangle)^y \\ &= 1^y \\ &= 1\,.\end{aligned}$$

Moreover, if $\operatorname{sgn}(\alpha_1/\alpha_2) \in \mathbf{F}_r^*$, then it is easy to check that the above root of unity is the *same* for α_1 and α_2.

EXAMPLE 2.1. Let $\pi^{1/d} \in \overline{\mathbf{K}}$ be a fixed root of the equation $y^d - \pi = 0$. Let $n \in \mathbf{A}$ be positive and let $i \in \mathbf{Z}$. One sees that

$$n^i = n^s, \text{ for } s = (\pi^{-i/d}, i)\,.$$

We let "i" denote any such s when no confusion will result.

Next, we recall, and slightly generalize, the notions of "analytic" and "essentially algebraic" functions in S_∞ as defined in [6]: Let $J \subset \overline{\mathbf{K}}$ be a finite extension of $\mathbf{K}$. Let $h(x) = \sum_{n \gg -\infty} a_n x^n \in J[[x, x^{-1}]]$ be a power series with a finite number of negative terms and which converges for all non-zero x. We set

$$\|h(x)\| = \sup\{|a_n|_\infty\}\,.$$

We call $h(x)$ a "finite-tailed entire power series".

DEFINITION 2.6. (a) Let $F : S_\infty \to \overline{\mathbf{K}}$ be a continuous function. We say F is *entire* if and only if

(1) $g_y(x) = F(x, y)$ is given by a finite-tailed entire power series in x^{-1};
(2) The coefficients of all such $g_y(x)$ belong to a fixed finite extension of $\mathbf{K}$;
(3) Let $\{\alpha_i\}$ be a collection of elements of $\mathbf{Z}_p$ converging to y. We then require that

$$\|g_y(x^{-1}) - g_{\alpha_i}(x^{-1})\|$$

tends to zero.

(b) We say an entire F is *essentially algebraic* if for i a non-negative integer, $i \gg 0$, and $x \in \overline{\mathbf{K}}^*$, the function

$$h_F(x, -i) = F(x\pi^{i/d}, -i)$$

is given by a polynomial in $\{x, x^{-1}\}$ with algebraic coefficients. The collection of all such coefficients are required to belong to a fixed finite extension of $\mathbf{k}$.

Let L, $\mathbf{O}$, etc., be as in Chapter 1. Recall that we set $\mathbf{O}$ to be the ring of $\mathbf{A}$-integers of L.

DEFINITION 2.7. Let ψ be an admissible grössencharakter on J_L/L^* with conductor f. We formally define its L-series by

$$L(\psi, s) = \prod_{\substack{\gamma \text{ a prime of } \mathbf{O} \\ \gamma \nmid f}} (1 - \tilde{\psi}(\gamma) N\gamma^{-s})^{-1}$$

where $s \in S_\infty$ and N denotes the norm from L to $\mathbf{k}$.

Let $\varrho_1, \ldots, \varrho_e$ be chosen representatives of $I(f)/\mathcal{P}(f)$ with $\mathcal{P}(f)$ as before. Then we have

$$L(\psi, s) = \sum_{\epsilon=1}^{e} \tilde{\psi}(\varrho_\epsilon) N\varrho_\epsilon^{-s} \left(\sum_{\substack{\varrho \in \mathcal{P}(f) \\ \varrho\varrho_\epsilon \text{ integral}}} \tilde{\psi}(\varrho) N\varrho^{-s} \right).$$

For $\epsilon = 1, \ldots, e$, let $L_\epsilon(\psi, s)$ be the above sum in ( ). (The reader will have no difficulty in distinguishing between L_ϵ, a function, and L_j, a field.)

Our main result of this chapter is

THEOREM 2.1.

(a) *For $\epsilon = 1, \ldots, \varrho$ the sums $L_\epsilon(\psi, s)$ converge for all $s = (x, y)$ in a "half-plane" of S_∞, i.e., where $|x|_\infty \gg 1$.*
(b) *Each L_ϵ can be analytically continued to an entire function on S_∞.*

COROLLARY 2.1.

(a) *$L(\psi, s)$ converges on a half-plane of S_∞. The Euler product for $L(\psi, s)$ converges to $L(\psi, s)$ on the same half-plane.*
(b) *$L(\psi, s)$ can be analytically continued to an entire function on S_∞.*

Before establishing Theorem 2.1, we need to further decompose the sums for $L(\psi, s)$. This is necessary since ∞ is *not*, in general, rational.

Let $\theta_1, \ldots, \theta_u$, $u = \frac{r^d - 1}{r - 1}$ be the cosets of $\mathbf{F}_\infty^*/\mathbf{F}_r^*$. For any principal ideal ϱ of $\mathbf{A}$, its sgn, $\operatorname{sgn}(\varrho)$, is invariantly given as an element of $\mathbf{F}_\infty^*/\mathbf{F}_r^*$, i.e., as an element of $\{\theta, \ldots, \theta_u\}$. Thus we can write

$$L_\epsilon(\psi, s) = \sum_{\rho=1}^{u} L_{\epsilon,\rho}(\psi, s),$$

where

$$L_{\epsilon,\rho}(\psi, s) = \sum_{\substack{\varrho \in \mathcal{P}(f) \\ \varrho\varrho_\epsilon \text{ integral} \\ \operatorname{sgn} N\varrho \in \theta_\rho}} \tilde{\psi}(\varrho) N\varrho^{-s}.$$

We now recall the statement of our main estimates. For the proof, which is elementary in that one uses only the multinomial theorem, we refer the reader to [7, 2.3, 4.3].

LEMMA 2.1.

(a) *Let J, J_1 be two fields over $\mathbf{F}_r$. Let $W \subseteq J$ be a finite dimensional $\mathbf{F}_r$-vector space of dimension α and let $\{\mathcal{L}_1, \dots, \mathcal{L}_t\}$ be $\mathbf{F}_r$-linear maps of J into J_1. Finally, let $x \in J$ and $i_1, \dots, i_t$ be non-negative integers such that*

$$\sum_{h=1}^{t} i_h < (r-1)\alpha .$$

Then

$$\sum_{w \in W} \left(\prod_{h=1}^{t} \mathcal{L}_h(x+w)^{i_h} \right) = 0 .$$

(b) *Assume now that J_1 comes equipped with an additive valuation ν and that $\nu(\mathcal{L}_h(w)) > 0$ for all h and w. Let $\{i_h\}$ now be an arbitrary collection of non-negative integers, and for $j > 0$ put*

$$W_j = \{w \in W | \nu(\mathcal{L}_h(w)) \geq j \text{ for all } h\} .$$

Then

$$\nu \left(\sum_{w \in W} \prod_{h=1}^{t} \mathcal{L}_h(w)^{i_h} \right) \geq (r-1)Q ,$$

for

$$Q = \sum_j \dim_{\mathbf{F}_r}(W_j) .$$

□

For applications of Lemma 2.1 (e.g., Theorem 2.1) it is crucial to note that the estimate of part b is *invariant* of our choice of $\{i_h\}$.

For simplicity, fix ϵ, ρ and put

$$L(s) = L_{\epsilon,\rho}(\psi, s) = \sum_{\substack{\varrho \in \mathcal{P}(f) \\ \varrho\varrho_\epsilon \text{ integral} \\ \operatorname{sgn}(N\varrho) \in \theta_\rho}} \tilde{\psi}(\varrho) N\varrho^{-s} .$$

Let $\hat{\varrho}_\epsilon$ be the inverse fractional ideal of ϱ_ϵ. Let $\alpha, \beta \in \hat{\varrho}_\epsilon$. We say $\alpha \sim \beta$ if and only if there exists $\delta \in \mathbf{O}^*$ with $\alpha = \delta\beta$. We let $[\hat{\varrho}_\epsilon]$ denote the equivalence classes of non-zero elements of $\hat{\varrho}_\epsilon$ under $\sim$. Let $[\alpha] \in [\hat{\varrho}_\epsilon]$. Clearly the norm of $[\alpha]$, $N[\alpha]$, is determined up to an element of $\mathbf{F}_r^*$. Thus the sign of $N[\alpha]$, $\operatorname{sgn} N[\alpha]$, makes sense as an element of $\mathbf{F}_\infty^*/\mathbf{F}_r^*$.

Let $\alpha \in \hat{\varrho}_\epsilon$ with $\sum_{i=1}^{\lambda} -C_i^\alpha \infty_i$ being the infinite part of its divisor. It is trivial to see that the infinite part of the divisor of $N\alpha$ is

$$\left(\sum -C_i^\alpha f_i \right) \cdot \infty ,$$

where f_i is the residue class degree of L_i/K. As the norm of an element in $\mathbf{O}^*$ is a constant,

$$d\left(\sum C_i^\alpha f_i \right) = \deg N\alpha$$

depends only on $[\alpha]$. Set

$$D_\infty(\alpha) = \sum_{i=1}^{\lambda} C_i^\alpha \infty_i .$$

We now use the symbols "f", "$\hat{\varrho}_\epsilon$", etc., to also denote the associated divisors. Let $\alpha \in \hat{\varrho}_\epsilon$ and set D_α to be the divisor

$$D_\alpha = \sum (C_i^\alpha - 1) \cdot \infty_i .$$

Put

$$\hat{\varrho}_\epsilon(D_\alpha) = \{\beta \in \hat{\varrho}_\epsilon | (\beta) + D_\alpha - f \geq 0\} ,$$

which, in standard language of divisors [4], equals

$$L(\varrho_\epsilon + D_\alpha - f) .$$

Of course, these spaces are finite dimensional over $\mathbf{F}$ (and $\mathbf{F}_r$).

LEMMA 2.2.

(a) *Let $\alpha \in \hat{\varrho}_\epsilon$ with $(\alpha) \in \mathcal{P}(f)$. Let $\beta \in \hat{\varrho}_\epsilon(D_\alpha)$. Then $\alpha + \beta \in \hat{\varrho}_\epsilon$ and $(\alpha + \beta) \in \mathcal{P}(f)$.*

(b) $D_\infty(\alpha) = D_\infty(\alpha + \beta)$.

(c) *Let ω be the unit character of any factor of any infinite component of ψ. Let α and β be as above. Then*

$$\omega(\alpha + \beta) = \omega(\alpha) .$$

(d) *Let $\lambda \in \hat{\varrho}_\epsilon$ with $(\lambda) \in \mathcal{P}(f)$ also. Then there exists $\delta \in \hat{\varrho}_\epsilon(D_\lambda)$ with $[\alpha+\beta] = [\lambda+\delta]$ if and only if there exists β_1 in $\hat{\varrho}_\epsilon(D_\alpha)$ with $[\alpha+\beta_1] = [\lambda]$.*

(e) *Let $\beta_1, \beta_2 \in \hat{\varrho}_\epsilon(D_\alpha)$ with $\beta_1 \neq \beta_2$. Then $(\alpha + \beta_1)/(\alpha + \beta_2)$ is not in $\mathbf{O}^*$.*

PROOF. Parts a, b, c follow directly from the definitions. To see part d, suppose $[\alpha + \beta] = [\lambda + \delta]$. Then there exists $u \in \mathbf{O}^*$ with

$$\alpha + \beta = u\lambda + u\delta .$$

By part b, $D_\infty(\alpha + \beta) = D_\infty(\alpha) = D_\infty(u\lambda + u\delta) = D_\infty(u\lambda)$. Set $\beta_1 = \beta - u\delta$. It is trivial to see that $\beta_1 \in \hat{\varrho}_\epsilon(D_\alpha)$.

To see part e, note that if $(\alpha + \beta_1)/(\alpha + \beta_2) \in \mathbf{O}^*$, then part b implies it is in $\mathbf{F}^*$. Solving for α gives a contradiction. □

LEMMA 2.3. *Let $\alpha \in \hat{\varrho}_\epsilon$ with $(\alpha) \in \mathcal{P}(f)$. Let $\beta \in \hat{\varrho}_\epsilon(D_\alpha)$. Then*

$$sgn(N\alpha) = sgn(N(\alpha + \beta))$$

as elements of $\mathbf{F}_\infty^*/\mathbf{F}_r^*$.

PROOF. One checks easily that the definition of $\hat{\varrho}_\epsilon(D_\alpha)$ implies that $N\alpha$ and $N(\alpha + \beta)$ differ by elements of higher order in π. Thus the result follows. □

Recall that

$$L(s) = \sum_{\substack{\varrho \in \mathcal{P}(f) \\ \varrho\varrho_\epsilon \text{ integral} \\ \operatorname{sgn} N\varrho \in \theta_\rho}} \tilde{\psi}(\varrho) N\varrho^{-s}$$

$$= \sum_{\substack{[\alpha] \in [\hat{\varrho}_\epsilon] \\ \alpha \equiv 1(f) \\ \operatorname{sgn}(N\alpha) \in \theta_\rho}} \tilde{\psi}((\alpha)) N(\alpha)^{-s} .$$

Pick $t_0 \in \mathbf{Z}$. Let $\{[\alpha_1], \ldots, [\alpha_m]\}$ be the (possibly empty) collection of those elements $[\alpha_z]$ such that $\alpha_z \in \hat{\varrho}_\epsilon$, $\alpha_z \equiv 1(f)$, $\operatorname{sgn} N\alpha_z \in \theta_\rho$ and $\deg(N\alpha_z) = t_0$. By using Lemmas 2.2 and 2.3 we can decompose $\{[\alpha_1], \ldots, [\alpha_m]\}$ as follows: Let $\{[\alpha_1], \ldots, [\alpha_j]\}$ be the classes generated by the translates of $[\alpha_1]$ under $\hat{\varrho}_\epsilon(D_{\alpha_1})$. Let $[\alpha_{j+1}]$ not be in this collection and repeat the process, etc. We thus obtain a decomposition of $\{[\alpha_1], \ldots, [\alpha_j]\}$ into disjoint subsets. Note that Lemma 2.2.e assures us that the translates of α_1 are in one to one correspondence with the elements $\hat{\varrho}(D_{\alpha_1})$, etc..

Finally recall that we set $t = [\mathbf{F} : \mathbf{F}_r]$, $d = \deg(\infty)$, $d_i = \deg_1(\infty_i)$ and $f_i =$ residue class degree of ∞_i over ∞. As pointed out in Section 1, $td_i = df_i$. We can now turn to the proof of Theorem 2.1.

PROOF OF THEOREM 2.1. It suffices to look at the function $L(s)$. We begin by rewriting it as

$$L(s) = \sum_{\substack{[\alpha] \in [\hat{\varrho}_\epsilon] \\ \alpha \equiv 1(f) \\ \operatorname{sgn}(N\alpha) \in \theta_\rho}} x^{-\deg N(\alpha)} \tilde{\psi}((\alpha)) \langle N(\alpha) \rangle^{-y} .$$

As before, set

$$D_\infty(\alpha) = \sum_{i=1}^{\lambda} C_i^\alpha \infty_i ;$$

so

$$\deg N\alpha = d \left(\sum C_i^\alpha f_i \right) .$$

Recall that, in Chapter 1, we associated to ψ a rational number g_ψ so that

$$\psi_i(\alpha) = \pi^{g_\psi \deg_1(\alpha_i)} \cdot \{\text{unit in } \overline{\mathbf{K}}\} ,$$

where α_i is the injection of α into L_i. Thus, by Lemma 1.1,

$$x^{-\deg N\alpha} \tilde{\psi}((\alpha)) \langle N(\alpha) \rangle^{-y} = x^{-\deg N\alpha} \pi^{g_\psi \sum C_i^\alpha d_i} \cdot \{\text{unit}\} .$$

Furthermore, if we set $x = \pi^{\nu_\infty(x)} \cdot \{\text{unit}\}$, then the above equals

$$\pi^{-(\nu_\infty(x) \deg N\alpha - g_\psi \sum C_i^\alpha d_i)} \cdot \{\text{unit}\} .$$

Note that, since $td_i = df_i$, $\sum C_i^\alpha d_i = \frac{d}{t} \sum C_i^\alpha f_i = t^{-1} \deg N\alpha$. Thus, if x is chosen so that $\nu_\infty(x) < g_\psi / t$, then the series for $L(s)$ will converge. Such x form a half-plane and the first part is thus established.

We turn now to the second part of the theorem. The idea of proof (as in [7]) is to put the sum for $L(s)$ in such a form that we may use Lemma 2.1.b. Thus pick $t_0 \in \mathbf{Z}$ and decompose into equivalence classes, as above, the collections of elements $[\alpha_z]$ with $\alpha_z \in \hat{\varrho}_\epsilon, \alpha_z \equiv 1 \pmod f$, $\operatorname{sgn} N\alpha_z \in \theta_\rho$ and $\deg N\alpha_z = t_0$. Choose the equivalence class, X, of some $[\alpha]$ in this decomposition. Define $\mathbf{F}_r$-linear maps $M_h : \hat{\varrho}_\epsilon(D_\alpha) \to \overline{\mathbf{K}}$ as follows: let $\lambda \in \hat{\varrho}_\epsilon(D_\alpha)$ and let α_i, λ_i be their injections into $L_i \subseteq \overline{\mathbf{K}}$. As in Definition 1.2, set

$$\alpha_i = \pi_i^{n_i} \cdot \{\text{unit}\}.$$

Put,

$$m_i(\lambda) = \lambda_i \pi_i^{-n_i}.$$

By construction $\nu_\infty(m_i(\lambda)) > 0$. Finally let $\{M_h\}$ be the finite set of linear maps obtained by composing $\{m_i\}$ with $\operatorname{Gal}(\overline{\mathbf{K}}/\mathbf{K})$.

Recall that the admissibility of ψ means that its infinite components are of a particularly simple form (Definition 1.2). Using this together with Lemmas 1.1.a and 2.2.c, the argument given before Example 2.1, the proof of the first part given above, the construction of $\hat{\varrho}_\epsilon(D_\alpha)$ and the binomial theorem, one sees that

$$\sum_{[\alpha] \in X} x^{-\deg N\alpha} \tilde{\psi}((\alpha)) \langle (N\alpha) \rangle^{-y}$$

can be expressed as

$$x^{-t_0} \pi^{g_\psi t_0/t} \cdot \sum_{\lambda \in \hat{\varrho}_\epsilon(D_\alpha)} H_y(\lambda);$$

where $H_y(\lambda)$ is obtained by composing

$$\lambda \mapsto (M_1(\lambda), \ldots, M_e(\lambda)), \; e = [L : \mathbf{k}],$$

and a bounded power series in e variables.

The Riemann-Roch Theorem, [4], applied to L, D_α, etc., and the second part of Lemma 2.1, gives a lower bound for

$$\nu_\infty \left(\sum_{\lambda \in \hat{\varrho}_\epsilon(D_\alpha)} H_y(\lambda) \right)$$

which depends only on t_0. Because of the nature of Lemma 2.1.b, one sees that for $t_0 \gg 0$, this lower bound grows quadratically in t_0. This forces convergence and the rest of the result follows directly. □

3. The Theory of Special-values for Grössencharakters of type A_0

In this chapter we examine $L(\psi, s)$ for ψ of type A_0. Let "v" now denote an element of $\operatorname{Spec}(\mathbf{A})$. We will see that $L(\psi, s)$ has a *very* rich theory of special-values that comes equipped with strong v-adic interpolations. This construction parallels that of $\psi \mapsto \psi_{\bar{\gamma}}$ given in Section 1.

We begin by discussing the v-adic version of the group $\hat{\Im}$ described in Section 2.

Let $\mathbf{k}_v$ be the completion of $\mathbf{k}$ at v.

DEFINITION 3.1. Let $\alpha \in \mathbf{k}_v$ be a unit. Write

$$\alpha = \beta u$$

where β is a root of unity and u is a unit. We set

$$\beta = \omega_v(\alpha)$$

and

$$u = \langle \alpha \rangle_v .$$

Clearly, both β and u are unique.

DEFINITION 3.2.

(a) Let $\Im_v$ be the group of A-fractional ideals of k *prime to* v.

(b) Let $m \in \mathbf{N}$. We let $\mathcal{P}_0(v^m)$ be the group of principal ideals generated by *positive* $\alpha \in k$ prime to v and $\alpha \equiv 1 \pmod{v^m}$. Put $\hat{\mathcal{P}}_v = \mathcal{P}_0(v^0)$. (Note: We use the symbol "$\mathcal{P}_0(?)$" to distinguish between the above group and $\mathcal{P}(v^m)$ of Chapter 1. The group $\mathcal{P}(v^m)$ does *not* use positivity.)

(c) We set

$$\hat{\Im}_v = \varprojlim_m \Im_v / \mathcal{P}_0(v^m) .$$

There is an obvious exact sequence

$$0 \to \varprojlim_m \hat{\mathcal{P}}_v / \mathcal{P}_0(v^m) \to \hat{\Im}_v \to \Im_v / \hat{\mathcal{P}}_v \to 0 .$$

As before, we see that $\Im_v / \hat{\mathcal{P}}_v$ is an extension of $\mathrm{Pic}(\mathbf{A})$ by $\mathbf{F}_\infty^* / \mathbf{F}_r^*$. Moreover, one see that

$$\varprojlim_m \hat{\mathcal{P}}_v / \mathcal{P}_0(v^m) \xrightarrow{\sim} U ,$$

where $U \subseteq \mathbf{k}_v$ is the group of units.

DEFINITION 3.3. (a) Let $c = \deg(v)$. We set

$$S_v = \varprojlim_j \mathbf{Z}/p^j(r^c - 1)) \simeq \mathbf{Z}/(r^c - 1) \times \mathbf{Z} .$$

(b) Let $s_v = (s_0, s_1) \in S_v$. Let $\alpha \in \mathbf{k}$ be a v-adic unit. Then we set

$$\alpha^{s_v} = \omega_v(\alpha)^{s_0} \langle \alpha \rangle_v^{s_1} .$$

As in Section 2, we can extend, in a possibly non-canonical fashion, α^{s_v} to a $\overline{\mathbf{k}}_v^*$-valued mapping of $\hat{\Im}_v$. For $\varrho \in \hat{\Im}_v$, we denote this mapping by

$$\varrho \mapsto \varrho^{s_v} .$$

Now let ψ be a fixed grössencharakter of type A_0 for L and let $\hat{\psi}$ be the associated Hecke-character. Let $-\omega$ be the infinity type of $\hat{\psi}$. Let f be the conductor of ψ so that for $\alpha \equiv 1(f)$,

$$\tilde{\psi}((\alpha)) = \alpha^{-\omega}\tilde{\chi}(\alpha),$$

where $\tilde{\chi}$ is the product of all characters of finite order at the infinite primes of L. Let $\bar{v}$ be a fixed prime of $\overline{\mathbf{k}}$ above v; so via $\bar{v}$ we can consider $\overline{\mathbf{k}}$ as a subfield of $\overline{\mathbf{k}}_v$.

Let $L(\psi, s)$ be as in Section 2. Our first result is

THEOREM 3.1. *$L(\psi, s)$ is an essentially algebraic function.*

PROOF. We need only check the second part of Definition 2.6. Moreover, we can clearly restrict ourselves to looking at the function $L(s) = L_{\epsilon,\rho}(s)$ of the main theorem of Section 2.

Let i be a non-negative integer. As in Definition 2.6.2, set

$$\begin{aligned} h_L(x, -i) &= L(x\pi^{i/d}, -i) \\ &= \sum_{\substack{\varrho \in \mathcal{P}(f) \\ \varrho\varrho_\epsilon \text{ integral} \\ \operatorname{sgn}(N\varrho) \in \theta_\rho}} \tilde{\psi}(\varrho) x^{-\deg N\varrho} (N\varrho)^i. \end{aligned}$$

Let

$$-\omega = \sum_{\substack{\sigma \ \mathbf{k}\text{-injection} \\ \text{of } L \text{ into } \overline{\mathbf{K}}}} a_\sigma \sigma \quad , \{a_\sigma\} \subseteq \mathbf{Z},$$

and, as is standard, $N = \sum \sigma$. Let i_o now be the smallest integer so that

$$i_0 N - \omega \geq 0,$$

i.e., is a non-negative sum. Let $i \geq i_0$. We need to establish that $h_L(x, -i)$ is a polynomial in x^{-1} with algebraic coefficients.

As in the proof of Theorem 2.1, we fix t_0, then decompose into equivalence classes and express the sums obtained in terms of $\hat{\varrho}_\epsilon(D_\alpha)$ by the binomial theorem, etc. By our choice of i, one sees directly that what appears is a *finite* sum of *finite* products of the injections σ. The result now follows from the Riemann-Roch Theorem, Lemma 2.1.a, and Proposition 1.2. □

Let i_0 be as above and g_ψ as in Theorem 2.1. Then the set of all $s = (x, y) \in S_\infty$ with

$$g_\psi/t \leq \nu_\infty(x) \leq i_0/d,$$

is a sort of "critical strip" for $L(\psi, s)$. Indeed it represents those (x, y) between the places where the Euler products *converge* and where the $h_L(x, -i)$ are *polynomials*.

Let $j \in \mathbf{Z}$. Let ϱ be a fractional ideal of $\mathbf{k}$ prime to v. In chapter 2 we showed how to define "ϱ^j" for j considered in S_∞ and above we saw how to define "ϱ^j" for $j \in S_v$. It is a simple exercise to see that, via $\bar{v}$, we can make the definitions *compatible* in the sense that both definitions are the image of the *same* element of

$\overline{\mathbf{k}}^*$. Moreover, via $\bar{v}$, it is clear that we may view $L(\psi(x\pi^{i/d}, -i))$ as a $\overline{\mathbf{k}}_v$-*valued polynomial* in x^{-1} for $i \geq i_0$.

For $i \geq i_0$ put

$$L_v(\psi, (x, -i)) = L(\psi, (x\pi^{i/d}, -i))E_v(\psi, (x\pi^{i/d}, -i)),$$

where E_v is the product of of all Euler v-factors; so we are just removing them from $L(\psi, (x\pi^{i/d}, -i))$. Clearly $L_v(\psi, (x, -i))$ is still a polynomial in x^{-1} with algebraic coefficients.

Finally, we leave the obvious definition of "entire function" on $\overline{\mathbf{k}}_v \times S_v$ to the reader. We consider $L_v(\psi(x, -i))$ as a function on $\overline{\mathbf{k}}_v^*$ via the substitution $x_v \to x$, $x_v \in \mathbf{k}_v^*$.

THEOREM 3.2. *The polynomials $L_v(\psi, (x_v, -i)), i \geq i_0$, interpolate to an entire function $L_v(\psi, (x_v, y_v))$ on $\overline{\mathbf{k}}_v^* \times S_v$.*

PROOF. The proof mimics the proof of Theorem 2.1; the main difference being that now one also needs to take into account congruences modulo v. So one takes representatives of $I(f)/\mathcal{P}(f)$ prime to v, etc., and replaces D_α with

$$D_\alpha - \sum v_i ;$$

where the v_i are the primes of L above v. One now uses the Riemann-Roch Theorem and our main estimates *v-adically* to conclude the proof. □

It is easy to see that $L_v(\psi, (x_v, y_v))$ has an Euler-product on the "half-plane" of $\overline{\mathbf{k}}_v^* \times S_v$ of all (x_v, y_v) with $\nu_v(x_v) > 0$.

DEFINITION 3.4. Let F be an essentially algebraic function on S_∞ as in Definition 2.6. We say F has "strong v-adic interpolations" if the polynomials $h_F(x, -i)$, $i \gg 0$, interpolate to entire functions on $\overline{\mathbf{k}}_v^* \times S_v$ for all v.

Thus, by Theorem 3.2, we see that the L-series of Drinfeld modules with complex multiplication have strong v-adic interpolations. As remarked in the introduction, it is not unreasonable to expect that this will be a general phenomenon.

REFERENCES

1. G. Anderson, *t-motives*, Duke Math. J. **53** (1986), 457-502.
2. P. Deligne, *Valeurs de Fonctions L et périodes d'Intégrales*, Proc. Symposia in Pure Math., vol. 33, 1979, pp. 313-346.
3. G. Faltings, *Endlichkeitssätze für abelsche Varietäten über Zahlkörpern*, Invent. Math. **73** (1983), 349-366; English transG. Cornell and J. Silverman (ed.), *Arithmetic Geometry*, Springer, 1986.
4. W. Fulton, *Algebraic Curves*, Benjamin, 1969.
5. E.-U. Gekeler, *On Finite Drinfeld Modules*, The Journal of Algebra **141** (1991), 187-203.
6. D. Goss, *On the Holomorphy of Certain Non-abelian L-series*, Math. Annalen **272** (1985), 1-9.
7. D. Goss, *The Theory of Totally-real Function Fields*, Applications of Algebraic K-Theory to Algebraic Geometry and Number Theory, vol. 55, AMS, 1986, pp. 449-477.
8. D. Goss, *Analogies Between Global Fields*, Conf. Proc. Canadian Math. Soc., vol. 7, 1987, pp. 83-114.

9. D. Hayes, *Hecke Characters and Eisenstein Reciprocity.*
10. G. Henniart, *Représentations l-adiques abéliennes*, Séminaire de Théorie des Nombres, Paris 1980-81, vol. 22, Birkhäuser, 1982, pp. 107-126.
11. S. Lang, *Complex Multiplication*, Springer, 1983.
12. N. Schappacher, *Periods of Hecke Characters*, Lecture Notes in Mathematics #1301, Springer, 1988.
13. T. Takahashi, *Good Reduction of Elliptic Modules*, J. Math. Soc. Japan **34** (1982), 475-487.
14. Y. Taguchi, *Endomorphisms of Drinfeld Modules over Fields of "Infinite Characteristic"*.
15. J. Tate, *Number Theoretic Background*, Proc. A.M.S. Symp. Pure Math, vol. 33, 1979, pp. 3-26.
16. D. Thakur, *Gauss Sums for* $\mathbf{F}_q[T]$, Inv. Math. **94** (1988), 105-112.
17. A. Weil, *On a Certain Type of Character of the Idèle-Class Group of an Algebraic Number-Field*, Proc. Int. Symp. on Algebraic Number Theory, Tokyo-Nikko, 1955, pp. 1-7.
18. A. Weil, *Basic Number Theory*, Springer, New York, 1967.

DEPARTMENT OF MATHEMATICS, THE OHIO STATE UNIVERSITY, COLUMBUS, OHIO 43210
E-mail address: goss@coltrane.mps.ohio-state.edu

This page intentionally left blank

Contemporary Mathematics
Volume 133, 1992

A p-Adic Cohomological Method for the Weierstrass Family and Its Zeta Invariants

GORO KATO

Dedicated to my parents.

ABSTRACT. After a survey of the Weierstrass family and cohomology, we compute the lifted homology of the Weierstrass family with compact supports so that explicit formulae for the zeta function of each fibre of the Weierstrass family may be obtained. The (co-)homology theory that we use is found in [L_1], [L_2] and [L_3]. Therefore, this article can be regarded as an application of Lubkin's p-adic theory of cohomologies to an algebraic family called the Weierstrass scheme over the ring $(\mathbb{Z}/p\mathbb{Z})[g_2,g_3]$. The cohomological background for the computation will be rather carefully exploited.

1. Introduction

One of the most fundamental motivations in algebraic geometry is to study the common zero points of a finite number of polynomials in several variables, and one of the main goals in number theoretic algebraic geometry is to count the number of common zero points of the set of polynomials.

The congruence zeta function associated with the polynomials provides the number of zeros. That is, let $k = \mathbb{F}_q$ be a finite field of $q = p^a$ elements, and let V be an algebraic variety over k and a complete scheme embedded in the projective space $\mathbb{P}^N(\Omega)$ over the universal domain Ω.

The final version of this paper will appear in [K].
1991 Mathematics Subject Classification. Primary 14F30, 14G10.
Partially suported by CARE Grant 5917.

For each natural number $m = 1, 2, \ldots$, let $k_m \subset \Omega$ be the unique extension of degree m over k, i.e. $k_m = \mathbb{F}_{q^m}$. The finite set N_m denotes the number of points on V whose coordinates are in k_m, i.e. k_m-rational points on V. That is,

$$N_m = |V \cap \mathbb{P}^N(k_m)|.$$

First consider the infinite series in u:

$$N_1 + N_2 u + N_3 u^2 + \ldots,$$

whose integral is given by

$$N_1 u + \frac{N_2}{2} u^2 + \frac{N_3}{3} u^3 + \ldots.$$

The congruence zeta function of V is defined by

$$Z_V(u) = \exp\left(N_1 u + \frac{N_2}{2} u^2 + \frac{N_3}{3} u^3 + \ldots\right).$$

Notice that the first few terms look like

$$Z_V(u) = 1 + N_1 u + \left(\frac{N_2}{2} + \frac{1}{2} N_1^2\right) u^2 + \left(\frac{N_3}{3} + \frac{N_1 N_2}{2} + \frac{N_1^3}{3!}\right) u^3 + \ldots.$$

Therefore, it is sufficient to know the explicit form of the zeta function to know numbers of zeros, $N_1, N_2, \ldots$ in $\mathbb{P}^N(k_1)$, $\mathbb{P}^N(k_2)$, ... respectively. An invariant called a p-adic cohomology group of V will be the device for us to compute the zeta function of V in what follows.

We are mainly concerned with the Weierstrass family and its zeta function (zeta matrix). This is one of few cases where exact computation is possible, and applications to factoring integers via elliptic curves may be possible. See [Le].

The Weierstrass family $\mathcal{W}_R$ corresponding to any ring R with 1_R is obtained by a normalization by linear changes of coordinates of "the square root of a general cubic family" $Y^2 = ax^3 + bx^2 + cx + d$, $a \neq 0$. Geometrically speaking, the Weierstrass family $\mathcal{W}_R$ is the pull-back of the square root of the general cubic family

$$\text{Proj}([a, a^{-1}, b, c, d, X, Y, Z] \,/\, (\text{homogeneous ideal generated by}$$

$$-Y^2Z + aX^3 + bX^2Z + CXZ^2 + dZ^3))$$

under the closed immersion

$$\text{Spec } (R[g_2, g_3]) \hookrightarrow \text{Spec } (R(a, a^{-1}, b, c, d]).$$

Here, the closed immersion is defined by the ideal generated by $\{a - 4, b, c + g_2, d + g_3\}$ and such that 2 is invertible in R. Explicitly,

$$\mathcal{W}_R = \text{Proj}(R[g_2, g_3, X, Y, Z]) \,/\, (\text{homogeneous ideal generated by}$$

$$-Y^2Z + 4X^3 - g_2XZ^2 - g_3Z^3)),$$

where each of X, Y and Z has degree $+1$ and all the elements of $R[g_2, g_3]$ have degree zero in the $R[g_2, g_3]$-algebra $R[g_2, g_3, X, Y, Z]$. On the other hand, the above linear changes provide a map over spec R

$$\mathrm{Spec}\,(R[g_2, g_3]) \leftarrow \mathrm{Spec}\,(R[a, a^{-1}, b, c, d])$$

such that the pull-back of $\mathcal{W}_R$ under this map is canonically isomorphic to the general cubic family, assuming 6 is invertible in R.

By applying the Jacobian criterion to the corresponding affine algebraic Weierstrass family over $\mathrm{Spec}\,(R[g_2, g_3])$, we find that the set of points of the base $\mathrm{Spec}\,(R[g_2, g_3])$ over which the fibre is singular is the set of all the points on the hypersurface $\Delta = g_2^3 - 27g_3^2 = 0$.

Note that for a point $\mathfrak{p} \in \mathrm{Spec}\,(R[g_2, g_3])$ on the base of $\mathcal{W}_R$ the fibre of $\mathcal{W}_R$ over $\mathfrak{p}$ is singular if and only if $\Delta = g_2^3 - 27g_3^2$ vanishes at $\mathfrak{p}$, i.e. Δ goes into zero in $\mathbb{K}(\mathfrak{p})$, the residue class field at $\mathfrak{p}$. At such a point $\mathfrak{p}$, observe that all the singular points lie on the affine open $\mathrm{Spec}(\mathbb{K}(\mathfrak{p})[X, Y] / (Y^2 - 4X^3 + g_2X + g_3))$. There is one and only one singular point on this affine open, i.e. (0,0) or $\left(-\frac{3}{2}\frac{g_3^{(0)}}{g_2^{(0)}}, 0\right)$ for the images $g_2^{(0)}, g_3^{(0)}$ of g_2, g_3 in $\mathbb{K}(\mathfrak{p})$: $g_2^{(0)} = g_3^{(0)} = 0$, or $g_3^{(0)} \neq 0$ (hence $g_2^{(0)} \neq 0$), respectively. Notice that the singular point on the singular fibre is a rational point over $\mathbb{K}(\mathfrak{p})$. Furthermore, if $\Delta = 0$ but $g_2^{(0)} \neq 0$ (hence $g_3^{(0)} \neq 0$), then exactly

two of the roots of the cubic $4X^3 - g_2X - g_3 = 0$ are equal, i.e. we have a projective line with an ordinary double point over $\mathbb{K}(\mathfrak{p})$. If $\Delta = 0$ and $g_2^{(0)} = 0$ (hence $g_3^{(0)} = 0$), then all the three roots are equal, i.e., the fibre is the cusp $Y^2 = 4X^3$.

2. Cohomology

Let X be a complex algebraic variety which is embeddable over the field of complex numbers $\mathbb{C}$. Then there exists a canonical isomorphism between the homology of X with compact supports $H_h^c(X, \mathbb{C})$ (see [L_2] for its definition) and the usual singular homology of the classical topological space X_{top} with compact supports. That is, let X_{top} be the closed points of X with the classical Hausdorff topology. Then by the definition of $H_h^c(X, \mathbb{C})$, one shows $H_h^c(X, \mathbb{C}) \approx H_h^c(X_{\text{top}}, \mathbb{C})$, the usual singular homology with compact supports. The proof is given essentially by definition of $H_h^c(X, \mathbb{C})$: Take Y, which is simple over $\mathbb{C}$, so that X may be closed in Y. Then $H_h^c(X, \mathbb{C}) = H^{2N-h}(Y, Y - X, \Omega_{\mathbb{C}}^*)$. Since Y is simple over $\mathbb{C}$, we have canonically $H^{2N-h}(Y, Y - X, \Omega_{\mathbb{C}}^*) \xrightarrow{\approx} H^{2N-h}(Y_{\text{top}}, Y_{\text{top}} - X_{\text{top}}, \mathbb{C})$, the classical singular cohomology. By Lefschetz duality being applied to the oriented $2N$-dimensional topological manifold Y_{top} and the subspace X_{top}, we have $H_{2N-h}(Y_{\text{top}}, Y_{\text{top}} - X_{\text{top}}, \mathbb{C}) \approx$

$\check{H}^h_c(X_{\text{top}}, \mathbb{C})$. Here $\check{H}^h_c(X_{\text{top}}, \mathbb{C})$ denotes the classical Čech cohomology. Since X is an algebraic variety, we have $\check{H}^h_c(X_{\text{top}}, \mathbb{C}) \approx H^h_c(X_{\text{top}}, \mathbb{C})$. All the groups are finitely generated over $\mathbb{C}$. Hence, taking the duality, we have

$$H^{2N-h}(Y_{\text{top}}, Y_{\text{top}} - X_{\text{top}}, \mathbb{C}) \approx H^c_h(X_{\text{top}}, \mathbb{C}),$$

completing the proof.

Particularly, if X is an embeddable complete complex algebraic variety, then $H^c_h(X, \mathbb{C}) \approx H_h(X_{\text{top}}, \mathbb{C})$.

On a compact topological space, singular homology with compact supports is the same as ordinary singular homology. In particular for a fibre X of the Weierstrass family over a point $\mathfrak{p}$ in the base where $\mathbb{K}(\mathfrak{p}) = \mathbb{C}$, we have the following:

$$H^c_0(X, \mathbb{C}) \approx \mathbb{C}$$

$$H^c_1(X, \mathbb{C}) \approx \begin{cases} \mathbb{C} \oplus \mathbb{C} & \text{for an elliptic curve } X \\ \mathbb{C} & \text{for a projective line with an ordinary double point} \\ 0 & \text{for a projective line with a cusp} \end{cases}$$

$$H^c_2(X, \mathbb{C}) \approx \mathbb{C}$$

and $H^c_h(X, \mathbb{C}) = 0$ for $h \neq 0, 1, 2$.

Another general principle for varieties over characteristic zero fields is the following theorem.

THEOREM. *Let K be a field of characteristic zero, let L be an extension field of K. Let X be an algebraic variety over K which is embeddable over K. Then $X \underset{K}{\times} L$ is an algebraic variety over L which is embeddable over L, and we have canonically*

$$H_h^c(X, K) \underset{K}{\otimes} L \approx H_h^c(X \underset{K}{\times} L, L)$$

as vector spaces over L.

The proof of this theorem goes as follows. Let Y be simple over K containing X as a closed subvariety. By the facts $Y \underset{K}{\times} L$ is affine over Y and the direct image of $\Omega_L^*(Y \underset{K}{\times} L)$ is $(\Omega_K^*(Y)) \underset{K}{\otimes} L$, we have

$$H^h(Y, Y - X, \Omega_K^*) \underset{K}{\otimes} L \approx H^h(Y \underset{K}{\times} L,\ Y \underset{K}{\times} L - X \underset{K}{\times} L,\ \Omega_L^*).$$

NOTE If we have such a hypothesis as: K and L are rings containing $\mathbb{Q}$ and there is a ring homomorphism from K to L, the conclusion of the above theorem becomes a right half plane spectral sequence

$$\mathrm{Tor}_p(H_q^c(X, K), L) \Rightarrow H_n^c(X, L).$$

See the forthcoming [K]. A generalization of the spectral sequence to the non-constant characteristics is difficult, since the dagger completion is involved.

Namely, one can compute the homology with compact supports of any embeddable variety X over a field K of characteristic zero by the following Lefschetz Principle:

For a field K embeddable in the field of complex numbers $\mathbb{C}$, by the theorem above,

$$\text{(LP)} \qquad H_h^c(X, K) \otimes \mathbb{C} \approx H_h^c(X \underset{K}{\times} \mathbb{C}, \mathbb{C})$$

holds. The right-hand side is the classical complex homology with compact supports of the complex algebraic variety $X \underset{K}{\times} \mathbb{C}$. If K is an arbitrary field of characteristic zero, $H_h^c(X, K)$ can still be computed by an embeddable algebraic variety X_0 over a finitely generated field K_0 over $\mathbb{Q}$ with $X_0 \underset{K_0}{\times} K \approx X$. That is, $H_h^c(X, K) \approx H_h^c(X_0, K_0) \underset{K_0}{\otimes} K$, $H_h^c(X_0, K_0)$ can be handled by (LP) above.

We will apply what we have mentioned to the fibres of the Weierstrass family.

PROPOSITION. *Let X be a fibre of the Weierstrass family $\mathcal{W}_R$, corresponding to any commutative ring R with identity, over a point*

$\mathfrak{p} \in \operatorname{Spec}(R[g_2, g_3])$ *such that* $k = \mathbb{K}(\mathfrak{p})$ *is of characteristic zero.*

Then we have:

$$H_0^c(X, k) \approx k$$

$$H_1^c(X, k) \approx$$

$$\begin{cases} k \oplus k, & \text{if } X \text{ is non-singular, i.e., } X \text{ is an elliptic curve} \\ k, & \text{if } X \text{ is a projective line with an ordinary double point} \\ 0, & \text{if } X \text{ is a projective line with a cusp} \end{cases}$$

$$H_2^c(X, k) \approx k, \text{ and}$$

$$H_h^c(X, k) \approx 0 \text{ for } h = 3, 4, \ldots .$$

Next we will consider the non-zero characteristic case. Let R be a ring and let X be a fibre of $\mathfrak{W}_R$ over some $\mathfrak{p} \in \operatorname{Spec}(R[g_2, g_3])$. Let $k = \mathbb{K}(\mathfrak{p})$. Suppose the characteristic p of k is not zero. Let $\mathfrak{O}$ be a complete discrete valuation ring with mixed characteristics with k as its residue class field and K as the quotient field of $\mathfrak{O}$. We have the following facts.

PROPOSITION. *If* X *is non-singular, then the lifted* K*-adic homology with compact supports behaves as follows:*

$$H_h^c(X, K) = \begin{cases} K & \text{for } h = 0 \text{ or } 2 \\ K \oplus K & \text{for } h = 1 \\ 0 & \text{for } h \neq 0, 1, 2 . \end{cases}$$

PROOF. If X is liftable over $\mathfrak{O}$ by a simple and proper lifting over K, then X_K is an elliptic curve. Then $H^c_h(X, K) = H^{2N-h}(X, K)$ by taking $Y = X$ in the definition. Then, by $[L_1]$, $H^{2N-h}(X, K)$ is isomorphic to the hypercohomology $H^{2N-h}(X_K, K)$. We have $H^{2N-h}(X_K, K) \approx H^c_h(X_K, K)$, where $H^c_h(X_K, K)$ is computed in the previous proposition for $h = 0, 1, 2, \ldots$.

If X is a singular fibre, then one can prove $H^c_0(X, K) \approx K$, $H^c_2(X, K) \approx K$ and $H^c_h(X, K) \approx 0$ for $h = 3, 4, \ldots$. In the following section, we will prove the following by direct computation:

$$H^c_1(X, K) \approx \begin{cases} K, & \text{if } X \text{ is a projective line with an ordinary double point} \\ 0, & \text{if } X \text{ is a projective line with a cusp.} \end{cases}$$

NOTE Each fibre X of the Weierstrass scheme $\mathfrak{W}_R$ over $\mathfrak{p} \in \operatorname{Spec}(R[g_2, g_3])$ has a rational and simple point, called the point at infinity with homogeneous coordinates (0,1,0). We denote the affine curve obtained from X - (0,1,0) by U. The long exact sequence of the homology with compact supports is induced as:

$$\cdots \to H^c_{h-2n}((0,1,0), K) \to H^c_h(X, K) \to H^c_h(U, K) \to \cdots,$$

where $n = \dim X = 1$ in our case. The first homology group is trivial unless $h = 2$. Therefore, it suffices to compute $H_1^c(U, K)$. Note also that U is closed in $\mathbb{A}^2(k) = \text{Spec}\,(k[X, Y])$, whose lifting is given by $\mathbb{A}^2(\mathcal{O}) = \text{Spec}\,(\mathcal{O}[X, Y])$. Then $H_h^c(U, K) \approx H^{4-h}\big(\mathbb{A}^2(k), \mathbb{A}^2(k) - U, \Omega^*_{\mathcal{O}}((\mathbb{A}^2(\mathcal{O}))^\dagger \underset{\mathcal{O}}{\otimes} K\big)$, $h \in \mathbb{Z}$. Since $\mathbb{A}^2(k)$ and $\mathbb{A}^2(k) - U$ are both affine open sets, the covering $\{\mathbb{A}^2(k), \mathbb{A}^2(k) - U\}$ may be used to compute the homology group.

3. The Universal Coefficient Spectral Sequence

Let $\underline{A}$ be an $\mathcal{O}$-algebra with an endomorphism F on $\underline{A}$ such that F induces the p-th power endomorphism of $A = \underline{A}\,/\,p\underline{A}$, where p is the characteristic of $k = \mathbb{K}(\mathcal{O})$. Then there exists a unique ring homomorphism

$$\underline{A} \to W(A^{p^{-\infty}}),$$

where $W(A)$ is the Witt vector of $A = \underline{A}/p\underline{A}$, such that the above ring homomorphism is compatible with the endomorphism F of $\underline{A}$, inducing the identity of A. The construction of the ring homomorphism $\underline{A} \to W(A)$ is as follows: let $\underline{A}^{F^{-\infty}}$ be the direct limit of the sequence

$$\underline{A} \xrightarrow{F} \underline{A} \xrightarrow{F} \underline{A} \xrightarrow{F} \cdots .$$

Then, $\underline{A}^{F^{-\infty}\wedge}$ obeys the universal characterization of the Witt vector $W(A)$ of A. Next let $\mathfrak{p}$ be a prime ideal of A. Then we have a natural map from $\underline{A}$ into $W(\mathbb{K}(\mathfrak{p})^{p^{-\infty}})$, which is compatible with F. For example, if $\mathbb{K}(\mathfrak{p})$ is perfect, there is induced a unique ring homomorphism from $\underline{A}$ into $W(\mathbb{K}(\mathfrak{p}))$ of $\mathbb{K}(\mathfrak{p})$. If $\mathbb{K}(\mathfrak{p})$ is a finite field, then there is a natural homomorphism

$$\underline{A} \to W(\mathbb{K}(\mathfrak{p})),$$

where $W(\mathbb{K}(\mathfrak{p}))$ is the unique Witt vector of $\mathbb{K}(\mathfrak{p})$ that is the mixed characteristic complete discrete valuation ring with its residue class field $\mathbb{K}(\mathfrak{p})$. In the case of the Weierstrass family, we let $\underline{A} = \mathbb{Z}_p[g_2, g_3]$. Then for any given closed point $\mathfrak{p} \in \mathrm{Spec}\,(A)$, the residue class field $\mathbb{K}(\mathfrak{p})$ is finite. Then the images $g_2^{(0)}$ and $g_3^{(0)}$ of g_2 and g_3 in $\mathbb{K}(\mathfrak{p})$ generate $\mathbb{K}(\mathfrak{p})$ over the prime field, i.e. $\mathbb{K}(\mathfrak{p}) = (\mathbb{Z}/p\mathbb{Z})[g_2^{(0)}, g_3^{(0)}]$. The Witt vector $W(\mathbb{K}(\mathfrak{p}))$ of $\mathbb{K}(\mathfrak{p})$ can be described as follows: each of $g_2^{(0)}$ and $g_3^{(0)}$ is either a root of unity of order prime to p, or else zero. Choose an element $\rho \in \mathbb{K}(\mathfrak{p})$ which is a multiplicative generator of the multiplicative cyclic group $\mathbb{K}(\mathfrak{p}) - \{0\}$.

Then each element of $\mathbb{K}(\mathfrak{p})$, e.g. $g_2^{(0)}$ and $g_3^{(0)}$, is either a power of ρ or else zero. Let a be the multiplicative order of ρ. Then embed $\hat{\mathbb{Z}}_p$ as a subring of $\mathbb{C}$, and let ρ' be any fixed root of unity in $\mathbb{C}$ of order exactly a. Then, $W(\mathbb{K}(\mathfrak{p})) = \hat{\mathbb{Z}}_p[\rho']$, i.e., the subring of $\mathbb{C}$ generated by $\hat{\mathbb{Z}}_p$ and ρ'. Let $(g_2^{(0)})' = (\rho')^i$, where $g_2^{(0)} = \rho^i$ and similarly for $(g_3^{(0)})$. If $g_2^{(0)} = 0$, then define $(g_2^{(0)})' = 0$. They are Teichmüller representatives of $g_2^{(0)}$ and $g_3^{(0)}$ in $W(\mathbb{K}(\mathfrak{p}))$.

Let $\underline{A}$ be an $\mathfrak{O}$-algebra, and let F be a ring endomorphism of $\underline{A}$ which induces the p-th power endomorphism of $\underline{A}/p\underline{A}$. Then, for any prime ideal $\mathfrak{p} \in \operatorname{Spec}(\underline{A} \underset{\mathfrak{O}}{\otimes} k)_{\mathrm{red}}$, we have shown in the above how one obtains a natural homomorphism from $\underline{A}$ to $W(\mathbb{K}(\mathfrak{p}))$. For example, for $\underline{A} = \hat{\mathbb{Z}}_p[g_2, g_3]$, if $\mathfrak{p}$ is a maximal ideal of $A = (\underline{A}/p\underline{A})_{\mathrm{red}} = (\underline{A} \underset{\mathfrak{O}}{\otimes} k)_{\mathrm{red}}$ so that $\mathbb{K}(\mathfrak{p}) = (\mathbb{Z}/p\mathbb{Z})[g_2^{(0)}, g_3^{(0)}]$, then $W(\mathbb{K}(\mathfrak{p})) = \hat{\mathbb{Z}}_p[(g_2^{(0)})', (g_3^{(0)})']$. In this case, the natural map $\underline{A} \to W(\mathbb{K}(\mathfrak{p}))$ is given by $g_2 \mapsto (g_2^{(0)})'$ and $g_3 \mapsto (g_3^{(0)})'$.

Let $\underline{B} = W(\mathbb{K}(\mathfrak{p})^{p^{-\infty}})$. Then $\underline{B}$ is a complete discrete valuation ring and $\underline{B} \underset{\mathbb{Z}}{\otimes} \mathbb{Q}$ is a field of characteristic zero. If X is a scheme over $\operatorname{Spec}(A)$ that is embeddable over $\underline{A}$, then the fibre $X_{\mathfrak{p}}$ over

$\mathbb{K}(\mathfrak{p})$ is an algebraic variety over the field $\mathbb{K}(\mathfrak{p})$. Let $Y_{\mathfrak{p}} = X_{\mathfrak{p}} \underset{\mathbb{K}(\mathfrak{p})}{\times} \mathbb{K}(\mathfrak{p})^{p^{-\infty}}$. Then the zeta matrices have coefficients in the quotient field $K_{\mathfrak{p}} = \underline{B} \underset{\mathbb{Z}}{\otimes} \mathbb{Q}$ of the complete discrete valuation ring $\underline{B} = W(\mathbb{K}(\mathfrak{p})^{p^{-\infty}})$. The universal coefficient spectral sequence is:

$$E^2_{p,q} = \mathrm{Tor}_p^{\underline{A}^\dagger \underset{\mathbb{Z}}{\otimes} \mathbb{Q}}(H^c_q(X, \underline{A}^\dagger \underset{\mathbb{Z}}{\otimes} \mathbb{Q}), K_{\mathfrak{p}})$$

with the abutment $H^c_n(Y_{\mathfrak{p}}, K_{\mathfrak{p}})$. Principally speaking, this universal spectral sequence shows how the lifted p-adic homology of the scheme X over Spec (A) with compact supports determines the lifted p-adic homology of all the fibres $Y_{\mathfrak{p}}$ in this algebraic family. Furthermore, the zeta endomorphisms of $H^c_*(X, \underline{A}^\dagger \underset{\mathbb{Z}}{\otimes} \mathbb{Q})$ compute the zeta endomorphisms of $H^c_*(Y_{\mathfrak{p}}, K_{\mathfrak{p}})$ of each fibre $Y_{\mathfrak{p}}$.

Suppose that $\mathbb{K}(\mathfrak{p})$ is a finite field. If the term $E^2_{p,q}$ of the above universal coefficient spectral sequence is finite dimensional over the quotient field $K_{\mathfrak{p}}$ of the complete discrete valuation ring $W(\mathbb{K}(\mathfrak{p}))$ for all the p and q, we can compute the zeta function of each fibre

$Y_{\mathfrak{p}} = X_{\mathfrak{p}}$ as follows. Let $P_{p,q}$ be the reverse characteristic polynomial of the endomorphism of the $E^2_{p,q}$-term, which is induced by p^r-th power map, $p^r = \mathrm{card}(\mathbb{K}(\mathfrak{p}))$. Then, the zeta function of the fibre $X_{\mathfrak{p}}$ is provided by the formula

$$Z_{X_{\mathfrak{p}}}(T) = \frac{\prod_{p+q=\text{odd}} P_{p,q}(T)}{\prod_{p+q=\text{even}} P_{p,q}(T)},$$

where we assume that $E^2_{p,q} = 0$ for all but finitely many pairs of p and q. When the lifted p-adic homology with compact supports is a free module, the zeta endomorphism is said to be the zeta matrix.

4. Zeta Endomorphisms and Zeta Matrices

We will compute the zeta endomorphism of the first homology group with compact supports of the finite points of the Weierstrass family $H_1^c(U, \underline{A}^\dagger \underset{\mathbf{Z}}{\otimes} \mathbb{Q})$. That is:

$$U = \mathbb{W}_{\mathbf{Z}/p\mathbf{Z}} \cap \mathbb{A}^2(\mathrm{Spec}((\mathbf{Z}/p\mathbf{Z})[g_2, g_3])).$$

Note that $H_1^c(U, \underline{A}^\dagger \underset{\mathbf{Z}}{\otimes} \mathbb{Q}) \overset{\approx}{\leftarrow} H_1^c(\mathbb{W}_{\mathbf{Z}/p\mathbf{Z}}, \underline{A}^\dagger \underset{\mathbf{Z}}{\otimes} \mathbb{Q})$ as seen, e.g., in [K$_2$]. Let $\mathbb{K}$ be the quotient field of $\underline{A} = \hat{\mathbf{Z}}_p[g_2, g_3]$, and let $\mathbb{K}^\dagger$ be

the quotient field of $\underline{A}^{\dagger} = \hat{\mathbb{Z}}_p[g_2, g_3]^{\dagger}$. Even though $H_1^c(U, \underline{A}^{\dagger} \underset{\mathbb{Z}}{\otimes} \mathbb{Q})$ is not finitely generated over $\underline{A}^{\dagger} \underset{\mathbb{Z}}{\otimes} \mathbb{Q}$, $H_1^c(U, \underline{A}^{\dagger} \underset{\mathbb{Z}}{\otimes} \mathbb{Q}) \otimes \mathbb{K}$ is a vector space of dimension two over $\mathbb{K}$. From the universal spectral sequence corresponding to any non-zero-diver $t \in \underline{A}$, we have the long exact sequence:

$$\cdots \to H_h^c(U, \underline{A}^{\dagger} \underset{\mathbb{Z}}{\otimes} \mathbb{Q}) \xrightarrow{\text{“}t\text{”}} H_h^c(U, \underline{A}^{\dagger} \underset{\mathbb{Z}}{\otimes} \mathbb{Q})$$

$$\to H_h^c(U', \underline{A}^{\dagger} \underset{\mathbb{Z}}{\otimes} \mathbb{Q} / t \cdot \underline{A}^{\dagger} \underset{\mathbb{Z}}{\otimes} \mathbb{Q}) \to \cdots .$$

One can extract the short exact sequence

$$0 \to H_1^c(U, \underline{A}^{\dagger} \underset{\mathbb{Z}}{\otimes} \mathbb{Q}) \xrightarrow{\text{“}t\text{”}} H_1^c(U, \underline{A}^{\dagger} \underset{\mathbb{Z}}{\otimes} \mathbb{Q}) \to$$

$$H_1^c(U', \underline{A}^{\dagger} \underset{\mathbb{Z}}{\otimes} \mathbb{Q} / t \cdot \underline{A}^{\dagger} \underset{\mathbb{Z}}{\otimes} \mathbb{Q}) \to 0.$$

Hence, $H_1^c(U, \underline{A}^{\dagger} \underset{\mathbb{Z}}{\otimes} \mathbb{Q})$ is torsion free, i.e., we have

$$H_1^c(U, \underline{A}^{\dagger} \underset{\mathbb{Z}}{\otimes} \mathbb{Q}) \hookrightarrow H_1^c(U, \underline{A}^{\dagger} \underset{\mathbb{Z}}{\otimes} \mathbb{Q}) \otimes \mathbb{K}.$$

The zeta matrix of the free module of rank two

$$H_1^c(U, \underline{A}^{\dagger} \underset{\mathbb{Z}}{\otimes} \mathbb{Q}) \underset{\mathbb{K}^{\dagger}}{\otimes} (A_{\Delta}^{\dagger} \underset{\mathbb{Z}}{\otimes} \mathbb{Q})$$

is computed in [K-L]. This free module is isomorphic to the $A_\Delta^\dagger \otimes_{\mathbb{Z}} \mathbb{Q}$-adic cohomology of the open subfamily U_Δ of non-singular fibres, denoted as $H^1(U_\Delta, A_\Delta^\dagger \otimes_{\mathbb{Z}} \mathbb{Q})$. The above isomorphism may be given between generators by

$$C^{-1}\, dX \wedge dY \mapsto Y dX$$

$$X\, C^{-1}\, dX \wedge dY \mapsto XY dX,$$

where $C = Y^2 - 4X^3 + g_2X + g_3$. (See what will follow.)

The zeta matrix $W^1 \in \mathrm{Mat}_{2\times 2}(A_\Delta^\dagger \otimes_{\mathbb{Z}} \mathbb{Q})$ on the free $A_\Delta^\dagger \otimes_{\mathbb{Z}} \mathbb{Q}$-module $H^1(U_\Delta, A_\Delta^\dagger \otimes_{\mathbb{Z}} \mathbb{Q})$ is given as follows: let $F : \underline{A} \to \underline{A}$ be an endomorphism of $\hat{\mathbb{Z}}_p$-algebra, inducing p-th power map on A, defined by

$$F(g_2) = g_2^p \text{ and } F(g_3) = g_3^p.$$

Let $f : U_\Delta \to U_\Delta$ be the p-th power endomorphism of the scheme U_Δ over A_Δ. Then define

$$H^1(F, f)(Y\, dX) = p\, X^{p-1} \sqrt{4X^{3p} - g_2^p X^p - g_3^p}\; dX$$

and

$$H^1(F,f)(XY\,dX) = pX^{2p-1}\sqrt{4X^{3p} - g_2^pX^p - g_3^p}\;dX,$$

where

$$\sqrt{4X^{3p} - g_2^pX^p - g_3^p} =$$

$$\sqrt{4X^{3p} - g_2^pX^p - g_3^p - (4x^3 - g_2X - g_3)^p + (4X^3 - g_2X - g_3)^p}$$

$$= \sqrt{(4X^3 - g_2X - g_3)^p - pT} = \sqrt{Y^{2p} - pT}$$

$$= Y^p\left(\sum_{i\geq 0}\binom{\frac{1}{2}}{i}\left(\frac{-pT}{(4X^3 - g_2X - g_3)^p}\right)^i\right),$$

here $pT = (4X^3 - g_2X - g_3)^p - 4X^{3p} + g_2^p X^p + g_3^p$. See [K-L] for necessary recursive formulae.

Therefore, the zeta endomorphism of $H_1^c(U, \underset{\mathbb{Z}}{A^\dagger} \otimes \mathbb{Q})$ is induced from this zeta matrix by the above restriction

$$H_1^c(U, \underset{\mathbb{Z}}{A^\dagger} \otimes \mathbb{Q}) \hookrightarrow H_1^c(U, \underset{\mathbb{Z}}{A^\dagger} \otimes \mathbb{Q}) \otimes \mathbb{K}.$$

The actual construction is as follows: by the definition of the lifted p-adic homology with compact supports of the Weierstrass family over $A = (\mathbb{Z}/p\mathbb{Z})[g_2, g_3]$,

$$H_1^c(U, \underline{A}^\dagger \underset{\mathbb{Z}}{\otimes} \mathbb{Q}) = H^3\big(\mathbb{A}^2(\mathrm{Spec}((\mathbb{Z}/p\mathbb{Z})[g_2, g_3])),$$

$$\mathbb{A}^2(\mathrm{Spec}((\mathbb{Z}/p\mathbb{Z})[g_2, g_3])) - U,$$

$$\Omega_{\underline{A}}^{*}(\mathbb{A}^2(\mathrm{Spec}(\hat{\mathbb{Z}}_p[g_2, g_3])))^\dagger \underset{\mathbb{Z}}{\otimes} \mathbb{Q}\big).$$

By the lemma in [K_1] on spectral sequences, one shows this third relative hypercohomology is isomorphic to

$$\mathrm{Coker}\,(\Omega_{\underline{A}}^1(\underline{A}[X, Y, C^{-1}])^\dagger \underset{\mathbb{Z}}{\otimes} \mathbb{Q} \overset{d_1^{1,0}}{\to} \Omega_{\underline{A}}^2(\underline{A}[X, Y, C^{-1}])^\dagger \underset{\mathbb{Z}}{\otimes} \mathbb{Q}).$$

We obtain the recursive cohomologous relations among the generators of $H_1^c(U, \underline{A}^\dagger \underset{\mathbb{Z}}{\otimes} \mathbb{Q})$ as

$$\begin{cases} 2(i\text{-}1)\Delta C^{-i}dX \wedge dY \sim (6i\text{-}13)6g_2XC^{-(i-1)}dX \wedge dY\text{-}(6i\text{-}11)9g_3C^{-(i-1)}dX \\ \quad \wedge dY, \text{ and} \\ \\ 4(i\text{-}1)\Delta XC^{-i}dX \wedge dY \sim (6i\text{-}11)g_2^2C^{-(i-1)}dX \\ \quad \wedge dY\text{-}(6i\text{-}13)18g_3XC^{-(i-1)}dX \wedge dY \end{cases}$$

Therefore, we let

$$H_1^c(F, f)(C^{-1}dX \wedge dY) = \frac{p^2X^{p-1}Y^{p-1}}{C^p - pT}\, dX \wedge dY$$

and

$$H_1^c(F, f)(XC^{-1}dX \wedge dY) = \frac{p^2X^{2p-1}Y^{p-1}}{C^p - pT}\, dX \wedge dY,$$

where $\frac{1}{C^p - pT} = C^{-p}(1 + pTC^{-p} + p^2T^2C^{-2p} + p^3T^3C^{-3p} + ...)$. See $[K_2]$ for necessary recursive formulae. Notice that the above recursive cohomologous relations compute the lifted homology groups of various singular fibres, e.g., the fiber over $\mathfrak{p} = (g_2, g_3)$ has the trivial homology group. See page 10.

NOTES

1. The universal spectral sequences are treated in the book $[L_4]$, Chapter 5. As for zeta invariants, see $[L_1]$, $[L_2]$, [K-L] and $[K_2]$.

2. A similar computation of p-th power map of Fermat curves may be given as follows: let U be the affine Fermat curve given by $X^l + Y^l = 1$ over $\mathbf{Z}/p\mathbf{Z}$. Then the associated $\hat{\mathbf{Z}}_p \underset{\mathbf{Z}}{\otimes} \mathbf{Q}$-adic cohomology $H^1(U, \hat{\mathbf{Z}}_p \underset{\mathbf{Z}}{\otimes} \mathbf{Q})$ is generated by $(l - 1) \cdot (l - 2)$ elements $\{X^\alpha Y^{\beta-l+1}dX\}$, where $\alpha = 0, 1, ..., l - 3$ and $\beta = l - 1, l, ..., 2l - 3$. There is a cohomologous relation:

$$X^{i+l-2}Y^{\beta-l+1}dX \sim X^{i+2l-2}Y^{\beta-l+1}dX$$

for $i = 0, 1, ..., l - 1$. Then the $(l - 1) \cdot (l - 2)$ square matrix $H^1(f, \hat{\mathbf{Q}}_p)$ on $H^1(U, \hat{\mathbf{Q}}_p)$ is defined as

$$H^1(f, \hat{\mathbf{Q}}_p)(X^\alpha Y^{\beta-l+1}dX)$$

$$= pX^{\alpha p+p+1}\, Y^{p(\beta-l+1)} \left(\sum_{k\geq 0} \binom{\frac{1}{l}}{k} \left(\frac{-pT}{u}\right)^k \right)^{\beta-l+1} dX,$$

where $u = (1 - X^l)^p$, $pT = (1 - X^l)^p - (1 - X^{lp})$.

See I. V. Volovich, p-adic string, Class. Quantum Grav. 4 (1987), 83, for a connection to a p-adic string theory. See also [Ko].

References

[D] P. Deligne, Letter to G. Kato, 1982.

[D_1] B. Dwork, A Deformation Theory for the Zeta Function of a Hypersurface, Proc. Int. Cong. Math. (1962), 247-259.

[D_2] B. Dwork, p-Adic Cycles, Pub. Math. I.H.E.S. 37 (1969), 27-116.

[D'] B. Dwork, Letters to G. Kato, 1980 and 1985.

[Ka] N. Katz, Travaux de Dwork, Séminaire Bourbaki, 1971/72, n. 409.

[Ko] N. Koblitz, *p-Adic Numbers, p-Adic Analysis and Zeta Functions*, Springer-Verlag, 1977.

[K-L] G. Kato, S. Lubkin, Zeta Matrices of Elliptic Curves, J. of Number Theory 15 (1982), 318-330.

[K_1] G. Kato, On the Generators of the First Homology with Compact Supports of the Weierstrass Family in Characteristic Zero, Trans., AMS, 278 (1983), 361-368.

[K_2] G. Kato, Lifted p-Adic Homology with Compact Supports of the Weierstrass Family and its Zeta Endomorphism, J. of Number Theory, 35, No. 2 (1990), 216-223.

[K] G. Kato, *Zeta Matrices of the Weierstrass Family*, in preparation.

[L_1] S. Lubkin, A p-Adic Proof of Weil's Conjectures, Ann. of Math. (2) 87 (1968), 105-255.

[L_2] S. Lubkin, Finite Generation of Lifted p-Adic Homology with Compact Supports. Generalization of the Weil Conjectures to Singular, Non-complete Algebraic Varieties, J. of Number Theory, 11 (1979), 412-464.

[L_3] S. Lubkin, Generalization of p-Adic Cohomology; Bounded Witt Vectors. A Canonical Lifting of a Variety in Characteristic $p \neq 0$ Back to Characteristic Zero, Compositio Mat. 34 (1977).

[L_4] S. Lubkin, *Cohomology of Completions*, Notas de Matemática, Vol. 42, North Holland, Amsterdam, 1980.

[Le] H. W. Lenstra, Factoring Integers with Elliptic Curves, Ann. of Math. (2) 126 (1987), no. 3, 649-673.

[T] J. Tate, The Arithmetic of Elliptic Curves, Inventiones Math. 23 (1974), 179-289.

[W] A Weil, Number of Solutions of Equations in Finite Fields, Bull. AMS, 55 (1949), 497-508.

MATHEMATICS DEPARTMENT, CALIFORNIA POLYTECHNIC STATE UNIVERSITY, SAN LUIS OBISPO, CALIFORNIA 93407

Contemporary Mathematics
Volume 133, 1992

Two-Dimensional Systems of Galois Representations

MICHAEL LARSEN

ABSTRACT. This paper is concerned with the maximality of the image of the Galois group of a number field in a compatible system of 2-dimensional, semisimple Galois representations. We show that in some sense the image is "as large as possible" for a set of primes ℓ of density 1.

0. Introduction

Let E be an elliptic curve defined over a number field K. For each prime ℓ, define

$$T_\ell = \lim_{\leftarrow n} E_{\ell^n}, \quad V_\ell = T_\ell \otimes_{\mathbb{Z}_\ell} \mathbb{Q}_\ell$$

where E_{ℓ^n} denotes the group of ℓ^n-torsion points of E. The Galois group $G_K = Gal(\bar{K}/K)$ acts continuously on T_ℓ and therefore on V_ℓ. In a series of works ([5], [6], [8], [9]), Serre investigated the image, Γ_ℓ, of G_K in $GL(V_\ell)$. If E has complex multiplication over $\mathbb{C}$, Γ_ℓ is contained in a Cartan subgroup of $GL(V_\ell)$ (resp. the normalizer of a Cartan) if K contains (resp. does not contain) the endomorphism ring of E. If E does not admit complex multiplication, Γ_ℓ is Zariski-dense in $GL(V_\ell)$ and for all $\ell \gg 0$, $\Gamma_\ell = GL(T_\ell)$. In [11], an analogous result is proved for compatible systems of 2-dimensional Galois representations arising from elliptic modular forms.

Fix a finite set S of primes of K. For the purposes of this paper, a *compatible system of Galois representations, unramified outside* S will be a system of continuous representations

$$\rho_\ell : G_K \to GL_n(\mathbb{Q}_\ell),$$

one for each rational prime $\ell \nmid \prod_{\wp \in S} \|\wp\|$, such that

(i) Each ρ_ℓ is unramified outside $S_\ell = S \cup \{\wp : \wp | \ell\}$.

1991 *Mathematics Subject Classification.* Primary 11R32; Secondary 11F33.

The author was supported in part by the National Science Foundation.

This paper is in final form and no version of it will be submitted for publication elsewhere

0271-4132/92 $1.00 + $.25 per page

(ii) If $\wp \notin S_\ell$, the characteristic polynomial of the Frobenius element $\rho_\ell(Frob_\wp)$ has coefficients in $\mathbb{Q}$.

(iii) This characteristic polynomial does not depend on ℓ.

Note that our notion of compatible system is Serre's notion of *strictly compatible* system, [**8**, I-11]. This paper analyzes the dependence of Γ_ℓ on ℓ for two-dimensional representations, using the general philosophy developed jointly with R. Pink in [**3**] and [**4**]. By restricting to the two-dimensional case, we can avoid the group-theoretic difficulties which arise in stating and proving analogous statements in higher dimension, and the final results are somewhat stronger than in the general case.

The result is the following:

THEOREM 1. *If ρ_ℓ is a compatible system of semisimple 2-dimensional representations, one of the following statements must be true:*

1): *The projective representations $\bar{\rho}_\ell : G_K \to PGL_2(\mathbb{Q}_\ell)$ all factor through a fixed finite quotient $\Gamma \cong A_4$ or $\Gamma \cong S_4$ of G_K.*

2): *There exists an extension L/K with $[L : K] \leq 2$ such that the restriction of ρ_ℓ to G_L forms a compatible system of abelian ℓ-adic representations.*

3): *For a set of primes ℓ of Dirichlet density 1, there exists a lattice $\Lambda \cong \mathbb{Z}_\ell^2 \subset \mathbb{Q}_\ell^2$ and a quadratic extension L/K, such that $\rho_\ell(G_L)$ is isomorphic to $SL(\Lambda)$.*

4): *For all ℓ in a set of primes of Dirichlet density 1, Γ_ℓ is Zariski-dense in $GL_2(\mathbb{Q}_\ell)$, and $\Gamma_\ell/Z(\Gamma_\ell)$ is either $PSL(\Lambda)$ or $PGL(\Lambda)$.*

Note that case (1) is trivial, while case (2) is well understood ([**8**]). Note also that in case (4), Γ_ℓ need not be isomorphic to $GL_2(\mathbb{Z}_\ell)$ for a set of primes of density 1. Consider, for instance, the system obtained from the Tate twist of an elliptic curve:

$$H^1_{\acute{e}t}(\bar{E}, \mathbb{Z}_\ell(1)).$$

This system has determininant $\mathbb{Z}_\ell(3)$, so

$$\Gamma_\ell \subset \{M \in GL_2(\mathbb{Z}_\ell) | \det(M) \in \mathbb{Z}_\ell^{*3}\}.$$

When $\ell \equiv 1 \pmod 3$, Γ_ℓ is properly contained in the stabilizer of a lattice Λ; it cannot be equal to the full stabilizer of a different lattice because $GL(\Lambda') \subset GL(\Lambda)$ if and only if Λ' is homothetic to Λ.

1. Preliminary Lemmas

PROPOSITION 1. *Every reductive subgroup G of PGL_2 is either a finite group, a Cartan subgroup, the normalizer of a Cartan subgroup, or the full group.*

PROOF. The identity component of G has rank ≤ 1 and is therefore either the trivial group, a rank-1 torus (which is automatically maximal), or PGL_2.

As G is contained in the normalizer of G°, in the torus case, $G^\circ \subset G \subset N(G^\circ)$. The proposition follows from the fact that G° is of index 2 in $N(G^\circ)$. □

Let $\bar{\Gamma}_\ell$ denote the image of Γ_ℓ under the projection $GL_2(\mathbb{Q}_\ell) \to PGL_2(\mathbb{Q}_\ell)$. Let G_ℓ denote the Zariski closure of Γ_ℓ in $GL_2(\mathbb{Q}_\ell)$, and $\bar{G}_\ell$ the Zariski closure of $\bar{\Gamma}_\ell$ in PGL_2, which is also the image of G_ℓ under $GL_2 \to PGL_2$. If U is an open and closed subset of $PGL_2(\mathbb{Q}_\ell)$ not containing the identity, the set of elements of $PGL_2(\mathbb{Q}_\ell)$ conjugate to an element of U is open and closed and does not contain the identity. Therefore, if $\bar{\Gamma}_{\ell_0}$ is infinite, we can find an infinite sequence U_i of open subsets of $\bar{\Gamma}_{\ell_0}$, such that no element of U_i is conjugate in $PGL_2(\mathbb{Q}_{\ell_0})$ to any element of U_j, $i \neq j$. By the Cebotarev density theorem, there exists a sequence $\wp_i$ such that $\bar{\rho}_{\ell_0}(\wp_i) \in U_i$. The compatibility of the representations ρ_ℓ implies that for all ℓ, $\bar{\rho}_\ell(\wp_i)$ are all distinct, and $\bar{\Gamma}_\ell$ is infinite for all ℓ.

PROPOSITION 2. *If $\bar{\Gamma}_{\ell_0}$ is finite for some ℓ_0, then case (1) or case (2) of the main theorem holds.*

PROOF. We have seen that all $\bar{\Gamma}_\ell$ must be finite. A semisimple element of $GL_2(\mathbb{Q}_\ell)$ maps to the identity in $PGL_2(\mathbb{Q}_\ell)$ if and only if both eigenvalues are equal and are of the form ± 1. For Frobenius elements this criterion is independent of ℓ. By the Cebotarev density theorem, the $\bar{\rho}_\ell$ must share a common kernel G_L, an open subgroup of G_K of finite index. The quotient $\Gamma = G_K/G_L$ is a subgroup of $PGL_2(\mathbb{Q}_\ell)$ for all $\ell \notin S$. The kernel of $PGL_2(\mathbb{Q}_\ell) \to PGL_2(\mathbb{F}_\ell)$ is a pro-ℓ-group, so if $\ell \notin S$ is larger than $|\Gamma|$, we can conclude Γ is a subgroup of $PGL_2(\mathbb{F}_\ell)$. By [2, Ch. XI, Th. 2.3], Γ must be cyclic, dihedral, or isomorphic to one of A_4, S_4, or A_5. Since there are arbitrarily large primes congruent to $\pm 2 \pmod 5$, by [2, Ch. XI, App., Th. 1], A_5 is ruled out. The cyclic and dihedral cases are subsumed in case (2) of the main theorem; the remaining cases constitute case (1). □

We assume henceforth that $|\bar{\Gamma}_\ell| = \infty$ for all ℓ. The ℓ-adic Lie algebra associated to Γ_ℓ is of the form $s+c$, where s is the semisimple part of the Lie algebra of $G_\ell(\mathbb{Q}_\ell)$ ([7, Prop. 2]). It follows that if $\bar{G}_\ell$ is PGL_2, then $\bar{\Gamma}_\ell$ is an open subgroup of $PGL_2(\mathbb{Q}_\ell)$. Let $\Phi : PGL_2 \to SL_3$ denote the adjoint homomorphism.

PROPOSITION 3. *Let E_λ be any quadratic extension of $\mathbb{Q}_\ell$. Every open subgroup Γ of $PGL_2(\mathbb{Q}_\ell)$ contains an open subset X such that for all $x \in X$, the characteristic polynomial of $\Phi(x)$ has simple roots and generates E_λ.*

PROOF. By Krasner's lemma, the splitting field of a polynomial over $\mathbb{Q}_\ell$ with simple roots is a locally constant function of the coefficients. It suffices, therefore, to find one $x \in \Gamma$ with the desired property. Since the eigenvalues of any $\Phi(x)$ are of the form 1, α, and α^{-1}, the roots will automatically be simple if the splitting field is unequal to $\mathbb{Q}_\ell$. If E_λ is a quadratic extension of $\mathbb{Q}_\ell$,

$$PGL_2(\mathbb{Q}_\ell) \cong (Aut_{\mathbb{Q}_\ell} E_\lambda)/\mathbb{Q}_\ell^*,$$

so $PGL_2(\mathbb{Q}_\ell)$ contains a torus T isomorphic to $E_\lambda^*/\mathbb{Q}_\ell^*$. Since the regular elements of T are dense in T, every neighborhood of 1 in $PGL_2(\mathbb{Q}_\ell)$ contains a regular element of T. Any such element x will do. □

PROPOSITION 4. *If for some ℓ_0, $\bar{G}_{\ell_0}$ is the normalizer of a Cartan subgroup H, then case (2) holds in the main theorem, with L a quadratic extension of K.*

PROOF. Let $\Gamma_{\ell_0}^+$ denote the open index-2 subgroup of Γ_{ℓ_0} whose image $\bar{\Gamma}_{\ell_0}$ lies in the Cartan subgroup H. Then $\rho_{\ell_0}^{-1}(\Gamma_{\ell_0}^+)$ is a an open subgroup of G_K of index 2. Let L be the fixed field. Then for all $Frob(\wp) \in G_K \setminus G_L$, $\rho_{\ell_0}(Frob(\wp))$ has trace 0. Since the characteristic polynomial of $\rho_\ell(Frob(\wp))$ doesn't depend on ℓ, it always has trace 0 when $\wp$ is a prime ideal inert in L. By continuity, $\Gamma_\ell^- = \rho_\ell(G_K \setminus G_L)$ is contained in the variety of matrices of trace 0. The Zariski closure of its image in PGL_2 is a coset of the Zariski closure of $\bar{\Gamma}_\ell^+$ in PGL_2. The closure of $\bar{\Gamma}_\ell^+$ is either the full group $\bar{G}_\ell$ or an index-2 subgroup thereof. By Prop. 1, $\bar{G}_\ell$ is a Cartan subgroup, PGL_2, or the normalizer of a Cartan. In the first two cases, there are no algebraic subgroups of index two, so $\bar{\Gamma}_\ell^+$ is Zariski-dense in $\bar{G}_\ell$. It follows that $\bar{\Gamma}_\ell^-$ is Zariski-dense as well. This is inconsistent with Γ_ℓ^- consisting only of matrices of trace 0. We conclude that $\bar{G}_\ell$ is always the normalizer of a Cartan. □

PROPOSITION 5. *If H/Q_ℓ is a torus of GL_2, there exists $\delta \in \mathbb{Q}_\ell$ such that for all $h \in H(\mathbb{Q}_\ell)$,*

$$\operatorname{disc}(\operatorname{char}(h)) \in \delta\mathbb{Q}_\ell^{*2} \cup \{0\},$$

where $\operatorname{char}(h)$ *denotes the characteristic polynomial of h.*

PROOF. Let $H' = HZ$, where Z is the group of diagonal matrices in GL_2. Then H' is a torus. Its rank can be no greater than 2, the reductive rank of GL_2. Therefore, the rational character group, $X^*(H) \otimes_{\mathbb{Z}} \mathbb{Q}$ is a vector space of dimension ≤ 2 with $Gal(\overline{Q}_\ell/\mathbb{Q}_\ell)$ action. We have

$$X^*(H) \otimes_{\mathbb{Z}} \mathbb{Q} = (X^*(Z) \otimes_{\mathbb{Z}} \mathbb{Q}) \oplus M,$$

where the Galois action on $X^*(Z)$ is trivial, and $\dim(M) \leq 1$. As a subgroup of $Aut_{\mathbb{Q}}(M)$, the image of the Galois group is compact and discrete, hence finite, hence of order ≤ 2. Therefore, H splits either over $\mathbb{Q}_\ell$ or over $\mathbb{Q}_\ell(\sqrt{\delta})$ for some $\delta \in \mathbb{Q}_\ell$. The eigenvalues of $h \in H(\mathbb{Q}_\ell)$ are respectively two elements of $\mathbb{Q}_\ell$ or a conjugate pair of elements of $\mathbb{Q}_\ell(\sqrt{\delta})$. In either case, the discriminant of $\operatorname{char}(h)$ is δ times a perfect square. □

COROLLARY. *Under the hypotheses of Proposition 5, $\Phi(h)$ has eigenvalues $\{1, \alpha, \alpha^{-1}\}$, where $\alpha \in \mathbb{Q}_\ell$ if $\delta \in \mathbb{Q}_\ell^{*2}$, and otherwise*

$$\alpha \in \{-1, 1\} \cup (\mathbb{Q}_\ell(\sqrt{\delta}) \setminus \mathbb{Q}_\ell).$$

PROOF. We know that α is the ratio of the two eigenvalues of h. If H is $\mathbb{Q}_\ell$-split, the ratio lies in $\mathbb{Q}_\ell$ and δ can be taken to be 1. If not, the eigenvalues are conjugate elements of $\mathbb{Q}_\ell(\sqrt{\delta})$. The ratio has the property that its conjugate is its reciprocal, so if it lies in $\mathbb{Q}_\ell$, it must be ± 1. □

PROPOSITION 6. *Let $\Gamma \subset PGL_2(\mathbb{Q}_\ell)$ be a compact subgroup with elements γ_1 and γ_2 such that modulo ℓ, $\Phi(\gamma_1)$ has 3 distinct eigenvalues in $\mathbb{F}_\ell$, while $\Phi(\gamma_2)$ has at least one eigenvalue not in $\mathbb{F}_\ell$. Suppose also that $\ell > 3$ and the (mod ℓ) reduction of either $\Phi(\gamma_1)$ or $\Phi(\gamma_2)$ has order > 5 in $SL_3(\mathbb{F}_\ell)$. Then Γ contains $PSL_2(\mathbb{Z}_\ell)$.*

PROOF. Let $\Gamma' \supset \Gamma$ be a maximal compact subgroup of $PGL_2(\mathbb{Q}_\ell)$. We claim that, up to conjugation, there are two possibilities for Γ': $PGL_2(\mathbb{Z}_\ell)$ and

$$\mathcal{M} = \left\{\bar{M} | M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \det(M) \in \mathbb{Z}_\ell^*,\ \ell | c\right\}$$
$$\cup \left\{\bar{M} | M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \det(M) \in \ell\mathbb{Z}_\ell^*,\ \ell | a,\ \ell | c,\ \ell | d\right\}.$$

This is a consequence of the general theory of semisimple groups over local fields [1], but it is not difficult to prove directly: Let $GL_2^+(\mathbb{Q}_\ell)$ denote the subgroup of $GL_2(\mathbb{Q}_\ell)$ with unit determinants, and let

$$\pi : GL_2^+ \to PGL_2^+ = \{\bar{M} | M \in GL_2(\mathbb{Q}_\ell),\ \det(M) \in \ell^{2\mathbb{Z}}\mathbb{Z}_\ell^*\}$$

denote the projection map. As π has compact kernel, the inverse image of

$$\Gamma^+ = \Gamma' \cap PGL_2^+,$$

is a compact subgroup of GL_2 and therefore stabilizes a rank-2 $\mathbb{Z}_\ell$-lattice Λ. If Γ' stabilizes Λ up to homothety, it is contained in $PGL(\Lambda)$, so by maximality it equals $PGL_2(\mathbb{Z}_\ell)$. Otherwise, there exists $\gamma' \in \Gamma' \setminus \Gamma^+$ such that $\gamma'\Lambda \not\sim \Lambda$. By the normality of index-2 subgroups, Γ^+ stabilizes $\gamma\Lambda$ as well as Λ. Now, the homothety classes of $\mathbb{Z}_\ell$-lattices in $\mathbb{Q}_\ell^2$ are the vertices of the *Bruhat-Tits tree*, where edges connect two lattices when one is of index ℓ in the other. The group $PGL_2(\mathbb{Q}_\ell)$ acts on this tree. If $\gamma^+ \in \Gamma^+$ fixes two vertices, it must also fix every vertex on the unique path connecting them. On the other hand, γ' exchanges Λ and $\gamma'\Lambda$ (as $\gamma'^2 \in \Gamma^+$), so it reverses the order of the intermediate vertices. If the number of edges between Λ and $\gamma'\Lambda$ is even, there is a central vertex which is fixed by Γ^+ and γ', hence by Γ. If the number is odd, there is a pair of adjacent vertices which are fixed by Γ^+ and exchanged by γ'. Choosing e_i so that these vertices correspond to the lattices $\mathbb{Z}_\ell e_1 + \mathbb{Z}_\ell e_2$ and $\mathbb{Z}_\ell e_1 + \ell\mathbb{Z}_\ell e_2$, we obtain the group $\mathcal{M}$ defined above.

The image of $\mathcal{M}$ under Φ is contained in the Iwahori subgroup of $SL_3(\mathbb{Z}_\ell)$, that is, the subgroup of elements which are upper-triangular (mod ℓ). Therefore, $\gamma_2 \notin \mathcal{M}$. It follows that $\Gamma \subset PGL_2(\mathbb{Z}_\ell)$. Let $\tilde{\Gamma}$ denote the pre-image of Γ in the projection $GL_2(\mathbb{Z}_\ell) \to PGL_2(\mathbb{Z}_\ell)$ and $\tilde{\gamma}_i$ an element of $\tilde{\Gamma}$ lying over γ_i. By [10,

Lemma 2], every proper subgroup of $GL_2(\mathbb{F}_\ell)$ which fails to contain $SL_2(\mathbb{F}_\ell)$ lies either in a Borel subgroup, in the normalizer of a Cartan subgroup, or in the pre-image in $GL_2(\mathbb{F}_\ell)$ of one of A_4, S_4, $A_5 \subset PGL_2(\mathbb{F}_\ell)$. As every element of A_4, S_4, and A_5 has order ≤ 5, these possibilities are ruled out for the (mod ℓ) reduction of $\tilde{\Gamma}$. If the trace of $\tilde{\gamma}_i$ were divisible by ℓ, the ratio of the eigenvalues would be congruent to -1 (mod ℓ), so two of the eigenvalues of $\Phi(\gamma_i)$ would be congruent to -1, contrary to hypothesis. This means that if $\tilde{\gamma}_i$ lies in the normalizer of a (mod ℓ) Cartan, it lies in the Cartan itself. By hypothesis, $\tilde{\gamma}_1$ can lie only in a split Cartan, while $\tilde{\gamma}_2$ can lie only in a non-split Cartan. Moreover, $\tilde{\gamma}_2$ is not contained in a Borel. Therefore, the reduction (mod ℓ) of $SL_2(\mathbb{Z}_\ell) \cap \tilde{\Gamma}$ is $SL_2(\mathbb{F}_\ell)$. By [**10**, Lemma 1], $\tilde{\Gamma} \supset SL_2(\mathbb{Z}_\ell)$, so $\Gamma \supset PSL_2(\mathbb{Z}_\ell)$. □

2. Proof of Main Theorem

We can now prove the main theorem. By Prop. 4, we may assume that for all ℓ, $\bar{G}_\ell$ is either a Cartan subgroup of PGL_2 or the full group. Suppose that for some ℓ_0, it is all of PGL_2. Then there exist primes $\wp_1$ and $\wp_2$ such that the eigenvalues of $\Phi(\rho_{\ell_0}(Frob(\wp_i)))$ generate distinct quadratic extensions $\mathbb{Q}_{\ell_0}(\sqrt{\delta_i})$. Let α_i denote an eigenvalue which generates $\mathbb{Q}_{\ell_0}(\sqrt{\delta_i})$. In particular, $\alpha_i \neq \pm 1$. For $S, T \subset \mathbb{Q}^{ab}$, the maximal abelian extension of $\mathbb{Q}$, let

$$L(S,T) = \{\ell \text{ prime } | S \subset \mathbb{Q}_\ell \text{ and } T \subset \overline{Q}_\ell \setminus \mathbb{Q}_\ell\}.$$

Note that this does not depend on the embedding $\overline{Q} \subset \overline{Q}_\ell$ because the elements of S and T lie in abelian extensions of $\mathbb{Q}$. If $\ell \in L(\{\alpha_1\}, \{\alpha_2\})$, then by the corollary to Prop. 5, $\bar{G}_\ell = PGL_2$. Indeed, no single Cartan subgroup of $PGL_2/\mathbb{Q}_\ell$ can contain elements whose Φ-images have eigenvalues in $\mathbb{Q}_\ell \setminus \{\pm 1\}$ and $\overline{Q}_\ell \setminus \mathbb{Q}_\ell$. By the Ceboratev density theorem, $L(S,T)$ is infinite whenever S and T are finite. Therefore, if $\bar{G}_\ell = PGL_2$ holds for one ℓ, it holds for infinitely many.

Now assume that $\bar{G}_{\ell_i} = PGL_2$ for primes $\ell_1, \ell_2, \ldots$. Applying Prop. 3, we obtain for each i a prime $\wp_i$ and an eigenvalue β_i of $\Phi(Frob(\wp_i))$ which generates a ramified extension of $\mathbb{Q}_{\ell_i}$. The total extension

$$\Omega = \mathbb{Q}(\beta_1, \beta_2, \ldots)/\mathbb{Q}$$

is infinite because it is ramified at every prime ℓ_i. We extract an infinite subsequence $\gamma_1 = \beta_{f(1)}, \gamma_2 = \beta_{f(2)}, \ldots$ such that $f(1) < f(2) < \cdots$,

$$[\mathbb{Q}(\gamma_1, \ldots, \gamma_{n+1}) : \mathbb{Q}(\gamma_1, \ldots, \gamma_n)] = 2,$$

and $\gamma_i \notin \mathbb{Q}(\zeta_{60})$. We apply Prop. 6 and the corollary of Prop. 5 to the set of $Frob(\gamma_i)$, $i = 1, 2, \ldots, n$. Every rational prime ℓ not dividing

$$N_n = N_{K/\mathbb{Q}}\left(\prod_{i=1}^{n} \wp_{f(i)}\right) \prod_{i=1}^{n} \mathrm{disc}(\gamma_i^{60} - 1),$$

and not contained in

$$\ell \notin L(\{\gamma_1, \dots, \gamma_n\}, \emptyset) \cup L(\emptyset, \{\gamma_1, \dots, \gamma_n\})$$

satisfies $\bar{G}_\ell = PGL_2$. By the Cebotarev density theorem (in this case, actually, the Dirichlet density theorem is enough), the density of $L(\{\gamma_1, \dots, \gamma_n\}, \emptyset)$ is 2^{-n} and the density of $L(\emptyset, \{\gamma_1, \dots, \gamma_n\})$ is 2^{-n}. As $N_n \neq 0$, the set of primes dividing N_n is finite and therefore of density zero. For every prime ℓ not dividing N_n, the eigenvalues of $Frob(\wp_{f(i)})$ are distinct and of order > 5 (mod ℓ). By Prop. 6, the set of primes such that $\bar{G}_\ell \supset PSL_2(\mathbb{Z}_\ell)$ is at least $1 - 2^{1-n}$. Taking the limit as $n \to \infty$, we conclude immediately that $\bar{G}_\ell = PSL_2(\mathbb{Z}_\ell)$ or $\bar{G}_\ell = PGL_2(\mathbb{Z}_\ell)$ for a set of primes ℓ of density 1.

Now, $\det(\rho_\ell)$ forms a compatible system of 1-dimensional Galois representations. Its image is finite for some ℓ if and the whole system factors through a homomorphism $G_K \to \mu_n$, where μ_n denotes the group of n^{th} roots of unity. For this to happen, $\mu_n \subset \mathbb{Q}_\ell$ for all ℓ, which means $n = 1$ or $n = 2$. Letting L be the fixed field of the kernel of $G_K \to \mu_n$, we see that condition (3) of the theorem is satisfied. If $\det(\rho_\ell)$ has infinite image, the image is Zariski dense in $GL_1(\mathbb{Q}_\ell)$, so if $\bar{G}_\ell \supset PSL_2(\mathbb{Z}_\ell)$, $G_\ell = GL_2$, $Z(\Gamma_\ell) = \Gamma_\ell \cap Z(GL_2(\mathbb{Q}_\ell))$, and condition (4) is satisfied. □

References

1. F. Bruhat and F. Tits. Groupes réductifs sur un corps local, *Publ. Math. IHES*, **41** (1971), 5-252.
2. S. Lang, *Introduction to Modular Forms*. Springer-Verlag, Berlin (1976).
3. M. Larsen and R. Pink, On ℓ-independence of algebraic monodromy groups in compatible systems of representations, Preprint.
4. M. Larsen, Maximality of Galois Action on Cohomology, Preprint.
5. J.-P. Serre, Groupes de Lie ℓ-adiques attacheés aux courbes elliptiques, *Colloque CNRS*, **143** (1966), 239-256.
6. J.-P. Serre, Résumé des cours de 1965-1966 in *Annuaire du Collège de France* (1966), 49-58.
7. J.-P. Serre, Sur les groupes de Galois attachées aux groupes p-divisibles, in *Proc. Conf. Local Fields, Driebergen* Springer-Verlag (1966), 118-131.
8. J.-P. Serre, *Abelian ℓ-adic representations and elliptic curves*. W. A. Benjamin, Inc., New York (1968).
9. J.-P. Serre, Points d'ordre fini des courbes elliptiques, *Invent. Math.* **15** (1972) 259-331.
10. H. P. F. Swinnerton-Dyer, On ℓ-adic representations and congruences of coefficients of modular forms, in *Modular Functions of One Variable III*, Lecture Notes 350, Springer-Verlag, New York (1973), 1-55.
11. H. P. F. Swinnerton-Dyer, On ℓ-adic representations and congruences of coefficients of modular forms (II), in *Modular Functions of One Variable V*, Lecture Notes 601, Springer-Verlag, New York (1976), 63-90.

Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104
E-mail address: larsen@math.upenn.edu

This page intentionally left blank

Contemporary Mathematics
Volume 133, 1992

Algebraic Identities Useful In The Computation Of Igusa Local Zeta Functions

MARGARET M. ROBINSON

ABSTRACT. This paper describes three related algebraic identities which are useful in the computation of Igusa local zeta functions.

Introduction

In the simplest case, Igusa's local zeta function for any polynomial $f(x)$ in $K[x_1, \ldots, x_n] - \{0\}$ and for $Re(s) > 0$ is defined as

$$Z_K(s) = \int_{O_K^n} |f(x)|_K^s dx.$$

In this definition, K is taken to be a p-adic completion of a number field k, O_K is the ring of integers in K, $\pi_K O_K$ is the ideal of non-units in O_K, and q_K is the cardinality of the residue field $O_K/\pi_K O_K$. Furthermore, dx is the Haar measure on K^n normalized such that $|\pi_K|_K = q_K^{-1}$ and measure$(O_K^n) = 1$. By a more general theorem, $Z_K(s)$ was shown to be a rational function of q_K^{-s} [2].

$Z_K(s)$ has been determined explicitly for a large number of polynomials $f(x)$, primarily by Igusa [3,4,5]. In the course of several of these calculations, algebraic identities are used to convert a sum of partial integrals into a product that represents the full integral. The most famous of these identities – which Gauss employed to convert an elliptic theta series into an infinite product and also (in a special case) to compute the sign of the Gauss sum – was used in [3,7] to compute particular integrals. In [6], a second identity was introduced for a similar purpose. A third identity was used to find the closed form expression for the Igusa local zeta function when $f(x)$ is the determinant or generic norm in the vector space of Hermitian matrices with coefficients in the quaternion division algebra over K [7].

1991 Mathematics Subject Classification. Primary 11R52, 14G10.
The final detailed version of this paper will be submitted for publication elsewhere.

0271-4132/92 $1.00 + $.25 per page

To state the Gauss identity, we let

$$F_{m,n}(x) = \prod_{i=1}^{n} \frac{1 - x^{m+i}}{1 - x^i},$$

for any non-negative integers m, n, where we let $F_{m,0}(x) = 1$. Then

(i) $F_{m,n}(x) = F_{n,m}(x)$

(ii) $F_{i,j+k}(x) \cdot F_{j,k}(x) = F_{i,j}(x) \cdot F_{i+j,k}(x)$

(iii) $F_{m,n}(x) = F_{m,n-1}(x) + x^n F_{m-1,n}(x)$, if $n, m \geq 1$.

From (iii), the Gauss identity follows [1]:

$$\sum_{i+j=k} F_{i,j}(x) x^{i(i-1)/2} t^i = \prod_{i=1}^{k} (1 + x^{i-1} t).$$

To define a generalization of Gauss' identity, let

$$F_{m,n}(x, y) = \prod_{i=1}^{n} \frac{1 - x^{m+i} y}{1 - x^i}$$

$$F_{m,n}(x) = F_{m,n}(x, 1)$$

for m, n integers and $n \geq -1$. With the understanding that $F_{m,0}(x, y) = 1$ and $F_{m,-1}(x, y) = 0$,

$$F_{m,n}(x, y) = F_{m-1,n}(x, xy) = F_{m,n-1}(x, y) + x^n F_{m-1,n}(x, y).$$

In [6], Igusa shows that

$$\sum_{0 \leq k \leq n} F_{m-k,k}(x) \cdot F_{k,n-k}(x, y) x^{k^2} y^k = F_{m,n}(x, y).$$

An Identity

Lemma. For an integer $k > 0$, let $G_k(x, t)$ equal the right-hand side of the Gauss identity, i.e.

$$G_k(x, t) = \prod_{i=1}^{k} (1 + x^{i-1} t).$$

For $k = 0$, let $G_k(x, t) = 1$. Then the following identity holds:

$$(1)\qquad x^{k(2k-1)/2}G_k(x,t)$$

$$= \sum_{j=0}^{k}(-1)^j x^{j(j-1)/2}F_{k-j,j}(x)G_j(x,-x^{k-j+1/2})G_{k-j}(x,x^{k-1/2}t).$$

Proof. Compare coefficients of t^i. Using Gauss' identity, the left-hand side of (1) is

$$x^{k(2k-1)/2}\sum_{i=0}^{k}F_{i,k-i}(x)x^{i(i-1)/2}t^i.$$

We see that the coefficient of t^i in the left-hand side is

$$x^{k(2k-1)/2}F_{i,k-i}(x)x^{i(i-1)/2}.$$

The Gauss identity allows the right-hand side of (1) to be rewritten as

$$\sum_{j=0}^{k}B_j\sum_{i=0}^{k-j}A_{i,j}\cdot t^i$$

where

$$B_j = (-1)^j x^{j(j-1)/2}F_{k-j,j}(x)G_j(x,-x^{k-j+1/2})$$

$$A_{i,j} = F_{i,k-j-i}(x)x^{i^2/2-i+ki}$$

Changing the order of summation, the right-hand side becomes

$$(2)\qquad \sum_{i=0}^{k}\left(\sum_{j=0}^{k-i}B_jA_{i,j}\right)t^i$$

in which the coefficient of t^i is $\sum_{j=0}^{k-i}B_jA_{i,j}$. Using the second property of the Gauss identity,

$$(3)\qquad F_{j,k-j}(x)\cdot F_{k-j-i,i}(x) = F_{j,k-j-i}(x)\cdot F_{k-i,i}(x).$$

Using (1) and substituting (3) in (2), the coefficient of t^i becomes precisely

$$F_{i,k-i}(x)\sum_{j=0}^{k-i}(-1)^j x^{j(j-1)/2+\frac{1}{2}i^2-i+ki}F_{k-j-i,j}(x)\cdot G_j(x,-x^{k-j+1/2}).$$

Equating coefficients of t^i from both sides of (1), we have left to show

$$x^{k(2k-1)/2}x^{i/2-ki} = \sum_{j=0}^{k-i}(-1)^j x^{j(j-1)/2}F_{k-j-i,j}(x)\cdot G_j(x,-x^{k-j+1/2}).$$

Using Gauss' identity and (ii) once more,

$$(4) \qquad x^{k(2k-1)/2}x^{i/2-ki}$$

$$= \sum_{j=0}^{k-i} \sum_{p+q=j} F_{k-p-q-i,p+q}(x)F_{p,q}(x)x^{kp-p/2+q(q-1)/2}(-1)^q$$

$$= \sum_{p=0}^{k-i} F_{p,k-i-p}(x)x^{kp-p/2} \cdot \left[\sum_{q=0}^{k-i-p} F_{q,k-i-p-q}(x)x^{q(q-1)/2}(-1)^q \right]$$

where the bracketed expression is precisely

$$G_{k-i-p}(x,-1) = \begin{cases} 1 & \textit{if } k\text{-}i\text{-}p=0 \\ 0 & \textit{otherwise} \end{cases}$$

The terms in the outer sum of (4) are, therefore, all 0 except when $p = k - i$ and the sum above reduces to $x^{k(k-i)-(k-i)/2}$ which is exactly the right-hand side. ■

REFERENCES

1. Gauss, C.F., Hundert Theoreme über die neuen Transscendenten, Werke III (1876), 461-469.

2. Igusa, J., Complex powers and asymptotic expansions II, J. Reine angew. Math. **278/279** (1975) 307-321.

3. Igusa, J., Some results on p-adic complex powers, Amer. J. Math. **106** (1984) 1013-1032.

4. Igusa, J., On functional equations of complex powers, Invent. Math. **85** (1986) 1-29.

5. Igusa, J., B-functions and p-adic integrals, Algebraic Analysis **1**, Academic Press, (1988) 231-241.

6. Igusa, J., Universal p-adic zeta functions and their functional equations, Amer. J. Math. **111** (1989) 671-716.

7. Robinson, M.M., On the complex powers associated with the twisted cases of the determinant and the Pfaffian, Doctoral Dissertation, The Johns Hopkins University, Baltimore, Maryland, 1986.

Department of Mathematics, Statistics and Computer Science, Mount Holyoke College, South Hadley, Massachusetts 01075.

E-mail address: robinson@mhc.mtholyoke.edu

Contemporary Mathematics
Volume 133, 1992

Points of Finite Order on Abelian Varieties

A. SILVERBERG

ABSTRACT. This paper includes a survey of the problem of finding explicit bounds on orders of torsion points on abelian varieties. In addition, we prove an explicit lower bound on ramification in extensions of number fields obtained by adjoining torsion on abelian varieties of CM-type. We include tables giving upper bounds on orders of torsion points on CM elliptic curves, and prime orders of torsion points on CM abelian varieties, which are defined over number fields of given degree.

1. Introduction

In this paper we discuss some effectivity questions concerning torsion points on abelian varieties. The motivating question is the Torsion Conjecture, which we state in §2, along with a brief summary of some of the known results. In §3 we give explicit upper bounds, in terms of the dimension of the abelian variety and the degree of the number field of definition, on orders of torsion subgroups of abelian varieties having potential good reduction. We define what we mean by an abelian variety of CM-type in §4. In §5 we give explicit lower bounds from [25] on degrees of field extensions obtained by adjoining torsion points on abelian varieties of CM-type, and deduce improvements on the upper bounds on orders of torsion points of §3, for the subclass of abelian varieties of CM-type, from which we obtain the tables in §10. In §6 we discuss the particular case of torsion on CM elliptic curves defined over number fields, improving on earlier results. In §7 and §8 we obtain explicit lower bounds on ramification indices of primes in field extensions obtained by adjoining torsion on abelian varieties of CM-type. In §9 we state some results which, when combined with the bounds of §5, severely restrict the possibilities for torsion subgroups of Mordell-Weil groups of abelian varieties of CM-type having a prime of purely additive reduction over

1991 *Mathematics Subject Classification.* Primary 11G10; Secondary 11G15, 14K22, 14K15, 14Q20.

The author thanks the Alfred P. Sloan Foundation, NSF, NSA, and the Ohio State University for generous financial support.

This paper is in final form and no version of it will be submitted for publication elsewhere.

a number field. We then speculate on the relative scarcity of abelian varieties of CM-type which do not have any primes of purely additive reduction.

Acknowledgments. I would like to thank Dino Lorenzini and Bill McCallum for helpful conversations, and Harvard University for its hospitality. I would also like to thank Marc Hindry for pointing out that combining the results of Kamienny and of Kenku and Momose yields Theorem 2.2.

Notation. We write G_{tors} for the subgroup of elements of finite order in an abelian group G, and $|H|$ for the order of a finite group H.

2. Torsion Conjecture

We begin by stating two forms of the Torsion Conjecture for abelian varieties over number fields.

TORSION CONJECTURE. *If A is an abelian variety of dimension d defined over a number field M, then $|A(M)_{tors}|$ is bounded above by a constant depending only on d and M.*

STRONG TORSION CONJECTURE. *If A is an abelian variety of dimension d defined over a number field M of degree m, then $|A(M)_{tors}|$ is bounded above by a constant depending only on d and m.*

In [1], Cassels stated the Torsion Conjecture for elliptic curves as "part of the folklore". In [15], Manin proved that for an elliptic curve E defined over a number field M, the p-primary part of $E(M)_{\text{tors}}$ is bounded above by a constant depending only on the field M and the prime p. Manin's result, which is ineffective, now follows also from the Mordell Conjecture, which has been proved by Faltings. The Strong Torsion Conjecture has been settled for $d = 1$ and $m = 1$ by Mazur, and for $d = 1$ and $2 \leq m \leq 8$ by Kamienny.

THEOREM 2.1 (MAZUR, [16]). *If E is an elliptic curve over $\mathbf{Q}$, then $E(\mathbf{Q})_{tors}$ is isomorphic to one of the following 15 groups:*

$$\begin{array}{ll} \mathbf{Z}/N\mathbf{Z} & \text{for } N = 1, \ldots, 10 \text{ or } 12, \\ \mathbf{Z}/2\mathbf{Z} \times \mathbf{Z}/2N\mathbf{Z} & \text{for } N = 1, 2, 3, \text{ or } 4. \end{array}$$

Further, it is known that each of these groups occurs infinitely often.

THEOREM 2.2 (KAMIENNY, [10]). *If E is an elliptic curve defined over a quadratic number field M, then $E(M)_{tors}$ is isomorphic to one of the following 26 groups:*

$$\begin{array}{ll} \mathbf{Z}/N\mathbf{Z} & \text{for } N = 1, \ldots, 16, \text{ or } 18, \\ \mathbf{Z}/2\mathbf{Z} \times \mathbf{Z}/2N\mathbf{Z} & \text{for } N = 1, \ldots, 6, \\ \mathbf{Z}/3\mathbf{Z} \times \mathbf{Z}/3N\mathbf{Z} & \text{for } N = 1 \text{ or } 2, \\ \mathbf{Z}/4\mathbf{Z} \times \mathbf{Z}/4\mathbf{Z}. & \end{array}$$

This result was conjectured by Kenku and Momose in [11]. Kamienny proved that, for M quadratic, the order of a torsion point in $E(M)$ must divide $2^4 \cdot 3^2 \cdot 5 \cdot 7 \cdot 11 \cdot 13$, and in [11] it had already been determined exactly which groups can then occur.

For $d > 1$, there are no number fields for which the Torsion Conjecture is known to hold.

For other results on torsion on elliptic curves over number fields, see, for example, [4], [8], and [9]. In particular, the proof of Theorem 7.1 of [8] shows that if E is an elliptic curve defined over a number field M of degree m, $\sigma = \sigma(E, M)$ is Szpiro's constant, and N is the order of a point in $E(M)_{\text{tors}}$, then

$$N \leq (20\sigma)^{4m}(100)^{\sigma}.$$

Therefore, Szpiro's Conjecture (Conjecture 0.4 of [8]) implies the Torsion Conjecture for elliptic curves.

When finding bounds on orders of torsion points on abelian varieties, one would like to know what rate of growth to expect, in terms of the dimension of the abelian variety or the degree of the number field of definition. Flynn has constructed sequences of hyperelliptic curves over $\mathbf{Q}$ with Jacobians that have $\mathbf{Q}$-rational torsion points whose orders grow quadratically in the genus of the curve.

THEOREM 2.3 (FLYNN, [5]). *The one-parameter space of curves of genus g (g even, $t \neq 0$):*

$$y^2 + t(x-1)y = \left(\sum_{i=1}^{g-r+1} x^{r+i}\right)^2 - t(x^{g+2} + x^{r+1})$$

where $0 \leq r \leq g$, has a divisor of exact order $g^2 + 3g + 1 - r$. Therefore in even genus g, there exist $\mathbf{Q}$-rational torsion divisors of all orders in the interval $[g^2 + 2g + 1, g^2 + 3g + 1]$. Moreover, there is a sequence $A(n)$ of abelian varieties over $\mathbf{Q}$ of strictly increasing dimension $d(n)$, each containing a $\mathbf{Q}$-rational torsion point of order $N(n)$, such that

$$N(n) > (1.6)^{(d(n)\log d(n))^{2/3}}.$$

One could ask whether, for an abelian variety A of dimension d defined over an arbitrary field M, $|A(M)_{\text{tors}}|$ is bounded above by a constant depending only on d and M. As remarked by Cassels, this is false for local fields (Lemma 17.1 and p. 264 of [1]). However, for abelian varieties (with no constant part) over function fields, one expects a Torsion Conjecture to hold. For a non-constant elliptic curve E defined over M, a function field in one variable, Theorem 1 of [13] gives an explicit upper bound on the order of the torsion subgroup of $E(M)$, depending only on the genus of M. See also Theorem 7.2 of [8], which shows

that if M has genus g and characteristic zero, then

$$|E(M)_{\text{tors}}| \leq 144(g+1)^{2/3}.$$

3. Potential good reduction

In this section we give upper bounds on orders of torsion subgroups of abelian varieties having potential good reduction everywhere, of fixed dimension and defined over number fields of fixed degree. These results are certainly well-known, though I failed to find a reference in the literature.

If A is an abelian variety defined over a number field M, write A_N for the group of points of order dividing N in $A(\bar{\mathbf{Q}})$, and write $M(A_N)$ for the smallest extension of M in $\bar{\mathbf{Q}}$ over which the elements of A_N are defined.

PROPOSITION 3.1. *Suppose A is an abelian variety defined over a number field M, and A has potential good reduction everywhere. If $N \geq 3$, then over $M(A_N)$, A has good reduction outside the primes dividing N.*

PROOF. This follows directly from Corollary 2, part (b), p. 497 of [**21**]. □

COROLLARY 3.2. *If A is an abelian variety defined over a number field M, and A has potential good reduction everywhere, then A has good reduction everywhere over $M(A_{12})$.*

When c and d are positive integers, let

$$a_c(d) = |GL_{2d}(\mathbf{Z}/c\mathbf{Z})|/2.$$

THEOREM 3.3. *If A is a d-dimensional abelian variety defined over a number field M of degree m, and A has potential good reduction everywhere, then*

$$|A(M)_{tors}| \leq [(1+2^{a_3(d)m})(1+3^{a_4(d)m})]^{2d}$$

$$< (1+10^{-11}) \cdot (2^{3^{4d^2}} \cdot 3^{4^{4d^2}})^{md} < (1+10^{-11}) \cdot 6^{md \cdot 4^{4d^2}}.$$

PROOF. Over $M(A_3)$, A has good reduction at the primes above 2, by the Proposition. Let $\mathcal{O}$ be the ring of integers in $M(A_3)$ and let $\mathfrak{p}$ be a prime above 2. The prime-to-two torsion in $A(M(A_3))$ injects into $\tilde{A}(\mathcal{O}/\mathfrak{p})$ under reduction modulo $\mathfrak{p}$, where $\tilde{A}$ is the reduction of A modulo $\mathfrak{p}$. Now

$$|\mathcal{O}/\mathfrak{p}| \text{ divides } 2^{[M(A_3):\mathbf{Q}]} \text{ which divides } 2^{|GL_{2d}(\mathbf{F}_3)|m}.$$

Therefore

$$|A(\mathcal{O}/\mathfrak{p})| \leq |A(\mathbf{F}_q)| \text{ where } q = 2^{|GL_{2d}(\mathbf{F}_3)|m}.$$

The well-known Weil bound

$$|A(\mathbf{F}_q)| \leq (1+\sqrt{q})^{2d}$$

then implies that the prime-to-two torsion in $A(M(A_3))$ has order at most

$$(1+2^{a_3(d)m})^{2d}.$$

Over $M(A_4)$, A has good reduction at the primes above 3, and we conclude similarly that the prime-to-three torsion in $A(M(A_4))$ has order at most

$$(1+3^{a_4(d)m})^{2d}.$$

Since $a_c(d) < (c^{4d^2})/2$, we obtain the result. □

Remark 1. In the above proof, $|\mathcal{O}/\mathfrak{p}|$ divides 2^{mf} where f is the order of Frobenius at $\mathfrak{p}$ in $\mathrm{Gal}(M(A_3)/M)$ (respectively, $\mathrm{Gal}(M(A_4)/M)$). The estimate can be improved by replacing $|GL_{2d}(\mathbf{F}_3)|$ (respectively, $|GL_{2d}(\mathbf{Z}/4\mathbf{Z})|$) by an upper bound on the orders of all elements in $GL_{2d}(\mathbf{F}_3)$ (respectively, $GL_{2d}(\mathbf{Z}/4\mathbf{Z})$).

4. Abelian varieties of CM-type

In this section we recall some definitions concerning abelian varieties and complex multiplication. Below, all abelian varieties are over the complex numbers.

DEFINITION 4.1. *An abelian variety A of dimension d has* sufficiently many complex multiplications, *or* smCM, *if there is an injective homomorphism*

$$\theta : K \hookrightarrow \mathrm{End}(A) \otimes \mathbf{Q}$$

of $\mathbf{Q}$*-algebras, with K a CM-field of degree* $2d$.

DEFINITION 4.2. *An abelian variety A is of* CM-type *if it is isogenous to a product*

$$A_1^{n_1} \times \ldots \times A_r^{n_r}$$

where the A_i are abelian varieties with smCM.

We will refer to a simple abelian variety of CM-type as a "simple CM abelian variety". When a field of definition is involved, we will sometimes say absolutely simple to emphasize that we mean simple over $\mathbf{C}$.

Suppose $K_1, \ldots, K_r$ are CM-fields, and d, $n_1, \ldots, n_r$ are positive integers such that

$$2d = \sum_{i=1}^{r} n_i [K_i : \mathbf{Q}].$$

Let

$$Z = M_{n_1}(K_1) \times \ldots \times M_{n_r}(K_r),$$
$$W = K_1^{n_1} \times \ldots \times K_r^{n_r},$$

and

$$W_{\mathbf{R}} = W \otimes_{\mathbf{Q}} \mathbf{R}.$$

Suppose Ψ is a faithful complex representation of Z of dimension d, A is an abelian variety of dimension d, θ is an embedding of Z into $\mathrm{End}(A) \otimes \mathbf{Q}$, C is

a polarization on A, and ρ is the (Rosati) involution of $\mathrm{End}(A) \otimes \mathbf{Q}$ determined by C. We will view W and $K_1 \times \ldots \times K_r$ as embedded in Z diagonally.

DEFINITION 4.3. *(A,θ) is of* type *(Z,Ψ) if there exist a $\mathbf{Z}$-lattice $\mathcal{F}$ in W, a lattice D in $\mathbf{C}^d$, and a commutative diagram*

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \mathcal{F} & \longrightarrow & W_{\mathbf{R}} & \longrightarrow & W_{\mathbf{R}}/\mathcal{F} & \longrightarrow & 0 \\ & & \downarrow & & \downarrow u & & \downarrow & & \\ 0 & \longrightarrow & D & \longrightarrow & \mathbf{C}^d & \xrightarrow{\xi} & A & \longrightarrow & 0 \end{array}$$

where the vertical arrows are isomorphisms and the rows are exact, such that

(i) $\theta(a) \circ \xi = \xi \circ \Psi(a)$ *for every* $a \in Z$, *and*

(ii) $u(ax) = \Psi(a)(u(x))$ *for every* $a \in Z$ *and* $x \in W_{\mathbf{R}}$.

Remark 2. If (A,θ) is of type (Z,Ψ), then A is isogenous to a product of abelian varieties $A_1^{n_1} \times \ldots \times A_r^{n_r}$ where θ induces maps

$$\theta_i : K_i \hookrightarrow \mathrm{End}(A_i) \otimes \mathbf{Q}$$

for $i = 1, \ldots r$ such that (A,θ_i) is of type (K_i,Φ_i), where $\Psi|_{K_i}$ is equivalent to the sum of $n_i\Phi_i$ and a zero representation. An abelian variety A is of CM-type if and only if there exist θ, Z, and Ψ such that (A,θ) is of type (Z,Ψ).

DEFINITION 4.4. *C is an* admissible polarization *for (A,θ) if*

$$\theta(a)^\rho = \theta(\bar{a})$$

whenever $a \in K_1 \times \ldots \times K_r$.

Remark 3. If (A,θ) is defined over a field k, then (A,θ) has an admissible polarization defined over k.

5. Torsion on abelian varieties of CM-type

In this section we recall some results proved in [24] and [25], stating them in a more precise form, from which we obtain the tables in §10. Retaining the notation of the previous section, let $(\tilde{K}_i,\tilde{\Phi}_i)$ be the reflex type of (K_i,Φ_i) (see p. 126 of [22] for the definition), let $\tilde{K}$ be the compositum of the fields $\tilde{K}_1, \ldots, \tilde{K}_r$, let

$$b = [\tilde{K} : \mathbf{Q}]/2,$$

let

$$\mathcal{O} = \theta^{-1}(\mathrm{End}(A) \cap \theta(K_1 \times \ldots \times K_r)),$$

and let

$$\mu = \#\mu(\mathcal{O}),$$

where $\mu(\mathcal{O})$ denotes the set of roots of unity in $\mathcal{O}$. For c and N positive integers, let

$$r_c(N) = \#\{m \in (\mathbf{Z}/N\mathbf{Z})^\times : m^c \equiv 1 \pmod{N}\}.$$

Write ϕ for the Euler ϕ function. In Theorem 1 of [25] we showed:

THEOREM 5.1. *Suppose (A,θ) is of type (Z,Ψ), C is an admissible polarization for (A,θ), and t is a point on A of order N. Let k_0 and k_t be the fields of moduli for (A,C,θ) and (A,C,θ,t), respectively. Then*

$$[k_t : k_0] \geq \frac{\phi(N)}{r_b(N)\mu}.$$

Write L for the field generated by the Galois closures of the CM-fields $K_1,\ldots,K_r$. As a consequence of Theorem 5.1 we have (see §5 of [25]):

COROLLARY 5.2. *Suppose A is a d-dimensional abelian variety of CM-type, A is defined over a number field M, and t is a point on A of order N.*

(i) *Then*

$$[M(t) : \mathbf{Q}] \geq \frac{\phi(N)}{r_b(N)\mu[L : \mathbf{Q}]}.$$

(ii) *If A is simple, then*

$$[M(t) : \mathbf{Q}] \geq \frac{\phi(N)}{r_b(N)\mu[\tilde{K} : \mathbf{Q}]}.$$

Let $\nu(N)$ be the number of distinct primes dividing N, and let

$$\delta(N) = \begin{cases} 0 & \text{if 8 does not divide } N, \\ 1 & \text{if 8 divides } N. \end{cases}$$

We have (see Lemma 3, (4.7), (4.8), and Corollary 4 of [25]):

$$\mu \leq 6^d,$$

$$r_b(N) \leq 2^{(d-1)\nu(N)+\delta(N)},$$

$$[\tilde{K} : \mathbf{Q}] \leq 2^d,$$

and

$$[L : \mathbf{Q}] \leq 2^d \cdot d!.$$

Remark 4. If A has smCM, then $\phi(\mu) \leq 2d$.

COROLLARY 5.3. *Suppose A is a d-dimensional abelian variety of CM-type, A is defined over a number field M, and t is a point on A of order N.*

(i) *Then*

$$[M(t) : \mathbf{Q}] \geq \frac{\phi(N)}{2^{(d-1)\nu(N)+2d+\delta(N)} \cdot 3^d \cdot (d!)}.$$

(ii) *If A is simple, then*

$$[M(t) : \mathbf{Q}] \geq \frac{\phi(N)}{2^{(d-1)\nu(N)+d+\delta(N)} \cdot \mu}$$

where $\phi(\mu) \leq 2d$.

For related results, see for example [19], [3], and [20]. We can now give an improvement on the bounds of Theorem 3.3, for the subclass of abelian varieties of CM-type (see §5 of [25]):

COROLLARY 5.4. *Suppose A is a d-dimensional abelian variety of CM-type, A is defined over a number field M of degree m, and $t \in A(M)$ is a point of order N.*

(i) *Then*

$$\phi(N) \leq r_b(N)\mu[L : \mathbf{Q}]m \leq 2^{(d-1)\nu(N)+d+\delta(N)}(d!)\mu m$$
$$\leq 2^{(d-1)\nu(N)+2d+\delta(N)} \cdot 3^d(d!)m.$$

(ii) *If N is prime, then*

$$N \leq 1 + 2^{2d-1}(d!)\mu m \leq 1 + 2^{3d-1} \cdot 3^d(d!)m.$$

(iii) *If A is simple, then*

$$\phi(N) \leq 2^{(d-1)\nu(N)+d+\delta(N)}\mu m$$

where $\phi(\mu) \leq 2d$.

(iv) *If A is simple and N is prime, then*

$$N \leq 1 + 2^{2d-1}\mu m$$

where $\phi(\mu) \leq 2d$.

Remark 5. Since $\phi(n) \gg n/\log\log n$ and $\nu(n) \ll \log\log n$ with explicit constants, the above Corollary gives explicit upper bounds on the order N in terms of d and m.

6. Elliptic Curves

In this section we prove a strengthening of Corollary 7 of [25].

COROLLARY 6.1. *Suppose E is an elliptic curve defined over a number field M of degree m, E has complex multiplication by an order $\mathcal{O}$ in an imaginary quadratic field K, and $t \in E(M)$ has order N. Let $\mu = \#\mu(\mathcal{O})$. Then:*

(i) $\phi(N) \leq \mu m \leq 6m$.

(ii) *If $K \subseteq M$ then $\phi(N) \leq \mu m/2 \leq 3m$.*

(iii) $N \ll m\log\log m$.

PROOF. Let j be the j-invariant of E. Then

$$K(j) = k_0 \subseteq k_t \subseteq K(j,t)$$

(see §5.5B of [22]). We have $\tilde{K} = K$, and so $b = 1$ and $r_b(N) = 1$. Applying Theorem 5.1 gives

$$[K(j,t) : K(j)] \geq \phi(N)/\mu.$$

Since $[K(j) : \mathbf{Q}(j)] = 2$, we have

$$[\mathbf{Q}(j,t) : \mathbf{Q}(j)] \geq \phi(N)/\mu.$$

Since E and t are defined over M, $M \supseteq \mathbf{Q}(j,t)$. Therefore

$$\phi(N) \leq \mu[M : \mathbf{Q}(j)] \leq \mu m,$$

giving (i). If $K \subseteq M$, then $K(j,t) \subseteq M$, so

$$[M : K(j)] \geq \phi(N)/\mu.$$

Therefore

$$m = 2[M : K] \geq 2\phi(N)/\mu,$$

giving (ii). From (i) and the estimate $\phi(n) \gg n/\log\log n$ we deduce (iii). □

Remark 6. For CM elliptic curves defined over $\mathbf{Q}$, Olson gave complete information on torsion in $E(\mathbf{Q})$ in [18]. In particular, the only possible values for $|E(\mathbf{Q})_{\text{tors}}|$ are 1, 2, 3, 4, or 6 (i.e., $\phi(|E(\mathbf{Q})_{\text{tors}}|) \leq 2$).

Remark 7. By Theorem 4 of [17], if E is an elliptic curve with integral j-invariant over a quadratic number field M, then $E(M)_{\text{tors}}$ is isomorphic to one of the following 13 groups:

$$\begin{array}{ll} \mathbf{Z}/N\mathbf{Z} & \text{for } N = 1, \ldots, 8 \text{ or } 10, \\ \mathbf{Z}/2\mathbf{Z} \times \mathbf{Z}/2N\mathbf{Z} & \text{for } N = 1, 2, \text{ or } 3, \\ \mathbf{Z}/3\mathbf{Z} \times \mathbf{Z}/3\mathbf{Z}. & \end{array}$$

Theorem 1 of [7] gave complete information on rational torsion for elliptic curves over $\mathbf{Q}$ with integral j-invariant.

7. Ramification

We will now use Shimura's Main Theorem of Complex Multiplication (Corollary 5.16 of [22]) to obtain explicit lower bounds on ramification in extensions of number fields obtained by adjoining torsion points on abelian varieties, paralleling the proofs of the theorems in [25].

Retaining the notation of §4, §5, and Theorem 5.1, we will assume for simplicity that the abelian variety A has sufficiently many complex multiplications, i.e., is of type (K,Φ) where K is a CM-field of degree $2d$. The pair $(\tilde{K},\tilde{\Phi})$ now denotes the reflex of (K,Φ), and

$$\mathcal{O} = \theta^{-1}(\text{End}(A) \cap \theta(K)).$$

The torsion on A can be identified with $(\mathcal{F} \otimes \mathbf{Q})/\mathcal{F}$ (see Definition 4.3).

Let T be a finite subset of $\mathcal{F} \otimes \mathbf{Q}$. Writing $t_1, \ldots, t_k$ for the torsion points of A corresponding to the elements of T, let k_T be the field of moduli of $(A, C, \theta, t_1, \ldots, t_k)$, and as before let k_0 be the field of moduli of (A, C, θ). Then k_T and k_0 are finite abelian extensions of $\tilde{K}$. Let $\mathfrak{p}_T$ be a prime ideal of k_T, let $\mathfrak{p}_0$ (respectively $\mathfrak{p}$, respectively p) be the prime ideal of k_0 (respectively $\tilde{K}$,

respectively $\mathbf{Q}$) below $\mathfrak{p}_T$, and let e ($= e(T, \mathfrak{p}_T)$) be the ramification index of $\mathfrak{p}_T$ over $\mathfrak{p}_0$.

Let

$$\mathfrak{a} = \{x \in \mathcal{O} : xT \subseteq \mathcal{F}\},$$

let

$$\eta : \tilde{K}_{\mathbf{A}}^{\times} \to K_{\mathbf{A}}^{\times}$$

be the continuous extension of the homomorphism

$$\det\tilde{\Phi} : \tilde{K}^{\times} \to K^{\times}$$

to the idèles of $\tilde{K}$ and K, and let

$$P = \{\text{prime ideals } \pi \text{ of } K : \pi \text{ divides } \mathfrak{a} \text{ and } \pi \text{ divides } \eta(\mathfrak{p})\}.$$

For $\pi \in P$ define $s(\pi) \in \mathbf{Z}^+$ by

$$p^{s(\pi)}\mathbf{Z} = \mathbf{Z} \cap \pi^{\mathrm{ord}_{\pi}(\mathfrak{a})},$$

and let

$$s = \max_{\pi \in P} s(\pi).$$

THEOREM 7.1.

$$\textit{If } p \textit{ is odd, then } e \geq \frac{\phi(p^s)}{2^{d-1}\mu}.$$

$$\textit{If } p = 2, \textit{ then } e \geq \frac{2^{s-1-d}}{\mu}.$$

8. Proof of Theorem 7.1

For $c \in \tilde{K}_{\mathbf{A}}^{\times}$ write $N(c)$ for the absolute norm of the ideal of $\tilde{K}$ associated to c. Let

$$S_T = \{c \in \tilde{K}_{\mathbf{A}}^{\times} : \exists q \in K^{\times} \text{ s.t. } q\bar{q}N(c) = 1, q\eta(c)\mathcal{F} = \mathcal{F}, (q\eta(c) - 1)T \subseteq \mathcal{F}\},$$

and let

$$S_0 = \{c \in \tilde{K}_{\mathbf{A}}^{\times} : \exists q \in K^{\times} \text{ s.t. } q\bar{q}N(c) = 1, q\eta(c)\mathcal{F} = \mathcal{F}\}.$$

By Corollary 5.16 of [22], k_T (respectively k_0) is the finite abelian extension of $\tilde{K}$ corresponding under class field theory to the subgroup S_T (respectively S_0) of the idèles of $\tilde{K}$. Let $\tilde{\mathcal{O}}$ be any order in $\tilde{K}$ such that $\eta(\tilde{\mathcal{O}} - 0) \subseteq \eta(\mathcal{O} - 0)$ and let $\tilde{\mathcal{O}}_{\mathfrak{p}}$ be the closure of $\tilde{\mathcal{O}}$ in $\tilde{K}_{\mathfrak{p}}$.

LEMMA 8.1. $e \geq [\tilde{\mathcal{O}}_{\mathfrak{p}}^{\times} : S_T \cap \tilde{\mathcal{O}}_{\mathfrak{p}}^{\times}]$.

PROOF. Let $\mathcal{O}_{\tilde{K}_\mathfrak{p}}$ be the maximal order in $\tilde{K}_\mathfrak{p}$. Then

$$e = [S_0 \cap \mathcal{O}^\times_{\tilde{K}_\mathfrak{p}} : S_T \cap \mathcal{O}^\times_{\tilde{K}_\mathfrak{p}}] \geq [S_0 \cap \tilde{\mathcal{O}}^\times_\mathfrak{p} : S_T \cap \tilde{\mathcal{O}}^\times_\mathfrak{p}].$$

Now $S_0 \cap \tilde{\mathcal{O}}^\times_\mathfrak{p} = \tilde{\mathcal{O}}^\times_\mathfrak{p}$, since

$$N(c) = 1 \text{ and } \eta(c)\mathcal{F} = \mathcal{F}$$

for every $c \in \tilde{\mathcal{O}}^\times_\mathfrak{p}$. □

Let

$$F = \tilde{K}^\times_\infty \cdot \prod_\mathfrak{p} \tilde{\mathcal{O}}^\times_\mathfrak{p},$$

$$E_\mathfrak{p} = \{c \in \tilde{\mathcal{O}}^\times_\mathfrak{p} : (\eta(c) - 1)T \subseteq \mathcal{F}\},$$

and

$$E = \tilde{K}^\times_\infty \cdot \prod_\mathfrak{p} E_\mathfrak{p}.$$

LEMMA 8.2 (SEE LEMMA 2 OF [25]). *Sending $c \in S_T$ to any $q \in K^\times$ such that*

$$q\bar{q}N(c) = 1, q\eta(c)\mathcal{F} = \mathcal{F}, \text{ and } (q\eta(c) - 1)T \subseteq \mathcal{F}$$

induces an injection

$$f : (S_T \cap F)/E \hookrightarrow \mu(\mathcal{O})/\mu(\mathcal{O}) \cap (1 + \mathfrak{a}).$$

PROOF. If $c \in F$ and $q \in K^\times$, then

$$q\eta(c)\mathcal{F} = \mathcal{F} \iff q\mathcal{F} = \mathcal{F} \iff q \in \mathcal{O}^\times.$$

Thus, for $c \in S_T \cap F$ and q as in the statement of the Lemma, we have $q \in \mathcal{O}^\times$ and $q\bar{q} = 1$, and therefore $q \in \mu(\mathcal{O})$. The Lemma now follows easily. □

COROLLARY 8.3. *The map f of Lemma 8.2 induces an injection*

$$(S_T \cap \tilde{\mathcal{O}}^\times_\mathfrak{p})/E_\mathfrak{p} \hookrightarrow \mu(\mathcal{O})/\mu(\mathcal{O}) \cap (1 + \mathfrak{a}).$$

Define

$$\eta_\mathfrak{p} : \mathbf{Q}^\times \to K^\times_\mathbf{A}$$

by viewing

$$\mathbf{Q}^\times \subset \tilde{K}_\mathfrak{p} \subset \tilde{K}^\times_\mathbf{A}$$

and letting

$$\eta_\mathfrak{p}(m) = \eta(m) \text{ for } m \in \mathbf{Q}^\times \subset \tilde{K}_\mathfrak{p} \subset \tilde{K}^\times_\mathbf{A}.$$

The representation $\tilde{\Phi}$ is equivalent to a direct sum of b embeddings of $\tilde{K}$ into $\mathbf{C}$. Let K^c denote the Galois closure of K. For π a prime of K, let λ be any prime of K^c which divides π, and let

$$c(\pi) = c(\lambda) = \#\{\sigma \in \tilde{\Phi} : \lambda \text{ divides } \sigma(\mathfrak{p})\}.$$

Composing $\eta_{\mathfrak{p}}$ with the embedding $K_{\mathbf{A}}^{\times} \subset K_{\mathbf{A}}^{c\times}$, we can view $\eta_{\mathfrak{p}}$ as a map from $\mathbf{Q}^{\times}$ into $K_{\mathbf{A}}^{c\times}$, and consider the λ-component of $\eta_{\mathfrak{p}}(m)$, for λ a prime of K^c. We see that $\eta_{\mathfrak{p}}(m)_{\lambda} = m^{c(\lambda)}$. Since $\eta_{\mathfrak{p}}(m) \in K_{\mathbf{A}}^{\times}$, we see that $c(\pi)$ is independent of the choice of λ, justifying the notation.

Define $N \in \mathbf{Z}^+$ by

$$N\mathbf{Z} = \mathbf{Z} \cap \mathfrak{a},$$

and let

$$R = \{m \in (\mathbf{Z}/N\mathbf{Z})^{\times} : m^{c(\pi)} \equiv 1 \pmod{p^{s(\pi)}} \text{ for all } \pi \in P\}.$$

LEMMA 8.4. *The inclusion $\mathbf{Z} \subset \tilde{\mathcal{O}}_{\mathfrak{p}}$ induces an injection*

$$(\mathbf{Z}/N\mathbf{Z})^{\times}/R \hookrightarrow \tilde{\mathcal{O}}_{\mathfrak{p}}^{\times}/E_{\mathfrak{p}}.$$

PROOF. For $m \in (\mathbf{Z}/N\mathbf{Z})^{\times}$, write $\bar{m}$ for any lift of m to $\mathbf{Z}$. Let

$$\hat{\mathfrak{a}} = \{x \in K_{\mathbf{A}}^{\times} : xT \subseteq \mathcal{F}\}.$$

Then

$$E_{\mathfrak{p}} = \tilde{\mathcal{O}}_{\mathfrak{p}}^{\times} \cap \eta^{-1}(1 + \mathfrak{a}),$$

and the map

$$(\mathbf{Z}/N\mathbf{Z})^{\times} \to \tilde{\mathcal{O}}_{\mathfrak{p}}^{\times}/E_{\mathfrak{p}}$$

has kernel

$$\{m \in (\mathbf{Z}/N\mathbf{Z})^{\times} : \eta_{\mathfrak{p}}(\bar{m}) \in 1 + \hat{\mathfrak{a}}\}$$

$$= \{m \in (\mathbf{Z}/N\mathbf{Z})^{\times} : \bar{m}^{c(\pi)} - 1 \in \mathfrak{a}_{\pi} \text{ for all primes } \pi \text{ of } K\}$$

$$= \{m \in (\mathbf{Z}/N\mathbf{Z})^{\times} : \bar{m}^{c(\pi)} - 1 \in \mathfrak{a}_{\pi} \text{ for all primes } \pi \text{ of } K \text{ s.t. } c(\pi) > 0\}$$

$$= \{m \in (\mathbf{Z}/N\mathbf{Z})^{\times} : \bar{m}^{c(\pi)} - 1 \in \mathfrak{a}_{\pi} \text{ for all primes } \pi \text{ of } K \text{ s.t. } \pi \text{ divides } \eta(\mathfrak{p})\}$$

$$= R.$$

□

For $c, k \in \mathbf{Z}^+$, p prime, and $0 \leq t \in \mathbf{Z}$ such that p^t divides k, let

$$r_c(k, p^t) = \#\{m \in (\mathbf{Z}/k\mathbf{Z})^{\times} : m^c \equiv 1 \pmod{p^t}\}.$$

Let $\mu_{\mathfrak{a}} = \#(\mu(\mathcal{O}) \cap (1 + \mathfrak{a}))$. Combining Lemma 8.1, Corollary 8.3, and Lemma 8.4, we now have:

PROPOSITION 8.5.

$$e \geq \frac{\phi(N)\mu_a}{r_{c(\pi)}(N, p^{s(\pi)})\mu} \text{ for every } \pi \in P.$$

Write (a, c) for the greatest common divisor of integers a and c. The following lemma is elementary.

LEMMA 8.6. *Suppose $t \leq u$. Then:*

(i) $r_c(2^u, 2^t) = \begin{cases} 2^{u-t}(c,2)(c, 2^{t-2}) & \text{if } t \geq 2, \\ 2^{u-1} & \text{if } t = 1 \text{ or } 0. \end{cases}$

(ii) *If p is an odd prime, then*

$$r_c(p^u, p^t) = (c, \phi(p^t))p^{u-t}.$$

(iii) *If p is a prime and p does not divide n, then*

$$r_c(np^u, p^t) = \phi(n) r_c(p^u, p^t).$$

PROPOSITION 8.7.

If p is odd, then $e \geq \dfrac{\phi(p^s)\mu_a}{2^{d-1}\mu}$.

If $p = 2$, then $e \geq 2^{s-1-d}\mu_a/\mu$.

PROOF. Combine Proposition 8.5, Lemma 8.6, the upper bound

$$c(\pi) \leq |\tilde{\Phi}| = b \leq 2^{d-1},$$

and the fact that $\operatorname{ord}_p(N) \geq s$. □

Theorem 7.1 now follows from Proposition 8.7.

Remark 8. Suppose A is an elliptic curve with complex multiplication by the ring of integers $\mathcal{O}$ in an imaginary quadratic field K, $\mathfrak{p}$ is a prime of K of residue characteristic p, T corresponds to the $\mathfrak{p}^t$-torsion on A, j is the j-invariant of the elliptic curve A, and e is the ramification index of primes above $\mathfrak{p}$ in the field extension $K(j, \mathfrak{p}^t\text{-torsion})$ over $K(j)$. Then Theorem 7.1 implies that

$$e \geq \frac{\phi(p^s)}{\mu} \text{ if } p \text{ is odd, and}$$

$$e \geq \frac{2^{s-2}}{\mu} \geq \frac{2^s}{24} \text{ if } p = 2,$$

where

$$s = \begin{cases} t & \text{if } p \text{ is inert or split in } K, \\ [\frac{t}{2}] & \text{if } p \text{ ramifies in } K. \end{cases}$$

However, the Main Theorem of Complex Multiplication shows in this case that

$$e \geq \frac{|(\mathcal{O}/\mathfrak{p}^t)^\times|}{\mu} = \begin{cases} \phi(p^t)/\mu & \text{if } p \text{ splits or ramifies in } K, \\ p^{2(t-1)}(p^2-1)/\mu & \text{if } p \text{ is inert in } K. \end{cases}$$

Remark 9. For simple CM abelian varieties, (K,Φ) coincides with the reflex of its reflex ([23],p. 72). Therefore $\det\Phi$ induces a continuous homomorphism

$$\tilde{\eta} : K_{\mathbf{A}}^\times \to \tilde{K}_{\mathbf{A}}^\times$$

which in turn gives a homomorphism

$$(\mathcal{O}'/N\mathcal{O}')^\times \to \tilde{\mathcal{O}}_{\mathfrak{p}}^\times / E_{\mathfrak{p}}$$

for $\mathcal{O}'$ a suborder of $\mathcal{O}$. The kernel of this latter homomorphism is

$$\{\alpha \in (\mathcal{O}'/N\mathcal{O}')^\times : \eta \circ \tilde{\eta}(\bar{\alpha}) \in 1 + \hat{\mathfrak{a}}\}$$

where $\bar{\alpha}$ is any lift of α to $\mathcal{O}'$. By studying the composition $\eta \circ \tilde{\eta}$, and utilizing the lower bound $b \geq 1 + \log_2 d$, it should be possible to improve the bounds in Theorems 5.1 and 7.1 in the case of simple CM abelian varieties.

9. Purely Additive Reduction

We state some results on torsion on abelian varieties with a prime of purely additive reduction.

THEOREM 9.1 (LENSTRA-OORT, THEOREM 1.13 OF [12]). *Suppose K is a field, v is a discrete valuation of K with perfect residue class field k, and A is an abelian variety defined over K of dimension d, with purely additive reduction at v. Then for every prime $\ell \neq \mathrm{char}(k)$ the number*

$$b(\ell) \in \{0, 1, 2, \ldots, \infty\}$$

defined by

$$\sup_{N \geq 0} |A[\ell^N](K)| = \ell^{b(\ell)}$$

is finite, and

$$\sum_{\substack{\ell \text{ prime} \\ \ell \neq \mathrm{char}(k)}} (\ell - 1)b(\ell) \leq 2d.$$

With the additional hypothesis that A have potential good reduction at v, Lorenzini (see [14]) has recently shown that in fact:

$$\sum_{\substack{\ell \text{ prime} \\ \ell \neq \mathrm{char}(k)}} [\ell^{e(\ell)} - 1 + (\ell - 1)(b(\ell) - e(\ell))] \leq 2d,$$

where $\ell^{e(\ell)}$ is the exponent of the group $A[\ell^\infty](K)$.

THEOREM 9.2 (FLEXOR-OESTERLÉ, [4]). *Suppose E is an elliptic curve defined over a number field M of degree m.*

(i) *If E has bad reduction of additive type at at least one finite place of M, then*

$$|E(M)_{\text{tors}}| \leq 48m.$$

(ii) *If E has bad reduction of additive type at at least two finite places of M, of different residue characteristics, then*

$$|E(M)_{\text{tors}}| \text{ divides } 12,$$

and if one of the residue characteristics is at least five, then

$$|E(M)_{\text{tors}}| \leq 4.$$

(iii) *If E has good reduction at a place of residue characteristic two, then*

$$|E(M)_{\text{tors}}| \leq 5 \cdot 2^m.$$

By supplementing (i) of the Flexor-Oesterlé theorem with the Lenstra-Oort theorem, one further restricts the possibilities for the orders of torsion subgroups of Mordell-Weil groups of elliptic curves defined over number fields and having a prime of bad, additive reduction. Similarly, by supplementing the theorems of §5 and §6 with the Lenstra-Oort or Lorenzini result, we can improve the restrictions on the possibilities for the torsion subgroups of Mordell-Weil groups of abelian varieties of CM-type with a prime of purely additive reduction. For example, if A is an abelian variety of dimension d and of CM-type, defined over a number field M, and A has purely additive reduction at a prime of residue characteristic p, then

$$|A(M)_{\text{tors}}| = p^{\alpha}c$$

where

$$c \in \mathbf{Z}^{+}, p \nmid c, \text{ and } \sum_{\substack{\ell \text{ prime} \\ \ell \neq p}} (\ell - 1)\text{ord}_{\ell}(c) \leq 2d,$$

and where upper bounds on p and p^{α}, and additional restrictions on $p^{\alpha}c$, can be obtained from Corollary 5.4.

By [21], if A is a simple CM abelian variety and M is a number field of definition for A and all its endomorphisms, then A has either good or purely additive reduction at the primes of M. Consequently, for any subfield of M over which A is defined, the reductions will be either good, purely additive, or mixed good and additive. By a result of Fontaine (Corollaire, p. 517 of [6]), every abelian variety (of dimension ≥ 1) defined over $\mathbf{Q}$, $\mathbf{Q}(\sqrt{-1})$, $\mathbf{Q}(\sqrt{-3})$, or $\mathbf{Q}(\sqrt{-5})$, has a prime of bad reduction. In all the examples I have looked at of absolutely simple CM abelian varieties defined over $\mathbf{Q}$, all rational primes of bad reduction are primes of purely additive reduction.

PROBLEM 9.3. *Find absolutely simple CM abelian varieties defined over* $\mathbf{Q}$ *which have no rational primes of purely additive reduction.*

In particular, one can ask whether such examples exist which are also Jacobians of curves. By results of Coleman and McCallum, such examples do not exist for quotients of Jacobians of Fermat curves (see Theorem 5.3 of [2] and the Remark following).

10. Tables

The tables in this section follow from Corollaries 5.4 and 6.1, the result of Olson stated in Remark 6 (for the case $d = 1$ and $m = 1$), the result of Müller-Ströher-Zimmer stated in Remark 7 (for the case $d = 1$ and $m = 2$, since CM elliptic curves have integral j-invariants), and the results of van Mulbregt in [26] (for the case $d = 2$ and $m = 1$ and the case $d = 3$, $m = 1$, and A is simple). Except for the Olson, Müller-Ströher-Zimmer, and van Mulbregt cases just mentioned, the tables give:

Table 1 : $\max\{\text{primes } p : p \leq 6m+1\}$ and $\max\{N : \phi(N) \leq 6m\}$,
Table 2 : $\max\{\text{primes } p : p \leq 1 + 2^{2d-1}\mu m$ and $\phi(\mu) \leq 2d\}$,
Table 3 : $\max\{\text{primes } p : p \leq 1 + 2^{2d-1}(d!)\mu m$ and $\phi(\mu) \leq 2d\}$,
Table 4 : $\max\{\text{primes } p : p \leq 1 + 2^{3d-1} \cdot 3^d(d!)m\}$.

TABLE 1. Upper bounds on orders N and prime orders p of torsion points on CM elliptic curves defined over number fields of degree m.

m	1	2	3	4	5	6	7	8	9	10	11
p	3	7	19	23	31	37	43	47	53	61	67
N	6	12	60	90	90	126	150	210	210	210	240
m	12	13	14	15	16	17	18	19	20	21	22
p	73	79	83	89	97	103	109	113	113	127	131
N	270	270	330	330	420	420	420	420	462	462	510
m	23	24	25	26	27	28	29	30	31	32	33
p	139	139	151	157	163	167	173	181	181	193	199
N	510	630	630	630	660	660	660	690	690	840	840
m	34	35	36	37	38	39	40	41	42	43	44
p	199	211	211	223	229	233	241	241	251	257	263
N	840	840	840	840	870	870	1050	1050	1050	1050	1050
m	45	46	47	48	49	50	51	52	53	54	55
p	271	277	283	283	293	293	307	313	317	317	331
N	1050	1050	1050	1260	1260	1260	1260	1260	1260	1320	1320

TABLE 2. Upper bounds on prime orders of torsion points on absolutely simple CM abelian varieties of dimension d defined over number fields of degree m.

		d				
		1	2	3	4	5
	1	3	79	319	3833	15361
	2	7	157	1153	7681	30713
	3	19	241	1723	11519	46073
	4	23	317	2297	15361	61441
m	5	31	401	2879	19183	76801
	6	37	479	3457	23041	92153
	7	43	557	4027	26881	107509
	8	47	641	4603	30713	122869
	9	53	719	5179	34549	138241
	10	61	797	5749	38393	153589

		d				
		6	7	8	9	10
	1	86017	344053	1966079	7864301	34602991
	2	172031	688111	3932153	15728611	69206017
	3	258031	1032193	5898209	23592937	103809011
	4	344053	1376257	7864301	31457269	138412033
m	5	430081	1720321	9830393	39321599	173015033
	6	516091	2064379	11796469	47185907	207618031
	7	602111	2408437	13762549	55050217	242221051
	8	688111	2752513	15728611	62914549	276824033
	9	774143	3096571	17694709	70778861	311427073
	10	860143	3440627	19660799	78643199	346030067

TABLE 3. Upper bounds on prime orders of torsion points on abelian varieties of dimension d with smCM defined over number fields of degree m.

		d					
		1	2	3	4	5	6
	1	3	79	3457	92153	1843201	61931491
	2	7	157	6911	184321	3686401	123863041
	3	19	241	10369	276467	5529581	185794463
	4	23	317	13807	368633	7372753	247726051
m	5	31	401	17257	460793	9215971	309657599
	6	37	479	20731	552917	11059201	371589121
	7	43	557	24181	645097	12902401	433520639
	8	47	641	27647	737281	14745559	495452161
	9	53	719	31091	829399	16588801	557383637
	10	61	797	34549	921601	18431999	619315199

TABLE 4. Upper bounds on prime orders of torsion points on abelian varieties of CM-type of dimension d defined over number fields of degree m.

		d				
		1	2	3	4	5
	1	3	193	41467	3981301	477757439
	2	7	383	82939	7962607	955514851
	3	19	577	124367	11943929	1433272279
	4	23	769	165887	15925241	1911029753
m	5	31	953	207343	19906553	2388787187
	6	37	1153	248827	23887849	2866544633
	7	43	1327	290249	27869173	3344302037
	8	47	1531	331777	31850491	3822059519
	9	53	1723	373231	35831779	4299816931
	10	61	1913	414721	39813119	4777574381

REFERENCES

1. J. W. S. Cassels, *Diophantine equations with special reference to elliptic curves*, J. London Math. Soc. **41** (1966), 193–291.
2. R. Coleman and W. McCallum, *Stable reduction of Fermat curves and Jacobi sum Hecke characters*, J. reine angew. Math. **385** (1988), 41–101.
3. S. David, *Fonctions theta et points de torsion des variétés abéliennes*, Comp. Math. **78** (1991), 121–160.
4. M. Flexor and J. Oesterlé, *Sur les points de torsion des courbes elliptiques*, Astérisque **183** (1990), 25–36.
5. E. V. Flynn, *Sequences of rational torsions on abelian varieties*, Invent. math. **106** (1991), 433–442.
6. J-M. Fontaine, *Il n'y a pas de variétés abéliennes sur* **Z**, Invent. math. **81** (1985), 515–538.
7. G. Frey, *Some remarks concerning points of finite order on elliptic curves over global fields*, Ark. Mat. **15** (1977), 1–19.
8. M. Hindry and J. Silverman, *The canonical height and integral points on elliptic curves*, Invent. math. **93** (1988), 419–450.
9. S. Kamienny, *On the torsion subgroups of elliptic curves over totally real fields*, Invent. math **83** (1986), 545–551.
10. S. Kamienny, *Torsion points on elliptic curves*, Proceedings of the conference on number theory and arithmetical geometry, March 12–15, 1991, Institut für Experimentelle Mathematik, Universität Essen **18** (1991), 19–22.
11. M. Kenku and F. Momose, *Torsion points on elliptic curves defined over quadratic fields*, Nagoya Math. J. **109** (1988), 125–149.
12. H. W. Lenstra, Jr. and F. Oort, *Abelian varieties having purely additive reduction*, J. Pure and Applied Alg. **36** (1985), 281–298.
13. M. Levin, *On the group of rational points on elliptic curves over function fields*, Amer. J. Math. **90** (1968), 456–462.
14. D. Lorenzini, *On the group of components of a Néron model*, preprint.
15. Y. Manin, *The p-torsion of elliptic curves is uniformly bounded*, Math USSR – Izvestija **3** (1969), 433–438.
16. B. Mazur, *Modular curves and the Eisenstein ideal*, Publ. math. IHES **47** (1977), 33–186.
17. H. Müller, H. Ströher, and H. Zimmer, *Torsion groups of elliptic curves with integral j-invariant over quadratic fields*, J. reine angew. Math. **397** (1989), 100–161.
18. L. Olson, *Points of finite order on elliptic curves with complex multiplication*, Manuscripta Math. **14** (1974), 195–205.
19. K. Ribet, *Division fields of abelian varieties with complex multiplication*, Mem. Soc. Math. France **2** (1980), 75–94.
20. J-P. Serre, *Résumé de Cours de 1985-86*, Collège de France.
21. J-P. Serre and J. Tate, *Good reduction of abelian varieties*, Ann. Math. **88** (1968), 492–517.
22. G. Shimura, *Introduction to the arithmetic theory of automorphic functions*, Princeton Univ. Press, Princeton, N. J. (1971).
23. G. Shimura and Y. Taniyama, *Complex multiplication of abelian varieties and its applications to number theory*, Math. Soc. Japan, Tokyo (1961).
24. A. Silverberg, *Points de torsion des variétés abéliennes de type CM*, Problèmes Diophantiens 1985–86, Publications mathématiques de l'Université de Paris VI **79** (1986), exp. no. 7, 12 pp.
25. A. Silverberg, *Torsion points on abelian varieties of CM-type*, Comp. Math. **68** (1988), 241–249.
26. P. van Mulbregt, *Torsion-points on low dimensional abelian varieties with complex multiplication*, this volume.

DEPARTMENT OF MATHEMATICS, OHIO STATE UNIVERSITY, COLUMBUS, OHIO 43210
E-mail address: silver@function.mps.ohio-state.edu

This page intentionally left blank

Contemporary Mathematics
Volume 133, 1992

The Arithmetic and Geometry of Elliptic Surfaces

PETER F. STILLER

ABSTRACT. We survey some aspects of the theory of elliptic surfaces and give some results aimed at determining the Picard number of such a surface. For the surfaces considered, this will be equivalent to determining the Mordell-Weil rank of an elliptic curve defined over a function field in one variable. An interesting conjecture concerning Galois actions on the relative de Rham cohomology of these surfaces is discussed.

This paper focuses on an important class of algebraic surfaces called elliptic surfaces. The results while geometric in character are arithmetic at heart, and for that reason we devote a fair portion of our discussion to those definitions and facts that make the arithmetic clear. Later in the paper, we will explain some recent results and conjectures. This is a preliminary version, the detailed version will appear elsewhere.

There are a number of natural routes leading to the definition of the class of elliptic surfaces. Let E denote a compact connected complex manifold with $\dim_{\mathbb{C}} E = 2$.

THEOREM 1. *(Siegel) The field of meromorphic functions on E has transcendence degree ≤ 2 over $\mathbb{C}$, i.e. the field of meromorphic functions is either:*

1) $\mathbb{C}$ constant functions,

2) a finite separable extension of $\mathbb{C}(x)$, or

3) a finite separable extension of $\mathbb{C}(x, y)$. □

Case 3) is precisely the set of algebraic surfaces, i.e. those admitting an embedding into $\mathbb{P}^N_{\mathbb{C}}$. Case 2) was studied by Kodaira in a series of three papers:

Kodaira, K., "On complex analytic surfaces I, II, III," Annals of Math. 77 and 78, 1963, which expound on elliptic surfaces. Kodaira makes the following definition:

1991 *Mathematics Subject Classification.* 14D05, 14J27.

Partially supported by ARO grant DAAL 03-88-K-0019. The detailed version of this paper will be submitted for publication elsewhere.

DEFINITION 2. *E is elliptic if:*

1) *there exists a smooth curve (read Riemann surface) X and a proper holomorphic map $\pi: E \to X$ of E onto X such that*
2) *$\pi^{-1}(x)$ (with multiplicity) is a non-singular curve E_x of genus one, i.e. a torus, for general $x \in X$. ("General" means for all but finitely many $x \in X$.)*

THEOREM 3. *(Kodaira) Transcendence degree $= 1$ (case 2) above) implies E is elliptic.* □

There are of course many elliptic surfaces which are algebraic and so have transcendence degree $= 2$.

From now on E will denote an elliptic surface. One immediate question is to determine the nature of the singular fibers of π: $E \to X$ at the finite set of points $S = \{x_1, \ldots, x_n\} \subset X$ where the fiber $\pi^{-1}(x_i) = E_{x_i}$ is something other than a non-singular curve (occuring with multiplicity one) of genus one. In the second of the above mentioned papers of Kodaira, a complete description of the singular fiber types is given. (We will ignore multiple fibers, as these can't occur for the type of elliptic surface defined below.)

For our purposes, we will narrow the definition of elliptic surface as follows:

DEFINITION 4. *A compact connected complex surface E will be called elliptic if:*

1) *there exists a smooth curve X (read Riemann surface) and a proper holomorphic map π: $E \to X$ mapping E onto X such that*
2) *$\pi^{-1}(x)$ is a non-singular curve of genus one for general $x \in X$, and*
3) *π: $E \to X$ has a section, i.e. there is a holomorphic map $\mathcal{O}$: $X \to E$ such that $\pi \circ \mathcal{O} = 1_X$,*
4) *E/X is relatively minimal, i.e. there are no exceptional curves of the first kind in the fibers, and*
5) *E/X is not isotrivial.*

A few comments are in order. First condition 3) forces E to be algebraic and devoid of multiple singular fibers. The section $\mathcal{O}$: $X \to E$ furnishes a $K(X)$-rational point on the generic fiber E^{gen} viewed as a curve over $K(X)$. Thus the generic fiber E^{gen} is an elliptic curve over the field $K(X)$ of meromorphic/rational functions on X. Assumption 5) means that we have a non-trivial variation of complex structure in the "good" fibers. Simply put, this means that the J-invariant of the fibers, which can be viewed as a meromorphic/rational function on X, is non-constant. We denote this function by $\mathcal{J} \in K(X)$. The fact that $\mathcal{J}$ is non-constant allows us, via the Mordell-Weil theorem, to conclude that the group $E^{\text{gen}}(K(X))$ of $K(X)$-rational points on the generic fiber, or what is the same, the group of sections of π: $E \to X$, is a finitely generated abelian group. We denote its rank by r_E. Finally, condition 4) in the definition implies that we have blown down all the exceptional curves in the fibers of π: $E \to X$. E is then

the unique minimal compactification of the so-called Néron model of the elliptic curve $E^{gen}/K(X)$. We remark that the map π and the curve X are essentially uniquely determined by the fact that the Jacobian of X must be the Albanese of E.

This places us in a situation analogous to the common arithmetic situation where E is an elliptic curve over a number field K, where the Néron model is an arithmetic surface over the "curve" $\mathrm{Spec}(\mathcal{O}_K)$, $\mathcal{O}_K$ being the ring of algebraic integers in K, and where the fiber over a point in $\mathrm{Spec}\ \mathcal{O}_K$ is the reduction of E modulo a nonzero prime ideal $\varphi \subset \mathcal{O}_K$. The finitely generated abelian group $E(K)$ describes the solutions in K to the Diophantine equation(s) defining E.
Classical Examples:

Legendre (Level 2)	$Y^2 = X(X-1)(X-\lambda)$	avoid characteristic 2 singular fibers at $\lambda = 0, 1, \infty$
Level 3	$X^3 + Y^3 + 1 = \mu XY$	avoid characteristic 3 singular fibers at $\mu^3 = 27$ or ∞

Further Examples (Elliptic Modular Surfaces):

In a paper entitled "On elliptic modular surfaces", which appears in the Journal Math. Soc. Japan, Vol. 24, No. 1 (1972), T. Shioda constructs an important class of elliptic surfaces. Given a subgroup of finite index $\Gamma \subset SL_2(\mathbb{Z})$ with $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \notin \Gamma$, Shioda constructs a family of elliptic curves E_Γ over X_Γ the modular curve $\Gamma\backslash\mathbb{H} \cup \mathbb{Q} \cup \{\infty\}$ with the obvious monodromy representation given by Γ and where the lattice of periods of the fiber over $x \in X_\Gamma$ is homothetic to $\mathbb{Z}\tau + \mathbb{Z}$ where $\tau \in \mathbb{H}$, the complex upper half-plane, corresponds to $x \in X_\Gamma$.

This leads us naturally into the world of classical automorphic forms. We will allude to this in several other places. For now, we content ourselves with recalling one of Shioda's results, namely that the space of cusp forms of weight three for Γ, $S_3(\Gamma)$, is naturally isomorphic to the space of holomorphic two forms on E, $H^0(E, \Omega_E^2)$.

We now turn to the main object of interest in this paper:

DEFINITION 5. *The Néron-Severi group $NS(E)$ of E is defined to be the group of divisors of E modulo algebraic equivalence (as opposed to rational or linear equivalence):*

$$NS(E) = \frac{\textit{divisors}}{\textit{alg. equiv. to } 0} \subset H^2(E, \mathbb{Z}).$$

We remark that for these surfaces, algebraic is the same as homological equivalence, so the Néron-Severi group sits inside $H^2(E, \mathbb{Z})$, which can be shown to be torsion-free. The Picard number is defined to be

$$\rho_E = \textit{rank } NS(E).$$

By the Lefschetz Theorem on (1,1)-classes one also has:

$$NS(E) = H^{1,1} \cap H^2(E, \mathbb{Z}) \subset H^2(E, \mathbb{C})$$

or

$$NS(E) = \text{the group of topological } \mathbb{C}\text{-line bundles which admit analytic structure.}$$

PROBLEM: Calculate ρ_E.

Because we are interested only in the rank of the Néron-Severi group, it is reasonable to tensor with $\mathbb{Q}$ and work with

$$NS(E) \otimes_{\mathbb{Z}} \mathbb{Q}.$$

Now away from the "bad" fibers over $S = \{x_1 \dots x_n\} \subset X$ the family $E|_{X-S} = \pi^{-1}(X - S)$ is locally differentiably trivial, so it is natural to use the Leray spectral sequence

$$E_2^{p,q} = H^p(X, R^q\pi_*\mathbb{Q}) \Rightarrow H^{p+q}(E, \mathbb{Q})$$

to understand $H^2(E, \mathbb{Q})$ in terms of the base X and the fibers which are tori. The Leray spectral sequence degenerates at E_2 and yields a filtration

$$0 \subset F^2_{\mathbb{Q}} \subset F^1_{\mathbb{Q}} \subset F^0_{\mathbb{Q}} = H^2(E, \mathbb{Q})$$

where

$$F^1_{\mathbb{Q}} = \ker(H^2(E, \mathbb{Q}) \to H^0(X, R^2\pi_*\mathbb{Q}))$$

consists of classes which restrict to zero on each fiber, and

$$F^2_{\mathbb{Q}} = \operatorname{im}(H^2(X, \mathbb{Q}) \xrightarrow{\pi^*} H^2(E, \mathbb{Q})) = \mathbb{Q}[E_{x_0}]$$

is generated by the cohomology class of a fiber.

The filtration quotient is

$$F^1_{\mathbb{Q}}/F^2_{\mathbb{Q}} \cong H^1(X, R^1\pi_*\mathbb{Q}).$$

Now the Hodge decomposition on $H^*(E, \mathbb{C})$ induces a Hodge structure on the filtration quotient above

$$H^1(X, R^1\pi_*\mathbb{Q}) \otimes_{\mathbb{Q}} \mathbb{C} = H^1(X, R^1\pi_*\mathbb{C}) = H^{2,0} \oplus H^{1,1} \oplus H^{0,2}$$

where $H^{2,0}$ is all of $H^0(E, \Omega^2_E)$ (because the restriction of a holomorphic two form to a fiber is necessarily zero when the fiber is a curve).

Now there is a well-known

THEOREM 6. *(Shioda)* $\rho_E = r_E + 2 + \sum_{s \in S}(m_s - 1)$ *where* m_s *is the number of irreducible components making up the fiber* $E_s = \pi^{-1}(s)$. *Thus the geometric quantity* ρ_E *is essentially the arithmetic quantity* r_E, *the rank of the Mordell-Weil group of the generic fiber* E^{gen} *treated as an elliptic curve over the field* $K(X)$. □

In practice the numbers m_s are easy to determine, it is r_E that is in general impossible to compute. What should we focus on?

Let $V_{\mathbb{Q}}^i$ be the span of the algebraic cycles in $F_{\mathbb{Q}}^i$ so

$$V_{\mathbb{Q}}^i = (NS(E) \otimes_{\mathbb{Z}} \mathbb{Q}) \cap F_{\mathbb{Q}}^i$$

and let $W_{\mathbb{Q}} = V_{\mathbb{Q}}^1 / V_{\mathbb{Q}}^2 = \frac{(NS(E)\otimes_{\mathbb{Z}}\mathbb{Q}) \cap F_{\mathbb{Q}}^1}{F_{\mathbb{Q}}^2} \subset H^1(X, R^1\pi_*\mathbb{Q})$.

THEOREM 7. $\dim_{\mathbb{Q}} W_{\mathbb{Q}} = r_E$. □

PROOF. See Stiller [5].

It therefore behooves us to look at $R^1\pi_*\mathbb{Q}$ or $R^1\pi_*\mathbb{C}$ which over $X - S$ is a locally constant sheaf of rank two, so that

$$R^1\pi_*\mathbb{C}|_{X-S} \otimes_{\mathbb{C}} \mathcal{O}_{X-S}$$

is a rank two holomorphic vector bundle on $X - S$. Let's study this bundle.

To get quickly to the heart of the matter we adopt a naive point of view.

Pick a base point $x_0 \in X - S$. Here S will contain the support of the "bad" fibers and some additional points to be named later.

In a sufficiently small neighborhood U of x_0

$$\begin{array}{c} \pi^{-1}(U) = E|_U \\ \downarrow \pi|_{\pi^{-1}(U)} \\ U \end{array}$$

$E|_U$ is a $\mathcal{C}^\infty$-trivial fiber bundle, i.e.

$$\begin{array}{ccc} \pi^{-1}(U) & \cong & U \times \pi^{-1}(x_0) \\ \downarrow \pi|_{\pi^{-1}(U)} & & \downarrow pr_1 \\ U & = & U \end{array} .$$

Write E_{x_0} for $\pi^{-1}(x_0)$ and choose an oriented basis $\gamma_1, \gamma_2 \in H_1(E_{x_0}, \mathbb{Z}) \cong \mathbb{Z}^2$ for the homology of the fiber and consider

$$\omega_i(x) = \int_{\gamma_i} \Omega|_{E_x}$$

where Ω is an appropriate meromorphic 1-form on E with poles only on the vertical fibers, i.e. $\Omega|_{E^{\text{gen}}}$ is a $K(X)$-rational differential of the 1^{st} kind on the curve $E^{\text{gen}}/K(X)$. Notice that for an appropriate finite set of points S which

includes the support of the singular fibers, the functions $\omega_i(x)$ can be analytically continued as holomorphic non-vanishing functions throughout $X - S$. Moreover,

$$\operatorname{Im} \omega_1(x)/\omega_2(x) > 0.$$

For a path $\gamma \in \pi_1(X - S, x_0)$, analytic continuation of the pair ω_1, ω_2 around γ yields

$$\begin{pmatrix}\omega_1\\ \omega_2\end{pmatrix} \longmapsto M_\gamma \begin{pmatrix}\omega_1\\ \omega_2\end{pmatrix}$$

where $M_\gamma \in SL_2(\mathbb{Z})$. This is called the *monodromy representation* of E/X and the image of the fundamental group in $SL_2(\mathbb{Z})$ is a subgroup Γ of finite index in $SL_2(\mathbb{Z})$ which is unique up to conjugation in $SL_2(\mathbb{Z})$. This group does not depend on the choice of Ω or on S (provided S contains the support of the singular fibers). Γ is called the *monodromy group* of E/X.

Recall (see Deligne SLN 163) that the following notions are equivalent:

1) a representation of $\pi_1(X - S, x_0) \longrightarrow GL_2(\mathbb{C})$
2) a local system (locally constant sheaf) V of rank 2 on $X - S$
3) a rank 2 holomorphic vector bundle $\mathcal{E}_0 = V \otimes_{\mathbb{C}} \mathcal{O}_{X-S}$ over $X - S$ with integrable holomorphic connection D_0

$$\mathcal{E}_0 \xrightarrow{D_0} \mathcal{E}_0 \otimes_{\mathcal{O}_{X-S}} \Omega^1_{X-S}$$

having regular singular points.

We recall that $\mathcal{E}_0$ can be uniquely extended to a holomorphic (algebraic) rank 2 bundle $\mathcal{E}$ on X together with a meromorphic (rational) connection D having regular singular points. This is known as the *Gauss-Manin connection*.

4) A second order linear differential operator Λ rational/$K(X)$ with regular singular points. In our case $\Lambda\omega_i = 0$ for $i = 1, 2$, i.e. Λ annihilates the periods of Ω as functions on the base. This is the *Picard-Fuchs equation* of E/X and ω_1, ω_2 form a basis for the two dimensional space of solutions at x_0, and elsewhere via analytic continuation.

Our point of view now shifts to this differential equation ($R^1\pi_*\mathbb{C}|_{X-S}$ is the V above). We ask "How much information can we recover from Λ?".

THEOREM 8. *E is determined by Λ up to generic isogeny. We remark that Λ depends on the choice of Ω but any other choice is $g\Omega$ for $g \in K(X)$ and this transforms Λ in the obvious simple way. (See Stiller [2]).* □

THEOREM 9. *If E/X and E'/X are generically isogeneous elliptic surfaces over a fixed base curve X then*

1) $[PSL_2(\mathbb{Z}):\overline{\Gamma}] = [PSL_2(\mathbb{Z}):\overline{\Gamma}']$

2) $b_i(E) = b_i(E')$ $i = 0, \ldots, 4$ *Betti numbers*

3) $p_g(E) = p_g(E')$ *and* $q(E) = q(E') =$ *genus* X

4) $\rho_E = \rho_{E'}$

5) $r_E = r_{E'}$

Remark on the proof: 5) is immediate and is used to prove 4) via the formula

$$\rho_E = r_E + 2 + \sum_{s \in S}(m_s - 1).$$

What is interesting is that m_s is not preserved by generic isogeny – only the sum is. In particular a generic isogeny

$$\begin{array}{ccc} E & \xrightarrow{\phi} & E' \\ & {\scriptstyle\pi}\searrow \quad \swarrow{\scriptstyle\pi'} & \\ & X & \end{array}$$

as a rational map which is an isogeny of the fibers almost everywhere, may not extend to all of E as a regular map, and the same may hold for the dual isogeny

$$\begin{array}{ccc} E' & \xrightarrow{\phi'} & E \\ & {\scriptstyle\pi'}\searrow \quad \swarrow{\scriptstyle\pi} & \\ & X & \end{array} .$$

The sum of the m_s can be determined using the exponents at s of Λ (which is easily obtainable local information) independent of the isogeny class. For details see Stiller [3]. □

Now how do we capture $H^1(X, R^1\pi_*\mathbb{C})$ in terms of Λ? We take our clue from Manin. Given any section s: $X \to E$, $\pi \circ s = 1_X$, we can locally (say near $x_0 \in X - S$) take a family of paths γ_x between the points $s(x)$ and $\mathcal{O}(x)$ on the fiber E_x and compute

$$f(x) = \int_{\gamma_x} \Omega|_{E_x}$$

which is defined up to the periods

$$f + m\omega_1 + n\omega_2.$$

Since Λ annihilates the periods,

$$\Lambda f = Z$$

turns out to be a well-defined rational function. f is thus annihilated by a 3rd order operator $\tilde{\Lambda}$ and the rank 3 local system $V_{\tilde{\Lambda}}$ arises an element of $\mathrm{Ext}^1(\underline{\mathbb{C}}, V_\Lambda)$ where $\underline{\mathbb{C}}$ is the trivial local system. The monodromy representation for $\tilde{\Lambda}$ takes the form

$$\gamma \in \pi_1(X - S, x_0) \longmapsto \begin{pmatrix} 1 & m_\gamma & n_\gamma \\ 0 & & M_\gamma \end{pmatrix} \in SL_3(\mathbb{Z}).$$

The motivation here is that the cohomology class of the algebraic cycle $s - \mathcal{O}$ is essentially zero on the fibers, and so provides an element in the Leray filtration quotient

$$F^1_{\mathbb{C}}/F^2_{\mathbb{C}} = H^1(X, R^1\pi_*\mathbb{C}).$$

We make the following definitions:

DEFINITION 10. $Z \in K(X)$ *is* exact *if* $\Lambda f = Z$ *has a global single-valued meromorphic solution. Thus* $Z \in \Lambda K(X) \subset K(X)$.

DEFINITION 11. $Z \in K(X)$ *is* locally exact *if* $\forall\, p \in X$ *the equation* $\Lambda f = Z$ *restricted to a small neighborhood* U_p *of* p *has a single-valued meromorphic solution. We denote the set of locally exact* $Z \in K(X)$ *by* L_Λ^{para}. *(The notation comes from the theory of automorphic forms and the notion of parabolic cohomology).*

DEFINITION 12. H^1_{IDR} *is defined to be* $L_\Lambda^{\text{para}}/\Lambda K(X)$ *and is called the* inhomogeneous de Rham cohomology.

Some remarks are in order. First we have slid over the non-intrinsic nature of Λ which depends on the choice of Ω and on the choice of a derivation $\frac{d}{dx}$ on $K(X)$. A more intrinsic formulation would treat Z as $Z(dx)^2$ a meromorphic quadratic differential, i.e. a meromorphic section of $(\Omega^1_X)^{\otimes 2}$. In any event H^1_{IDR} is independent of any choices. Secondly, local exactness can be formulated as a residue condition.

THEOREM 13. H^1_{IDR} *is canonically isomorphic to* $H^1(X, R^1\pi_*\mathbb{C})$.

The proof is achieved by showing that both groups are naturally isomorphic to the subgroup of locally split extensions in $\text{Ext}^1(\underline{\mathbb{C}}, V_\Lambda)$. Note that

$$\text{Ext}^1(\underline{\mathbb{C}}, V_\Lambda) \cong H^1(\pi_1(X - S, x_0), (V_\Lambda)_{x_0}) \cong H^1(X - S, R^1\pi_*\mathbb{C}|_{X-S})$$

with the middle group being the usual group cohomology. The locally split classes correspond to parabolic cohomology and $H^1(X, R^1\pi_*\mathbb{C})$. Here

$$H^1(X, R^1\pi_*\mathbb{C}) \hookrightarrow H^1(X_0, R^1\pi_*\mathbb{C}|_{X_0})$$

where $X_0 = X - S$. This last inclusion comes from the exact sequence of low order terms in the Leray spectral sequence for i: $X_0 \hookrightarrow X$ and the sheaf $R^1\pi_*\mathbb{C}|_{X_0}$. (See Stiller [4], [5]). □

Now that we have identified H^1_{IDR} with $H^1(X, R^1\pi_*\mathbb{C})$ what about the Hodge decomposition on the latter.

THEOREM 14. *There are two divisors* $\mathcal{A}_0 < \mathcal{A}$ *on* X, *easily computable in terms of the local behavior of* Λ, *such that every element of* $L(\mathcal{A}_0)$ *is locally exact but never exact (unless 0) and such that no locally exact element in* $L(\mathcal{A}) \cap L_\Lambda^{\text{para}}$ *is ever exact (except 0), and such that*

$$L(\mathcal{A}_0) \hookrightarrow H^1_{IDR}$$

corresponds to $H^{2,0}$ *in* $H^1(X, R^1\pi_*\mathbb{C})$ *and*

$$L(\mathcal{A}) \cap L_\Lambda^{\text{para}} \hookrightarrow H^1_{IDR}$$

corresponds to $H^{2,0} \oplus H^{1,1}$. □

The point of this result is that we now have unique representatives of the form $\Lambda f = Z$ for elements of $H^{2,0}$ and $H^{2,0} \oplus H^{1,1}$ in $H^1(X, R^1\pi_*\mathbb{C})$. (See Stiller [5]).

APPLICATION.

We assume for the moment that $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \notin \Gamma \subset SL_2(\mathbb{Z})$ where Γ is the monodromy group of E/X. We then have a diagram

$$\begin{array}{ccccc} E & \xrightarrow{\sim} & E_\Gamma \times_{X_\Gamma} X & \longrightarrow & E_\Gamma \\ {\scriptstyle\pi}\searrow & & \downarrow & & \downarrow {\scriptstyle\pi_\Gamma} \\ & & X & \xrightarrow{\omega} & X_\Gamma \end{array}$$

where $\omega = \omega_1/\omega_2$ is the so-called *period map*. If we suppose X is Galois over the modular curve X_Γ then $G = \mathrm{Gal}(X/X_\Gamma)$ acts on $H^1(X, R^1\pi_*\mathbb{Q})$ and preserves Hodge type in $H^1(X, R^1\pi_*\mathbb{C})$.

PROBLEM. Let V be an irreducible rational or complex representation of G. What is the multiplicity of V in $H^1(X, R^1\pi_*\mathbb{Q})$ or $H^1(X, R^1\pi_*\mathbb{C})$?

We have been able to show in many cases where G is cyclic, i.e. $K(X)$ is a cyclic extension of the field of modular functions $K(X_\Gamma)$, that all multiplicities are one. This can't be true in general, but we conjecture that it holds when $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \notin \Gamma$ under suitable hypotheses on G (modulo the obvious trivial constituents from $H^0(X_\Gamma, \Omega^2_{X_\Gamma})$ etc.). When the multiplicities are all 1, we can explicitly decompose the G-modules $H^{2,0}$ and $H^{2,0} \oplus H^{1,1}$ using our unique representatives. Since Hodge type is preserved and the multiplicities are one, $H^{1,1}$ is the sum of those irreducible constituents of $H^{2,0} \oplus H^{1,1}$ not in $H^{2,0}$. If in turn all the complex irreducible constituents (say G is abelian – so that over $\mathbb{C}$ all irreducible constituents V are one dimensional, and over $\mathbb{Q}$ we want eigenvalues which are all primitive d^{th} roots of one) of a given irreducible rational representation (dimension $\phi(d)$ in the abelian case) lie in the $H^{1,1}$ part, we get a contribution (of $\phi(d)$ in the abelian case) to ρ_E. (See Stiller [5] for examples.)

One approach to the multiplicity problem is suggested by a similar looking multiplicity problem that goes back to

C. Chevalley and A. Weil, "Über das Verhalten der Integrale ersten Gattung bei Automorphismen des Funktionenkörpers," Abh. Math. Sem. Univ. Hamburg 10 (1934), 358-361.

A. Weil, "Über Matrizenringe auf Riemannschen Flächen und den Riemann-Rochschen Satz," Abh. Math. Sem. Univ. Hamburg 11 (1936) 110-115.

and in modern exposition:

J.F. Glazebrook and D.R. Grayson, "Galois representations on holomorphic differentials," preprint.

The set-up is

$\widetilde{X}/\mathbb{C}$ a curve of genus $\tilde{g}$

G acts faithfully on $\widetilde{X}$ so $G \hookrightarrow \mathrm{Aut}(\widetilde{X})$

G acts on $H^0(\widetilde{X}, (\Omega^1_{\widetilde{X}})^{\otimes q})$.

The results describe $H^0(\widetilde{X}, (\Omega^1_{\widetilde{X}})^{\otimes q})$ as a representation of G for $q \geq 1$. Namely given an irreducible complex representation V of G, a formula is given, in terms of local ramification invariants, for the multiplicity of V in $H^0(\widetilde{X}, (\Omega^1_{\widetilde{X}})^{\otimes q})$.

BIBLIOGRAPHY

1. J. Manin, *Algebraic curves over fields with differentiation*, AMS Transactions **50** (1966).
2. P. Stiller, *Differential equations associated with elliptic surfaces*, Journal Math. Soc. Japan **32, No. 2** (1981), 203–233.
3. P. Stiller, *Monodromy and invariants of elliptic surfaces*, Pacific Journal of Math. **92, No. 2** (1981), 433–452.
4. P. Stiller, *Automorphic forms and the Picard number of an elliptic surface*, Aspects of Mathematics **E5** (1984), Vieweg Verlag.
5. P. Stiller, *The Picard numbers of elliptic surfaces with many symmetries*, Pacific Jour. of Math. **128, No. 1** (1987).

Department of Mathematics, Texas A&M University, College Station, TX 77843-3368

DEPARTMENT OF MATHEMATICS, TEXAS A&M UNIVERSITY, COLLEGE STATION, 77843-3368

E-mail: stiller@alggeo.tamu.edu

Contemporary Mathematics
Volume 133, 1992

Torsion-points on low dimensional abelian varieties with Complex Multiplication

P. VAN MULBREGT

ABSTRACT. This paper finds explicit bounds on orders of torsion points on abelian varieties with complex multiplication. We include tables giving upper bounds for the order of a rational torsion point when the dimension is 2 or 3.

1. Introduction

The results in Section 5 of Silverberg's paper [4] give bounds for the order of a rational torsion point on a d-dimensional abelian variety with Complex Multiplication. These bounds are based on asymptotic estimates for two quantities. While the estimates are good for large values of d, neither is best possible for small values of d. This is especially true if the variety is simple, or more generally, if the algebra of endomorphisms $\mathrm{End}_{\mathbf{Q}}(A)$ contains a field K of dimension $2d$, the case of "sufficiently many complex multiplications" (smCM). In this situation, A is isogenous to a product $B \times B \times \ldots \times B$ of one simple abelian variety. This paper analyzes the estimates for two- and three-dimensional abelian varieties, and improves these bounds.

2. Two-dimensional Abelian Varieties

The notation is that of Silverberg in [3] and [4]. We have an abelian variety A of dimension d of CM-type, i.e. $K_1 \ldots K_m$ are CM fields, $\sum_{i=1}^{m} n_i [K_i : \mathbf{Q}] = 2d$, $Z = M_{n_1}(K_1) \times \cdots \times M_{n_m}(K_m)$, and θ is an embedding of Z into $\mathrm{End}(A) \otimes \mathbf{Q} = \mathrm{End}_{\mathbf{Q}}(A)$. $K = K_1 \times \cdots \times K_m$, $\tilde{K}_i$ is the reflex of K_i induced by θ and $\tilde{K}$ is the compositum of the $\tilde{K}_i$. Let $[\tilde{K} : \mathbf{Q}] = 2b$. Note that there is no requirement

---

1991 *Mathematics Subject Classification.* 11G10; Secondary 11G15, 14K22, 14K15, 14Q20.
I would like to thank Alice Silverberg for many helpful conversations.
This paper is in final form and no version of it will be submitted for publication elsewhere.

that the fields K_i be distinct, nor that Z be the full endomorphism algebra. In what follows, m is usually 1, so that θ reduces to an embedding of K_1, or $M_{n_1}(K_1)$, into $\mathrm{End}_{\mathbf{Q}}(A)$.

The theorem we start with appears as Corollary 8 in [3].

PROPOSITION 2.1. *Suppose A is a two-dimensional abelian variety of CM-type defined over $\mathbf{Q}$, and N is the order of a torsion point of $A(\mathbf{Q})$. Then:*

(i) $\phi(N) \leq 2^{6+\nu(N)} \cdot 3^2$.

(ii) $\phi(N) \leq 2^{5+\nu(N)} \cdot 3^2$ *if* $8 \nmid N$.

(iii) $\nu(N) \leq 6$ *($\nu(N)$ = the number of prime divisors of N.)*

(iv) $N \leq 185640$.

(v) *If N is prime then* $N \leq 577$.

(vi) $|A(\mathbf{Q})_{torsion}| \leq (185640)^4$.

By looking at the estimates made in the proof of this result, we can bring these bounds down. The two quantities referred to in the introduction are μ, the number of roots of unity in the product of the fields occurring in the CM-type, and $r_b(N)$, the number of solutions to the congruence $m^b \equiv 1 \pmod{N}$. When one of these quantities is large then the other is small, and vice-versa. For if $Z = K_1$ is a field, and μ is large, then K_1 contains a large degree cyclotomic field, which is abelian over the rationals $\mathbf{Q}$. Hence, the degree of the reflex field over $\mathbf{Q}$, $[\tilde{K}_1 : \mathbf{Q}]$, is not much larger than d. On the other hand, if b is large, then the degree $[\tilde{K}_1 : \mathbf{Q}]$ is large relative to d. This means that any abelian subfields of K_1 are of small degree. By Kronecker-Weber, this implies that K_1 contains very few roots of unity. Thus, in the end, different CM fields actually have similar bounds for the order of a rational torsion point.

PROPOSITION 2.2. *Let A, N be as above.*

(i) *Then* $\nu(N) \leq 5$.

Suppose, in addition, that A is simple.

(ii) *Then* $N \leq 840$ *unless* $K = \mathbf{Q}(\zeta_{10})$, *in which case* $N \leq 10920$.

(iii) *If N is prime then* $N \leq 17$ *unless* $K = \mathbf{Q}(\zeta_{10})$, *in which case* $N \leq 79$.

Suppose that A is not simple.

(iv) *Then* $N \leq 28560$.

(v) *If N is prime then* $N \leq 193$.

PROOF. (i) This is an *a fortiori* observation.

(ii) & (iii) Suppose A is simple. Then $\mathrm{End}_{\mathbf{Q}}(A) = K$ for some quartic CM field. The crucial inequality (Corollary 3 of [3]) for A simple is

$$\phi(N) \leq [\tilde{K} : \mathbf{Q}] r_b(N) \mu$$

where $\tilde{K}$ is the reflex field of K. We now analyze the various cases.

If K is a quartic CM-field which is not a biquadratic field and has only two roots of unity, then we are looking for N satisfying $\phi(N) \leq 4 \cdot r_2(N) \cdot 2$. This process requires $N \leq 17$ if N is prime, and $N \leq 840$ if N is composite.

The biquadratic fields all have as their Galois group the Klein-4 group. This group has the property that any subset of two elements is a coset of some subgroup. This fact implies that the reflex fields are not big; rather they are only quadratic extensions of the rationals. In particular, the CM-type is not simple (see §8.4 of [2].) Note that the two cyclotomic fields $\mathbf{Q}(\zeta_8)$ and $\mathbf{Q}(\zeta_{12})$ are biquadratic.

It remains to consider the field $\mathbf{Q}(\zeta_{10})$. Here the Galois group is a cyclic group of 4 elements, which does not possess the above property, so simple CM types do exist for this field. By computation we get an upper bound of 79 for the order of a prime N, and 10920 for composite N. These results are summarized in Table 1.

(iv) & (v) Now suppose that A is not simple. The appropriate inequality for non-simple CM type is

$$\phi(N) \leq [L : \mathbf{Q}] r_b(N) \mu$$

where L is the compositum of the Galois closures of the K_i.

First consider the case where $\mathrm{End}_{\mathbf{Q}}(A)$ contains a quartic CM field K. By §5.1 of [2], A is isogenous to $B \times B$, and $\mathrm{End}_{\mathbf{Q}}(A) = M_2(\mathrm{End}_{\mathbf{Q}}(B))$, where B is an elliptic curve with CM by an imaginary quadratic field L. L is Galois over $\mathbf{Q}$, so $b = [L : \mathbf{Q}]/2 = 1$, and $r_1(N)$ is identically 1. The number of roots of unity, μ, in the field L, is either 2, 4 or 6. We look for N with $\phi(N) \leq 2 \cdot 1 \cdot \mu$. For $\mu = 6$, the maximum such N is 42; for $\mu = 4$, 30; and for $\mu = 2$, 12. If K happens to be a biquadratic field $\mathbf{Q}(\sqrt{-D_1}, \sqrt{-D_2})$, then L is one of the two imaginary quadratic subfields. For the two cyclotomic fields $\mathbf{Q}(\zeta_8)$ and $\mathbf{Q}(\zeta_{12})$ in particular, we obtain bounds for N of 30 and 42 respectively.

We are left with the case where A is isogenous to a product of two elliptic curves, both with CM. So we check $\phi(N) \leq [L : \mathbf{Q}] r_b(N) \mu$. If the two fields K_1 and K_2 are distinct, then $[L : \mathbf{Q}] = 4$ and $b = 2$, but if the fields are the same, then $[L : \mathbf{Q}] = 2$ and $b = 1$. The various possibilities are listed in Table 2. □

Remark 1. It is interesting to note that for an abelian variety isogenous to a product of two elliptic curves with different CM fields the upper bound for the order of a torsion point is 28560, but if the two CM fields are the same then the upper bound is only 42. Also note that the numbers in Table 2 are rather large compared to those in Table 1, especially considering the known bounds for rational torsion points on elliptic curves. Whether the differences in size are due to A not being simple or due to the method remains to be determined.

Remark 2. For computational purposes, it is useful to note that $r_b(p^t) = \gcd(b, \phi(p^t))$ for p an odd prime, and $r_b(2^t) = \gcd(b, 2)\gcd(b, 2^{t-2})$ if $t \geq 2$, $r_b(2) = 1$, and that $\phi(N)/r_b(N)$ is a multiplicative function of N. Since $\phi(N)$ grows much faster than $r_b(N)$, the possibilities for N quickly run out.

3. Three-dimensional Abelian Varieties

If the abelian variety is three-dimensional, then the number of roots of unity in Z is potentially larger than can be obtained for two-dimensional abelian varieties. This allows the the possibility of a higher order torsion point. Applying Corollary 6 of [3] gives

PROPOSITION 3.1. *Suppose A is a three-dimensional abelian variety of CM-type defined over $\mathbf{Q}$, and that N is the order of a torsion point of $A(\mathbf{Q})$. Then:*

(i) $\phi(N) \leq 2^{8+2\nu(N)} \cdot 3^4$.

(ii) $\phi(N) \leq 2^{7+2\nu(N)} \cdot 3^4$ *if* $8 \nmid N$.

(iii) $\nu(N) \leq 11$.

(iv) $N \leq 2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 19 \cdot 23 \cdot 29 \cdot 43 = 278196808890$.

(v) *If N is prime then* $N \leq 41467$.

Here again, the bounds for simple varieties are a lot lower. The number of roots of unity plays a significant role in allowing a point of high order. In particular with $\mathbf{Q}(\zeta_{14})$, the upper bound on the order of a torsion point of prime order is 241, and when the field is $\mathbf{Q}(\zeta_{18})$, the bound increases to 319. Both are much smaller than 41467.

PROPOSITION 3.2. *Suppose A is a simple three-dimensional abelian variety of CM-type defined over $\mathbf{Q}$, and N is the order of a torsion point of $A(\mathbf{Q})$. Then*

(i) $N \leq 31122$.

(ii) *If N is prime then* $N \leq 319$.

PROOF. $\mathrm{End}_{\mathbf{Q}}(A) = K$, for K some sextic CM field. If M is the Galois closure of K, then $\mathrm{Gal}(M/\mathbf{Q})$ is quite often an extension of $(\mathbf{Z}/2\mathbf{Z})^3$ by S_3, a group of order 48. Generally the CM type will be simple, since there is no *a priori* reason for all the CM types to be non-simple.

For any CM-field K, with reflex $\tilde{K}$, there is the inequality $b \leq 2^{d-1}$ (1.9 of [1]). Since the reflex of the reflex is the original field for a simple CM type, we can apply the inequality twice to obtain $d \leq 2^{b-1}$ or, equivalently, $b \geq \log_2(d)+1$. If $d = 3$, then we only have to check $b = 3, 4$.

So we consider the inequality

$$\phi(N) \leq [\tilde{K} : \mathbf{Q}] r_b(N) \mu$$

Once again treat each field separately, starting with the two sextic cyclotomic fields. Since K is Galois over $\mathbf{Q}$, the reflex is then a subfield of K, so b is at most 3. By the remark above it must be equal to 3. Alternatively, K is abelian over $\mathbf{Q}$ so every subfield is a normal extension of $\mathbf{Q}$, and hence the reflex field must equal K. The inequality becomes $\phi(N) \leq 6 \cdot r_3(N) \cdot \mu$.

If $K = \mathbf{Q}(\zeta_{18})$, then $\mu = 18$ and one can check that the upper bound for N is 31122. The biggest prime that could possibly divide N is 319.

Repeat for $K = \mathbf{Q}(\zeta_{14})$ to obtain 8190 as a upper bound for composite N, and 241 for prime N.

The possible prime divisors of N come in two groups, depending on the congruence class modulo 3. If $p \equiv 1 \pmod 3$, then $r_3(p) = 3$ and much larger primes are allowed than if $p \not\equiv 1 \pmod 3$, when $r_3(p) = 1$.

If K is any other CM field, then we know that $b = 3$ or 4 and checking the various possibilities leads to the information in Table 3.

In the last three lines of Table 3 it is necessary to check both $b = 3$ and $b = 4$. This check is required because $b = 3$ imposes stronger restrictions on primes congruent to 2 (mod 3) than it does on the other primes, and $b = 4$ imposes restrictions according to whether or not $p \equiv 1 \pmod 4$. □

Remark 3. If Z is a sextic field and A is isogenous to a product $B \times B \times B$, then the full endomorphism algebra for A is $M_3(L)$, where L is a quadratic imaginary subfield of Z. Then the bounds listed in the first three lines of Table 2 for $M_2(L)$, also apply for $M_3(L)$, since b and μ depend only on the subfield L. For instance, $\mathbf{Q}(\zeta_{14})$ contains $\mathbf{Q}(\sqrt{-7})$ and $\mathbf{Q}(\zeta_{18})$ contains $\mathbf{Q}(\sqrt{-3})$.

Remark 4. The methods used here can be modified to calculate bounds on the order of torsion points on abelian varieties defined over an arbitrary number field. If k is the field of definition of A, then the only modification needed is apply the following inequality

$$\phi(N) \leq [k : \mathbf{Q}][\tilde{K} : \mathbf{Q}] r_b(N) \mu$$

This differs from our original inequality in that there is an extra factor of $[k : \mathbf{Q}]$ on the right hand side.

Remark 5. The method also applies to higher values of d. In this remark and the previous one, the only limiting factor is computational power.

4. Tables

TABLE 1. Upper bounds on orders of torsion points on simple 2-dimensional abelian varieties of CM-type defined over $\mathbf{Q}$.

Z	μ	b	Upper Bound for Prime N	Upper Bound for N
$K \supset \mathbf{Q}(\sqrt{-D})$	2	2	17	840
$\mathbf{Q}(\zeta_{10})$	10	2	79	10920

TABLE 2. Upper bounds on orders of torsion points on non-simple 2-dimensional abelian varieties of CM-type defined over $\mathbf{Q}$.

Z	μ	b	Upper Bound for Prime N	Upper Bound for N
$M_2(\mathbf{Q}(\sqrt{-D}))$	2	1	5	12
$M_2(\mathbf{Q}(\sqrt{-1}))$	4	1	7	30
$M_2(\mathbf{Q}(\sqrt{-3}))$	6	1	13	42
$\mathbf{Q}(\sqrt{-D_1}) \times \mathbf{Q}(\sqrt{-D_2})$	4	2	31	2040
$\mathbf{Q}(\sqrt{-1}) \times \mathbf{Q}(\sqrt{-D_2})$	8	2	61	9240
$\mathbf{Q}(\sqrt{-3}) \times \mathbf{Q}(\sqrt{-D_2})$	12	2	97	14280
$\mathbf{Q}(\sqrt{-1}) \times \mathbf{Q}(\sqrt{-3})$	24	2	193	28560

TABLE 3. Upper bounds on orders of torsion points on simple 3-dimensional abelian varieties of CM-type defined over $\mathbf{Q}$.

Z	μ	b	Upper Bound for Prime N	Upper Bound for N
$\mathbf{Q}(\zeta_{18})$	18	3	319	31122
$\mathbf{Q}(\zeta_{14})$	14	3	241	8190
$K \supset \mathbf{Q}(\zeta_6)$	6	3 or 4	193	371280
$K \supset \mathbf{Q}(\zeta_4)$	4	3 or 4	113	127920
K	2	3 or 4	61	53040

REFERENCES

1. G. Shimura, *Canonical Models of Arithmetic Quotients of Bounded Symmetric Domains*, Ann. Math. **94** (1970), 144–222.
2. G. Shimura and Y. Taniyama, *Complex multiplication of abelian varieties and its applications to number theory*, Math. Soc. Japan, Tokyo (1961).
3. A. Silverberg, *Torsion points on abelian varieties of CM-type*, Comp. Math. **68** (1988), 241–249.
4. A. Silverberg, *Points of Finite Order on Abelian Varieties*, this volume.

DEPARTMENT OF MATHEMATICS, WELLESLEY COLLEGE, WELLESLEY, MASSACHUSETTS 02181
E-mail address: pvanmulbregt@lucy.wellesley.edu

Contemporary Mathematics
Volume 133, 1992

Prime-like Subsets of a Commutative Ring

MARIE A. VITULLI

ABSTRACT. In this note we study certain prime-like subsets of a ring called weak CMC subsets. The prototype for a weak CMC subset is the unit interval with respect to an ordering on a field.

Introduction

In [5] the author and D.K. Harrison introduced a valuation theory for commutative rings that, in turn, gave rise to a prime-like subset of a ring. These subsets were called CMC subsets because their complements are multiplictively closed in the containing ring. The V-valuation theory of Harrison and Vitulli extends the Artin notion of a generalized valuation on a field as follows.

In [1] E. Artin defined a valuation on a field F as a map v from F into the nonnegative reals satisfying three properties: $v(x) = 0$ if and only if $x = 0$; $v(xy) = v(x)v(y)$; and, $v(x+y) \leq c \cdot max\{v(x), v(y)\}$ for some positive constant c. The possible constants satisfying the third condition determine whether or not the valuation is Archimedean; the Archimedean valuations give rise to the classical infinite primes of the field whereas the non-Archimedean or Krull valuations correspond to the valuation subrings of the field. The third condition in the Artin definition of a valuation v can be reinterpreted as follows. Consider the set $A = \{x \in F | v(x) \leq 1\}$. Then, there exists a unit e in F such that for all x, y in A, $e(x+y) \in A$. Observe that since v has values in the nonnegative reals, for all z in $F \setminus A$, there exists a positive integer n such that ez^n is in $F \setminus A$.

The V-valuations of Harrison and Vitulli take values in an additive V-monoid, which is a generalization of the notion of an extended totally ordered abelian group. To a V-valuation v on a ring R one associates the CMC subset $A = \{x \in R | v(x) \geq 0\}$. A V-valuation admits an "exponent" e which is a unit in the ring R and has the two above-noted properties with respect to A. The

1991 *Mathematics Subject Classification.* Primary 13A18, 13F30 Secondary 11N80.
This paper is in final form and no version of it will be submitted for publication elsewhere

0271-4132/92 $1.00 + $.25 per page

possible values of e determine whether or not the CMC subset is a subring. The extended absolute values introduced by R. Brown in [2] correspond to nonring CMC subsets of a field. The second technical requirement on the exponent, that for all z in $R \setminus A$, there exists a positive integer n such that ez^n is in $R \setminus A$, was needed to prove that the nonring CMC subsets of an algebraic number field are precisely the infinite primes of the number field (see [**5, Theorems 3.26, 5.3, and 5.6**]). It is known that every nonring CMC subset of a field arises as the inverse image of the complex unit disc under a complex-valued place on the field (see [**6, Theorem 5**]). In particular, only fields of characteristic zero admit nonring CMC subsets.

The rich interplay between the theories of valuations and orderings of field motivates our interest in the connections between these theories in the commutative ring setting. Classically, given an ordering on a field, one associates a valuation ring by considering all elements in the field that are bounded by some integer with respect to the ordering. The residue class field of the valuation ring inherits an Archimedean ordering, so there is a real-valued place on the field whose domain is the valuation ring. In turn, the inverse image of the real unit interval under this place is a nonring CMC subset of the field. One can take a more direct approach in associating a nonring CMC subset to an ordering. With this end in mind, we first consider the unit interval with respect to the ordering. It satisfies all the defining properties of a CMC subset except the second technical condition on the exponent; such is called a weak CMC subset. We then obtain a nonring CMC subset and a CMC subring by enlarging the weak CMC subset in a manner that is independent of the original ordering. We can then recover the real-valued place that is classically associated with the ordering. We give an example (see Example 3) that shows that the unit interval with respect to an ordering may fail to be a CMC subset. For further connections between orderings of a commutative ring and the V-valuation theory the reader should consult [**8**]. In related work, R. Brown has studied the extended absolute values that arise from orderings of fields (see [**3**] and [**4**]).

We are thus interested in the structure of the weak CMC subsets of an arbitrary ring. We would like to point out that if one deletes the analog of the second condition on the exponent in the definition of a V-valuation ([**5, Definition 2.10**]) and defines isomorphism of these weak V-valuations as in [**5**], one again gets a bijective correspondence between the set of weak CMC subsets of a ring R and the class of all isomorphism classes of weak V-valuations on R. We will see that nonring weak CMC subsets can occur in fields of positive characteristic (Example 15).

In this note we characterize the weak CMC subsets of an arbitrary ring. We show that for every proper weak CMC subset A of a ring R there is a proper CMC subset $(D, P(D))$ of R, where $P(D) = \{d \in R | \exists x \in R \setminus D, dx \in A\}$, and a submonoid V of the multiplicative monoid $U := D \setminus P(D)$ with $(U \setminus V) \cdot (U \setminus V) \subseteq U \setminus V$ and such that $A = P(D) \cup V$. Conversely, one can use this decomposition

to construct weak CMC subsets (see Theorem 12).

By "ring" we mean a commutative ring with identity. By "ring homomorphism" we mean a map which preserves addition, multiplication, and the identity. By a "unit" in a ring R we mean an element that has a multiplicative inverse. We write $X \subset Y$ to denote that X is a proper subset of Y and let $X \setminus Y$ denote $\{x \in X | x \notin Y\}$. For subsets X, Y of a ring F we let $X + Y$ denote $\{x + y | x \in X, y \in Y\}$ and $X \cdot Y = \{xy | x \in X, y \in Y\}$. If $X = \{x\}$, we write xY in lieu of $\{x\} \cdot Y$. We frequently identity $\mathbb{Z}$ with its image in R.

Main Results

We now recall the notion of a CMC subset of a ring. For additional details the reader should consult [5].

DEFINITIONS 1. We say a subset A of a ring R is a *CMC subset* of R if:

(1) $A \cdot A \subseteq A, (R \setminus A) \cdot (R \setminus A) \subseteq (R \setminus A), 0, 1 \in A$; and

(2) There exists a unit e in R, called an *exponent* for A, such that:

(i) $\forall a, b \in A, e(a + b) \in A$, and

(ii) $\forall s \in R \setminus A, \exists n \in \mathbb{N}$ such that $es^n \notin A$.

If $e = 1$ is an exponent for A we say that A is a *CMC subring* of R. Otherwise we say that A is a *nonring CMC subset* of R. If all but the last technical condition on the exponent hold, we say that A is a *weak CMC subset* of R and that e is a *weak exponent* for A. If $e = 1$ is not a weak exponent for the weak CMC subset we call it a *nonring weak CMC subset*.

We point out that $e = 1$ is a (weak) exponent for the (weak) CMC subset A of R if and only if A is a subring of R. For a field R the CMC subrings are precisely the valuation subrings of R. A weak CMC subset of R with weak exponent $e = 1$ is precisely a CMC subring of R. Our primary focus in this note is on the nonring weak CMC subsets of a ring.

EXAMPLE 2. Let $\phi : B \to \mathbb{C}$ be a place on the field F, i.e., B is a valuation subring of F, ϕ is a ring homomorphism and $\phi(B)$ is a subfield of $\mathbb{C}$. Define A_ϕ by

$$A_\phi = \{x \in F | \, |\phi(x)| \leq 1\}.$$

One checks that A_ϕ is a nonring CMC subset of F with exponent $\frac{1}{2}$. In fact, every nonring CMC subset of a field F is of the form A_ϕ for some complex-valued place ϕ on F and ϕ is essentially unique (see [**6, Theorem 5**]). In particular, a CMC subset A of a field F is a valuation subring if and only if $2 \in A$. Every nonring CMC subset of a field F has exponent $\frac{1}{2}$. Furthermore, for A a nonring CMC subset of a field F and an element e of A, e is an exponent for A if and only if $2e \in A$ and $e \notin \bigcap_{n \in \mathbb{N}} (\frac{1}{2})^n A$ [**6, Corollary 9**].

Similarly, if A is a nonring CMC subset of a ring R there exists a Connell preplace $\phi : B \to \mathbb{C}$ (i.e., B is a CMC subring of R, ϕ is a ring homomorphism,

and there is a containment $\{b \in B | \exists x \in R \setminus B, bx \in B\} \subseteq ker\phi$ such that $A = A_\phi$ [**7, Theorem 2.7**]. Thus a CMC subset A of a ring R is a subring if and only if $2 \in A$. For a nonring CMC subset A of a ring R with exponent e, a unit f of R is an exponent for A if and only if $2f \in A$ and $f \notin \bigcap_{n\in\mathbb{N}} e^n A$ [**7, Corollary 2.8**].

EXAMPLE 3. Consider the ordering T_0 on $\mathbb{Q}[x]$ defined by

$$T_0 = \{a_m x^m + \cdots + a_n x^n | m \leq n \text{ and } a_m > 0\} \cup \{0\},$$

and let T denote the induced ordering on the field of rational functions $\mathbb{Q}(x)$. Define A by

$$A = \{\rho \in \mathbb{Q}(x) | 1 + \rho \in T \text{ and } 1 - \rho \in T\}$$

and notice that A is the unit interval with respect to T. One checks that A is a weak CMC subset of $\mathbb{Q}(x)$ with weak exponent $\frac{1}{2}$. Since $1 + x \notin A$ and $\frac{1}{2}(1 + x)^n \in A$ for all $n \geq 1$, A is not a CMC subset of $\mathbb{Q}(x)$.

NOTATION 4. For a weak CMC subset A of a ring R with weak exponent e and subsets X and Y of R let

$$(X : Y) = \{x \in R | xY \subseteq X\},$$

$$P(A) = \{x \in R | \exists y \in R \setminus A, xy \in A\} \cup \{0\},$$

$$U(A) = \{x \in A | \exists y \in A, xy = 1\},$$

$$A(e) = \{x \in A | \forall n \in \mathbb{N}, ex^n \in A\},$$

$$Q(e) = \bigcap_{n\in\mathbb{N}} e^n A,$$

$$C(e) = \bigcup_{n\in\mathbb{N}} e^{-n} A, \text{ and}$$

$$B(e) = (Q(e) : Q(e)).$$

Note that for a proper weak CMC subset A of a ring R the set $R \setminus P(A)$ is multiplicatively closed. Furthermore, if A is a nonring weak CMC subset of R with weak exponent e, then $R \setminus Q(e)$ is multiplicatively closed. The proof of the next result follows easily from the proof of Proposition 2.4 of [7] .

PROPOSITION 5. *Let A be a nonring weak CMC subset of R with weak exponent e. Then:*

(1) *$C(e) \subseteq B(e)$ are CMC subrings of R and $Q(e)=(A:C(e))=(A:B(e))$ is a prime ideal of both $C(e)$ and $B(e)$;*
(2) *$A(e)$ is a proper CMC subset of R with exponent e and $A(e) \subset C(e)$;*
(3) *If e and f are weak exponents for A such that $A(e)$ and $A(f)$ are nonring CMC subsets of R, then $Q(e)=Q(f)$, $A(e)=A(f)$, $C(e)=C(f)$, and $B(e)=B(f)$;*

(4) $P(C(e)) + P(B(e)) \subseteq Q(e) \subset P(A(e)) \subseteq P(A)$. *If F is a field, then $C(e)=B(e)$ is a valuation subring of F and $P(C(e))=Q(e)$ is its unique maximal ideal;*

(5) $A(e) + Q(e) = A(e)$ *and* $A(e)/Q(e) := \{a + Q(e) | a \in A(e)\}$ *is a proper CMC subset of the quotient ring $C(e)/Q(e)$.*

REMARK 6. Example 15 below demonstrates that without the assumption that $A(e)$ is a nonring CMC subset of R, the ideal $Q(e)$ may vary with e. For weak CMC subsets A, A' of a field F, $A \subseteq A' \Leftrightarrow P(A') \subseteq P(A)$. We also have $A = (P(A) : P(A))$. These remarks follow from observing that in the field case $P(A) = \{0\} \cup \{x \in F \setminus \{0\} | x^{-1} \notin A\}$. One further checks that the weak CMC subsets of a field F containing a fixed weak CMC subset A are linearly ordered by inclusion (see [**6, Remark 11**]).

DEFINITION/NOTATION 7. Let e be a weak exponent for a nonring weak CMC subset A of the ring R. We say that e is a *special weak exponent* if $A(e)$ is not a ring and that A is a *special weak CMC subset* if A admits a special weak exponent. If A is a special weak CMC subset with special weak exponent e, we let A^* denote $A(e)$ and write $C(A)$ and $B(A)$ in lieu of $C(e)$ and $B(e)$.

PROPOSITION 8. *Let A be a special weak CMC subset of a ring R with special weak exponent e.*

(1) *A^* is minimal among the nonring CMC subsets of R that contain A;*

(2) *A unit f of R is a special weak exponent for A if and only if $f \notin Q(A)$;*

(3) *Let S be a subring of R that contains A. Then, A is a weak CMC subset of S if and only if $e^{-1} \in A$;*

(4) *$C(A)$ is the smallest subring of R that contains A as a special weak CMC subset;*

(5) *A^* is the unique nonring CMC subset D of R with $P(D) \subseteq A \subseteq D$.*

PROOF. (1) Let D be a nonring CMC subset of R with $A \subseteq D \subseteq A^*$. Since e is an exponent for A^* there exists $m \in \mathbb{N}$ such that e^m is an exponent for D [**7, Proposition 3.3**] . Suppose $x \in A^*$ and let $n \in \mathbb{N}$. Then,

$$e^m x^n = e^{m-1} e x^n \in A \cdot A \subseteq A \subseteq D.$$

Since e^m is an exponent for D this implies $x \in D$. Hence $D = A^*$.

(2) Suppose $f \in Q(A)$. We then have $f(A + A) \subseteq Q(A) + Q(A) = Q(A) \subseteq A$. So f is a weak exponent for A. Since $f2^n \in Q(A) \subseteq A$ for all $n \in \mathbb{N}$ we may deduce that $A(f)$ is a subring of R. Now assume that f is a nonspecial weak exponent for A, that is, $f2^n \in A$ for all $n \in \mathbb{N}$. Just suppose that $e^{-m} f \notin A$ for some $m \in \mathbb{N}$. Then, $e^m = fa$ with $a \in A$. Hence $e^m 2^{mn} \in A$, which implies $e2^n \in A$, for all $n \in \mathbb{N}$. This is the desired contradiction.

(3) Suppose A is a special weak CMC subset with special weak exponent f. Since f is a unit of S,

$$\{x \in S | \forall n \in \mathbb{N}, fx^n \in A\} = \{x \in R | \forall n \in \mathbb{N}, fx^n \in A\}.$$

So f is a special weak exponent for A viewed as a weak exponent CMC subset of the ring R and

$$\{x \in S | \forall n \in \mathbb{N}, fx^n \in A\} = \{x \in R | \forall n \in \mathbb{N}, fx^n \in A\} = A^*.$$

Since A^* is not a ring $f^{-1} \notin A^*$. Hence there exists an element $m \in \mathbb{N}$ with $ef^{-m} \notin A$. Then $e^{-1}f^m \in A \subseteq S$ and $e^{-1} \in S$, as asserted. The converse is immediate as $e^{-1} \in S$ implies e is a special weak exponent for the weak CMC subset A of the ring S.

(4) Suppose S is a subring of R that contains A as a special weak CMC subset. Then $e^{-1} \in S$ by part (3). One checks that this implies $C(A) \subseteq S$.

(5) Suppose D is a nonring CMC subset of R with $P(D) \subseteq A \subseteq D$. Let f be an exponent for D. Then,

$$f^2(A + A) \subseteq f \cdot f(D + D) \subseteq f \cdot D \subseteq P(D) \subseteq A,$$

which demonstrates that f^2 is a special weak exponent for A in R. So $A^* = A(f^2)$ and again using the assumption that $P(D) \subseteq A$, one checks that $D \subseteq A(f^2)$. By part (1), $D = A(f^2) = A^*$.

COROLLARY 9. *Let F be a field and A be a special weak CMC subset of F. Then, $C(A)$ is the smallest valuation subring of F that contains A and A^* is the unique nonring CMC subset of F that contains A.*

PROOF. Recall that if $A \subset D$ are weak CMC subsets of a field, then we must have $P(D) \subset P(A)$. The assertions now easily follow from parts (4) and (5) of Proposition 8.

COROLLARY 10. *Let $A \subset A'$ be nonring CMC subsets of a ring R. Then, $P(A') \nsubseteq P(A)$. In particular, there are no containment relations among the nonring CMC subsets of a field.*

PROOF. Just suppose $P(A') \subseteq P(A)$. Viewing A as a special weak CMC subset of R, we must have $A' = A^*$ by Proposition 8. But $A^* = A$ since A is already a CMC subset of R, a contradiction. The second assertion now follows from Remark 6.

We point out that it is possible to have containments among the nonring CMC subsets of a ring [**7, Example 3.2**]. Another immediate consequence of the last assertion of Proposition 8 is the following.

COROLLARY 11. *If A is a nonring CMC subset of a ring R, then every special weak exponent for A is an exponent for A.*

We will now characterize the weak CMC subsets of an arbitrary ring. First, notice that for a submonoid V of an abelian group U the complement of V in U is multiplicatively closed if and only if $\{x \in U | x^{-1} \notin V\} \subseteq V$. Also, recall that if D is a proper CMC subset of a ring R, then $D \setminus P(D)$ is a multiplicative submonoid of R.

THEOREM 12. *Let R be a ring.*

(1) *If A is a proper weak CMC subset of R with weak exponent e, then $D:=A(e)$ is a proper CMC subset of R and $V := A \setminus P(D)$ is a multiplicative submonoid of $D \setminus P(D)$ whose complement in $D \setminus P(D)$ is multiplicatively closed and $A = P(D) \cup V$.*

(2) *Conversely, suppose that D is a proper CMC subset of R and V a submonoid of $D \setminus P(D)$ with multiplicatively closed complement. Then $A := P(D) \cup V$ is a multiplicative submonoid of R and $R \setminus A$ is multiplicatively closed.*

(2)(i) *If D is a nonring CMC subset with exponent f, then A is a special weak CMC subset with special weak exponent f^2 and $A^* = D$.*

(2)(ii) *Now assume that D is a CMC subring of R and that e is in $P(D) \cap U(R)$. Then, A is a weak CMC subset of R with weak exponent e and $D \subseteq A(e)$. In particular, $A(e)$ is a ring.*

PROOF. (1) Since A and $R \setminus P(D)$ are submonoids of R whose complements are closed under multiplication, we have that $A \setminus P(D)$ is a submonoid of $D \setminus P(D)$ with multiplicatively closed complement.

(2) Since

$$P(D) \cdot A \subseteq P(D) \cdot D \subseteq P(D) \subseteq A \text{ and } V \cdot V \subseteq V \subseteq A,$$

A is multiplicatively closed. Suppose $x, y \in R \setminus A$. We distinguish 3 cases: both elements are in $R \setminus D$; one is in $R \setminus D$; and, both elements are in D. Suppose $x, y \in R \setminus D$. Then $xy \in R \setminus D \subseteq R \setminus A$. Suppose $x \in R \setminus D$ and $y \in D$. Since $x \notin D$ and $y \notin P(D)$, we may conclude that $xy \in R \setminus D \subseteq R \setminus A$. Suppose $x, y \in D$. Then $x, y \in D \setminus P(D)$ and $xy \in D \setminus P(D) \subseteq R \setminus A$. So A is a submonoid of R with $(R \setminus A) \cdot (R \setminus A) \subseteq R \setminus A$. Notice that $0 \in P(D) \subseteq A$.

(2)(i) Suppose D is a nonring CMC subset of R with exponent f. As in the proof of Proposition 8, part (5), A is a special weak CMC subset with special weak exponent f^2 and $D = A^*$.

(2)(ii) Now suppose D is a subring of R and that $e \in U(R) \cap P(D)$. In this case $P(D)$ is a prime ideal of D and, in particular, is closed under addition. Then,

$$e(A + A) \subseteq P(D) + P(D) = P(D) \subseteq A.$$

Hence e is a weak exponent for the weak CMC subset A of R. Let $x \in D$ and $n \in \mathbb{N}$. Then $ex^n \in eD \subseteq P(D) \subseteq A$. Hence $D \subseteq A(e)$.

Notice that if A is a nonring weak CMC subset of a ring R with weak exponent e such that $D := A(e)$ is a ring, then $e \in P(D) \cap U(R)$. So the condition in (2)(ii) above is a natural one. We point out that the weak CMC subset A of the ring R arising from the construction of part (2)(ii) above may well be a CMC subring of R. For example, if $A \subset D$ are valuation subrings of the field F, then $A = P(D) \cup V$ is a ring, where $V = A \setminus P(D)$.

COROLLARY 13. *Let A be a nonring CMC subset of a field F and let C be the intersection of all valuation subrings of F that contain A. Then C is a valuation subring of F. Furthermore, A is a special weak CMC subset of F if and only if A admits a weak exponent e that is a unit in C.*

PROOF. Remark 6 implies that C is a valuation subring of F and that $P(C) \subseteq A$. Notice that if $e \in U(C)$ is a weak exponent for A, then $A(e) \subset C$ and hence $A(e)$ is a nonring CMC subset of F. The latter assertion now follows from part (2)(ii) of Theorem 12.

COROLLARY 14. *Suppose that D is a proper valuation subring of a field F with unique maximal ideal P. Let V be a submonoid of $U(D)$ whose complement is multiplicatively closed, set $A = P \cup V$, and suppose that $e \in P \setminus \bigcap_{n\in\mathbb{N}} P^n$. Then, A is a weak CMC subset of F with weak exponent e and $D = A(e)$.*

PROOF. By part (2)(ii) of Theorem 12, $A(e)$ is a valuation subring of F that contains D. Just suppose that $D \subset A(e)$. Choose $x \in A(e) \setminus D$. Since $x \notin D$ we must have $x^{-1} \in P$. Thus $e = ex^n(x^{-1})^n \in A \cdot P^n = P^n$ for all $n \in \mathbb{N}$, a contradiction.

EXAMPLE 15. Let k be a field, x and y be indeterminates and let $F = k(x, y)$. Let $B = k[x, y]_{(x)} = k(y)[x]_{(x)}$, $M = xB$, and identify the residue field of B with $k(y)$. Write $\bar{b}$ for the image of b under the canonical homomorphism $\pi : B \to k(y)$. Let $D = \{b \in B | \bar{b} \in k[y]_{(y)}\}$. Then D is a discrete rank two valuation ring with unique maximal ideal yD. Let $V = \{c \in U(D) | v(\bar{c}) \leq 0\}$ where v is the valuation on $k(y)$ corresponding to the valuation ring $k[y]_{(y-1)}$.

Clearly V is a multiplicative submonoid of $U(D)$. If $c \in U(D) \setminus V$ then $v(\bar{c}) > 0$ so that $v(\bar{c}^{-1}) < 0$; this implies $c^{-1} \in V$. Thus $A := yD \cup V$ is a weak CMC subset of F. As $y - 2 \in A$, but $1 + (y - 2) = y - 1 \notin A$, A is not a ring. Any nonzero element of yD is a weak exponent for A. One checks that $Q(y) = \bigcap_{n\in\mathbb{N}} y^n D = xB$ is the unique height one prime ideal of D. Now $Q(x) = \bigcap_{n\in\mathbb{N}} x^n A \subseteq \bigcap_{n\in\mathbb{N}} x^n B = (0)$ as (B, xB) is a discrete rank one valuation ring.

We claim that $A(y) = D$ whereas $A(x) = B$. Now $A(y) = D$ follows from Theorem 12, part (2)(ii). Also, $A(x)$ is a proper valuation subring of F and $D \subseteq A(x)$ by Theorem 12, part (2)(ii), and Proposition 5, part (2). Since $y^{-1} \in A(x) \setminus A(y)$, we have $A(y) \subset A(x)$. Hence $A(x) = B$ as $A(y)$ has rank 2.

One checks that $2 \in A$ (so unlike with CMC subsets, a nonring weak CMC subset may contain 2). Hence $A(e)$ is a valuation subring of F for every weak exponent e for A. Thus A is a weak CMC subset that is not special.

DEFINITION 16. Let F be a field of finite transcendence degree d over its prime field K. We define the absolute dimension of F to be $d + \epsilon$ where $\epsilon = 0$ if K has positive characteristic and $\epsilon = 1$ is K has characteristic zero.

We have just seen that if e is a weak exponent for a nonspecial weak CMC subset A of a field F, the proper CMC subring $A(e)$ of F depends on the weak

exponent e for A. When F has absolute dimension one, $A(e)$ is independent of e as shown below. A field of positive characteristic has no nonring CMC subsets. A field F of positive characteristic and absolute dimension 0 has no proper CMC subsets since a CMC subset of F is integrally closed in F. So Proposition 5, part (2), implies that a field of positive characteristic admits a nonring weak CMC subset only if it has absolute dimension at least one.

COROLLARY 17. *Suppose F is a field of absolute dimension one and that A is a nonring weak CMC subset of F with weak exponent e. Then $Q(e) = \{0\}$ and $C(e) = F$. Furthermore, $A(e)$ is the unique proper CMC subset of F containing A.*

PROOF. By Proposition 5, $A(e) \subset C(e)$ are CMC subsets of F. To conclude that $C(e) = F$ and $Q(e) = \{0\}$ it suffices to show that there are no containments between the proper CMC subsets of F. Suppose $char(F)$ is positive. The proper CMC subsets of F are precisely the proper valuation subrings of F. Since F has transcendence degree one over its prime subfield K and every valuation subring of F contains K, every proper valuation subring of F has rank one [**9, Theorem 3, page 8**]. Hence there are no inclusions between the proper CMC subsets of F. Now assume that F has characteristic zero. Just suppose $B \subset C$ are proper CMC subsets of F. Choose $c \in C \setminus B$ and $x \in F \setminus C$. Let $K = \mathbb{Q}(c, x)$, $B' = B \cap K$, and $C' = C \cap K$. Then $B' \subset C'$ are proper CMC subsets of the finite-dimensional algebraic number field K by [**6, Corollary 12**], contradicting the fact that the CMC subsets of K correspond to the classical primes of K by [**5, Theorem 5.3**] and the proof of [**5, Theorem 5.6**]. The latter assertion follows from Remark 6 and the fact that there are no containments between the proper CMC subsets of F.

REFERENCES

1. E. Artin, *Theory of Algebraic Numbers*, Striker, Gottingen, Germany, 1959.
2. R. Brown, *An approximation theorem for extended prime spots*, Can. J. Math. **XXIV** (1972), 167–184.
3. R.Brown, *Automorphisms and isomorphisms of real henselian fields*, Trans. Amer. Math. Soc. **307** (1988), 675–703.
4. R. Brown, *Orderings and order closures of not necessarily formally real fields*, preprint.
5. D.K. Harrison and M.A. Vitulli, *V-valuations of a commutative ring I*, J. of Algebra **126** (1989), 264–292.
6. D.K. Harrison and M.A. Vitulli, *Complex-valued places and CMC subsets of a field*, Comm. in Alg. **17** (1989), 2529–2537.
7. K.G. Valente and M.A. Vitulli, *Complex-valued preplaces and the nonring CMC subsets of a ring*, Comm. in Alg. **18** (1990), 3743–3758.
8. K.G. Valente and M.A. Vitulli, *Orderings and prime-like subsets of a commutative ring*, J. of Pure and Appl. Algebra (to appear).
9. O. Zariski and P. Samuel, *Commutative Algebra, Volume II*, Van Nostrand, Princeton, NJ, 1960.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF OREGON, EUGENE, OR 97403

E-mail address: vitulli@bright.math.uoregon.edu

This page intentionally left blank

Contemporary Mathematics
Volume 133, 1992

Newton Polygons and Congruence Decompositions of L-Functions over Finite Fields

DAQING WAN

Abstract. In this paper, we study L-functions associated with exponential sums over a finite field. A congruence decomposition theorem is proved for such L-functions. The result is then used to study the conjecture of Adolphson and Sperber about generic Newton polygons of exponential sums.

1. Introduction

Let p be a prime, $q = p^a$, $\mathbf{F}_q$ the finite field of q elements, and $\mathbf{F}_{q^m}$ its extension of degree m. Fix a nontrivial additive character Ψ of $\mathbf{F}_p$. For any Laurent polynomial $f(x_1, \cdots, x_n) \in \mathbf{F}_q[x_1, x_1^{-1}, \cdots, x_n, x_n^{-1}]$, we form the exponential sum

$$S_m^*(f) = \sum \Psi(\mathrm{Tr}_{\mathbf{F}_{q^m}/\mathbf{F}_p}(f(x_1, \cdots, x_n))), \tag{1.1}$$

where the sum is over all $(x_1, \cdots, x_n) \in (\mathbf{F}_{q^m}^*)^n$. The associated L-function is

$$L^*(f, t) = \exp(\sum_{m=1}^{\infty} S_m^*(f)\frac{t^m}{m}). \tag{1.2}$$

1991 Mathematics Subject Classification. Primary 11L03; Secondary 11L40 and 11L99.

This paper is in final form and no version of it will be submitted for publication elsewhere.

The Laurent polynomial f is a sum of monomials and as such has a well defined Newton polyhedron $\Delta(f)$ at infinity. This is the convex closure in $\mathbf{R}^n$ of the lattice points which occur as exponents of the terms of f together with the origin. Assuming that f is nondegenerate (see [2]), Adolphson and Sperber [2] proved that $L^*(f,t)^{(-1)^{n-1}}$ is a polynomial of a certain degree, depending only on the polyhedron $\Delta(f)$. Furthermore [2], the Newton polygon of $L^*(f,t)^{(-1)^{n-1}}$ lies above a certain explicit lower bound $P(\Delta)$ depending only on $\Delta(f)$. Based on earlier results of Sperber [6-8] and Carpentier [4], they conjectured [2] that if $p \equiv 1 (\mathrm{mod}\ D)$ for a certain integer D determined by Δ and if p is a large prime, then the Newton polygon of $L^*(f,t)^{(-1)^{n-1}}$ coincides generically with the lower bound $P(\Delta)$. For simplicity, this conjecture will be referred to the AS conjecture in the following.

Fix an n-dimensional Newton polyhedron Δ in $\mathbf{R}^n$. Let $M_p(\Delta)$ be the moduli space of all Laurent polynomials over $\bar{\mathbf{F}}_q$ with a fixed Δ. It is clear that $M_p(\Delta)$ is isomorphic to $(\bar{\mathbf{F}}_q^*)^a \times (\bar{\mathbf{F}}_q)^b$ for certain positive integers a and b. Hence, $M_p(\Delta)$ is an irreducible quasi-projective variety. Let $H_p^*(\Delta)$ be the moduli space of those $f \in M_p(\Delta)$ such that f is nondegenerate with respect to Δ and the Newton polygon of $L^*(f,t)^{(-1)^{n-1}}$ coincides with the given lower bound. The AS conjecture states that $H_p^*(\Delta)$ contains a dense Zariski open subset of $M_p(\Delta)$ if p lies in the residue class $p \equiv 1 (\mathrm{mod}\ D)$. In [9], we proved that $H_p^*(\Delta)$ is a Zariski open subset of $M_p(\Delta)$ (possibly empty). Thus, the AS conjecture is reduced to finding a single example (for each Δ) with desired Newton polygon. If Δ is a simplex, such examples can be constructed; see [9] for a special simplex Δ and see §5 of the present paper for a general simplex Δ.

Our main result here is to reduce the AS conjecture from a general Δ to a simpler Δ by decompositions of Δ. To describe this more precisely, let σ be an $(n-1)$-dimensional closed face of Δ not containing the origin, and let f_σ be the restriction of f to the closed cone generated by σ and the origin. If f is nondegenerate, by the definition of nondegenaracy, all f_σ are also nondegenerate. A theorem of Adolphson and Sperber shows that $L^*(f_\sigma,t)^{(-1)^{n-1}}$ is a polynomial, whose Newton polygon lies above its lower bound $P(\sigma)$. As a consequence of our congruence decomposition theorem, we have

Theorem 1.1. Let f be nondegenerate. Then the Newton polygon of $L^*(f,t)^{(-1)^{n-1}}$ coincides with its lower bound $P(\Delta)$ if and only if the Newton

polygon of $L^*(f_\sigma, t)^{(-1)^{n-1}}$ coincides with its lower bound $P(\sigma)$ for all $(n-1)$-dimenisonal closed faces σ not containing the origin.

If Δ is a simplex, we will construct examples of f with $\Delta(f) = \Delta$ which have the property that the Newton polygon of $L^*(f,t)^{(-1)^{n-1}}$ coincides with its lower bound for all large primes p in the residue class $p \equiv 1 (\mathrm{mod}\ D^*)$, where D^* is a certain integer determined by Δ (D^* may be larger than the D in the AS conjecture). This shows that $H_p^*(\Delta)$ is open and dense for such p. Combining with Theorem 1.1, we deduce

Theorem 1.2. Assume that all $(n-1)$-dimensional closed faces of Δ are simplices (this is always true if $n \leq 2$). Then there is an integer D^* determined by Δ such that if p is a large prime and $p \equiv 1 (\mathrm{mod}\ D^*)$, then $H_p^*(\Delta)$ is open and dense.

Theorem 1.2 shows that a sightly weaker form of the AS conjecture is true for $n \leq 2$. In [10], we will prove that a similar weaker form of the AS conjecture is true for all n, however, the full form of the conjecture is false. Theorem 1.2 also holds for a more general L-function $L(f,t)$ studied by Adolphson and Sperber [2], see §4 for more details. Adolphson and Sperber [3] have recently extended their conjecture to even more general twisted exponential sums. We expect that our results also generalize to these twisted sums.

Acknowledgment. I wish to thank Alan Adolphson, Neal Koblitz and the referee for many helpful comments and corrections.

2. Dwork's p-Adic Theory

In this section, we shall follow Adolphson and Sperber [1] [2] (with some paraphrasing) to present their version of Dwork's p-adic theory describing L-functions in terms of the Fredholm determinant of a certain nuclear infinite matrix.

Let $\mathbf{Q}_p$ be the field of p-adic numbers, and let Ω be the completion of an algebraic closure of $\mathbf{Q}_p$. Let K denote the unramified extension of $\mathbf{Q}_p$ in Ω of degree a, where $q = p^a$. Let $\pi \in \Omega$ satisfy $\pi^{p-1} = -p$. Then $\Omega_1 = \mathbf{Q}_p(\pi)$ is a totally ramified extension of $\mathbf{Q}_p$ of degree $p-1$, in fact, $\Omega_1 = \mathbf{Q}_p(\zeta_p)$, where ζ_p is a primitive p-th root of unity. Let Ω_a be the compositum of Ω_1 and K. Then Ω_a is an unramified extension of Ω_1 of degree a. The residue fields of Ω_a and K

are both $\mathbf{F}_q$, and the residue fields of Ω_1 and $\mathbf{Q}_p$ are both $\mathbf{F}_p$. The Frobenius automorphism $x \mapsto x^p$ of $\mathrm{Gal}(\mathbf{F}_q/\mathbf{F}_p)$ lifts to a generator τ of $\mathrm{Gal}(K/\mathbf{Q}_p)$ which is extended to Ω_a by requiring that $\tau(\pi) = \pi$. If ζ is a $(q-1)$-st root of unity in Ω_a, then $\tau(\zeta) = \zeta^p$. Denote by "ord" the additive valuation on Ω normalized by $\mathrm{ord}p{=}1$, and denote by "ord_q" the additive valuation on Ω normalized by $\mathrm{ord}_q q{=}1$.

Let $E(t)$ be the Artin-Hasse exponential series:

$$E(t) = \exp(\sum_{m=0}^{\infty} \frac{t^{p^m}}{p^m}) \in (\mathbf{Z}_p \cap \mathbf{Q})[[t]]. \tag{2.1}$$

Let $\gamma \in \Omega_1$ be a root of $\sum_{m=0}^{\infty} \frac{t^{p^m}}{p^m} = 0$ satisfying $\mathrm{ord}\gamma = \mathrm{ord}\pi = \frac{1}{p-1}$. The series

$$\theta(t) = E(\gamma t) = \sum_{m=0}^{\infty} \lambda_m t^m \tag{2.2}$$

is a splitting function in Dwork's terminology and its coefficients satisfy

$$\mathrm{ord}\lambda_m \geq \frac{m}{p-1}, \quad \lambda_m \in \Omega_1, \tag{2.3}$$

$$\mathrm{ord}\lambda_m = \frac{m}{p-1} \quad \text{for } 0 \leq m \leq p-1. \tag{2.4}$$

Write $j = (j_1, \cdots, j_n) \in \mathbf{Z}^n$, $x = (x_1, \cdots, x_n)$, $x^j = x_1^{j_1} \cdots x_n^{j_n}$. For a Laurent polynomial $f(x_1, \cdots, x_n) \in \mathbf{F}_q[x_1, x_1^{-1}, \cdots, x_n, x_n^{-1}]$, write

$$f = \sum_{j \in J} \bar{a}_j x^j, \tag{2.5}$$

where the sum is over a finite subset J of $\mathbf{Z}^n$ and the $\bar{a}_j$'s are non-zero elements of $\mathbf{F}_q$. Let a_j be the Teichmuller lifting of $\bar{a}_j$ in Ω, i.e., $a_j^q = a_j$. Set

$$F(f, x) = \prod_{j \in J} \theta(a_j x^j) \tag{2.6}$$

$$F_a(f, x) = \prod_{i=0}^{a-1} F^{\tau^i}(f, x^{p^i}). \tag{2.7}$$

Note that (2.3) implies that $F(f, x)$ and $F_a(f, x)$ are well defined as formal Laurent series in $x_1, \cdots, x_n$ with cofficients in Ω_a.

To describe the growth conditions satisfied by F, write

$$F(f,x) = \sum_{r\in\mathbf{Z}^n} F_r(f)x^r. \tag{2.8}$$

Let A be the $(n\times \mid J \mid)$-matrix whose columns are the $j = (j_1,\cdots,j_n) \in J$. Then from (2.2) and (2.6),

$$F_r(f) = \sum_u (\prod_{j\in J} \lambda_{u_j} a_j^{u_j}), \tag{2.9}$$

where the outer sum is over all $\mid J \mid$-tuples $u = (u_j)_{j\in J}$ of nonnegative integers such that (thinking of u and r as column vectors)

$$Au = r. \tag{2.10}$$

Thus $F_r = 0$ if (2.10) has no such solutions. Otherwise, (2.3) implies

$$\mathrm{ord}F_r(f) \geq \frac{1}{p-1}\inf_u\{\sum_{j\in J} u_j\}, \tag{2.11}$$

where the inf is taken over all such solutions.

Adolphson and Sperber [1] defined a weight function $w(r)$ by

$$w(r) = \inf_u\{\sum_{j\in J} u_j\} \tag{2.12}$$

where now the inf is taken over all $\mid J \mid$-tuples $u = (u_j)_{j\in J}$ of nonnegative rational solutions. Then (2.11) implies

$$\mathrm{ord}F_r \geq \frac{w(r)}{p-1}, \tag{2.13}$$

with the obvious convention that $F_r = 0$ if $w(r) = +\infty$.

The weight function can be described geometrically. Let $\Delta(f)$ be the convex hull in $\mathbf{R}^n$ of the set $J\cup\{(0,\cdots,0)\}$. It is called the Newton polyhedron of f. Let $\mathbf{R}_+\langle\mathbf{J}\rangle$ denote the subset of $\mathbf{R}^n$ consisting of all linear combinations of elements of J with nonnegative real coefficients. Then $\mathbf{R}_+\langle\mathbf{J}\rangle$ is the union of all rays emanating from the origin and passing through $\Delta(f)$. Equation (2.10) has a solution u whose components are nonnegative rational numbers if and only

if $r \in \mathbf{R}_+\langle \mathbf{J} \rangle$. Thus, $w(r) = +\infty$ if and only if $r \notin \mathbf{R}_+\langle \mathbf{J} \rangle$. If $r \in \mathbf{R}_+\langle \mathbf{J} \rangle$, the ray emanating from the origin and passing through r intersects $\Delta(f)$ in a face that does not contain the origin. Let $\sum_{i=1}^{n} \alpha_i X_i = 1$ be the equation of a hyperplane passing through this face (this hyperplane is not uniquely determined unless the face has dimension $n-1$). Then by standard arguments in linear programming,

$$w(r) = \sum_{i=1}^{n} \alpha_i r_i, \tag{2.14}$$

Let D be the least common denominator of all the α_i's over all the faces of $\Delta(f)$, then

$$w(\mathbf{Z}^n) \subseteq \frac{1}{D}\mathbf{Z}_{\geq 0} \cup \{+\infty\} \tag{2.15}$$

where $\mathbf{Z}_{\geq 0}$ denotes the set of nonnegative integers. From the above description of $w(r)$, it is immediately seen that

Lemma 2.1. (a) For any nonnegative integer k, $w(kr) = kw(r)$.

(b) $w(r + r') \leq w(r) + w(r')$, with equality holding if and only if r and r' are cofacial, i.e., $w(r)^{-1}r$ and $w(r')^{-1}r'$ lie on the same closed face of Δ.

Let $A_1(f) = (a_{s,r}(f))$ be the infinite matrix defined by

$$a_{s,r}(f) = F_{ps-r}(f)\pi^{(w(r)-w(s))}. \tag{2.16}$$

Let $A_a(f)$ be the infinite matrix defined by

$$A_a(f) = \prod_{i=0}^{a-1} A_1(f)^{\tau^i} = A_1 A_1^{\tau^1} \cdots A_1^{\tau^{a-1}}. \tag{2.17}$$

By (2.13), (2.16) and Lemma 2.1, we obtain the estimate

$$\mathrm{ord} a_{s,r} \geq \frac{w(ps-r) + w(r) - w(s)}{p-1} \geq w(s). \tag{2.18}$$

Thus, $A_1(f)$ and $A_a(f)$ are nuclear matrices. They all have well defined traces. In addition, the Fredholm determinants $\det(I - tA_a)$ and $\det(I - tA_1)$ are well defined and p-adically entire (i.e., convergent for all $t \in \Omega$). The Dwork trace formula asserts that

$$S_m^*(f) = (q^m - 1)^n \mathrm{Tr}(A_a(f)^m). \tag{2.19}$$

The nontrivial additive character Ψ implicit in the left side of (2.19) (see (1.1)) is given by

$$\Psi(t) = \theta(1)^t \tag{2.20}$$

for $t \in \mathbf{F}_p$ ($\theta(1)$ is a primitive p-th root of unity). An equivalent form of (2.19) is

$$L^*(f,t)^{(-1)^{n-1}} = \prod_{i=0}^{n} \det(I - tq^i A_a(f))^{(-1)^i\binom{n}{i}}, \tag{2.21}$$

where

$$\det(I - tA_a(f)) = \exp(-\sum_{m=1}^{\infty} \frac{\mathrm{Tr}(A_a(f)^m)t^m}{m}). \tag{2.22}$$

Define a weight function $W(k)$ on non-negative integers as follows:

$$W(k) = \mathrm{card}\{r \in \mathbf{Z}^n \mid w(r) = \frac{k}{D}\}. \tag{2.23}$$

Let $P(\Delta)$ be the Newton polygon with vertices $(0,0)$ and

$$(\sum_{k=0}^{m} W(k), \frac{1}{D}\sum_{k=0}^{m} kW(k)), \quad m = 0,1,2,\cdots. \tag{2.24}$$

Let $\xi \in \Omega$ be such that $\xi^D = \pi^{p-1} = -p$; then $\mathrm{ord}\xi = 1/D$. By (2.18), $A_1(f)$ has the block form

$$A_1(f) = \begin{pmatrix} A_{00} & A_{01} & \cdots & A_{0i} & \cdots \\ \xi^1 A_{10} & \xi^1 A_{11} & \cdots & \xi^1 A_{1i} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \\ \xi^i A_{i0} & \xi^i A_{i1} & \cdots & \xi^i A_{ii} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \end{pmatrix}, \tag{2.25}$$

where the A_{ij}'s are matrices of $W(i)$ rows and $W(j)$ columns. The block form (2.25) and the lower bound of Adolphson and Sperber [1] imply.

Lemma 2.2. The Newton polygon of $\det(I - tA_a(f))$ computed with respect to q and the Newton polygon of $\det(I - tA_1(f))$ computed with respect to p lie above $P(\Delta)$.

Using the triangulation procedure in [9] and by induction, we obtain the following result. The detail is available in [10].

Proposition 2.3. The Newton polygon of $\det(I - tA_a(f))$ computed with respect to q coincides with its lower bound $P(\Delta)$ if and only if the Newton polygon of $\det(I - tA_1(f))$ computed with respect to p coincides with the same lower bound $P(\Delta)$.

3. The Congruence Decomposition of $L^*(f,t)$

We first describe a decomposition of Δ. Let $S(\Delta)$ be the set of all open faces σ of dimension at most $n-1$ such that the closure $\bar{\sigma}$ does not contain the origin. The origin will also be regarded as an element in $S(\Delta)$. For $\sigma \in S(\Delta)$, let $C(\sigma)$ be the open cone spanned by σ and the origin. The origin itself is regarded as an open cone. Then the closed cone $C(\Delta)$ is the disjoint union of the open cones $C(\sigma)$. We shall call $S(\Delta)$ the natural open decomposition of Δ. We say that $C(\sigma)$ is a simplex if σ is a simplex. The natural open decomposition is saided to be simplicial if all $C(\sigma)$ are simplicies.

Next, we define a congruence relation between entire functions. Recall that the Newton polygon of an entire function $g(t) = \sum_{k=0}^{\infty} a_k t^k$ is defined to be the convex closure in the plane of the vertices $(k, \mathrm{ord} a_k)$.

Definition 3.1. (a) Let $P_1, \cdots, P_m$ be the Newton polygons of p-adic entire functions $g_1(t), \cdots, g_m(t)$. We define the product polygon $\prod_i P_i$ to be the Newton polygon of the entire product function $\prod_i g_i$.

(b) Let P be a Newton polygon. Let g_1 and g_2 be two entire functions whose Newton polygons lie above P. We say that g_1 is congruent to g_2 modulo P, denoted

$$g_1 \equiv g_2 \pmod{P},$$

if either the Newton polygons of both g_i lie strictly above P (i.e., do not coincide with P) or else the Newton polygons of both g_i coincide with P, and in the latter case, P and the Newton polgon of $g_1 - g_2$ have no common points.

If g_1 is congruent to g_2 modulo P and the Newton polygons of both g_i coincide with P, the above definition implies that the reciprocal roots of g_1 and g_2 can be ordered as $\{\alpha_i\}$ and $\{\beta_i\}$ such that $\mathrm{ord}_p \alpha_i = \mathrm{ord}_p \beta_i$ and $\mathrm{ord}_p(\alpha_i - \beta_i) > \mathrm{ord}_p \alpha_i$ for all i. Thus, α_i and β_i have the same first non-zero "p-adic digit". The following lemma is easy to check.

Lemma 3.2. (a) If the Newton polygon of g_i lies above P_i for every i, then the Newton polygon of the product $\prod_i g_i$ lies above the product polygon $\prod_i P_i$.

(b) Let g_1 be congruent to g_2 modulo P. If the Newton polygon of g_1 coincides with P, then the Newton polygon of g_2 coincides with P.

(c) The congruence relation is an equivalence relation. Furthermore, let $g_1 \equiv g_2 \pmod{P(g)}$ and $h_1 \equiv h_2 \pmod{P(h)}$. Then,

$$g_1 h_1 \equiv g_2 h_2 \pmod{P(g)P(h)}.$$

Unless otherwise stated, we shall use σ to denote an element of $S(\Delta)$. Let $f_{\bar{\sigma}}$ be the restriction of the Laurent polynomial f to the **closed** cone spanned by σ and the origin. Then Dwork's theory as described in §2 applies to $f_{\bar{\sigma}}$. Let $A_1(\sigma)$ be the "open" piece of the matrix $A_1(f_{\bar{\sigma}}) = (a_{s,r}(f_{\bar{\sigma}}))$ consisting of all $a_{s,r}$ with s and r running through the open cone $C(\sigma)$. Define

$$A_a(\sigma) = \prod_{i=0}^{a-1} A_1(\sigma)^{r^i}, \quad A_a(\bar{\sigma}) = \prod_{i=0}^{a-1} A_1(f_{\bar{\sigma}})^{r^i}.$$

These are nuclear matrices whose Fredholm determinants are entire.

Define a weight function $W_\sigma(k)$ on non-negative integers as follows:

$$W_\sigma(k) = \operatorname{card}\{r \in \mathbf{Z}^n \cap C(\sigma) \mid w(r) = \frac{k}{D}\}. \tag{3.1}$$

Let $P(\sigma)$ be the Newton polygon with vertices $(0,0)$ and

$$(\sum_{k=0}^{m} W_\sigma(k), \frac{1}{D}\sum_{k=0}^{m} kW_\sigma(k)), \quad m = 0,1,2,\cdots. \tag{3.2}$$

One checks that

$$W(k) = \sum_{\sigma \in S(\Delta)} W_\sigma(k).$$

It follows that $P(\Delta)$ is the product polygon of the $P(\sigma)$. Similarly, one can define $W_{\bar{\sigma}}(k)$ and $P(\bar{\sigma})$ for a closed face $\bar{\sigma}$. The following proposition follows from the block form of $A_1(\sigma)$.

Proposition 3.3. The Newton polygon of $\det(I - tA_1(\sigma))$ computed with respect to p and the Newton polygon of $\det(I - tA_a(\sigma))$ computed with respect to q lie above $P(\sigma)$.

Our main result is the following open decomposition theorem.

Theorem 3.4. (a) With respect to p,

$$\det(I - tA_1(f)) \equiv \prod_{\sigma \in S(\Delta)} \det(I - tA_1(\sigma)) \pmod{P(\Delta)}.$$

(b) With respect to q,

$$\det(I - tA_a(f)) \equiv \prod_{\sigma \in S(\Delta)} \det(I - tA_a(\sigma)) \pmod{P(\Delta)}.$$

Corollary 3.5. The Newton polygon of $\det(I - tA_a)$ computed with respect to q coincides with $P(\Delta)$ if and only if the Newton polygon of $\det(I - tA_a(\sigma))$ computed with respect to q coincides with $P(\sigma)$ for all $\sigma \in S(\Delta)$.

Recall that F_r is defined in (2.9). We first prove

Lemma 3.6. Let $r \in C(\sigma)$. Then

$$\operatorname{ord}(F_r(f) - F_r(f_{\bar{\sigma}})) > \frac{w(r)}{p-1}.$$

Proof. Let $v(1), \cdots, v(J)$ be the non-zero exponents of f. Let $u_1, \cdots, u_J$ be non-negative integers satisfying $r = \sum_i u_i v(i)$. Without loss of generality, we may assume that $u_i > 0$ for $i \leq j$ and $u_i = 0$ for $i > j$. Assume that $\sum_i u_i = w(r)$. We need to prove that the $v(i)$ $(i \leq j)$ lie in the closed cone $C(\bar{\sigma})$. By Lemma 2.1, we have

$$w(r) = w(\sum_i u_i v(i)) \leq \sum_i u_i w(v(i)) \leq \sum_i u_i = w(u).$$

Since the inequalities are equalities, the $v(i)$ $(i \leq j)$ are cofacial. Since the $v(i)$ also span $r \in C(\sigma)$, it follows that the $v(i)$ $(i \leq j)$ must be in the closed cone $C(\bar{\sigma})$. The Lemma is proved.

Proof of Theorem 3.4. Fix an ordering of $S(\Delta) = \{\sigma_0, \cdots, \sigma_h\}$ such that $\dim(\sigma_i) \leq \dim(\sigma_{i+1})$. Let $i < j$, $s \in C(\sigma_i)$ and $r \in C(\sigma_j)$. Since $\dim(\sigma_i) \leq$

$\dim(\sigma_j)$, we can choose a hyperplane H which is generated by the origin and some $(n-2)$-dimensional face $\sigma \in S(\Delta)$, such that s and r lie on different sides of H (H may contain s but not r). Then $ps - r$ and r lie on different sides of H. Thus, they are not cofacial. By Lemma 2.1, we then have $a_{s,r} > w(s)$.

Let B_{ij} $(0 \leq i,j \leq h)$ be the nuclear submatrix $(a_{s,r})$ of $A_1(f)$, where $s \in \sigma_i$ and $r \in \sigma_j$. It is clear that the Newton polygon of the entire function $\det(I - tB_{ii})$ lies above $P(\sigma_i)$. Furthermore,

$$A_1(f) = \begin{pmatrix} B_{00} & B_{01} & \cdots & B_{0h} \\ B_{10} & B_{11} & \cdots & B_{1h} \\ \vdots & \vdots & \ddots & \vdots \\ B_{h0} & B_{h1} & \cdots & B_{hh} \end{pmatrix}. \tag{3.3}$$

For $i < j$, we showed in the above that each element $a_{s,r}$ in B_{ij} has greater p-adic valuation than the lower bound given in (2.18). This shows that the above block form for $A_1(f)$ is, in some sense, lower triangular. Let $P(x)$ be the piecewise linear function whose graph is the polygon $P(f)$. Then, by the determinant expansion of an matrix and the "lower triangular" form of the matrix $A_1(f)$, we deduce that with respect to p,

$$\det(I - tA_1(f)) = \prod_{i=0}^{h} \det(I - tB_{ii}) + \sum_{k=0}^{\infty} c_k t^k, \tag{3.4}$$

where the c_k are p-adic numbers in Ω satisfying

$$\operatorname{ord}_p c_k > P(k), k = 0, 1, 2, \cdots.$$

Thus,

$$\det(I - tA_1(f)) \equiv \prod_{i=0}^{h} \det(I - tB_{ii}) \pmod{P(\Delta)}. \tag{3.5}$$

Similarly, using the determinant expansion of an matrix and Lemma 3.6 we conclude that with respect to p,

$$\det(I - tA_1(\sigma_i)) \equiv \det(I - tB_{ii}) \pmod{P(\sigma_i)}. \tag{3.6}$$

These relations and Lemma 3.2 implies the first part of the theorem.

We now prove the second part. If the Newton polygon of $\det(I - tA_1(f))$ lies strictly above $P(\Delta)$, then the Newton polygon of $\det(I - tA_1(\sigma_i))$ lies

strictly above $P(\sigma_i)$ for some i. By Proposition 2.3, both the Newton polygon of $\det(I - tA_a(f))$ and the Newton polygon of $\prod_j \det(I - tA_a(\sigma_j))$ lie strictly above $P(\Delta)$, Thus, the second part is true in this case. In the other case, the Newton polygon of $\det(I - tA_1(f))$ (resp. $\det(I - tA_1(\sigma))$) coincides with $P(\Delta)$ (resp. $P(\sigma)$). By induction and the semi-linear triangulation procedure used in [9], we conclude that the second part is also true. The theorem is proved.

Applying Theorem 3.4 twice, we deduce the following closed decomposition theorem.

Corollary 3.7. The Newton polygon of $\det(I - tA_a)$ computed with respect to q coincides with $P(\Delta)$ if and only if the Newton polygon of $\det(I - tA_a(\bar{\sigma}))$ computed with respect to q coincides with $P(\bar{\sigma})$ for all $(n-1)$-dimensional faces $\sigma \in S(\Delta)$.

We are now in a position to prove Theorem 1.1. Presumably, Theorem 1.1 can also be proved by combining the ideas of the present paper together with the cohomological arguments in [2]. But this would repeat much of the work in [2]. Here we have used the chain level approach, because it is simpler, more elementary and self-contained.

Proof of Theorem 1.1. For an integer k, define the cohomological weight function by

$$W'(k) = \sum_{l \geq 0} (-1)^l \binom{n}{l} W(k - lD).$$

If $k > nD$, then $W'(k) = 0$. Let f be a non-degenerate Laurent polynomial. The theorem of Adolphson-Sperber implies that $L^*(f,t)^{(-1)^{n-1}}$ is a polynomial of degree $\sum_{k \geq 0} W'(k)$, whose Newton polygon lies above the convex polygon with vertices

$$(\sum_{i=0}^{k} W'(i), \sum_{i=0}^{k} \frac{i}{D} W'(i)), \ k = 0, 1, 2, \cdots. \tag{3.7}$$

The following lemma reduces Theorem 1.1 from cohomology level to chain level. Theorem 1.1 then follows from the chian level decomposition (Corollary 3.7).

Lemma 3.8. Let $L^*(f,t)^{(-1)^{n-1}}$ be a polynomial. The Newton polygon of $L^*(f,t)^{(-1)^{n-1}}$ coincides with its lower bound (3.7) if and only if the Newton polygon of $\det(I - tA_a)$ coincides with its lower bound $P(\Delta)$ (computed with respect to q).

Proof. We use formula (2.21) and the assumption that $L^*(f,t)^{(-1)^{n-1}}$ is a polynomial. If the Newton polygon of $\det(I - tA_a)$ coincides with $P(\Delta)$, then $\det(I - tA_a)$ has exactly $W(k)$ reciprocal roots of q-adic order k/D (for all k). Formula (2.21) implies that $L^*(f,t)^{(-1)^{n-1}}$ has exactly $W'(k)$ reciprocal roots of q-adic order k/D. That is, the Newton polygon of $L^*(f,t)^{(-1)^{n-1}}$ coincides with its lower bound (3.7). If the Newton polygon of $\det(I - tA_a)$ does not coincide with $P(\Delta)$, let k be the smallest non-negative integer such that $\det(I - tA_a)$ has less than $W(k)$ reciprocal roots of q-adic order k/D. Formula (2.21) implies that $L^*(f,t)^{(-1)^{n-1}}$ has less than $W'(k)$ reciprocal roots of q-adic order k/D. That is, the Newton polygon of $L^*(f,t)^{(-1)^{n-1}}$ does not coincide with its lower bound (3.7). Thus, the lemma is true.

4. The Congruence Decomposition of $L(f,t)$

Let $0 \leq h \leq n$, and let $f \in \mathbf{F}_q[x_1, \cdots, x_n, (x_1 \cdots x_h)^{-1}]$ be a Laurent polynomial. One can define the exponential sum

$$S_m(f) = \sum \Psi(\mathrm{Tr}_{\mathbf{F}_{q^m}/\mathbf{F}_p}(f(x_1, \cdots, x_n))), \tag{4.1}$$

where the sum is over all $(x_1, \cdots, x_n) \in (\mathbf{F}_{q^m})^n$ such that $x_j \neq 0$ $(1 \leq j \leq h)$. If $h = n$, $S_m(f)$ reduces to $S_m^*(f)$. The associated L-function is

$$L(f,t) = \exp(\sum_{m=1}^{\infty} S_m(f)\frac{t^m}{m}). \tag{4.2}$$

In this section, we show that Theorem 1.2 can be carried over to the function $L(f,t)^{(-1)^{n-1}}$. Since the proof is similar, we shall only give the necessary modifications needed in the present case.

Let T be a subset of $\{1, 2, \cdots, n\}$, $S_1 = \{1, 2, \cdots, h\}$ and $S_2 = \{h+1, \cdots, n\}$. Let $C(T) = \{r \in C(\Delta) \mid r_i > 0 \text{ if } i \in T \cap S_2\}$. Let $A_1(T)$ be the nuclear matrix $(a_{s,r})$, where s and r run through the lattice points in the cone $C(T)$. Let

$$A_a(T) = \prod_{i=0}^{a-1} A_1(T)^{\tau^i} = A_1(T)A_1^{\tau}(T)\cdots A_1^{\tau^{a-1}}(T). \tag{4.3}$$

The Dwork trace formula [2] asserts that

$$L(f,t)^{(-1)^{n-1}} = \prod_{l=0}^{n} \prod_{|T|=l} \det(I - tq^{n-l}A_a(T))^{(-1)^{n-l}}. \tag{4.4}$$

Let

$$W_T(k) = \text{card}\{r \in C(T) \cap \mathbf{Z}^n \mid w(r) = \frac{k}{D}\}. \tag{4.5}$$

Then standard arguments imply

Proposition 4.1. The Newton polygon of $\det(I - tA_a(T))$ lies above the polygon with vertices $(0,0)$ and

$$(\sum_{k=0}^{m} W_T(k), \frac{1}{D}\sum_{k=0}^{m} kW_T(k)), \quad m = 0,1,2,\cdots. \tag{4.6}$$

From §3, the closure of $C(T)$ has a natural open decomposition. Intersecting this decomposition with $C(T)$, we obtain the natural open decomposition of $C(T)$. As in section §3, the following proposition can be proved.

Theorem 4.2. In terms of the natural open decomposition of $C(T)$, $\det(I - tA_a(T))$ and $\det(I - tA_1(T))$ have a congruence decomposition (similar to the one in Theorem 3.4) with respect to the lower bound in (4.6).

This is the chain level (open) congruence decomposition theorem. If f is commode, the restriction f_σ of f to a closed face σ may not be commode. Thus, Theorem 1.1 does not seem to generalize to the function $L(f,t)^{(-1)^{n-1}}$. To prove the analogue of Theorem 1.2, the chain level decomposition theorem will be sufficient.

Let

$$W'(k) = \sum_{l \geq 0}(-1)^{n-l} \sum_{|T|=l} W_T(k - (n-l)D). \tag{4.7}$$

It is known that $W'(k) = 0$ for $k > nD$.

Proposition 4.3 (Adolphson and Sperber [2]). If f is nondegenerate with respect to Δ and commode with respect to S_2, then $L(f,t)^{(-1)^{n-1}}$ is a polynomial of degree $\sum_{k \geq 0} W'(k)$. Furthermore, the Newton polygon of $L(f,t)^{(-1)^{n-1}}$ lies above the polygon with vertices $(0,0)$ and

$$(\sum_{k=0}^{m} W'(k), \frac{1}{D}\sum_{k=0}^{m} kW'(k)), \quad m = 0,1,2,\cdots,nD. \tag{4.8}$$

Theorem 4.4. Assume that $L(f,t)^{(-1)^{n-1}}$ is a polynomial. Then the Newton polygon of $L(f,t)^{(-1)^{n-1}}$ coincides with its lower bound (4.8) if and

only if the Newton polygon of $\det(I - tA_a(\sigma))$ computed with respect to q coincides with $P(\sigma)$ for all element σ in the natural open decomposition of $C(\Delta)$.

Proof. We use (4.4) and the fact that $L(f,t)^{(-1)^{n-1}}$ is a polynomial. Following the proof of Lemma 3.8, we deduce that the Newton polygon of $L(f,t)^{(-1)^{n-1}}$ coincides with its lower bound (4.8) if and only if the Newton polygon of $\det(I - tA_a(T))$ coincides with its lower bound (4.6) (computed with respect to q) for all T. By Theorem 4.2, the last statement is true if and only if the coincidence is true for all the elements in the union of the natural open decompositions of $C(T)$ over all T. One checks that the natural open decomposition of $C(\Delta)$ is simply the union of the natural open decompositions of $C(T)$ over all T. The theorem is proved.

We now turn to the Adolphson-Sperber conjecture for $L(f,t)^{(-1)^{n-1}}$. As in the introduction, let $M_p(\Delta)$ be the moduli space of all Laurent polynomials f over $\bar{\mathbf{F}}_q$ with $\Delta(f) = \Delta$. One sees that $M_p(\Delta)$ is an irreducible quasi-projective variety. Let $H_p(\Delta)$ be the moduli space of those $f \in M_p(\Delta)$ such that f is nondegenerate and commode and such that the Newton polygon of $L(f,t)^{(-1)^{n-1}}$ coincides with the lower bound (4.8). The AS conjecture states that $H_p(\Delta)$ contains a dense Zariski open subset of $M_p(\Delta)$ if p is a large prime in the residue class $p \equiv 1 (\mathrm{mod}\ D)$. In [9], we proved that $H_p(\Delta)$ is Zariski open. Thus, the AS conjecture is reduced to proving that $H_p(\Delta)$ is non-empty.

We have the following analogue of Theorem 1.2.

Theorem 4.5. Assume that all $(n-1)$-dimensional closed faces of Δ which do not contain the origin are simplices (this is always true if $n \leq 2$). Then there is an integer D^* determined by Δ such that if p is a large prime and $p \equiv 1 (\mathrm{mod}\ D^*)$, then $H_p(\Delta)$ is open and dense.

Proof. Let $f = \sum_v x^v$, where v runs over the vertices of Δ. For an $(n-1)$-dimensional closed face σ of Δ not containing the origin, let f_σ be the restriction of f to the closed cone generated by σ and the origin. By assumption, each σ is a simplex. The computation in §5 (Theorem 5.2) shows that there is an integer D^* determined by Δ such that if p is a large prime and $p \equiv 1 (\mathrm{mod}\ D^*)$, then the Newton polygon of $L^*(f_\sigma,t)^{(-1)^{n-1}}$ coincides with its lower bound for all such σ. Applying Lemma 3.8 and Theorem 3.4 to each σ, we deduce that the

Newton polygon of $\det(I - tA_a(\sigma_1))$ computed with respect to q coincides with $P(\sigma_1)$ for all elements σ_1 in the natural open decomposition of $C(\Delta)$. Theorem 4.5 then follows from Theorem 4.4. The proof is complete.

5. A Computation of L-Functions

Let Δ be an n-dimensional simplex with integral vertices $0, v^{(1)}, \cdots, v^{(n)}$. The v^i are non-zero vertices. Let

$$f(x) = \sum_{j=1}^{n} \bar{a}_i x^{v^{(j)}} \tag{5.1}$$

be the diagonal polynomial, where $\bar{a}_j \in \mathbf{F}_q^* = \mathbf{F}_q - 0$. Let A be the $n \times n$ matrix $(v^{(1)}, \cdots, v^{(n)})$, where the $v^{(i)}$ are written as column vectors. It is easy to check that f is nondegenerate with respect to Δ if p is relatively prime to the determinat $\det A$ of the matrix A. We assume that p satisfies this condition. In this section, we compute the L-function $L^*(f,t)$. First, we compute the associated exponential sums.

Let T_q be the set of Teichmüller representatives of $\mathbf{F}_q^*$, and let a_j be the Teichmüller lifting of $\bar{a}_j$, i.e., $a_j^q = a_j \in T_q$. Define a polynomial of degree $q-2$:

$$P(t) = \sum_{m=0}^{q-2} c_m t^m, \tag{5.2}$$

satisfying

$$P(t) = \psi(\bar{t}) \quad \text{for all } t \in T_q. \tag{5.3}$$

One checks that

$$c_m = \frac{g_m(q)}{1-q} \ (0 \le m \le q-2), \tag{5,4}$$

where $g_m(q)$ is the Gauss sum defined by

$$g_0(q) = -1, \quad g_m(q) = -\sum_{t \in T_q} t^{-m} \psi(\bar{t}) \ (1 \le m \le q-2). \tag{5.5}$$

From (5.3)-(5.5), we have

$$\begin{aligned} S_q^*(f) &= \sum_{t\in T_q^n} \psi(f(\bar{t})) \\ &= \sum_{t\in T_q^n} \prod_{i=1}^{n} \psi(\bar{a}_i(\bar{t})^{v^{(i)}}) \\ &= \sum_{t\in T_q^n} \prod_{i=1}^{n} \sum_{m_i=0}^{q-2} c_{m_i} a_i^{m_i} t^{m_i v^{(i)}} \\ &= \sum_{m_1=0}^{q-2} \cdots \sum_{m_n=0}^{q-2} \Big(\prod_{i=1}^{n} c_{m_i} a_i^{m_i}\Big) \sum_{t\in T_q^n} t^{m_1 v^{(1)}+\cdots+m_n v^{(n)}} \\ &= (-1)^n \sum_{m_1 v^{(1)}+\cdots+m_n v^{(n)} \equiv 0 (\mathrm{mod} q-1)} \prod_{i=1}^{n} a_i^{m_i} g_{m_i}(q). \end{aligned} \tag{5.6}$$

This finishes our computation of the exponential sum $S^*(f)$.

Next, we compute $L^*(f,t)$. We first make some preliminary remarks. Consider the solutions of the following equation

$$A\begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix} \equiv 0(\mathrm{mod}\ 1),\ r_i \text{ rational},\ 0 \le r_i < 1. \tag{5.7}$$

The map $(r_1,\cdots,r_n) \to r_1 v^{(1)} + \cdots + r_n v^{(n)}$ clearly establishes a one-to-one correspondence between the solutions of (5.7) and the integral lattice points of the fundamental domain

$$\mathbf{R}v^{(1)} + \cdots + \mathbf{R}v^{(n)} (\mathrm{mod}\ \mathbf{Z}v^{(1)} + \cdots + \mathbf{Z}v^{(n)}). \tag{5.8}$$

Under this bijection, we identify the solution set of (5.7) and the set of integral lattice points in the above fundamental domain.

For a closed face σ of Δ containing the origin with vertices $0, v^{(i_1)}, \cdots, v^{(i_k)}$, let $I(\sigma)$ (resp. $I_0(\sigma)$) denote the set of integral lattice points (resp. interior integral lattice points) in the fundamental domain

$$\mathbf{R}v^{(i_1)} + \cdots + \mathbf{R}v^{(i_k)} (\mathrm{mod}\ \mathbf{Z}v^{(i_1)} + \cdots + \mathbf{Z}v^{(i_k)}). \tag{5.9}$$

It is well known that

$$| I(\sigma) | = (\dim\sigma)! V(\sigma), \tag{5.10}$$

where $V(\sigma)$ is the relative volume of the face σ. The inclusion-exclusion principle shows that

$$| I_0(\sigma) | = \sum_{0 \in \tau \in \sigma} (-1)^{\dim\sigma - \dim\tau} (\dim\tau)! V(\tau), \tag{5.11}$$

where τ runs over all closed faces of σ containing the origin.

To compute the L-function, for each integer $d > 0$, let $E_1(d)$ be the set of solutions $(r_1, \cdots, r_n)$ of (5.7) such that all denominators of the r_i's are factors of $q^d - 1$. Set $E(1) = E_1(1)$ and $E(d) = E_1(d) - \cup_{k|d, k<d} E_1(k) (d > 1)$. Then for large d, $E(d)$ is the empty set. It is easy to see that multiplication by q defines an equivalence relation on $E(d)$, i.e.,

$$(r_1, \cdots, r_n) \sim (qr_1 (\mathrm{mod}\ 1), \cdots, qr_n (\mathrm{mod}\ 1)). \tag{5.12}$$

Each equivalence class in $E(d)$ consists of exactly d elements. Let $E(d)/\sim$ be the set of equivalence classes of $E(d)$ under the relation (5.12). Using (5.6) and the well-known Hasse-Davenport relation:

$$g_{m_i(q^d-1)/(q-1)}(q^d) = g_{m_i}(q)^d, \tag{5.13}$$

one deduces the following formula for $L^*(f,t)$.

Theorem 5.1. Let p be relatively prime to $\det A$. Using the above notation, we have

$$L^*(f,t)^{(-1)^{n-1}} = \prod_{d \geq 1} \prod_{r \in E(d)/\sim} (1 - t^d \prod_{i=1}^{n} a_i^{r_i(q^d-1)} g_{r_i(q^d-1)}(q^d)), \tag{5.14}$$

where $r = (r_1, \cdots, r_n)$.

By (5.10), this is a polynomial of degree $n!V(\Delta)$.

We now determine the Newton polygon of $L^*(f,t)^{(-1)^{n-1}}$ for all primes p in a certain residue class.

Recall that A is the $n \times n$ integral matrix formed from the exponents of f (see (5.7)). From linear algebra, one knows that A is equivalent to a diagonal matrix with positive integers d_i $(1 \leq i \leq n)$ on the main diagonal, where $d_1 \mid d_2 \mid \cdots \mid d_n$. The d_i's are called the invariant factors of A. Let $d = d_n$ be the largest invariant factor of A. Then A^{-1} is a matrix with coefficients in $\frac{1}{d}\mathbf{Z}$.

Theorem 5.2. Let p be a prime number satisfying $p \equiv 1 (\mathrm{mod}\ d)$, then the Newton polygon of $L^*(f,t)^{(-1)^{n-1}}$ coincides with its lower bound.

Proof. We first describe the lower bound. For $u \in \mathbf{Z}^n$, the linear equation

$$A\begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix} = u,\ 0 \leq r_i, \tag{5.15}$$

has at most one solution. If $(r_1, \cdots, r_n)$ is the unique solution of (5.15), then one sees that $w(u) = r_1 + \cdots + r_n$, where $w(u)$ is the weight function introduced in §2. Otherwise, $w(u) = +\infty$. Thus, $W(k/D)$ is the number of solutions $(r_1, \cdots, r_n)$ of (5.15) such that

$$r_1 + \cdots + r_n = \frac{k}{D}. \tag{5.16}$$

Let

$$C(k) = W(\frac{k}{D}) - \binom{n}{1} W(\frac{k}{D} - 1) + \binom{n}{2} W(\frac{k}{D} - 2) - \cdots. \tag{5.17}$$

The lower bound for the Newton polygon of $L^*(f,t)^{(-1)^{n-1}}$ is the polygon with vertices (0,0) and $(\sum_{k=0}^{m} C(k), \sum_{k=0}^{m} \frac{k}{D} C(k))$. An inclusion-exclusion argument shows that $C(k)$ is the number of solutions $(r_1, \cdots, r_n)$ of (5.7) satisfying (5..16).

To prove Theorem 5.2, it suffices to prove that $L^*(f,t)^{(-1)^{n-1}}$ has exactly $C(k)$ reciprocal roots ρ such that $\mathrm{ord}_q \rho = k/D$. Let $q = p^a$. Since $p \equiv 1 (\mathrm{mod} d)$, the inverse matrix A^{-1} has entries in $\frac{1}{d}\mathbf{Z}$. Thus, the common denominator of any solution of (5.7) divides $p - 1$. This means $E(d)$ is empty for all $d \geq 2$. It follows from Theorem 5.1 that

$$L^*(f,t)^{(-1)^{n-1}} = \prod_{r} (1 - t \prod_{i=1}^{n} a_i^{r_i(q-1)} g^a_{r_i(p-1)}(p)), \tag{5.18}$$

where r runs through all solutions of (5.7). From this formula and the Stickelberger relation for Gauss sums, one deduces that $L^*(f,t)^{(-1)^{n-1}}$ has exactly $C(k)$ reciprocal roots ρ such that $\mathrm{ord}_q \rho = k/D$. Theorem 5.2 is proved.

Remark 5.3. Let D be the smallest positive integer such that $w(u) \in \frac{1}{D}\mathbf{Z}$. It can be easily proved that D is a factor of d. In general, $d \neq D$. For example,

take $f = x_1^k x_2^{1-k} + x_2$, where $k > 1$ is an integer. One finds that $d = k > 1$ and $D = 1$.

As a corollary of Theorem 1.1 and Theorem 5.2, we have

Corollary 5.4. Let f be a Laurent polynomial over $\mathbf{F}_q$ whose exponents are vertices of an n-dimensional Newton polyhedron Δ containing the origin. Let σ_i $(1 \leq i \leq m)$ be the faces of Δ of codimension 1. Assume that each σ_i is a simplex whose matrix of vertices has the inverse with entries in $\frac{1}{d}\mathbf{Z}$. Then the Newton polygon of $L^*(f,t)^{(-1)^{n-1}}$ coincides with its lower bound if $p \equiv 1 (\mathrm{mod}\ d)$.

Example 5.5. We give an application to generalized Kloosterman sums. The study of Newton polygons for such sums was suggested by Katz [5]. Many results in this direction were obtained by Sperber [6-8] and Carpentier [4]. For simplicity, we let $f(x) = a_1 x_1^{k_1} + \cdots + a_n x_n^{k_n} + x_1^{-b_1} \cdots x_n^{-b_n}$ be a Laurent polynomial, where the k_i and b_i are positive integers and the a_i are non-zero elements of $\mathbf{F}_q$. One sees that $\Delta(f)$ naturally decomposes into a union of n simplices of dimension n. Let

$$d = \mathrm{lcm}(k_1, \cdots, k_n, b_1, \cdots, b_n).$$

The matrices of vertices of these simplices have inverses with entries in $\frac{1}{d}\mathbf{Z}$. Then Corollary 5.4 reduces to the theorems of Sperber and Carpentier.

REFERENCES

1. A. Adolphson and S. Sperber, Newton polyhedra and the degree of the L-function associated to an exponential sum, Invent. Math., 88(1987), 555-569.
2. A. Adolphson and S. Sperber, Exponential sums and Newton polyhedra: Cohomology and estimates, Ann. of Math., 130(1989), 367-406.
3. A. Adolphson and S. Sperber, P-adic estimates for exponential sums, Lecture Notes in Mathematics, Vol. 1454(1990), 11-22.
4. M. Carpentier, p-Adic cohomology of generalized hyperkloosterman sums, Ph.D. thesis, University of Minnesota, 1985.
5. N.M. Katz, Sommes Exponentielles, Astérisque, 79(1980), 1-209.

6. S. Sperber, Congruence properties of the hyperkloosterman sum, Compositio Math., 40(1980), 3-33.
7. S. Sperber, Newton polygons for general hyperkloosterman sums, Asterisque, 119-120(1984), 267-330.
8. S. Sperber, On the p-adic theory of exponential sums. Amer. J. Math., 109(1986), 255-296.
9. D. Wan, The Newton polygon and p-adic analytic variation of L-functions over finite fields, preprint.
10. D. Wan, Newton polygons of zeta functions and L-functions, preprint.

Department of Mathematical Sciences, University of Nevada
Las Vegas, Nevada 89154
E-mail: dwan@nevada.edu

This page intentionally left blank

Recent Titles in This Series

(*Continued from the front of this publication*)

104 **Michael Makkai and Robert Paré,** Accessible categories: The foundations of categorical model theory, 1989

103 **Steve Fisk,** Coloring theories, 1989

102 **Stephen McAdam,** Primes associated to an ideal, 1989

101 **S.-Y. Cheng, H. Choi, and Robert E. Greene, Editors,** Recent developments in geometry, 1989

100 **W. Brent Lindquist, Editor,** Current progress in hyperbolic systems: Riemann problems and computations, 1989

99 **Basil Nicolaenko, Ciprian Foias, and Roger Temam, Editors,** The connection between infinite dimensional and finite dimensional dynamical systems, 1989

98 **Kenneth Appel and Wolfgang Haken,** Every planar map is four colorable, 1989

97 **J. E. Marsden, P. S. Krishnaprasad, and J. C. Simo, Editors,** Dynamics and control of multibody systems, 1989

96 **Mark Mahowald and Stewart Priddy, Editors,** Algebraic topology, 1989

95 **Joel V. Brawley and George E. Schnibben,** Infinite algebraic extensions of finite fields, 1989

94 **R. Daniel Mauldin, R. M. Shortt, and Cesar E. Silva, Editors,** Measure and measurable dynamics, 1989

93 **M. Isaacs, A. Lichtman, D. Passman, S. Sehgal, N. J. A. Sloane, and H. Zassenhaus, Editors,** Representation theory, group rings, and coding theory, 1989

92 **John W. Gray and Andre Scedrov, Editors,** Categories in computer science and logic, 1989

91 **David Colella, Editor,** Commutative harmonic analysis, 1989

90 **Richard Randell, Editor,** Singularities, 1989

89 **R. Bruce Richter, Editor,** Graphs and algorithms, 1989

88 **R. Fossum, W. Haboush, M. Hochster, and V. Lakshmibai, Editors,** Invariant theory, 1989

87 **Laszlo Fuchs, Rüdiger Göbel, and Phillip Schultz, Editors,** Abelian group theory, 1989

86 **J. Ritter, Editor,** Representation theory and number theory in connection with the local Langlands conjecture, 1989

85 **Bor-Luh Lin, Editor,** Banach space theory, 1989

84 **Stevo Todorcevic,** Partition problems in topology, 1989

83 **Michael R. Stein and R. Keith Dennis, Editors,** Algebraic K-theory and algebraic number theory, 1989

82 **Alexander J. Hahn, Donald G. James, and Zhe-xian Wan, Editors,** Classical groups and related topics, 1989

81 **Kenneth R. Meyer and Donald G. Saari, Editors,** Hamiltonian dynamical systems, 1988

80 **N. U. Prabhu, Editor,** Statistical inference from stochastic processes, 1988

79 **Ernst Kunz and Rolf Waldi,** Regular differential forms, 1988

78 **Joan S. Birman and Anatoly Libgober, Editors,** Braids, 1988

77 **Wang Yuan, Yang Chung-chun, and Pan Chengbiao, Editors,** Number theory and its applications in China, 1988

76 **David C. Hobby and Ralph McKenzie,** The structure of finite algebras, 1988

75 **Frank M. Cholewinski,** The finite calculus associated with Bessel functions, 1988

74 **William M. Goldman and Andy R. Magid, Editors,** Geometry of group representations, 1988

73 **Rick Durrett and Mark A. Pinsky, Editors,** Geometry of random motion, 1988

(See the AMS catalog for earlier titles)

This page intentionally left blank

ISBN 0-8218-5145-4