Pacific Journal of Mathematics

Volume 268 No. 2 April 2014

PACIFIC JOURNAL OF MATHEMATICS

msp.org/pjm

Founded in 1951 by E. F. Beckenbach (1906–1982) and F. Wolf (1904–1989)

See inside back cover or msp.org/pjm for submission instructions.

The subscription price for 2014 is US $410/year for the electronic version, and $535/year for print and electronic.
Subscriptions, requests for back issues and changes of subscribers address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163, U.S.A. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, PASCAL CNRS Index, Referativnyi Zhurnal, Current Mathematical Publications and Web of Knowledge (Science Citation Index).

The Pacific Journal of Mathematics (ISSN 0030-8730) at the University of California, c/o Department of Mathematics, 798 Evans Hall #3840, Berkeley, CA 94720-3840, is published twelve times a year. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices. POSTMASTER: send address changes to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163.

PJM peer review and production are managed by EditFLOW® from Mathematical Sciences Publishers.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

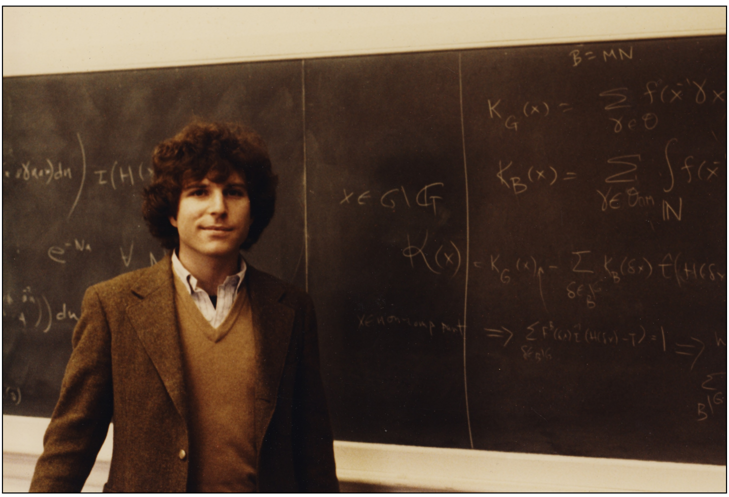
http://msp.org/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 268, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.268.257

The following three articles were intended
for the special issue entitled

IN MEMORIAM
JONATHAN ROGAWSKI (1955–2011)

(Volume 260, No. 2, 2012)



I do not ask to see the distant scene, one step enough for me.
John Henry Newman

We, the editors of that issue, are very happy to be able to publish them at this time.

DON BLASIUS
UNIVERSITY OF CALIFORNIA, LOS ANGELES

DINAKAR RAMAKRISHNAN
CALIFORNIA INSTITUTE OF TECHNOLOGY

V. S. VARADARAJAN
UNIVERSITY OF CALIFORNIA, LOS ANGELES

PACIFIC JOURNAL OF MATHEMATICS
Vol. 268, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.268.259

FORMES MODULAIRES SUR LA $\mathbb{Z}_p$-EXTENSION CYCLOTOMIQUE DE $\mathbb{Q}$

LAURENT CLOZEL

In memoriam Jon Rogawski

Soit F la $\mathbb{Z}_p$-extension cyclotomique de $\mathbb{Q}$. On peut se demander s'il existe une théorie non triviale des formes modulaires pour GL(2, F). On montre qu'une telle théorie existe en caractéristique p, au moins si GL(2) est remplacé par une algèbre de quaternions définie en (p, ∞). En particulier, une telle théorie réalise naturellement le changement de base de Saito–Shintani–Langlands.

Let F be the cyclotomic $\mathbb{Z}_p$-extension of the rationals. It is natural to ask whether there exists a nontrivial theory of modular forms on GL(2, F). We show that this is the case if the ring of coefficients has characteristic p and if GL(2) is replaced by the quaternion algebra ramified at (p, ∞). In particular, such a theory incorporates the base change of Saito, Shintani and Langlands.

1.

Fontaine m'a demandé s'il existait une théorie des formes modulaires modulo p sur GL(2, $\mathbb{Q}(p^\infty)$) où l'on désigne par $\mathbb{Q}(p^\infty)$ « la » $\mathbb{Z}_p$-extension cyclotomique de $\mathbb{Q}$. Le but de cet article est de montrer qu'une telle théorie existe en effet, avec quelques restrictions, pour $p \neq 2$. [1] [2]

La restriction essentielle porte sur le niveau en p. On fera ici une hypothèse de mauvaise réduction sur les formes classiques, qui nous permet de travailler sur (le groupe projectif d')une algèbre de quaternions ramifiée en p et l'infini. On verra aisément qu'une telle restriction n'est pas indispensable ; elle nous permet cependant ici d'obtenir une famille S_α ($\alpha \geq 0$) de « variétés de Shimura de dimension zéro »

L'auteur est membre de l'Institut Universitaire de France.
MSC2010 : 11F11, 11F99.
Mots-clefs : extension cyclotomique, forme modulaire, Saito–Shintani–Langlands, modular form, cyclotomic extension.

1. On est en fait amené à supposer $p \geq 5$ (§2) mais on pourrait éviter cette hypothèse en accroissant le niveau en p.

2. Voir également la note à la page 274.

essentiellement intrinsèque, en supposant au contraire que les formes modulaires considérées sont partout non ramifiées en-dehors de p.

Le test d'une théorie non triviale des formes modulaires est l'existence d'opérateurs de Hecke dont les valeurs propres devraient être liées aux valeurs propres d'opérateurs de Frobenius dans une représentation galoisienne associée. On montre l'existence d'une famille d'opérateurs de Hecke T_ℓ — sur cet espace infini de « formes » sur $\mathrm{GL}(2, \mathbb{Q}(p^\infty))$ — associées à tous les nombres premiers ℓ inertes dans $\mathbb{Q}(p^\infty)$. Sur cet espace infini, ces opérateurs sont donnés formellement par une somme infinie. Pour chaque extension finie $\mathbb{Q}(p^\alpha)$, leur expression devient finie et égale à l'expression classique. Ces opérateurs, conformément à l'intuition de Fontaine, n'ont un sens que si l'anneau des coefficients R des formes modulaires est de caractéristique p (ou, peut-être, s'il est complet pour la topologie p-adique).

Les opérateurs obtenus jouissent de la propriété usuelle d'autodualité (§ 4). La commutativité de l'algèbre des T_ℓ est plus délicate, car leur définition naturelle envoie un espace de fonctions à support fini vers un espace de fonctions arbitraires. On montre que les T_ℓ s'étendent naturellement à l'espace (linéairement compact) des fonctions invariantes par un sous-groupe ouvert de $\mathrm{Gal}(\mathbb{Q}(p^\infty)/\mathbb{Q})$, qu'il préservent. Dans ce cadre, ils forment une famille commutative (§ 4).

Enfin, on montre dans le § 5 que ces constructions donnent une nouvelle démonstration, naturelle, du changement de base de Saito–Shintani–Langlands. Celle-ci est limitée à la caractéristique p (et aux T_ℓ relatifs aux ℓ inertes) mais ne repose pas sur une identité de traces.

Dans leurs espaces linéairement compacts naturels, on peut évidemment se poser la question de la diagonalisation des T_ℓ ; en particulier on peut se demander si les valeurs propres classiques donnent lieu à des espaces propres dans ces espaces infinis. Nous n'obtenons qu'une réponse incomplète (proposition 5.2).

Notons deux perspectives naturelles. Tout d'abord, on peut sans doute faire des constructions similaires sur des anneaux p-adiques plutôt que de caractéristique p : ceci amènerait à définir un espace de Banach (de type L^∞) de formes p-adiques dont notre espace est la réduction modulo p. On peut aussi tenter de construire une théorie locale, qui n'apparaît pas ici puisque nous avons fixé la ramification (en particulier en p).

2.

On note $\mathbb{Q}(p^\infty)$ la $\mathbb{Z}_p$-extension cyclotomique de $\mathbb{Q}$. On écrira parfois

$$F_\infty = \mathbb{Q}(p^\infty) = \bigcup_{\alpha \geq 0} \mathbb{Q}(p^\alpha),$$

$F_\alpha = \mathbb{Q}(p^\alpha)$ étant la sous-extension de groupe $\mathbb{Z}/p^\alpha\mathbb{Z}$. Les corps F_α sont totalement réels.

On écrit de même

$$\mathbb{Q}_p(p^\infty) = \bigcup_\alpha \mathbb{Q}_p(p^\alpha), \quad F_{p,\alpha} = \mathbb{Q}_p(p^\alpha).$$

L'idéal (p) est totalement ramifié dans F_∞ ; les notations $\mathfrak{p}_\alpha$, $\mathfrak{p}_\infty$ sont évidentes. On note $\mathbb{O}_{p,\alpha}$ l'anneau des entiers de $F_{p,\alpha}$.

Considérons un nombre premier $\ell \neq p$, donc non ramifié dans F_∞.

Si $x \in \mathbb{Z}_p^\times$ a pour image $\bar{x} \in \mathbb{F}_p^\times$, on peut écrire $x = x_1\tau(\bar{x})$, $\tau(\bar{x}) \in \mathbb{Z}_p^\times$ étant le représentant de Teichmüller de x, et x_1 appartenant à $1 + p\mathbb{Z}_p$. Soit

$$\chi : (1 + p\mathbb{Z}_p)/(1 + p^2\mathbb{Z}_p) \to \mathbb{F}_p, \quad 1 + py \mapsto \bar{y}$$

l'isomorphisme naturel. On pose alors $\psi(x) = \chi(x_1) \in \mathbb{F}_p$. Si $\psi(\ell) \neq 0$, l'image de Frob_ℓ dans $\mathrm{Gal}(\mathbb{Q}(p^\infty)/\mathbb{Q}) \cong \mathbb{Z}_p$ en est un générateur topologique et il en résulte que ℓ est inerte dans chaque extension $\mathbb{Q}(p^\alpha)$. Il existe donc une unique place, aussi notée ℓ, de $\mathbb{Q}(p^\alpha)$ ou $\mathbb{Q}(p^\infty)$ au-dessus de la place rationnelle.

Soit B l'algèbre de quaternions sur $\mathbb{Q}$ ramifiée exactement en p et en la place archimédienne. Puisque p est impair, $B \otimes F_\alpha$ est ramifiée exactement en $\mathfrak{p}_\alpha$ et en les places réelles de F_α. On note G le groupe linéaire projectif associé à B : ainsi $G(\mathbb{Q}) = B^\times/\mathbb{Q}^\times$. On obtient de même un groupe G_α sur F_α, qui est obtenu par extension des scalaires à partir de G.

Considérons d'abord le groupe adélique $G(\mathbb{A})$ où $\mathbb{A}$ désigne les adèles rationnels. Nous considérerons des formes automorphes sur $G(\mathbb{A})$, donc des fonctions sur le quotient compact $G(\mathbb{Q})\backslash G(\mathbb{A})$.

Nous nous limitons dans cette note au cas le plus simple, en imposant des conditions de ramification minimale.

Fixons un ordre maximal $\mathbb{O}_B$ de B. Pour $\ell \neq p$, $(\mathbb{O}_B \otimes \mathbb{Z}_\ell)^\times$ est un sous-groupe compact maximal de $(B \otimes \mathbb{Q}_\ell)^\times \cong \mathrm{GL}(2, \mathbb{Q}_\ell)$. Soit K_ℓ son image dans $G(\mathbb{Q}_\ell)$.

En p, soit $\mathfrak{P}$ l'idéal maximal de $\mathbb{O}_B \otimes \mathbb{Z}_p$. Soit K_p l'image dans $G(\mathbb{Q}_p)$ du sous-groupe $1 + \mathfrak{P}$ de $(B \otimes \mathbb{Q}_p)^\times$. On pose $K = \prod_q K_q \subset G(\mathbb{A}_f)$.

Dans ce qui suit on appelle α le *degré* d'un objet relatif à F_α. En degré α, on définit de même en toute place $\lambda \nmid p$ de F_α un sous-groupe K_λ à partir de l'ordre local $\mathbb{O}_B \otimes \mathbb{O}_\lambda$.

Le discriminant (réduit) de l'ordre maximal $\mathbb{O}_B \otimes \mathbb{Z}_p \subset B_p$ est égal à p. En degré α, il en résulte par un calcul simple (voir [Vignéras 1980, p. 35]) que le discriminant de $\mathbb{O}_B \otimes \mathbb{O}_{p,\alpha} \subset B \otimes F_{p,\alpha}$ est l'idéal $(p) = \mathfrak{p}_\alpha^{p^\alpha}$. En particulier cet ordre n'est pas l'ordre maximal, le discriminant de celui-ci étant $\mathfrak{p}_\alpha$. Pour tout α, on munit au contraire $B_{p,\alpha}$ de son ordre maximal $\mathbb{O}_{B_{p,\alpha}}$. Ceux-ci vérifient les inclusions évidentes pour α variable. Pour tout α, la donnée de $\mathbb{O}_{B_{p,\alpha}}$ et des ordres précédents en $\lambda \nmid p$ définit un ordre maximal global $\mathbb{O}_{B,\alpha} \subset B \otimes F_\alpha$. Enfin, on définit $K_{p,\alpha} \subset G(F_{p,\alpha})$ comme précédemment à partir de l'idéal premier $\mathfrak{P}_\alpha \subset \mathbb{O}_{B_{p,\alpha}}$.

Soit

$$S_\alpha = G(F_\alpha)\backslash G(\mathbb{A}_{F_\alpha})/G(F_{\alpha,\infty})K_\alpha.$$

C'est un ensemble fini, et on pose

$$\mathcal{S}_\alpha = \mathcal{S}_\alpha(\bar{\mathbb{F}}_p) = \{f : S_\alpha \to \bar{\mathbb{F}}_p\}.$$

Noter que $S_\alpha = G(F_\alpha)\backslash G(\mathbb{A}_\alpha^f)/K_\alpha$ où on a noté $\mathbb{A}_\alpha^f$ les adèles finis de F_α.

On peut évidemment définir $\mathcal{S}_\alpha(R)$ pour toute algèbre commutative R. En particulier, $\mathcal{S}_\alpha(\mathbb{C})$ correspond, par la correspondance de Jacquet–Langlands, à un espace de formes modulaires sur PGL(2, $\mathbb{A}_\alpha$) (donc de caractère central trivial), de poids parallèle 2, et qui sont, ou des caractères d'Artin d'ordre 2 de $\mathbb{A}_\alpha^\times$, partout non ramifiés aux places ne divisant pas p, ou bien contenues dans des représentations cuspidales de PGL(2, $\mathbb{A}_\alpha$), non ramifiées hors p, et appartenant à la série discrète (avec une ramification bornée) en p.

Lemme 2.1. *Supposons $p > 3$. Alors $G(F_\alpha)$ opère librement sur $G(\mathbb{A}_\alpha^f)/K_\alpha$.*

Démonstration. Il suffit de vérifier que $K_{p,\alpha}$ ne contient aucun élément non trivial d'ordre fini. Puisque $1 + \mathfrak{P}_\alpha$ est un p-groupe, l'ordre d'un tel élément γ serait égal à p^β, $\beta \geq 1$. Soit $S \subset B^\times$ le groupe des éléments de norme 1, de sorte qu'on a une suite exacte

$$1 \to \{\pm 1\} \to S \to G \to 1$$

d'où pour les points sur $F = F_{p,\alpha}$:

$$1 \to \{\pm 1\} \to S(F) \to G(F) \to F^\times/(F^\times)^2.$$

L'élément γ provient donc de $S(F)$; l'ordre de son image inverse dans $S(F)$ est p^β ou $2p^\beta$; dans le dernier cas le produit par (-1) donne un élément d'ordre p^β.

On obtient donc un élément semi-simple non central $\gamma_1 \in B^\times(F)$, qui engendre une extension quadratique de $F_{p,\alpha}$, donc une extension de degré $2p^\alpha$ de $\mathbb{Q}_p$.

Mais γ_1 s'identifie à une racine primitive de 1 d'ordre p^β, et son degré sur $\mathbb{Q}_p$ est donc $(p-1)p^{\beta-1}$. Si $p > 3$, ceci est impossible. □

Les démonstrations qui suivent vont reposer sur des arguments galoisiens, concernant en particulier les groupes de Galois relatifs $\mathrm{Gal}(F_\beta/F_\alpha) \cong \mathbb{Z}/p^{\beta-\alpha}\mathbb{Z}$ $(\alpha \leq \beta)$. On notera que les sous-groupes compacts considérés sont stables par l'action galoisienne : c'est clair pour les K_λ (les places λ non inertes étant bien sûr permutées) et aussi pour les $K_{p,\alpha}$: l'ordre maximal $\mathcal{O}_{B_{p,\alpha}}$ étant unique, est invariant.

La suite des K_α étant croissante, on a des applications évidentes $S_\alpha \to S_\beta$ $(\alpha \leq \beta)$, qui sont donc équivariantes.

Lemme 2.2 ($p \geq 5$). *Les applications naturelles $\iota_\alpha^\beta : S_\alpha \to S_\beta$ sont injectives $(\alpha \leq \beta)$.*

Pour $\alpha \leq \beta$, on note j_α^β l'application $\mathcal{S}_\alpha \to \mathcal{S}_\beta$ obtenue par l'extension par zéro, et r_α^β la restriction $\mathcal{S}_\beta \to \mathcal{S}_\alpha$.

On définit alors, pour tout R

$$\mathcal{S}_\infty = \mathcal{S}_\infty(R) = \varinjlim \mathcal{S}_\alpha(R)$$

et $\mathcal{S}^\infty = \varprojlim \mathcal{S}_\alpha(R)$, les limites étant prises respectivement pour les j_α^β et r_α^β.

Le premier espace est l'espace des fonctions à support fini sur $S_\infty = \varinjlim S_\alpha$, alors que $\mathcal{S}^\infty$ est l'espace des fonctions $S_\infty \to R$.

Démonstration du lemme 2.2. Soit $x, x' \in S_\alpha$ d'image commune $y \in S_\beta$ et $g, h \in G(\mathbb{A}_\alpha^f)$ des représentants de x, x'. On a donc par hypothèse

$$g = \gamma h k \quad (\gamma \in G(F_\beta),\ k \in K_\beta). \tag{2-1}$$

Soit $\sigma \in \mathrm{Gal}(F_\beta/F_\alpha)$, alors

$$g = \sigma(g) = \sigma(\gamma) h \sigma(k)$$

donc

$$\sigma(\gamma)^{-1}\gamma h k = h\sigma(k), \quad k(\sigma(k) \in K_\beta).$$

D'après le lemme 2.1, on a donc $\sigma(\gamma) = \gamma$ pour tout σ, donc $\gamma \in G(F_\alpha)$; il résulte de (2-1) que $\sigma(k) \equiv k$ et donc $k \in K_\alpha$, soit $x = x'$. □

La dernière assertion résulte du lemme suivant :

Lemme 2.3. *Soit* $\Gamma = \mathrm{Gal}(F_\beta/F_\alpha)$. *Alors* $H^0(\Gamma, K_\beta) = K_\alpha$.

Démonstration. Le groupe K_α est décomposé, et l'assertion du lemme se voit place par place. Aux places $\nmid p$ elle est évidente.

Rappelons que $K_{p,\alpha}$ est l'image dans $G(F_{p,\alpha})$ de $1+\mathfrak{P}_\alpha$. On vérifie aisément que $(1+\mathfrak{P}_\alpha)\cap F_{p,\alpha}^\times = 1+\mathfrak{p}_\alpha$, le corps étant plongé centralement dans $B_{p,\alpha}$. Par ailleurs $(1+\mathfrak{P}_\beta)\cap B_{p,\alpha} = 1+\mathfrak{P}_\alpha$. Il suffit évidemment de vérifier que $\mathfrak{P}_\beta \cap B_{p,\alpha} = \mathfrak{P}_\alpha$. Notons N_α, N_β les normes réduites. Alors $\mathfrak{P}_\alpha$ (par exemple) est défini par $|N_\alpha(x)| < 1$, et pour $x \in B_{p,\alpha}$, $N_\beta(x) = N_\alpha(x)^{[F_\beta:F_\alpha]}$.

La suite exacte en cohomologie déduite de

$$1 \to (1+\mathfrak{p}_\beta) \to (1+\mathfrak{P}_\beta) \to K_\beta \to 1$$

donne donc

$$1 \to (1+\mathfrak{p}_\alpha) \to (1+\mathfrak{P}_\alpha) \to K_\beta^\Gamma \to H^1(\Gamma, 1+\mathfrak{p}_\beta) \to H^1(\Gamma, 1+\mathfrak{P}_\beta),$$

et il suffit de vérifier que la dernière flèche est injective. Mais la norme réduite donne de nouveau :

$$1 \to (1+\mathfrak{p}_\beta) \to (1+\mathfrak{P}_\beta) \xrightarrow[N_\beta]{} (1+\mathfrak{p}_\beta) \to 1,$$

la composée des deux applications étant l'application carré. Puisque $p \neq 2$, il en résulte que $H^1(\Gamma, 1+\mathfrak{p}_\beta) \to H^1(\Gamma, 1+\mathfrak{P}_\beta)$ est injective. □

La fin de ce paragraphe, qui n'est pas nécessaire pour la suite, est consacrée au calcul de la dimension des espaces $\mathcal{S}$.

Rappelons que

$$S_\alpha = G(F_\alpha)\backslash G(\mathbb{A}_\alpha)/G(F_{\alpha,\infty})K_\alpha. \tag{2-2}$$

Munissons $G(\mathbb{A}_\alpha)$ de la mesure de Tamagawa [Vignéras 1980]. D'après Weil, la mesure de $G(F_\alpha)\backslash G(\mathbb{A}_\alpha)$ est égale à 2. Puisque $G(F_\alpha)$ opère librement, on a donc

$$N_\alpha := \# S_\alpha = \frac{2}{\mathrm{vol}(G(F_{\alpha,\infty})K_\alpha)}$$

le volume étant calculé à l'aide de la mesure de Tamagawa τ.

Nous suivons pour ce calcul l'exposé de Vignéras [1980]. Écrivons pour simplifier $\mathbb{A} = \mathbb{A}_\alpha$. On a

$$G(\mathbb{A}) = B^\times(\mathbb{A})/\mathbb{A}^\times,$$

et la mesure τ est le quotient de $dX^\times_{\mathbb{A}}$ et $dx^\times_{\mathbb{A}}$, chacune multipliée par les facteurs $\mathrm{Res}_{s=1}\zeta_{F_\alpha}$ qui donc s'annulent [Vignéras 1980, p. 65]. On a :

$$dx^\times_{\mathbb{A}} = \prod_v dx^\times_v,$$

$$d^\times x_v = \frac{dx_v}{|x_v|} \quad (v \text{ réelle}),$$

$$d^\times x_v = (1-q_v^{-1})^{-1} D_v^{-1/2} \frac{dx_v}{|x_v|} \quad (v\ p\text{-adique}).$$

De même

$$d^\times X_v = \frac{dX_v}{|Nrd(X_v)|^2} \quad (v \text{ réelle}),$$

$$d^\times X_v = D_v^{-1/2}(1-q_v^{-1})^{-1} \frac{dX_v}{|NrdX_v|^2} \quad (v\ p\text{-adique}),$$

où D_v désigne maintenant le discriminant de B_v. Les mesures additives dx_v sont les mesures usuelles ; les mesures dX_v sont spécifiées dans [Vignéras 1980, pp. 49–50].

***Places non ramifiées* $(v \nmid p)$.** Dans ce cas, $K_v = \mathrm{PGL}(2, \mathbb{O}_v)$ et $\mathrm{vol}(K_v) = 1 - q_v^{-2}$ [Vignéras 1980, p. 49].

***Place* $v = \mathfrak{p}_\alpha \mid p$.** Dans ce cas, on a [ibid.]

$$\mathrm{vol}(\mathbb{O}_B^\times,\ d^\bullet X_v) = \frac{1-q_v^{-2}}{1-q_v^{-1}} \quad \text{et} \quad \mathrm{vol}(\mathbb{O}_{F_v}^\times, d^\bullet x_v) = 1,$$

où les mesures $d^\bullet$ ne contiennent pas le discriminant, donc pour le compact maximal :

$$\mathrm{vol}\left(K_v^0, \frac{d^\times X_v}{d^\times x_v}\right) = \left(\frac{1-q_v^{-2}}{1-q_v^{-1}}\right)\left(\frac{D(B_v)}{D(F_v)}\right)^{-1/2}.$$

On a $D(B_v) = D(F_v)^4\ N(d_B)^2$ [ibid., p. 65] où $N(d_B) = N\mathfrak{p}_\alpha = p$ [ibid., p. 35]. Donc

$$\mathrm{vol}(K_v^0) = D(F_v)^{-3/2}\left(\frac{1-p^{-2}}{p-1}\right)$$

puisque $q_v = p$.

Mais notre groupe K_v (en $v = \mathfrak{p}_\alpha$) est quotient de $1+\mathfrak{P}$, non $\mathbb{O}_B^\times$. Leurs intersections respectives avec le centre $F_{p,\alpha}^\times$ de $B^\times$ sont $1+\mathfrak{p}_\alpha \subset \mathbb{O}_{p,\alpha}^\times$. On vérifie aisément que $K_v^0/K_v \cong k_2^\times/k^\times$ où $k = k_\alpha$, k_2 est son extension quadratique, donc d'ordre $p+1$. Au total, la partie finie du volume est

$$\mathrm{vol}(K_f) = D_v^{-3/2}\zeta_{F_\alpha}(2)^{-1}\frac{1}{p^2-1}$$

où D_v est évidemment le discriminant de F_α.

Places archimédiennes. On doit calculer

$$\mathrm{vol}\left(B^\times/\mathbb{R}^\times, \frac{d^\times X_v}{d^\times x_v}\right),$$

où B est l'algèbre de Hamilton et $dX_v = 4\,dX_1 \dots dX_4$ avec les coordonnées usuelles. En identifiant B à $\mathbb{R}^4$, on voit qu'on doit calculer

$$\mathrm{vol}\left(S^3/\pm 1, \frac{4dX}{r^4}\Big/\frac{dr}{r}\right) = \frac{1}{2}\,\mathrm{vol}(S^3, 4d\omega) = 2\,\mathrm{vol}(S^3)$$

avec la mesure de surface $d\omega$, donc $4\pi^2$.

Il reste à calculer le discriminant, qui se déduit aisément de la Führerdiskriminantenproduktformel [3]

$$D_v = D_{F_\alpha} = \prod_\chi \mathfrak{f}_\chi$$

où χ parcourt les caractères d'ordre multiple de p de $(\mathbb{Z}/p^{\alpha+1}\mathbb{Z})^\times$. Il vient

$$D_{F_\alpha} = p^d,$$

où

$$\begin{aligned} d = d_\alpha &= (\alpha+1)(p^\alpha - p^{\alpha-1}) + \alpha(p^{\alpha-1} - p^{\alpha-2}) + \cdots + 2(p-1) \\ &= (p^\alpha - 1)\frac{p-2}{p-1} + \alpha p^\alpha. \end{aligned}$$

3. C'est un exercice amusant de calculer D_v à l'aide de la théorie p-adique.

On a donc démontré :

Proposition 2.4. $N_\alpha = \# S_\alpha = 2 \cdot (4\pi^2)^{-p^\alpha} p^{\frac{3}{2} d_\alpha} (p^2-1) \zeta_{F_\alpha}(2).$

D'après l'équation fonctionnelle, on a en fait

$$N_\alpha = -2 \cdot 2^{-p^\alpha} (p^2-1) \zeta_{F_\alpha}(-1) \tag{2-3}$$

ce qui montre la compatibilité avec les résultats connus sur la rationalité de $\zeta_{F_\alpha}(-1)$. (En fait, on sait d'après Serre, Harder et Hirzebruch [Serre 1971] que $24 \cdot 2^{-p^\alpha} \zeta_F(-1) \in \mathbb{Z}$, ce qui implique l'intégralité de N_α.)

Soit $\Gamma_\alpha = \mathrm{Gal}(F_\infty / F_\alpha)$. Puisque $\Gamma_{\alpha-1}/\Gamma_\alpha \cong \mathbb{Z}/p\mathbb{Z}$ opère sans point fixe sur $S_\alpha - S_{\alpha-1}$, on en déduit d'ailleurs :

Proposition 2.5. $\zeta_{F_\alpha}(-1) - \zeta_{F_{\alpha-1}}(-1)$ *est divisible par* p.

Noter par ailleurs que la condition $p > 3$ était nécessaire pour le lemme 2.1, au moins si $\alpha = 0$. En effet si $F = \mathbb{Q}$, $\zeta_F(-1) = -\frac{1}{12}$; d'après (2-3), N_α n'est entier que pour $p > 3$.

Pour comprendre la croissance de N_α, revenons à la proposition 2.4. On a $\zeta_{F_\alpha}(2) \geq 1$, donc

$$\log_p(N_\alpha) \geq \tfrac{3}{2} \alpha p^\alpha (1 + o(1)) - p^\alpha \log_p(4\pi^2) + O(1) \sim \tfrac{3}{2} \alpha p^\alpha .$$

En particulier, N_α croît très vite avec α.

Dans le § 4, nous serons amenés à considérer les fonctions sur S_β ($\beta \geq \alpha$) invariantes par Γ_α, α étant fixé. Le groupe Γ_α opère sur $S_\beta - S_{\beta-1}$ par son quotient $\Gamma_{\beta-1}/\Gamma_\beta \cong \mathbb{Z}/p\mathbb{Z}$, l'action de ce quotient étant libre. La dimension de l'espace des fonctions invariantes sur $S_\beta - S_{\beta-1}$ est donc $\frac{1}{p}(N_\beta - N_{\beta-1})$. Des estimées analogues — à l'aide d'une majoration de $\zeta_{F_\alpha}(2)$, par exemple par $(\zeta_{\mathbb{Q}}(2))^{p^\alpha}$ — montrent que cet espace croît très vite. Il en résulte que l'espace $\mathcal{S}^\alpha_\beta := \mathcal{S}_\beta^{\Gamma_\alpha}$ considéré au § 4 n'est pas constitué (disons sur $\mathbb{C}$) par des formes automorphes provenant par changement de base à partir de formes sur F_α.

3.

Dans ce paragraphe nous définissons *en caractéristique p* des opérateurs de Hecke T_ℓ (ℓ inerte) opérant sur $\mathcal{S}_\infty$.

Soit donc $\ell \neq p$ un nombre premier inerte : $\psi(\ell) \neq 0$. Le corps $\mathbb{Q}_\ell(p^\infty)$ a pour corps résiduel une $\mathbb{Z}_p$-extension $\mathbb{F} = \varinjlim \mathbb{F}_{\ell^{p^\alpha}}$ de $\mathbb{F}_\ell$. On écrit $\mathbb{F}_\alpha$ pour $\mathbb{F}_{\ell^{p^\alpha}}$.

En degré fini α, l'opérateur $T_\ell : \mathcal{S}_\alpha \to \mathcal{S}_\alpha$ est donné (quels que soient les coefficients) par

$$\begin{aligned} T_\ell f(g) &= \sum_{\xi \in \mathbb{F}_\alpha} f\left(g \begin{pmatrix} \ell & \xi \\ & 1 \end{pmatrix}\right) + f\left(g \begin{pmatrix} 1 & \\ & \ell \end{pmatrix}\right) \\ &=: U_\ell f(g) + V_\ell f(g). \end{aligned} \tag{3-1}$$

On choisit des représentants dans l'anneau d'entiers $\mathbb{O}(\mathbb{Q}_\ell(p^\alpha))$ des $\xi \in \mathbb{F}_\alpha$; la valeur de $T_\ell f(g)$ n'en dépend point. La fonction f doit être considérée comme une fonction de

$$g \in Y_\alpha := G(F_\alpha)\backslash G(\mathbb{A}_\alpha^f)/K_\alpha^\ell,$$

où K_α^ℓ est le produit des composantes de K_α aux places ne divisant pas ℓ. L'espace Y_α est donc réunion finie de quotients de $G(F_{\alpha,\ell})$ par des groupes de congruence.

On vérifie aisément, en imitant la démonstration du lemme 2.2, que les applications

$$Y_\alpha \to Y_\beta \qquad (\beta \geq \alpha)$$

sont injectives. En effet, $G(F_\alpha)$ opère librement sur $G(\mathbb{A}_\alpha^f)/K_\alpha^\ell$ (considérer les composantes en ℓ) ; la démonstration précédente s'applique en utilisant le lemme 2.3 aux places ne divisant pas p. Si $f \in \mathcal{S}_\infty$, on peut évidemment considérer f comme une fonction sur

$$Y_\infty = \varinjlim Y_\alpha.$$

On définit formellement T_ℓ par

$$T_\ell = U_\ell + V_\ell$$

l'expression de V_ℓ étant inchangée et U_ℓ étant donné par

$$U_\ell f(g) = \sum_\xi f\left(g\begin{pmatrix}\ell & \xi \\ & 1\end{pmatrix}\right). \tag{3-2}$$

La somme porte maintenant sur les $\xi \in \mathbb{F}$; $g \in Y_\infty$ donc $g \in Y_\beta$ pour un certain β.

Avant de poursuivre, notons T_ℓ^α l'opérateur (3-1), i.e. T_ℓ en degré fini. Le diagramme

$$\begin{array}{ccc} \mathcal{S}_\alpha & \longrightarrow & \mathcal{S}_\beta \\ {\scriptstyle T_\ell^\alpha}\downarrow & & \downarrow{\scriptstyle T_\ell^\beta} \\ \mathcal{S}_\alpha & \longrightarrow & \mathcal{S}_\beta \end{array}$$

$(\beta \geq \alpha)$ n'est certainement pas commutatif, par exemple si l'anneau de coefficients est égal à $\mathbb{C}$. La commutativité impliquerait en effet qu'une valeur propre a_α de T_ℓ^α (opérateur de Hecke en la place $(\ell) = \lambda_\alpha$ de F_α) serait une valeur propre de T_ℓ^β en la place λ_β. Mais les valeurs propres sont, au moins pour les formes issues de formes paraboliques sur GL(2), de la forme

$$a_\alpha = \ell^{p^{\alpha/2}}(u+v) \tag{3-3}$$

où $|u| = |v| = 1$, et on ne peut donc avoir $a_\alpha = a_\beta$ pour des raisons de poids. Noter que le changement de base de Langlands [1980] associe à (3-3) une valeur propre

$$a_\beta = \ell^{p^{\beta/2}}(u^{p^{\beta-\alpha}} + v^{p^{\beta-\alpha}}) \tag{3-4}$$

puisque $p^{\beta-\alpha}$ est le degré de l'extension locale, inerte. On peut vérifier directement, sur l'expression (3-1), que, pour $f \in \mathcal{S}_\alpha$, $T_\beta f$ n'est pas en général à support dans S_α. Pour les mêmes raisons, la somme (3-2) n'est pas évidemment convergente, i.e. finie.

Mais supposons l'anneau des coefficients R de caractéristique p. On considère l'ensemble $\mathcal{F} \subset \mathcal{P}(\mathbb{F})$ des complémentaires des $\mathbb{F}_\alpha$ ($\alpha \geq 0$). Ce n'est pas un filtre, mais il définit une notion de convergence analogue. En particulier, si R est muni de la topologie discrète, on dira qu'une somme $\sum_{\xi\in\mathbb{F}} f(\xi)$ est convergente pour $\mathcal{F}$ si la somme

$$\sum_{\mathbb{F}_0} f(\xi) + \sum_{\mathbb{F}_1-\mathbb{F}_0} f(\xi) + \cdots + \sum_{\mathbb{F}_\alpha-\mathbb{F}_{\alpha-1}} f(\xi) + \cdots$$

est convergente, i.e. si tous les termes sont nuls pour $\alpha \gg 0$.

Proposition 3.1. *Pour $f \in \mathcal{S}_\infty(R)$ la somme*

$$U_\ell f(g) = \sum_\xi f\left(g\begin{pmatrix} \ell & \xi \\ & 1\end{pmatrix}\right)$$

est convergente pour $\mathcal{F}$ si R est de caractéristique p. De plus $T_\ell f = U_\ell f + V_\ell f$ est un élément de $\mathcal{S}^\infty(R)$.

Démonstration. Supposons f à support dans S_α (donc dans Y_α pour le calcul de U_ℓ). Soit $x \in S_\beta$, image de $g \in Y_\beta$. Considérons un terme

$$f\left(g\begin{pmatrix} \ell & \xi \\ & 1\end{pmatrix}\right) \tag{3-5}$$

de (3-2) ; soit $\mathbb{F}_\gamma = \langle \mathbb{F}_\alpha, \mathbb{F}_\beta\rangle$ et supposons que $\xi \notin \mathbb{F}_\gamma$. Soit $\sigma \in \Gamma = \mathrm{Gal}(\mathbb{F}_\ell(\xi)/\mathbb{F}_\gamma)$. Si le terme (3-5) est non nul, $g\left(\begin{smallmatrix} \ell & \xi \\ & 1\end{smallmatrix}\right) \in Y_\alpha$ est invariant par σ. On a alors

$$f\left(g\begin{pmatrix} \ell & \xi \\ & 1\end{pmatrix}\right) = f\left(\sigma\left(g\begin{pmatrix} \ell & \xi \\ & 1\end{pmatrix}\right)\right) = f\left(g\begin{pmatrix} \ell & \sigma\xi \\ & 1\end{pmatrix}\right)$$

car σ fixe g. Puisque ξ n'est pas fixe par Γ, le cardinal de l'orbite de ξ est divisible par p. On a donc pour tout $\delta > \gamma$

$$\sum_{\xi\in\mathbb{F}_\delta-\mathbb{F}_\gamma} f\left(g\begin{pmatrix} \ell & \xi \\ & 1\end{pmatrix}\right) = 0,$$

ce qui démontre la convergence de la somme.

Pour la fin de la démonstration, on peut supposer $\beta \geq \alpha$. Alors (si $x \in S_\beta$) le calcul précédent montre que $T_\ell f(g) = T_\ell^\beta f(g)$; la théorie classique montre alors que $T_\ell f$ est invariante par $K_{\ell,\beta}$ donc définit par restriction un élément de $\mathcal{S}_\beta$. □

La démonstration précédente montre en fait que (pour R de caractéristique p) $T_\ell^\beta f(x) = T_\ell^\alpha f(x)$ si f est à support dans S_α, $x \in S_\alpha$ et $\beta \geq \alpha$. En termes de la restriction $r_\alpha^\beta : \mathcal{S}_\beta \to \mathcal{S}_\alpha$, ceci s'écrit

$$r_\alpha^\beta T_\ell^\beta j_\alpha^\beta = T_\ell^\alpha. \tag{3-6}$$

Si on considère une base de $\mathcal{S}_\infty$ obtenue à l'aide de bases des fonctions sur $S_{\alpha+1} - S_\alpha$, T_ℓ est donc représenté par une matrice doublement infinie dont les blocs diagonaux (pour $\alpha \geq 0$) sont les T_ℓ^α.

Supposons pour simplifier que $R = \bar{\mathbb{F}}_p$.

Écrivons une valeur propre de T_ℓ^α — par exemple venant par réduction modulo p d'une valeur propre (3-3) — sous la forme

$$a_\alpha = u + v \tag{3-7}$$

où nous avons changé de notation et u, v sont les valeurs propres de $\mathrm{Frob}_{\lambda_\alpha}$ dans la représentation galoisienne (de poids géométrique 1) associée, modulo p évidemment. La valeur propre associée pour T_ℓ^β est alors par changement de base, cf. (3-4) :

$$a_\beta = u^{p^{\beta-\alpha}} + v^{p^{\beta-\alpha}}.$$

Mais l'ensemble des valeurs propres (3-7) apparaissant en degré α est, par rationalité ($\mathcal{S}_\alpha$ est défini sur $\mathbb{F}_p$) invariant par $\mathrm{Gal}(\bar{\mathbb{F}}_p/\mathbb{F}_p)$ opérant sur les *coefficients*. On voit donc que $a_\beta = \Phi_p^{\beta-\alpha}(a_\alpha)$ (Frobenius Φ_p *arithmétique*) apparaît dans $\mathcal{S}_\alpha$.

Il en résulte, de même, que toute valeur propre a_α de T_ℓ^α apparaît dans $\mathcal{S}_\beta$. Nous y reviendrons.

4.

Nous considérons maintenant les deux propriétés fondamentales des opérateurs de Hecke dans la théorie classique : ils sont auto-adjoints et forment une famille commutative. On suppose toujours ℓ inerte.

On a construit $T_\ell : \mathcal{S}_\infty \to \mathcal{S}^\infty$. On dispose d'une dualité naturelle

$$\mathcal{S}_\infty \times \mathcal{S}^\infty \to R, \quad \langle f, F \rangle = \sum_{x \in S_\infty} f(x)F(x).$$

Pour fixer les idées, supposons désormais que R est un corps, et en fait que $R = \bar{\mathbb{F}}_p$. Alors $\mathcal{S}_\infty$ est limite inductive d'espaces de dimension finie et $\mathcal{S}^\infty$, son dual, est linéairement compact (compact si R est une extension finie de $\mathbb{F}_p$).

La relation d'adjonction est immédiate :

Proposition 4.1. *Pour $f, g \in \mathcal{S}_\infty$,*

$$\langle T_\ell f, g\rangle = \langle f, T_\ell g\rangle.$$

Démonstration. On peut en effet considérer $j_\alpha : \mathcal{S}_\alpha \to \mathcal{S}_\infty$ et $r_\alpha : \mathcal{S}^\infty \to \mathcal{S}_\alpha$. La relation (3-6) donne alors

$$r_\alpha T_\ell j_\alpha = T_\ell^\alpha.$$

On peut supposer que $f, g \in \mathcal{S}_\alpha$ et donc

$$\langle T_\ell f, g\rangle = \langle r_\alpha T_\ell j_\alpha f, g\rangle = \langle T_\ell^\alpha f, g\rangle_\alpha = \langle f, T_\ell^\alpha g\rangle_\alpha$$

(produit scalaire sur S_α), d'où le résultat par symétrie. □

Puisque T_ℓ ne préserve pas $\mathcal{S}_\infty$, la relation de commutation

$$T_\ell T_m = T_m T_\ell$$

(ℓ, m premiers inertes) n'a pas de sens pour l'instant. Nous devons étendre le domaine de T_ℓ. Noter que

$$\mathcal{S}^\infty = \varprojlim \mathcal{S}_\alpha$$

(limite projective pour la restriction) ; pour $f \in \mathcal{S}_\infty$, la relation (3-6) montre que $T_\ell f \in \mathcal{S}^\infty$ est la limite projective des $T_\ell^\alpha f$ pour $\alpha \geq \gamma$ et $f \in \mathcal{S}_\gamma$. Mais l'application $T_\ell : \mathcal{S}^\infty \to \varprojlim \mathcal{S}_\alpha$ ne s'étend pas continûment à $\mathcal{S}_\infty$.

Notons maintenant Γ le groupe $\mathrm{Gal}(F_\infty/\mathbb{Q})$; on a $\Gamma \cong \mathbb{Z}_p$. Soit également $\Gamma_\alpha = \mathrm{Gal}(F_\infty/F_\alpha)$; alors $\Gamma_\alpha = p^\alpha \mathbb{Z}_p$.

Proposition 4.2. *T_ℓ commute à l'action de* Γ.

Démonstration. Soit en effet $f \in \mathcal{S}_\alpha$ et $g \in Y_\infty$; on peut supposer $g \in Y_\beta$ avec $\beta \geq \alpha$. Alors

$$U_\ell f(g) = \sum_\xi f\left(g \begin{pmatrix} \ell & \xi \\ & 1 \end{pmatrix}\right) \qquad (\xi \in \mathbb{F}_\beta)$$

d'après la preuve de la proposition 3.1. Si $\sigma \in \Gamma$,

$$\begin{aligned} U_\ell f(\sigma g) &= \sum_\xi f\left(\sigma g \begin{pmatrix} \ell & \xi \\ & 1 \end{pmatrix}\right) \\ &= \sum_\xi f\left(\sigma\left(g \begin{pmatrix} \ell & \sigma^{-1}\xi \\ & 1 \end{pmatrix}\right)\right) \\ &= \sum_\xi f\circ\sigma\left(g \begin{pmatrix} \ell & \xi \\ & 1 \end{pmatrix}\right). \end{aligned}$$ □

Soit $\mathcal{S}^\alpha \subset \mathcal{S}^\infty$ le groupe des invariants pour Γ_α :

$$\mathcal{S}^\alpha = H^0(\Gamma_\alpha, \mathcal{S}^\infty)$$

de sorte que $\mathcal{S}^\alpha = \varprojlim \mathcal{S}^\alpha_\beta$ où $\mathcal{S}^\alpha_\beta = H^0(\Gamma_\alpha, \mathcal{S}_\beta)$. Noter que $\mathcal{S}^\alpha$ est le produit des $\mathcal{S}^\alpha_\gamma/\mathcal{S}^\alpha_{\gamma-1}$ ($\gamma \geq 0$), chacun de ces groupes s'identifiant à l'espace des fonction (à valeurs dans $\bar{\mathbb{F}}_p$) sur l'ensemble fini des orbites de Γ_α dans $S_\gamma - S_{\gamma-1}$.

On peut définir de même $\mathcal{S}^\alpha_\infty = \mathcal{S}_\infty \cap \mathcal{S}^\alpha$. D'après la proposition 4.2,

$$T_\ell : \mathcal{S}^\alpha_\infty \to \mathcal{S}^\alpha.$$

Proposition 4.3. *T_ℓ s'étend continûment en un opérateur $T_\ell : \mathcal{S}^\alpha \to \mathcal{S}^\alpha$.*

La topologie sur $\mathcal{S}^\infty$ est la topologie naturelle d'espace linéairement compact.

Démonstration de la proposition 4.3. Soit en effet $F \in \mathcal{S}^\alpha$, et $g \in Y_\beta$. On veut définir

$$U_\ell F(g) = \sum_\xi F\left(g \begin{pmatrix} \ell & \xi \\ & 1 \end{pmatrix}\right). \tag{4-1}$$

On peut supposer $\beta \geq \alpha$. Si $\sigma \in \Gamma_\beta$,

$$F\left(g \begin{pmatrix} \ell & \sigma\xi \\ & 1 \end{pmatrix}\right) = F\left(\sigma\left(\sigma^{-1} g \begin{pmatrix} \ell & \xi \\ & 1 \end{pmatrix}\right)\right) = F\left(g \begin{pmatrix} \ell & \xi \\ & 1 \end{pmatrix}\right),$$

puisque F et g sont fixes par Γ_β. On peut donc restreindre la sommation aux $\xi \in \mathbb{F}_\beta$, et l'opérateur ainsi défini ne dépend pas du choix de β. L'opérateur $T_\ell = U_\ell + V_\ell$ ainsi défini étend évidemment $T_\ell : \mathcal{S}^\alpha_\infty \to \mathcal{S}^\infty$. Son image est dans $\mathcal{S}^\alpha$ comme on le voit en imitant la démonstration de la proposition 4.2. (Ceci résulte aussi de la continuité.) Mais la topologie sur $\mathcal{S}^\infty$ est aussi la topologie de la convergence simple. Si $F \in \mathcal{S}^\alpha$, $F \to 0$,

$$U_\ell F(g) = \sum_\xi F\left(g \begin{pmatrix} \ell & \xi \\ & 1 \end{pmatrix}\right)$$

est donné, pour g fixé, par une somme finie d'après la démonstration précédente, et tend donc vers 0. □

On a alors :

Théorème 4.4. *Pour tout $\alpha \geq 0$, les T_ℓ (ℓ inerte) forment une famille commutative d'opérateurs continus $\mathcal{S}^\alpha \to \mathcal{S}^\alpha$.*

Ceci résulte de la démonstration précédente. Soit ℓ, m inertes et notons $\mathbb{F}_{\ell,\alpha}$, $\mathbb{F}_{m,\alpha}$ les p^α-extensions de $\mathbb{F}_\ell$ et $\mathbb{F}_m$. Pour $g \in Y_\beta$, l'expression (4-1) s'applique à U_ℓ et U_m, les sommes portant sur $\mathbb{F}_{\ell,\beta}$ et $\mathbb{F}_{m,\beta}$. Puisque les matrices $\left(\begin{smallmatrix} \ell & \xi \\ & 1 \end{smallmatrix}\right)$ et $\left(\begin{smallmatrix} m & \eta \\ & 1 \end{smallmatrix}\right)$ (dans $G(F_{\ell,\alpha})$ et $G(F_{m,\alpha})$) commutent le résultat est évident.

La propriété d'auto-adjonction reste vraie dans ce cadre, si l'on définit convenablement le produit scalaire dans $\mathcal{S}^\alpha$. Soit $F, G \in \mathcal{S}^\alpha = \varprojlim \mathcal{S}^\alpha_\beta$. Considérons $\beta \geq \alpha$ et les restrictions de F, G à S_β. On considère leur produit scalaire $\langle F, G\rangle_\beta = \sum_{x \in S_\beta} F(x)G(x)$.

Lemme 4.5. *Si H est une fonction invariante par Γ_α,*

$$\int_{S_\beta} H = \sum_{x\in S_\beta} H(x)$$

ne dépend pas de $\beta \geq \alpha$. En particulier $\langle F, G\rangle_\beta = \langle F, G\rangle_\alpha$ pour $\beta \geq \alpha$.

Démonstration. En effet les orbites de Γ_α sur $S_\beta - S_\alpha$ ont pour ordre des puissances non nulles de p. □

On définit donc, pour $F, G \in \mathcal{S}^\alpha$:

$$\langle F, G\rangle = \langle F, G\rangle_\alpha = \langle F, G\rangle_\beta \qquad (\beta \geq \alpha).$$

La démonstration de la proposition 4.3 montre par ailleurs :

Lemme 4.6. *Si $F \in \mathcal{S}^\alpha$ et $x \in S_\beta$ $(\beta \geq \alpha)$,*

$$T_\ell F(x) = T_\ell^\beta (r_\beta F)(x),$$

où $r_\beta F = F|_{S_\beta}$.

On a donc :

Proposition 4.7. *Pour $F, G \in \mathcal{S}^\alpha$,*

$$\langle T_\ell F, G\rangle = \langle F, T_\ell G\rangle.$$

Démonstration. En effet $\langle T_\ell F, G\rangle = \langle T_\ell F, G\rangle_\alpha = \langle T_\ell^\alpha F, G\rangle_\alpha$, d'où le résultat d'après la propriété classique. Noter que l'accouplement $\langle F, G\rangle$ est évidemment très dégénéré sur $\mathcal{S}^\alpha$. □

5.

Nous considérons maintenant la relation des constructions précédentes avec le changement de base. Fixons α, et soit $f \in \mathcal{S}_\alpha$ forme propre des opérateurs de Hecke :

$$T_\ell^\alpha f = a_\ell f \quad (\ell \text{ inerte}).$$

On a

$$r_\alpha^\beta\, T_\ell^\beta\, j_\alpha^\beta f = a_\ell f \quad (\beta \geq \alpha)$$

mais ceci n'implique pas, évidemment, que $j_\alpha^\beta f \in \mathcal{S}_\beta$ est forme propre. Soit $h \in \mathcal{S}_\beta$: alors

$$\langle T_\ell^\beta h, f\rangle_\beta = \langle h, T_\ell^\beta f\rangle_\beta,$$

où on identifie f à $j_\alpha^\beta f$.

Supposons h invariante par Γ_α ; f l'est évidemment. Alors $T_\ell^\beta f = T_\ell f|_{S_\beta}$ est Γ_α-invariante (proposition 4.2). On a donc

$$\langle T_\ell^\beta h, f\rangle = \langle h, r_\alpha^\beta T_\ell^\beta f\rangle_\alpha$$

(d'après le lemme 4.5). Ainsi

$$\langle T_\ell^\beta h, f\rangle = \langle h, T_\ell^\alpha f\rangle_\alpha = a_\ell\langle h, f\rangle_\alpha = a_\ell\langle h, f\rangle_\beta.$$

Alors $h \mapsto \lambda(h) = \langle h, f\rangle$ est une forme linéaire sur $\mathcal{S}_\beta^\alpha$ telle que $(T_\ell^\beta)^*\lambda = a_\ell\lambda$ pour tout ℓ inerte ; elle est non nulle car on peut choisir h dont la restriction à S_α n'est pas orthogonale à f.

On en déduit par dualité :

Théorème 5.1 (changement de base en degré fini). *Pour tout $\beta \geq \alpha$ il existe $f_\beta \in \mathcal{S}_\beta^\alpha$ telle que*

$$T_\ell^\beta f_\beta = a_\ell f_\beta \quad (\ell \textit{ inerte}).$$

Comme on l'a vu à la fin du §3, ceci résulte du changement de base de Langlands [1980] (et Saito, Shintani) ; on notera cependant que la démonstration présente (en caractéristique p, et pour les ℓ inertes) donne un argument *direct* pour l'existence de f_β.

Nous terminons sur le *problème* suivant. Partant de la forme propre $f \in \mathcal{S}_\alpha$, nous avons construit, pour tout β, une forme $f_\beta \in \mathcal{S}_\beta^\alpha$, forme propre des T_ℓ^β pour la famille de valeurs propres (a_ℓ). L'espace $\mathcal{S}^\alpha$ étant linéairement compact, une sous-suite des formes $(f_\beta)_\beta$ converge vers une forme propre $f \in \mathcal{S}^\alpha$ qui est forme propre des T_ℓ. Mais nous ne savons pas montrer qu'il existe une limite non nulle. Le problème est évidemment d'obtenir une suite (f_β) telle que (pour quelque β_0 fixe) $f_\beta|_{S_{\beta_0}} \neq 0$ pour tout β. Or l'argument de dualité utilisé dans la démonstration du théorème 5.1 ne garantit pas que la forme $f_\beta \in \mathcal{S}_\beta^\alpha$ obtenue a une restriction non nulle à S_α. On aimerait évidemment — peut-être sous des conditions convenables relatives à la famille (a_ℓ) — obtenir une forme propre $F \in \mathcal{S}^\alpha$ pour les opérateurs « infinis » T_ℓ.

On peut cependant obtenir ainsi, partant de $f \in \mathcal{S}_\alpha$, une forme propre généralisée $F \in \mathcal{S}^\alpha$, au sens suivant. Revenons à l'argument précédent. On a considéré f comme une forme linéaire sur $\mathcal{S}_\beta^\alpha$, associée à la valeur propre a_ℓ. Considérons l'espace propre généralisé

$$\bigcup_n \ker\big(((T_\ell^\beta)^* - a_\ell)^n\big)$$

de $(T_\ell^\beta)^*$ dans $(\mathcal{S}_\beta^\alpha)^*$. Il est en dualité parfaite avec l'espace propre généralisé de T_ℓ^β dans $\mathcal{S}_\beta^\alpha$. Il existe donc une forme propre généralisée $h_\beta \in \mathcal{S}_\beta^\alpha$ telle que $\langle h, f\rangle \neq 0$; en particulier $h|_{S_\alpha} \neq 0$.

Puisque l'espace $\mathcal{S}^\alpha = \varprojlim \mathcal{S}_\beta$ est linéairement compact, une sous-suite des h_β donne une forme $F \in \mathcal{S}^\alpha$ de même restriction à S_α. (Par ailleurs l'argument précédent permet d'obtenir une forme propre généralisée simultanée pour les T_ℓ). On a donc :

Proposition 5.2. *Il existe $F \in \mathcal{S}^\alpha$, non nulle, telle que, pour tout $\beta \geq \alpha$, $F_\beta = F|_{S_\beta}$ soit forme propre généralisée des T_ℓ^β, pour les valeurs propres (a_ℓ).*

D'après le lemme 4.6, $T_\ell - a_\ell$ est donc « localement nilpotent », mais le degré de nilpotence dépend a priori de β.

Note (ajoutée sur épreuves)

Après la rédaction de cet article, l'auteur a appris qu'une généralisation étendue de ces résultats (pour les extensions finies) avait été démontrée par D. Treumann et A. Venkatesh (en préparation). Par ailleurs J. Coates lui a montré comment la proposition 2.5 résultait naturellement de la théorie de la fonction zêta p-adique. Son argument est exposé dans une note de l'auteur à paraître (« Formes modulaires modulo p, changement de base et théorie d'Iwasawa »).

Bibliographie

[Langlands 1980] R. P. Langlands, *Base change for* GL(2), Annals of Mathematics Studies **96**, Princeton University Press, 1980. MR 82a:10032 Zbl 0444.22007

[Serre 1971] J.-P. Serre, "Cohomologie des groupes discrets", pp. 77–169 dans *Prospects in mathematics* (Princeton, 1970)), Ann. of Math. Studies **70**, Princeton University Press, 1971. MR 52 #5876 Zbl 0235.22020

[Vignéras 1980] M.-F. Vignéras, *Arithmétique des algèbres de quaternions*, Lecture Notes in Mathematics **800**, Springer, Berlin, 1980. MR 82i:12016 Zbl 0422.12008

Received March 13, 2013. Revised June 13, 2013.

LAURENT CLOZEL
MATHÉMATIQUES
UNIVERSITÉ PARIS-SUD
BÂT. 425
91405 ORSAY CEDEX
FRANCE
laurent.clozel@math.u-psud.fr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 268, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.268.275

WEIGHT ZERO EISENSTEIN COHOMOLOGY OF SHIMURA VARIETIES VIA BERKOVICH SPACES

MICHAEL HARRIS

In memory of Jon Rogawski

We present an alternative interpretation, based on a theorem of Berkovich, of the Eisenstein classes in the cohomology of Shimura varieties, used in forthcoming work of the author with K. W. Lan, R. Taylor, and J. Thorne.

Introduction

This paper represents a first attempt to understand a geometric structure that plays an essential role in my forthcoming paper with Lan, Taylor, and Thorne [Harris et al. 2013] on the construction of certain Galois representations by p-adic interpolation between Eisenstein cohomology classes and cuspidal cohomology. The classes arise from the cohomology of a locally symmetric space Z without complex structure — specifically, the adelic locally symmetric space attached to $\mathrm{GL}(n)$ over a CM field F. It has long been known, thanks especially to the work of Harder and Schwermer (see [Harder 1990], for example) that classes of this type often give rise to nontrivial Eisenstein cohomology of a Shimura variety S; in the case of $\mathrm{GL}(n)$ as above, S is attached to the unitary similitude group of a maximally isotropic hermitian space of dimension $2n$ over F. This is the starting point of the connection with Galois representations. The complete history of this idea will be explained in [Harris et al. 2013]; here I just want to explore a different perspective on the construction of these classes.

By duality, the Eisenstein classes of Harder and Schwermer correspond to classes in cohomology with compact support, and it turns out to be more fruitful to look at them in this way. One of Taylor's crucial observations was that certain of these classes are of weight zero and can therefore be constructed geometrically in any cohomology theory with a good weight filtration, in particular in rigid cohomology, which lends itself to p-adic interpolation. The geometric construction involves the

The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007–2013) / ERC Grant agreement no. 290766 (AAMOT) .
MSC2010: primary 11G18, 14G35; secondary 14G22.
Keywords: Shimura varieties, Eisenstein cohomology, Berkovich spaces.

abstract simplicial complex Σ defined by the configuration of boundary divisors of a toroidal compactification; this complex, which is homotopy equivalent to the original locally symmetric space Z, arises in the calculation of the weight filtration on the cohomology of the logarithmic de Rham complex. Under certain conditions (as I was reminded by Wiesia Nizioł, commenting on the construction in [Harris et al. 2013][1]) Berkovich has defined an isomorphism between the weight-zero cohomology with compact support of a scheme and the compactly supported cohomology of the associated Berkovich analytic space, which is a topological space. His results apply to both ℓ-adic and p-adic étale cohomology as well as to Hodge theory. In this paper we apply this isomorphism to the toroidal compactification S' of a Shimura variety S. Both S and S' are defined over some number field E; we fix a place v of E dividing the rational prime p and let $|S|$ and $|S'|$ be the associated analytic spaces over E_v in the sense of Berkovich. We observe that Σ is homotopy equivalent to $|S'| \setminus |S|$.[2] Moreover, when S and S' both have good reduction at v, $|S|$ and $|S'|$ are both contractible [Berkovich 1999], and it follows easily that the cohomology of Σ maps to $H_c^*(|S'|)$ in the theories considered in [Berkovich 2000].

These ideas will be worked out systematically in forthcoming work. The present note explains the construction in the simplest situation. We only consider cohomology with trivial coefficients of Shimura varieties with a single class of rational boundary components, assumed to be of dimension 0. We work with connected rather than adelic Shimura varieties and write the boundary as a union of connected quotients of a (nonhermitian) symmetric space by discrete subgroups. We also only work at places of good reduction, in order to quote Berkovich's theorems directly. In [Harris et al. 2013] it is crucial to consider arbitrary level, but the relevant target spaces are the ordinary loci of Shimura varieties. Perhaps Berkovich's methods apply to these spaces as well, but for the moment there would be no way to use such an application, since the results of [Berkovich 2000] have not been verified for rigid cohomology.

Berkovich gives a topological interpretation of the weight zero stage of the Hodge filtration, but it can also be used as a topological definition of this part of the cohomology. Since the cohomology of Σ has a natural integral structure, it's conceivable that the results of Berkovich provide some information about torsion in cohomology. This is one of the main motivations for reconsidering the construction of [Harris et al. 2013] in the light of Berkovich's theory.

---

[1] Since writing the first version of this article I have learned that Laurent Fargues had essentially the same idea independently.

[2] Our Shimura varieties are attached to groups of rational rank 1, whose toroidal boundary is the blowup of point boundary components in the minimal compactification. More general Shimura varieties are compactified by adding strata attached to different conjugacy classes of maximal parabolics, and then $|S'| \setminus |S|$ has several strata as well.

After this paper was written, but before [Harris et al. 2013] was made public, Peter Scholze found a different construction of the Galois representations studied there, based on the theory of perfectoid spaces [Scholze 2013]. His method is stronger in that it applies to torsion classes as well as to classes in characteristic zero; it also treats all systems of coefficients simultaneously, because they all become trivial in the perfectoid limit. Scholze's method does not use the weight arguments that were crucial in [Harris et al. 2013] and that are the motivation for the present paper. In particular, Scholze's theory provides affirmative answers to the analogues of Questions 2.4 and 2.6 below. The questions remain meaningful in the framework of Berkovich's theory.

I also thank my coauthors Kai-Wen Lan, Richard Taylor, and Jack Thorne, for providing the occasion for the present paper; Wiesia Nizioł, for pointing out the connection with Berkovich's work; and Sam Payne, for explaining the results of Berkovich and Thuillier.

Jon Rogawski was exceptionally generous in person. Although we met rarely, on at least two separate occasions he took the time to help me find my way around technical problems central to the success of my work. He will be greatly missed.

1. The construction

The standard terminology and notation for Shimura varieties will be used without explanation. Let (G, X) be the datum defining a Shimura variety $S(G, X)$, with G a connected reductive group over $\mathbb{Q}$ and X a union of copies of the hermitian symmetric space attached to the identity component of $G(\mathbb{R})$. Let D be one of these components and let $\Gamma \subset G(\mathbb{Q})$ be a congruence subgroup; then $S = \Gamma \backslash D$ is a connected component of $S(G, X)$ at some finite level K; here K is an open compact subgroup of $G(\boldsymbol{A}_f)$. Then S has a canonical model over some number field $E = E(D, \Gamma)$. We assume Γ is neat; then S is smooth and has a family of smooth projective toroidal compactifications, as in [Ash et al. 1975]. We make a series of simplifying hypotheses.

Hypothesis 1.1. *The group G has rational rank one. Let P be a rational parabolic subgroup of G* (*unique up to conjugacy*); *then P is the stabilizer of a point boundary component of D.*

It follows from the general theory in [Ash et al. 1975] that if $S' \supset S$ is a toroidal compactification then the complement $S' \backslash S$ is a union of rational divisors. We pick such an S', assumed smooth and projective, and assume that $S' \backslash S$ is a divisor with normal crossings. Let v be a place of E dividing the rational prime p.

Hypothesis 1.2. *The varieties S and S' have smooth projective models $\mathfrak{S}$ and $\mathfrak{S}'$ over the v-adic integer ring* $\mathrm{Spec}(\mathbb{O}_v)$.

This is proved by Lan for PEL type Shimura varieties at hyperspecial level in several long papers, starting with [Lan 2013].

Let $\bar{\mathfrak{S}}$ denote the base change of $\mathfrak{S}$ to the algebraic closure of F_v, $\bar{\mathfrak{S}}^{\mathrm{an}}$ the associated Berkovich analytic space, and $|S|$ (rather than $|\bar{\mathfrak{S}}^{\mathrm{an}}|$) the underlying topological space. We use the same notation for S'. Let $Z = |S'| \setminus |S|$.

Lemma 1.3. *The spaces $|S|$ and $|S'|$ are contractible and $|S'|$ is compact. In particular*:

(a) *The inclusion $|S| \hookrightarrow |S'|$ is a homotopy equivalence.*

(b) *There are canonical isomorphisms $H^i_c(|S|, A) \xrightarrow{\sim} H^i(|S'|, Z; A)$ for any ring A.*

(c) *The connecting homomorphism $H^i(Z, A) \to H^{i+1}(|S'|, Z; A) = H^{i+1}_c(|S'|, A)$ is an isomorphism for $i > 0$.*

Proof. Contractibility of $|S|$ and $|S'|$ follows from Hypothesis 1.2 by the results of [Berkovich 1999, Section 5] (though the contractibility of analytifications of spaces with good reduction seems only to be stated explicitly in the introduction). Since S' is proper, $|S'|$ is compact. Then (a) and (b) are clear and (c) follows from the long exact sequence for cohomology

$$\cdots \to H^i(|S'|, A) \to H^i(Z, A) \to H^{i+1}(|S'|, Z; A) \to H^{i+1}(|S'|, A) \to \cdots \qquad \square$$

Let $P = LU$ be a Levi decomposition, with L reductive and U unipotent, let L^0 denote the identity component of the Lie group $L(\mathbb{R})$, and let D_P denote the symmetric space attached to L^0 (or to its derived subgroup $(L^0)^{\mathrm{der}}$). The *minimal compactification* (or Satake compactification) S^* of S is a projective algebraic variety obtained by adding a finite set of points, say N points, which we can call "cusps," to S. The *toroidal compactifications* S^{tor} of [Ash et al. 1975], which depend on combinatorial data, are constructed by blowing up the cusps; each one is replaced by a configuration of rational divisors, to which we return momentarily. For appropriate choices of data S^{tor} is a smooth projective variety and $\partial S^{\mathrm{tor}} = S^{\mathrm{tor}} \setminus S$ is a divisor with normal crossings; $\partial S^{\mathrm{tor}}$ is a union of N connected components, one for each cusp. The *reductive Borel–Serre compactification* S^{rs} of S is a compact (nonalgebraic) manifold with corners (boundary in this case) containing S as dense open subset, and such that

$$S^{\mathrm{rs}} \setminus S = \coprod_{j=1}^{N} \Delta_j \backslash D_P \tag{1.4}$$

where, for each j, Δ_j is a cocompact congruence subgroup of $L(\mathbb{Q})$.

Details on S^{rs} can be found in a number of places, for example [Borel and Ji 2006]. We introduce this space only in order to provide an independent description of Z. Roughly speaking, Z is canonically homotopy equivalent to $S^{\mathrm{rs}} \setminus S$. More precisely, let $S \hookrightarrow S^{\mathrm{tor}}$ be a toroidal compactification as above and consider the *incidence complex* Σ of the divisor with normal crossings $\partial S^{\mathrm{tor}}$. This is a simplicial complex whose vertices are the irreducible components ∂_i of $\partial S^{\mathrm{tor}}$, whose edges are the nontrivial intersections $\partial_i \cap \partial_j$, and so on.

Proposition 1.5. *The incidence complex* Σ *is homeomorphic to a triangulation of* $S^{\mathrm{rs}} \backslash S$.

Proof. In [Harris and Zucker 1994, Corollary 2.2.10], it is proved that Σ is a triangulation of a compact deformation retract of $S^{\mathrm{rs}} \backslash S$; but under Hypothesis 1.1 $S^{\mathrm{rs}} \backslash S$ is already compact. In any case, Σ and $S^{\mathrm{rs}} \backslash S$ are homotopy equivalent. □

Theorem 1.6. (*Berkovich and Thuillier*) *There is a canonical deformation retraction of* $Z = S' \backslash S$ *onto* Σ.

This is proved but not stated in [Thuillier 2007], and can also be extracted from [Berkovich 1999]. A more precise reference will be provided in the sequel.

In what follows, $H_c^\bullet$ will be one of these cohomology theories, considered in [Berkovich 2000, Theorem 1.1]:

(a′) $H_{\ell,c}^\bullet$ (ℓ-adic étale cohomology, with $\ell \neq p$);

(a″) $H_{p,c}^\bullet$ (p-adic étale cohomology); or

(c) $V \mapsto H_c^\bullet(V(\mathbb{C}), \mathbb{Q})$ (Betti cohomology with compact support of the complex points of the algebraic variety V).

Corresponding to the choice of $H_c^\bullet$, the ring A is either (a′) $\mathbb{Q}_\ell$, (a″) $\mathbb{Q}_p$, or (c) $\mathbb{Q}$.

Corollary 1.7. *For* $i > 0$, *there is a canonical injection*

$$\phi : H^i(S^{\mathrm{rs}} \backslash S, A) = H^i\left(\coprod_{j=1}^{N} \Delta_j \backslash D_P, A\right) \hookrightarrow H_c^{i+1}(\bar{S}).$$

The image of ϕ *is the weight-zero subspace in cases* (a′) *and* (c) *and is the space of smooth vectors for the action of the Galois group* (see [Berkovich 2000, page 666] for the definition) *in case* (a″).

Proof. This follows directly from Proposition 1.5, Theorem 1.6 and [Berkovich 2000, Theorem 1.1]. □

The key word is *canonical*. This means that the retractions commute with change of discrete group Γ (provided the condition of Hypothesis 1.2 is preserved) and with Hecke correspondences, or (more usefully) the action of the group $G(\boldsymbol{A}_f)$ in the adelic Shimura varieties. In particular, the adelic version of the corollary asserts roughly that the induced representation from $P(\boldsymbol{A}_f)$ to $G(\boldsymbol{A}_f)$ of the topological cohomology of the locally symmetric space attached to L injects into the cohomology with compact support of the adelic Shimura variety $S(G, X)$, with image either the weight zero subspace or the smooth vectors for the Galois action.

2. Some extensions and questions

Comment 2.1. Hypothesis 1.1 is superfluous. The homotopy type of $|S'| \setminus |S|$ is more complicated but can be described along lines similar to Theorem 1.6.

Comment 2.2. The article [Harris et al. 2013] treats more general local coefficients by studying the weight-zero cohomology of Kuga families of abelian varieties over Shimura varieties. The analytic space of the boundary in this case is a torus bundle with fiber $(S^1)^d$ over the base Z, where d is the relative dimension of the Kuga family over S. The Leray spectral sequence identifies the cohomology of the total space as the cohomology of Z with coefficients in a sum of local systems attached to irreducible representations of L. In this way one can recover the Eisenstein classes of [Harris et al. 2013] for general coefficients.

Question 2.3. In [Harris and Zucker 1994] the combinatorial calculation of the boundary contribution to coherent cohomology is accompanied by a differential calculation, in which the Dolbeault complex near the toroidal boundary is compared to the de Rham complex on the incidence complex. Does this have analogues in other cohomology theories?

Question 2.4. Is there a version of Berkovich's theorem in [Berkovich 2000] for local systems that works directly with Z and S and avoids the use of Kuga families? For ℓ prime to p, local systems over S with coefficients in $\mathbb{Z}/\ell^n\mathbb{Z}$, attached to algebraic representations of G, become trivial when Γ is replaced by an appropriate subgroup of finite index. This suggests that the analogue of Corollary 1.7 for ℓ-adic cohomology with twisted coefficients can be proved directly on the adelic Shimura variety. It's not so clear how to handle cases (a$''$) and (c).

Question 2.5. Does Berkovich's theorem apply to rigid cohomology, which is the theory used in [Harris et al. 2013]? In particular, does it apply to the ordinary locus of the toroidal compactification?

Question 2.6. Most importantly, is there a version of Corollary 1.7 that keeps track of the torsion cohomology of $S^{\mathrm{rs}} \setminus S$? The possibility of assigning Galois representation to torsion cohomology classes is the subject of a series of increasingly precise and increasingly influential conjectures. Can the methods of [Harris et al. 2013] be adapted to account for these classes?

References

[Ash et al. 1975] A. Ash, D. Mumford, M. Rapoport, and Y. Tai, *Smooth compactification of locally symmetric varieties*, Lie Groups: History, Frontiers and Applications **4**, Mathematical Science Press, Brookline, MA, 1975. MR 56 #15642 Zbl 0334.14007

[Berkovich 1999] V. G. Berkovich, "Smooth p-adic analytic spaces are locally contractible", *Invent. Math.* **137**:1 (1999), 1–84. MR 2000i:14028 Zbl 0930.32016

[Berkovich 2000] V. G. Berkovich, "An analog of Tate's conjecture over local and finitely generated fields", *Internat. Math. Res. Notices* 13 (2000), 665–680. MR 2001h:14022 Zbl 1068.14502

[Borel and Ji 2006] A. Borel and L. Ji, "Compactifications of locally symmetric spaces", *J. Differential Geom.* **73**:2 (2006), 263–317. MR 2007d:22031 Zbl 1122.22005

[Harder 1990] G. Harder, "Some results on the Eisenstein cohomology of arithmetic subgroups of GL_n", pp. 85–153 in *Cohomology of arithmetic groups and automorphic forms* (Luminy-Marseille, 1989), edited by J.-P. Labesse and J. Schwermer, Lecture Notes in Math. **1447**, Springer, Berlin, 1990. MR 91j:11040 Zbl 0719.11034

[Harris and Zucker 1994] M. Harris and S. Zucker, "Boundary cohomology of Shimura varieties, I: Coherent cohomology on toroidal compactifications", *Ann. Sci. École Norm. Sup.* (4) **27**:3 (1994), 249–344. MR 95a:14024 Zbl 0860.11030

[Harris et al. 2013] M. Harris, K.-W. Lan, R. Taylor, and J. Thorne, "On the rigid cohomology of certain Shimura varieties", preprint, 2013, Available at http://www.math.umn.edu/~kwlan/articles/rigcoh.pdf.

[Lan 2013] K.-W. Lan, *Arithmetic compactifications of PEL-type Shimura varieties*, Princeton University Press, 2013. MR 2711676 Zbl 06151348

[Scholze 2013] P. Scholze, "On torsion in the cohomology of locally symmetric varieties", preprint, 2013. arXiv 1306.2070

[Thuillier 2007] A. Thuillier, "Géométrie toroïdale et géométrie analytique non Archimédienne: application au type d'homotopie de certains schémas formels", *Manuscripta Math.* **123**:4 (2007), 381–451. MR 2008g:14038 Zbl 1134.14018

Received July 6, 2012.

MICHAEL HARRIS
INSTITUT DE MATHÉMATIQUES DE JUSSIEU
U.M.R. 7586 DU CNRS
FRANCE

harris@math.jussieu.fr

and

UFR DE MATHÉMATIQUES
UNIVERSITÉ PARIS 7 DENIS DIDEROT
FRANCE

PACIFIC JOURNAL OF MATHEMATICS
Vol. 268, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.268.283

Λ-ADIC BARSOTTI–TATE GROUPS

HARUZO HIDA

To my dear friend Jon Rogawski, who created the UCLA number theory group.

We define Λ-BT groups as well-controlled ind-Barsotti–Tate groups under the action of the Iwasawa algebra and construct a prototypical example of such groups out of modular Jacobians. We then discuss the relation of these groups to Weil numbers of weight 1 and to the nonvanishing problem of the adjoint $\mathscr{L}$-invariant.

1. Introduction

Fix a prime $p \geq 5$ (throughout the paper). For a given valuation ring R, a Λ-BT group $\mathscr{G} = \mathscr{G}_R$ is by definition an inductive limit of (p-divisible) Barsotti–Tate groups $\mathscr{G}_n = \mathscr{G}_{n,R}$ $(0 < n \in \mathbb{Z})$ defined over R with an action of the Iwasawa algebra $\Lambda = \Lambda_W := W[\![x]\!]$ as endomorphisms. Here the limit is taken as an object of the ind-category of commutative group schemes over R or in the (bigger) abelian category of abelian fppf sheaves over R (see [Hida 2012, §1.12.1] for abelian fppf sheaves), W is a discrete valuation ring finite flat over $\mathbb{Z}_p$, and $W[\![x]\!]$ is the ring of power series in one variable. We write K for the quotient field of R and $\mathbb{F}$ for the residue field; $\overline{K}$, $\overline{\mathbb{F}}$ denote algebraic closures thereof. We assume R to be of mixed characteristic $(0, p)$ (so K has characteristic 0 and $\mathbb{F}$ has characteristic p),

This is an expanded version of the notes from lectures delivered in September 2005 at the Centre de Recherches Mathématiques, Montreal, with the addition of some comments on the results received after the lecture. The author is partially supported by the NSF grants DMS 0753991 and DMS 0854949.
MSC2010: primary 11F11, 11F25, 11F33, 11F80, 14L05; secondary 11F32.
Keywords: Hecke algebra, Galois representation, Galois deformation, Barsotti–Tate group.

though the definition is obviously valid over any valuation ring (and it is interesting to know if there is some good theory over a general R). We impose the following two conditions:

(CT′) Writing $\gamma = 1+x$, we have $\mathcal{G}_{n,R} = \mathcal{G}_R[\gamma^{p^{n-1}} - 1] := \mathrm{Ker}(\gamma^{p^{n-1}} - 1 : \mathcal{G}_R \to \mathcal{G}_R)$ (in particular, $\mathcal{G}_{n,R} \hookrightarrow \mathcal{G}_R$ is a closed immersion).

(DV) The geometric generic fiber $\mathcal{G}(\bar{K})$ is isomorphic to $(\Lambda^*)^n$ for the Pontryagin dual $\Lambda^* := \mathrm{Hom}_{\mathbb{Z}_p}(\Lambda, \mathbb{Q}_p/\mathbb{Z}_p)$, so $T\mathcal{G} = \mathrm{Hom}_\Lambda(\Lambda^*, \mathcal{G}(\bar{K}))$ is Λ-free of rank n.

In the text (see (CT) in Section 3), we often impose a slightly stronger condition than (CT′) here (though we have an example of $\mathcal{G}$ satisfying only (CT′) not (CT); see Remark 5.5). Such group schemes have been studied in [Hida 1986; Mazur and Wiles 1986; Tilouine 1987; Ohta 1995] and more recently in [Cais 2012, §5.4], primarily through the deformation theory of modular forms and Galois representations. In this note, we would like to give some basic facts of Λ-BT groups and to point out their relation to the $\mathcal{L}$-invariant of adjoint square L-functions of modular forms. In some sense, this note is a revisiting of the topics presented in [Hida 1986] through a new formulation via arithmetic geometry, and in near future, we hope to delve into deeper in this direction. We write $\Gamma = \gamma^{\mathbb{Z}_p}$ for the subgroup of $\Lambda^\times$ topologically generated by γ, which is isomorphic to $1 + p\mathbb{Z}_p$ (we fix the isomorphism $1 + p\mathbb{Z}_p \cong \Gamma$ sending $1 + p$ to $1 + x$).

Starting with a brief explanation of $U(p)$-isomorphisms in Section 2, in Section 3, we list expected properties (other than (CT′) and (DV)) of Λ-BT groups, and we discuss seemingly naive questions on Λ-BT groups. Some of the implications of the questions will be discussed in later sections. In Sections 4 and 5, we construct modular Λ-BT groups, and verify most of the properties listed in Section 3 for the modular Λ-BT groups. In Section 6, we prove the remaining properties in Section 3 related to reduction modulo p of the modular Λ-BT groups. In the last two Sections 7 and 8, we relate the theory to the problem of nonvanishing of the adjoint $\mathcal{L}$-invariants. We note that p-adic Hodge theory of the modular Λ-BT group, as well as other aspects, has been studied by Cais [2012], who may have been influenced by my lecture notes [Hida 2005] at CRM, which were the origin of the present paper.

2. $U(p)$-isomorphisms

For $\mathbb{Z}[U]$-modules X and Y, we call a $\mathbb{Z}[U]$-linear map $f : X \to Y$ a U-injection (resp. a U-surjection) if $\mathrm{Ker}(f)$ is killed by a power of U (resp. $\mathrm{Coker}(f)$ is killed by a power of U). If f is a U-injection and a U-surjection, we call f a U-isomorphism. In other words, f is a U-injection, a U-surjection, or a U-isomorphism if, after tensoring with $\mathbb{Z}[U, U^{-1}]$, it becomes an injection, a surjection, or an isomorphism.

In terms of U-isomorphisms, we describe briefly the facts we study in this article (and in later sections, we fill in more details in terms of the ordinary projector e).[1]

Let N be a positive integer prime to p. We consider the (open) modular curve $Y_r := Y_1(Np^r)_{/\mathbb{Q}}$ which classifies elliptic curves E with an embedding $\phi : \mu_{Np^r} \hookrightarrow E[Np^r] = \mathrm{Ker}(Np^r : E \to E)$. Let $R_i = \mathbb{Z}_{(p)}[\mu_{p^i}]$ and $K_i = \mathbb{Q}[\mu_{p^i}]$ ($i = 1, 2, \dots, \infty$). We fix an isomorphism $\mathbb{Z}_p(1) = \varprojlim_r \mu_{p^r}(R_\infty)$ choosing a coherent sequence of primitive roots of unity $\zeta_{p^r} \in \mu_{p^r}(R_r)$ such that $\zeta_{p^{r+1}}^p = \zeta_{p^r}$ for all r, and therefore R_i has a specific primitive root of unity denoted by ζ_{p^i}. Similarly, we fix an isomorphism $\mathbb{Z}/N\mathbb{Z} \cong \mu_N$ over $\mathbb{Q}(\mu_N)$ (given by $m \mapsto \zeta_N^m$) choosing a primitive root of unity ζ_N. We write $\zeta_{Np^r} = \zeta_N \zeta_{p^r}$. Let R be either a valuation ring or a field (over $\mathbb{Z}_{(p)}$) inside $\mathbb{Q}[\mu_{p^\infty}]$ with quotient field K. We write $X_{r,R}$ for the normalization of the j-line $\boldsymbol{P}(j)_{/R}$ in the function field of $Y_{r/K}$. The group $z \in (\mathbb{Z}/p^r\mathbb{Z})^\times$ acts on $Y_{r,K}$ by $\phi \mapsto \phi \circ z$ (and hence on its normalization $X_{r,R}$), as $\mathrm{Aut}(\mu_{Np^r}) \cong (\mathbb{Z}/Np^r\mathbb{Z})^\times$. Thus $\Gamma = 1 + p\mathbb{Z}_p = \gamma^{\mathbb{Z}_p}$ acts on X_r (and its Jacobian) through its image in $(\mathbb{Z}/Np^r\mathbb{Z})^\times$. For $s > r \geq 0$, we define another modular curve $Y^r_{s,K}$ by the geometric quotient of Y_s by $(1 + p^r\mathbb{Z}_p)/(1 + p^s\mathbb{Z}_p) \subset (\mathbb{Z}/Np^s\mathbb{Z})^\times$ and define $X^r_{s,R}$ to be the normalization of $\boldsymbol{P}(j)_{/R}$ in the function field $K(Y^r_{s,K})$. Then $X^r_{s,K}(\mathbb{C})$ is given by $\Gamma^r_s \backslash (\mathfrak{H} \sqcup \boldsymbol{P}^1(\mathbb{Q}))$ for $\Gamma^r_s = \Gamma_1(Np^r) \cap \Gamma_0(p^s)$ ($s > r \geq 0$). Hereafter we take $U = U(p)$ for the Hecke operator $U(p)$ (as defined in [Hida 2012, §3.2.3 and §4.2.1]).

As before, take a valuation ring $R \subset \mathbb{Q}[\mu_{p^\infty}]$ over $\mathbb{Z}_{(p)}$. Let $J_{r,R} = \mathrm{Pic}^0_{X_{r,R}/R}$ be the connected component of the Picard scheme. Then $J_{r,R}$ is the identity connected component of the Néron model of the Jacobian $J_{r/K}$ of $X_{r/K}$. Indeed, by the table of geometric multiplicities of irreducible components of $X_{r/\mathbb{F}_p} = X \times_R \mathbb{F}_p$ in [Katz and Mazur 1985, 13.5.6], the greatest common divisor D of the geometric multiplicities of irreducible components of $X_{r/\mathbb{F}_p}$ is equal to 1. Then by a result of Raynaud [Bosch et al. 1990, Theorem 9.5.4(b)], the identity $D = 1$ implies that $J_{r,R}$ is isomorphic to the identity connected component of the Néron model of $J_{r/K}$ over R. Similarly, $J^r_s := \mathrm{Pic}^0_{X^r_s/R}$ is isomorphic to the identity connected component of the Néron model over R of the Jacobian variety $J^r_{s/K}$ of the modular curve $X^r_{s/K}$. Note that

$$\text{(2-1)}\qquad \Gamma^r_s \backslash \Gamma^r_s \begin{pmatrix} 1 & 0 \\ 0 & p^{s-r} \end{pmatrix} \Gamma_1(Np^r) = \left\{ \begin{pmatrix} 1 & a \\ 0 & p^{s-r} \end{pmatrix} \,\middle|\, a \bmod p^{s-r} \right\}$$
$$= \Gamma_1(Np^r) \backslash \Gamma_1(Np^r) \begin{pmatrix} 1 & 0 \\ 0 & p^{s-r} \end{pmatrix} \Gamma_1(Np^r).$$

Now, write $U^s_r(p^{s-r}) : J^s_r \to J_r$ for the Hecke operator of $\Gamma^s_r \alpha_{s-r} \Gamma_1(Np^r)$ for

[1] This section about U-isomorphisms is from a conference talk at CRM in September in 2005 (see http://www.crm.umontreal.ca/Representations05/indexen.html).

$\alpha_m = \left(\begin{smallmatrix}1 & 0\\ 0 & p^m\end{smallmatrix}\right)$. As described in [Shimura 1971, Chapter 7], for a modular curve $X(\Gamma) := \Gamma\backslash(\mathfrak{H} \sqcup \boldsymbol{P}^1(\mathbb{Q}))$, each double coset $\Gamma\alpha\Gamma'$ gives rise to a correspondence $X(\Gamma \cap \alpha\Gamma'\alpha^{-1})$ embedded in $X(\Gamma) \times X(\Gamma')$ by $z \mapsto (z, \alpha(z))$. In our cases of $\Gamma = \Gamma^r_s, \Gamma_1(Np^r)$ and $\Gamma' = \Gamma_1(Np^r)$, the modular curve $X(\Gamma\cap\alpha\Gamma'\alpha^{-1})$ is known to be defined over $\mathbb{Q}$ [loc. cit.], and hence the correspondences are also defined over $\mathbb{Q}$. These correspondences defined over $\mathbb{Q}$ act on the Jacobians by morphisms defined over $\mathbb{Q}$ (by Picard and Albanese functoriality, respectively) and its composition relation verified over $\mathbb{C}$ remains valid over $\mathbb{Q}$ (and over any subfield in $\mathbb{C}$). Then we have the following commutative diagram from the above identity, first over $\mathbb{C}$, then over K (via the correspondences defined over $\mathbb{Q}$ and hence over K) and by functoriality (of Picard schemes or Néron models) over R:

$$\begin{array}{ccc} J_{r,R} & \xrightarrow{\pi^*} & J^r_{s,R} \\ {\scriptstyle u}\downarrow & \swarrow{\scriptstyle u'} & \downarrow{\scriptstyle u''} \\ J_{r,R} & \xrightarrow{\pi^*} & J^r_{s,R} \end{array} \tag{2-2}$$

where the middle u' is given by $U^s_r(p^{s-r})$ and u and u'' are $U(p^{s-r})$. Thus

(u1) $\pi^* : J_{r,R} \to J^r_{s,R}$ is a $U(p)$-isomorphism (for the projection $\pi : X^r_s \to X_r$).

Taking the dual $U^*(p)$ of $U(p)$ with respect to the Rosati involution induced by the canonical polarization on the Jacobians, we have a dual version of the above diagram for $s > r > 0$:

$$\begin{array}{ccc} J_{r,R} & \xleftarrow{\pi_*} & J^r_{s,R} \\ {\scriptstyle u^*}\uparrow & \nearrow{\scriptstyle u'^*} & \uparrow{\scriptstyle u''^*} \\ J_{r,R} & \xleftarrow{\pi_*} & J^r_{s,R} \end{array} \tag{2-3}$$

Here the superscript "$*$" indicates the Rosati involution corresponding to the canonical divisor on the Jacobians, and $u^* = U^*(p)^{s-r}$ for the level $\Gamma_1(Np^r)$ and $u''^* = U^*(p)^{s-r}$ for Γ^r_s. Without applying the duality, these morphisms come directly from Hecke correspondences associated to the following coset decomposition:

$$\begin{aligned} \Gamma^r_s\Big\backslash\Gamma^r_s\begin{pmatrix}p^{s-r}s & 0\\ 0 & 1\end{pmatrix}\Gamma_1(Np^r) &= \Gamma_1(Np^r)\Big\backslash\Gamma_1(Np^r)\begin{pmatrix}p^{s-r} & 0\\ 0 & 1\end{pmatrix}\Gamma^r_s \\ &= \left\{\begin{pmatrix}p^{s-r} & a\\ 0 & 1\end{pmatrix}\,\middle|\, a \bmod p^{s-r}\right\} \\ &= \Gamma_1(Np^r)\Big\backslash\Gamma_1(Np^r)\begin{pmatrix}p^{s-r} & 0\\ 0 & 1\end{pmatrix}\Gamma_1(Np^r). \end{aligned} \tag{2-4}$$

Alternatively, the diagram (2-3) follows from Albanese functoriality of Jacobians applied to the Hecke correspondence $U(p)$. In any case, we get

(u1*) $\pi_* : J_{r,R} \to J^r_{s,R}$ is a $U^*(p)$-isomorphism, where π_* is the dual of π^*.

In particular, if we take the ordinary and the coordinary projectors

$$e = \lim_{n\to\infty} U(p)^{n!} \quad \text{and} \quad e^* = \lim_{n\to\infty} U^*(p)^{n!}$$

on $J[p^\infty]$ for $J = J_r, J_s, J^r_s$, noting $U(p^m) = U(p)^m$, we have

$$\pi^* : J_r[p^\infty]^{\mathrm{ord}} \cong J^r_s[p^\infty]^{\mathrm{ord}} \quad \text{and} \quad \pi_* : J^r_s[p^\infty]^{\mathrm{coord}} \cong J_r[p^\infty]^{\mathrm{coord}},$$

where "ord" (resp. "coord") indicates the image of the projector e (resp. e^*). For simplicity, we write $\mathscr{G}_{r,R} := J_r[p^\infty]^{\mathrm{ord}}_{/R}$. The group scheme $J_r[p^n]$ is often neither flat nor finite (for example if $J_{r/\mathbb{F}_p}$ has additive part). Thus for the moment, as explained in the introduction, we take $J_r[p^\infty]^{\mathrm{ord}}_{/R}$ as defined either in the ind-category of group schemes over R or in the abelian category of fppf abelian sheaves. Our point is that we will show that the ordinary part behaves well under Picard functoriality and is represented by a Λ-BT group if $R \supset R_\infty$.

Pick a congruence subgroup Γ defining the modular curve $X = X(\Gamma)$, and write its Jacobian as J. We now identify $J(\mathbb{C})$ with a subgroup of $H^1(\Gamma, \boldsymbol{T})$ (for the Γ-module $\boldsymbol{T} := \mathbb{R}/\mathbb{Z} \cong \{z \in \mathbb{C} \mid |z| = 1\}$ with trivial Γ-action). Since $\Gamma^r_s \rhd \Gamma_1(Np^s)$, we may consider the finite cyclic quotient group $C := \Gamma^r_s/\Gamma_1(Np^s)$. By the inflation restriction sequence, we have the following commutative diagram with exact rows:

(2-5)
$$\begin{array}{ccccccc}
H^1(C,\boldsymbol{T}) & \hookrightarrow & H^1(\Gamma^r_s,\boldsymbol{T}) & \longrightarrow & H^1(\Gamma_1(Np^s),\boldsymbol{T})^{\gamma^{p^{r-1}}=1} & \longrightarrow & H^2(C,\boldsymbol{T}) \\
\uparrow & & \uparrow \cup & & \uparrow \cup & & \uparrow \\
? & \longrightarrow & J^r_s(\mathbb{C}) & \longrightarrow & J_s(\mathbb{C})[\gamma^{p^{r-1}}-1] & \longrightarrow & ?
\end{array}$$

Since C is a finite cyclic group of order p^{s-r} (with generator g) acting trivially on $\boldsymbol{T}$, we have $H^1(C,\boldsymbol{T}) = \mathrm{Hom}(C,\boldsymbol{T}) \cong C$ and

$$H^2(C,\boldsymbol{T}) = \boldsymbol{T}/(1+g+\cdots+g^{p^{s-r}-1})\boldsymbol{T} = \boldsymbol{T}/p^{s-r}\boldsymbol{T} = 0.$$

By the same token, replacing $\boldsymbol{T}$ by $\mathbb{T}_p := \mathbb{Q}_p/\mathbb{Z}_p$, we get $H^2(C,\mathbb{T}_p) = 0$. By computing explicitly the double coset action of $U(p)$ (see [Hida 1986, Lemma 6.1]), we confirm that $U(p)$ acts on $H^1(C,\boldsymbol{T})$ and $H^1(C,\mathbb{T}_p)$ via multiplication by its degree p, and hence $U(p)^{s-r}$ kill $H^1(C,\boldsymbol{T})$ and $H^1(C,\mathbb{T}_p)$. Hence $J^r_s \to J_s$ is a U-isomorphism over $\mathbb{C}$ and hence over K. We record what we have proved:

(2-6) $$U(p)^{s-r}(H^1(C,\mathbb{T}_p)) = H^2(C,\boldsymbol{T}) = H^2(C,\mathbb{T}_p) = 0.$$

Thus $J_s[\gamma^{p^{r-1}}-1](\mathbb{C})$ is connected, and hence $J_s[\gamma^{p^{r-1}}-1]$ is an abelian variety over $\mathbb{Q}$ (as it is the kernel of the $\mathbb{Q}$-rational endomorphism $\gamma^{p^{r-1}}-1$). By the diagram (2-5), we get an isogeny $i^s_r : J^r_{s,\mathbb{C}} \to J_{s,\mathbb{C}}[\gamma^{p^{r-1}}-1]$ whose kernel is a cyclic group of order p^{s-r}. Since this isogeny i^s_r is induced by the Picard functoriality from the covering map $X_s \twoheadrightarrow X^r_s$ defined over $\mathbb{Q}$, i^s_r is defined over $\mathbb{Q}$. Then it extends to $i^s_{r/R} : J^r_{s,R} \to J_{s,R}$ by the functoriality of Picard schemes (which factors through $J_{s,R}[\gamma^{p^{r-1}}-1] = \mathrm{Ker}(\gamma^{p^{r-1}}-1 : J_{s,R} \to J_{s,R})$). There is a more arithmetic proof of these facts valid for any k-points (for a general field k) in place of $\mathbb{C}$-points of the Jacobians which we hope to discuss in our future article.

Though we do not use this fact in this paper, there is a dual homology version of (2-5) coming from group homology [Brown 1982, VII.6.4]:

$$\text{(2-7)}\qquad \begin{array}{ccccccc}
H_2(C,\boldsymbol{T}) & \hookrightarrow & \dfrac{H_1(\Gamma_1(Np^s),\boldsymbol{T})}{(\gamma^{p^{r-1}}-1)H_1(\Gamma_1(Np^s),\boldsymbol{T})} & \longrightarrow & H_1(\Gamma^r_s,\boldsymbol{T}) & \longrightarrow & H_1(C,\boldsymbol{T}) \\
\downarrow & & \downarrow{\scriptstyle\text{onto}} & & \downarrow{\scriptstyle\text{onto}} & & \downarrow \\
? & \longrightarrow & (J_s/(\gamma^{p^{r-1}}-1)J_s)(\mathbb{C}) & \longrightarrow & J^r_s(\mathbb{C}) & \longrightarrow & ?
\end{array}$$

Since $H_j(C,\boldsymbol{T})$ is the Pontryagin dual of $H^j(C,\mathbb{Z})$, we have

$$H_1(C,\boldsymbol{T}) = \mathrm{Hom}(C,\mathbb{Z}) = 0 \quad\text{and}\quad H_2(C,\boldsymbol{T}) = H^2(C,\mathbb{Z}) = \mathbb{Z}/N^r_s\mathbb{Z} \cong C$$

for $N^r_s = 1+g+\cdots+g^{p^{s-r}-1}$ with $g=\gamma^{p^{r-1}}$. This shows that $J^r_{s,\mathbb{Q}}$ is a quotient of $(J_s/(\gamma^{p^{r-1}}-1)J_s)_{/\mathbb{Q}}$ by a finite cyclic group of order p^{s-r} killed by $U(p)^{s-r}$ and also $U^*(p)^{s-r}$.

As we have seen from (2-5), we have a morphism $i^s_{r/R} : J^r_{s,R} \to J_{s,R}[\gamma^{p^{r-1}}-1]$. This morphism composed with $J_{r,R} \to J^r_{s,R}$, induced by Picard functoriality from the covering map $X^r_{s,R} \to X_{r,R}$, gives rise to the morphism $I^s_{r/R} : J_{r,R}[p^\infty] \to J_{s,R}[p^\infty]$, which gives rise to a inductive system $\{\mathcal{G}_{r,R}, I^s_r \mid \mathcal{G}_{r,R} \to \mathcal{G}_{s,R}\}_{s>r}$ of ind-group schemes. Then $\mathcal{G}_R = \varinjlim_r \mathcal{G}_r$ is again a well-defined ind-group scheme. We want to study the control property as in (CT′) for $\mathcal{G}_R$ if either $R=K$ or $R=R_\infty$. Suppose $R=R_\infty=\mathbb{Z}_{(p)}[\mu_{p^\infty}]$. Through the diamond operators, the multiplicative group $\mathbb{Z}_p^\times$ acts on $\mathcal{G}_R$. For $a\in\mathbb{Z}/(p-1)\mathbb{Z}$, write $\mathcal{G}(a)_R$ for the eigenspace (that is, the maximal ind-Barsotti–Tate subgroup) on which $\zeta\in\mu_{p-1}(\mathbb{Z}_p)$ acts via the multiplication by ζ^a. In particular, $\mathcal{G}(0)$ is the μ-fixed part $\mathcal{G}^\mu$ for $\mu=\mu_{p-1}\subset\mathbb{Z}_p^\times$. Regarding $\mathcal{G}_R$ as an fppf p-abelian sheaf (meaning it has values in the category of p-abelian groups), we will show that the projector $x\mapsto 1/(p-1)\sum_{\zeta\in\mu_{p-1}\subset\mathbb{Z}_p^\times}\zeta^{-a}\langle\zeta\rangle(x)$ projects the sheaf $\mathcal{G}_R$ onto $\mathcal{G}(a)_R$. We put $\mathcal{G}^{(0)}_R=\bigoplus_{0<a<p-1}\mathcal{G}(a)_R$; thus $\mathcal{G}_R=\mathcal{G}^\mu_R\oplus\mathcal{G}^{(0)}_R$. Using the good reduction theorem of Langlands and Carayol combined with an analysis of the relation between $\mathcal{G}^{(0)}_R$ and the good abelian quotients from [Mazur and Wiles

1984, Section 3], we will show $\mathcal{G}^{(0)}_{R_\infty}$ is a Λ-BT group (see Sections 4 and 5 for the proof, and Remark 5.5 and Proposition 6.3 for the structure of the complement $\mathcal{G}^{\mu}$).

Roughly, X_{r,R_r} classifies degree p^r cyclic isogenies $\pi : E \to E'$ with some additional data (here "cyclicity" means the kernel of the isogeny is "cyclic" in the sense of Drinfeld as explained in [Katz and Mazur 1985, Chapter 6]). After a base change (tensoring $\mathbb{F}_p$ over R_r), π factors as

$$E \xrightarrow{F^a} E^{(p^a)} \cong E'^{(p^b)} \xrightarrow{V^b} E'$$

for the p-power relative Frobenius F and its dual V (the Verschiebung) for some nonnegative integers a, b with $a+b=r$. Thus $X_{r,\mathbb{F}_p} := X_{r,R_r} \otimes_{R_r} \mathbb{F}_p$ is a union $\bigcup_{a+b=r} X_{(a,b)}$ for $X_{(a,b)}$ classifying cyclic isogenies of type (a, b) as above (with additional data). We define $Y_r = X_{(0,r)} \cup X_{(r,0)}$ inside $X_{r,R_r} \otimes_{R_r} \mathbb{F}_p$ (good components in the sense of [Mazur and Wiles 1984, Section 3]). We will see in Corollary 6.1 that

(u) the projection $J_{r,R_r}[p^\infty] \otimes_{R_r} \mathbb{F}_p \to \mathrm{Pic}^0_{Y_r/\mathbb{F}_p}[p^\infty]$ is a $U(p)$-isomorphism,

where as a correspondence, $U(p) \cap Y_r := U(p) \times_{X_{r,R_r}} Y_r$ induces a correspondence on $Y_r \times Y_r$ and hence it acts on $\mathrm{Pic}^0_{Y_r/\mathbb{F}_p}$.

3. More structures on the modular Λ-BT groups

We list here some more good properties satisfied by the modular Λ-BT group $\mathcal{G}$ (to be constructed in the following two sections) whose proofs will be given in the later sections. We would like to know to what extent a general Λ-BT group $\mathcal{G}_R$ satisfies the following properties, though the only example we know is made out of modular Jacobians. In this section only, we denote by $R = R_\infty$ a general valuation ring with residue field $\mathbb{F}$ of mixed characteristic $(0, p)$ not necessarily in $\mathbb{Q}[\mu_{p^\infty}]$, and suppose $R_\infty = \bigcup_j R_j$ for an increasing sequence of discrete valuation subrings R_j (i.e., $R_j \subset R_{j+1} \subset R$ for all j). Again R_n is not necessarily in $\mathbb{Q}[\mu_{p^n}]$. We write K_n for the quotient field of R_n.

Recall the quotient field K of R, and fix an algebraic closure $\overline{K}$ of K. We have the geometric generic fiber $\mathcal{G}_r[p^n](\overline{K})$ of the (quasi)finite group scheme $\mathcal{G}_r[p^n]$ and put $\mathcal{G}(\overline{K}) = \varinjlim_r \varinjlim_n \mathcal{G}_r[p^n](\overline{K})$ (which we will call the geometric generic fiber of $\mathcal{G}$). We may regard $\mathcal{G}(\overline{K})$ as a discrete Λ-module. Similarly taking the special fiber $\mathcal{G}_r[p^n]_{/\mathbb{F}} := \mathcal{G}_r[p^n] \otimes_R \mathbb{F}$, we define $\mathcal{G}_{\mathbb{F}} = \varinjlim_r \varinjlim_n \mathcal{G}_r[p^n]_{/\mathbb{F}}$ as an ind-group scheme (we call $\mathcal{G}_{\mathbb{F}}$ the special fiber of $\mathcal{G}_R$). We will verify in the next section the following condition for the modular Λ-BT group:

(DV) $\mathcal{G}(\overline{K}) \cong \Lambda^{*n}$ (as Λ-modules) for $\Lambda^* := \mathrm{Hom}(\Lambda, \mathbb{Q}_p/\mathbb{Z}_p)$.

If R has a finite residue field $\mathbb{F} = \mathbb{F}_q$ of characteristic p, we further consider the following properties for $\mathcal{G}$:

(CT) Writing $\gamma = 1+x$, we have $\mathcal{G}_n = \mathrm{Ker}(\gamma^{p^{n-1}} - 1 : \mathcal{G}_R \to \mathcal{G}_R)$ (closed immersion) and $\mathcal{G}_{n,R}$ descends to a Barsotti–Tate group over the discrete valuation ring R_n (for each $0 < n \in \mathbb{Z}$). We have $\mathcal{G}_m \times_{R_m} R_n \cong \mathcal{G}_{m,R_n}$ if $n > m$ (compatibility).

(D) We have a Cartier self-duality $\mathcal{G}_n[p^m] \times \mathcal{G}_n[p^m] \to \mu_{p^m}$ over R_n which, after taking the limit, gives (Galois equivariant) Pontryagin duality $T\mathcal{G} \times \mathcal{G}(\bar{K}) \to \mu_{p^\infty}(\bar{K})$ for $T\mathcal{G} = \varprojlim_n T\mathcal{G}_n$ (for $T\mathcal{G}_n = \varprojlim_m \mathcal{G}_n[p^m](\bar{K})$) with respect to the map $T\mathcal{G}_{n+1} \twoheadrightarrow T\mathcal{G}_n$ dual to $\mathcal{G}_n \hookrightarrow \mathcal{G}_{n+1}$.

(Od) The connected component of $\mathcal{G}_r^\circ[p^n]$ for all $n > 0$ and $r > 0$ is a multiplicative locally free group over the strict henselization of R.

(U) On the special fiber, we have the Frobenius map F and its dual V with $FV = VF = q$. Thus we have a splitting $\mathcal{G}_{\mathbb{F}} = \mathcal{G}^\circ \times \mathcal{G}^{et}$ so that $\mathcal{G}^\circ = \mathrm{Ker}(e_F) = \mathrm{Im}(e_V)$ and $\mathcal{G}^{et} = \mathrm{Ker}(e_V) = \mathrm{Im}(e_F)$ for $e_F = \lim_{n\to\infty} F^{n!}$ and $e_V = \lim_{n\to\infty} V^{n!}$. Then we have a Λ-linear automorphism U of $\mathcal{G}$ such that $U|_{\mathcal{G}_r}$ is defined over R_r for all $r > 0$, U commutes with F and V, and U on $\mathcal{G}^{et}$ lifts $F|_{\mathcal{G}^{et}}$. Moreover, $e = \lim_{n\to\infty} U^{n!} = e_F|_{\mathcal{G}^{et}} + e_V|_{\mathcal{G}^\circ}$ on $\mathcal{G}_{\mathbb{F}}$.

A Λ-BT group satisfying the above properties will be called an ordinary Λ-BT group over R. We prove these properties for the modular Λ-BT groups in the following two sections for $R = \mathbb{Z}_{(p)}[\mu_{p^\infty}]$. We may replace the base ring R in the above conditions by a field of characteristic p (for example $\mathbb{F}_q$); so, the definition of Λ-BT group makes sense over a finite field and $\bar{\mathbb{F}}_p$ (see (Q1) below).

Pick a linear operator $L \in \mathrm{End}_{\Lambda[\mathrm{Gal}(\bar{K}/K)]}(T\mathcal{G})$ whose restriction to $T\mathcal{G}_r$ (for each $r > 0$) commutes with the action of $\mathrm{Gal}(\bar{K}/K_r)$ (the bigger Galois group). Since the Barsotti–Tate group $\mathcal{H}$ over a field k of characteristic 0 is an étale group and therefore is determined by its Galois module, we have $\mathrm{End}_{\mathbb{Z}_p[\mathrm{Gal}(\bar{k}/k)]}(T\mathcal{H}) \cong \mathrm{End}_{BT}(\mathcal{H})$. Then the restriction $L_r \in \mathrm{End}_{\Lambda[\mathrm{Gal}(\bar{K}/K_r)]}(T\mathcal{G}_r)$ gives rise to an endomorphism of $\mathcal{G}_{r,K_r}$ and extends uniquely to an endomorphism of the Barsotti–Tate group $\mathcal{G}_{r,R_r}$ defined over R_r [Tate 1967, Theorem 4]. We write this restriction as $L_r^{BT} \in \mathrm{End}_{BT}(\mathcal{G}_{r,R_r})$. Then we have $L^{BT} = \varprojlim_r L_r^{BT} \in \mathrm{End}_\Lambda(\mathcal{G}_R)$. If confusion is unlikely, we simply write L for L^{BT}.

Suppose that $\det(L) \neq 0$ in Λ as an endomorphism of $T\mathcal{G} \cong \Lambda^n$. Define

$$\mathcal{G}[L]_R = \mathrm{Ker}(L : \mathcal{G}_R \to \mathcal{G}_R),$$

which is a well-defined fppf abelian sheaf over R (as the category of fppf abelian sheaves is abelian; see [Milne 1980, Chapter 2]). We regard $\mathcal{G}[L](\bar{K})$ as an abelian group. Since $\det(L) \neq 0$, by the classification of Λ-modules, the maximal p-divisible subgroup $\mathcal{G}[L](\bar{K})^{\mathrm{div}}$ of $\mathcal{G}[L](\bar{K})$ has finite corank; that is, $\mathcal{G}[L](\bar{K})^{\mathrm{div}} \cong (\mathbb{Q}_p/\mathbb{Z}_p)^m$ as an abstract group for a finite $m > 0$, which is called the corank of $\mathcal{G}[L](\bar{K})$. We call $\mathcal{G}[L](\bar{K})^{\mathrm{div}}$ the p-divisible part of $\mathcal{G}[L](\bar{K})$, which is canonically determined inside $\mathcal{G}[L](\overline{K})$. Thus its p^n-torsion subgroup $\mathcal{G}[L](\overline{K})^{\text{div}}[p^n]$ is finite, and we can find a finite $r = r(n) > 0$ such that $\mathcal{G}[L](\overline{K})^{\text{div}}[p^n] \subset \mathcal{G}_r[p^n](\overline{K})$.

Since L_r^{BT} commutes with the action of $\mathrm{Gal}(\overline{K}/K_r)$ and the p-divisible part is unique in $\mathcal{G}[L](\overline{K})$, it follows that $\mathcal{G}[L](\overline{K})^{\text{div}}[p^n]$ is stable under the action of $\mathrm{Gal}(\overline{K}/K_r)$. Thus $\mathcal{G}[L](\overline{K})^{\text{div}}[p^n]$ is the group of geometric points of a finite flat subgroup $\mathcal{G}[L]^{\text{div}}[p^n]_{/K_r}$ of $\mathcal{G}_{r,K_r}[p^n]$ for sufficiently large r. Take the schematic closure G_{n/R_r} of $\mathcal{G}[L]^{\text{div}}[p^n]_{/K_r}$ in $\mathcal{G}_{r,R_r}[p^n]$ defined over R_r. Writing $\mathcal{G}_r[p^n] = \mathrm{Spec}(A)$ for a Hopf R_r-algebra $A = A_n$ with $\mathcal{G}[L]^{\text{div}}[p^n]_{/K_r} = \mathrm{Spec}(A \otimes_{R_r} K_r/I)$ for an ideal I of $A \otimes_{R_r} K_r$, we have $G_{n/R_r} = \mathrm{Spec}(A/A \cap I)$. Note that $A/(A \cap I) \cong (A+I)/I \hookrightarrow A \otimes_{R_r} K_r/I$ is p-torsion free (and hence flat over R_r). Thus G_{n/R_r} is a finite flat subgroup scheme of $\mathcal{G}_{r,R_r}[p^n]$.

Since $\mathcal{G}_{r,R_s} = \mathcal{G}_{r,R_r} \otimes_{R_r} R_s \hookrightarrow \mathcal{G}_{s,R_s}$ $(s > r)$ is a closed immersion, the schematic closure of $\mathcal{G}[L]^{\text{div}}[p^n]_{/K_s} = \mathcal{G}[L]^{\text{div}}[p^n]_{/K_r} \otimes_{K_r} K_s$ coincides with $G_{n/R_r} \otimes_{R_r} R_s$; thus the formation of the base change $G_{n/R} := G_{n/R_r} \otimes_{R_r} R$ is independent of the chosen r. Put $\mathcal{G}[L]^{BT}_{/R} = \varinjlim_n G_{n/R}$, which should be a Barsotti–Tate group over R with the identity $\mathcal{G}[L]^{BT}[p^n]_{/R} = G_{n/R}$. To discuss our naive questions, we just take $\mathcal{G}[L]^{BT}_{/R}$ to be a Barsotti–Tate group as a working hypothesis.

Of course, starting with a self-dual Λ-BT group H with a lift U, $TH \otimes_{\mathbb{Z}_p} \Lambda^*$ gives a constant Λ-BT group. Here we put $TH = \mathrm{Hom}_{\mathbb{Z}_p}(\mathbb{T}_p, H(\overline{K}))$ (the Tate module of H). We hereafter suppose that all Λ-BT groups we consider are nonconstant. Thus it could be said that the representation of $\mathrm{Gal}(\overline{K}/K)$ on $T\mathcal{G}$ is a nonconstant deformation of $T\mathcal{G}_1$ in the sense of Mazur (see [Mazur 1989] and [Hida 2000a]).

A p-ordinary Barsotti–Tate group H over R is called of GL$(2g)$-*type* if it is self-dual and there exists a local ring $E \subset \mathrm{End}_{\mathbb{Z}_p}(H_{/R})$ such that we have an isomorphism of E-modules for its Tate module: $TH \cong E^{2g}$. Here we say that H is *p-ordinary* if H satisfies (Od). We call H *minimal* if E is generated by $\mathrm{Tr}(\sigma) \in E$ for all $\sigma \in \mathrm{Gal}(\overline{K}/K)$, where $\mathrm{Tr}(\sigma) \in E$ is the trace of the action of σ on $TH \cong E^{2g}$. If we have a local $\Lambda[U]$-algebra $\mathbb{T}$ inside $\mathrm{End}_{\Lambda[U]}(\mathcal{G})$ for an ordinary Λ-BT group $\mathcal{G}_R$ such that $T\mathcal{G} \cong \mathbb{T}^{2g}$ and $\mathbb{T}$ is self-adjoint under the duality, we call $\mathcal{G}$ *of* GL$(2g)$-*type over* $\mathbb{T}$. In this Λ-adic case, we call $\mathcal{G}$ *minimal* if $\mathbb{T}$ is topologically generated by $\mathrm{Tr}(\sigma)$ and U. Suppose that there exists a nonconstant Λ-BT group $\mathcal{G}$ over a valuation ring R inside $\overline{\mathbb{Q}}$. Then we can ask a lot of simple questions:

(Q0) Does there exist R discretely valued?

(Q1) If we are given an ordinary Λ-BT group $\mathcal{G}$ over a finite field $\mathbb{F}$ of characteristic p, can one lift it to a Λ-BT group over R for a suitable R? (Deformation question to characteristic 0.)

(Q2) Is there any systematic way of constructing such an ordinary Λ-BT group $\mathcal{G}$ over a given R? If it exists, does it create all such Λ-BT groups over R of GL$(2g)$-type? (Construction.)

(Q3) If an ordinary Λ-BT group $\mathcal{G}$ is nonconstant, can $\det(U) \in \mathbb{T}^\times$ be algebraic over W? (Transcendency.)

(Q4) Let us give ourselves a Weil number $\alpha \in \overline{\mathbb{Q}} \cap W$ with $|\alpha| = \sqrt{p}$ of degree $2g$ over $\mathbb{Q}$. Supposing α *ordinary* (in the sense that the minimal polynomial in X of α modulo p is divisible by X^g but not by higher powers), is there a Barsotti–Tate subgroup $\mathcal{G}[U-\alpha]^{BT}_{/R} \subset \mathcal{G}_R$ whose geometric generic fiber is given by $\mathcal{G}[U-\alpha](\overline{K})^{\mathrm{div}}$? Suppose this is the case. Does $\mathcal{G}[U-\alpha]^{BT}_{/R}$ descend to a discrete valuation ring? (Descent.) Here, $\mathcal{G}[U-\alpha]^{BT}$ should be the maximal Barsotti–Tate subgroup in $\mathcal{G}[U-\alpha]_{/R} = \mathrm{Ker}(U-\alpha)$.

(Q5) Under the notation in (Q4), is it possible to embed the Barsotti–Tate part $\mathcal{G}[U-\alpha]^{BT}_{/R}$ of $\mathcal{G}[U-\alpha] = \mathrm{Ker}(U-\alpha)$ into an abelian scheme defined over a finite extension of R? (Relation to abelian varieties.)

(Q6) For a given minimal $\mathcal{G}_{1,R}$ of $\mathrm{GL}(2g)$-type whose Tate module $T\mathcal{G}_1(\overline{\mathbb{Q}}) \otimes_{\mathbb{Z}_p} \overline{\mathbb{Q}}_p$ is simple as a Galois module, is there a universal $\mathcal{G}$? (Universality.) Here the universality is defined as follows. If we have a minimal Barsotti–Tate group H of $\mathrm{GL}(2g)$-type with a morphism $i : \mathcal{G}_1 \to H$ having kernel represented by a finite group scheme (so, $i \circ \mathrm{Tr}(\sigma|_{\mathcal{G}_1}) = \mathrm{Tr}(\sigma|_H) \circ i$ for any $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/K)$), there exists a unique morphism $i_H : H_{/R} \hookrightarrow \mathcal{G}_{/R}$ of Barsotti–Tate groups with finite (group scheme) kernel making the following diagram commute:

$$\begin{array}{ccc} \mathcal{G}_1 & \xrightarrow{\ i\ } & H \\ \downarrow & & \downarrow{\scriptstyle i_H} \\ \mathcal{G} & = & \mathcal{G} \end{array}$$

Question (Q0) probably has a negative answer. Here is one reason why: Suppose that $\mathcal{G}$ is minimal of $\mathrm{GL}(2)$-type and suppose that $\mathcal{G}$ extends as a Λ-BT group to the integral closure of $\mathbb{Z}[1/N]$ in R. If $\mathcal{G}$ is defined over a discrete valuation ring $R = \mathbb{Z}_p$ or $\mathbb{Z}_{(p)}$, then, by the classification of p-ordinary divisible groups [Raynaud 1974, 4.2], the determinant of the Galois representation on $T\mathcal{G}$ has to be the p-adic cyclotomic character χ. Thus $T\mathcal{G}$ is a deformation of $T\mathcal{G}_1$ which is p-ordinary and of determinant χ. If $T\mathcal{G}_1$ is modular whose residual representation is irreducible over $\mathbb{Q}[\sqrt{p^*}]$ ($p^* = (-1)^{(p-1)/2}p$), by Wiles [1995], the universal Galois deformation ring for p-ordinary deformations unramified outside Np with fixed determinant χ is of finite rank over $\mathbb{Z}_p$. Thus $T\mathcal{G}$ has to be constant; therefore, $\mathcal{G}$ has to be constant. Thus if such a $\mathcal{G}$ exists, at least R contains the p-adic valuation ring of the cyclotomic $\mathbb{Z}_p$-extension $\mathbb{Q}_\infty/\mathbb{Q}$.

Suppose $g = 1$. Questions related to the ones given above have been studied in [Hida 1986; Mazur and Wiles 1986; Tilouine 1987; Ohta 1995] for this case.

In this paper, I will give an automorphic way of constructing such $\mathcal{G}$ over $R_\infty = \mathbb{Z}_{(p)}[\mu_{p^\infty}]$. By the solution of Galois deformation problems of ordinary type (Mazur, Wiles–Taylor) and by the solution of Serre's modulo p modularity conjecture (Khare–Wintenberger, Kisin), this gives almost all such Λ-BT groups of GL(2)-type, basically solving (Q2) and (Q6) for GL(2)-type groups. After giving the construction of this modular example in terms of the ordinary projector e (in place of $U(p)$-isomorphisms), we will make some comments on the other questions listed above for the modular Λ-BT group.

4. Construction over $\mathbb{Q}$ via the ordinary projector

Fix a prime $p \geq 5$ and a positive integer N prime to p. Here, we give a down-to-earth construction of the modular Λ-BT group $\mathcal{G}_{\mathbb{Q}}$ over $\mathbb{Q}$ via the ordinary projector e, though we follow the line explained in Section 2. Here we mean by a Λ-BT group over $\mathbb{Q}$ an ind-étale group defined over $\mathbb{Q}$ satisfying conditions (CT) and (DV) from the previous section (as modified by replacing the valuation ring R and R_n by the field $\mathbb{Q}$). Since the category of Barsotti–Tate groups over $\mathbb{Q}$ is equivalent to the category of p-divisible modules of finite corank with a continuous action of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ (as any finite flat group scheme over $\mathbb{Q}$ is étale), we are just dealing with such Galois modules. We also prove the corresponding properties (CT), (DV) and (D) over $\mathbb{Q}$ for this $\mathcal{G}$. Note that the conditions (DV) and (CT) concern only the Λ-module structure of the group not the Galois action (under the condition $R = \mathbb{Q}$).

As before, let $J_r = \mathrm{Pic}^0_{X_1(Np^r)/\mathbb{Q}}$ be the Jacobian variety. Similarly we take J^r_s to be the Jacobian variety associated to the modular curve with the congruence subgroup $\Gamma^r_s = \Gamma_1(Np^r) \cap \Gamma_0(p^s)$ $(0 \leq r \leq s)$. From (2-2) with $R = \mathbb{Q}$, since $e = \lim_{n\to\infty} U(p)^{n!}$ acts on $J[p^\infty]$ for $J = J_r, J_s, J^r_s$ (noting $U(p^m) = U(p)^m$), we have

$$\mathcal{G}_{r,\mathbb{Q}} := J_r[p^\infty]^{\mathrm{ord}} \cong J^r_s[p^\infty]^{\mathrm{ord}},$$

where "ord" indicates the image of e.

For the Jacobian J of $X = X(\Gamma)$ with $\Gamma = \Gamma^r_s$, we identify $J[p^\infty](\mathbb{C})$ with a subgroup of $H^1(\Gamma, \mathbb{T}_p)$ (here $\mathbb{T}_p := \mathbb{Q}_p/\mathbb{Z}_p$, on which Γ acts trivially). Applying the ordinary projector $e = \lim_{n\to\infty} U(p)^{n!}$ to the diagram (2-5) (replacing $\boldsymbol{T}$ there by $\mathbb{T}_p$), we get from (2-6) the controllability

$$\begin{aligned}\mathcal{G}_{s,\mathbb{Q}}[\gamma^{p^{r-1}} - 1] &= \mathrm{Ker}(\gamma^{p^{r-1}} - 1 : J_s[p^\infty]^{\mathrm{ord}} \to J_s[p^\infty]^{\mathrm{ord}}) \\ &= J_r[p^\infty]^{\mathrm{ord}} = \mathcal{G}_{r,\mathbb{Q}}.\end{aligned}$$

Define, as an ind-group scheme over $\mathbb{Q}$ (or, as a p-abelian fppf sheaf),

$$\mathcal{G}_{\mathbb{Q}} := J_\infty[p^\infty]^{\mathrm{ord}} = \varinjlim_r J_r[p^\infty]^{\mathrm{ord}}.$$

For each character $\varepsilon : \Gamma/\Gamma^{p^{r-1}} \to \mu_{p^\infty}$, by the inflation and restriction sequence, we get

$$\mathcal{G}_{\mathbb{Q}}[p^n](\overline{\mathbb{Q}}) \otimes_{\mathbb{Z}} \mathbb{Z}[\varepsilon][\gamma - \varepsilon(\gamma)] \cong J_r[p^n](\overline{\mathbb{Q}})^{\mathrm{ord}} \otimes_{\mathbb{Z}} \mathbb{Z}[\varepsilon][\gamma - \varepsilon(\gamma)] \\ \cong H^1(X_r^1, \mathbb{T}_p(\varepsilon))^{\mathrm{ord}},$$

where $\mathbb{T}_p(\varepsilon)$ is a Γ_r^1-module isomorphic to $\mathbb{T}_p$ on which Γ_r^1 acts by ε. Thus the group $\mathcal{G}_{\mathbb{Q}}(\overline{\mathbb{Q}}) \otimes \mathbb{Z}[\varepsilon][\gamma - \varepsilon(\gamma)]$ is a nontrivial p-divisible group. Taking the Pontryagin dual $T := \mathcal{G}_{\mathbb{Q}}^*$, by Nakayama's lemma applied to $T/\mathfrak{m}T \cong J_1[p]^{\mathrm{ord}}$ (for the maximal ideal $\mathfrak{m}$ of Λ), we find a surjection $\pi : \Lambda^{2j} \twoheadrightarrow T$ for $2j = \dim_{\mathbb{F}_p} J_1[p]^{\mathrm{ord}}$. Then for a prime $P = P_\varepsilon := (\gamma - \varepsilon(\gamma)) \cap \Lambda$, T/PT is the dual of $\mathcal{G}_{\mathbb{Q}}[P] \otimes \mathbb{Z}[\varepsilon][\gamma - \varepsilon(\gamma)]$, which is $\mathbb{Z}_p$-free of rank $2j$. Thus $\mathrm{Ker}(\pi) \subset P_\varepsilon \Lambda^{2j}$. Moving ε around, from $\bigcap_\varepsilon P_\varepsilon \Lambda^{2j} = \{0\}$, we find that $T \cong \Lambda^{2j}$; therefore, $\mathcal{G}_{\mathbb{Q}}$ is a Λ-BT group satisfying (CT) and (DV) (over $\mathbb{C}$ and hence over $\mathbb{Q}$).

As for (D), the canonical polarization of $J_{r/\mathbb{Q}}$ gives rise to the self-duality pairing $[\cdot,\cdot]$ of $J_r[p^r]$ and $J_r \cong {}^tJ_r$. Let $U^*(p)$ (resp. $T^*(n)$) be the image of $U(p)$ (resp. $T(n)$) under the canonical Rosati involution of J_r in $\mathrm{End}(J_r)$. The Weil involution τ over $\mathbb{Q}(\mu_{Np^r})$ associated to $\left(\begin{smallmatrix} 0 & -1 \\ Np^r & 0 \end{smallmatrix}\right)$ satisfies $\tau U(p)\tau^{-1} = U^*(p)$ and $\tau T(n)\tau^{-1} = T^*(n)$ inside $\mathrm{End}(J_{r/\mathbb{Q}[\mu_{Np^r}]})$ because τ is only defined over $\mathbb{Q}[\mu_{Np^r}]$. See [Hida 1986, Theorem 9.3] for more details. Thus, twisting the pairing by τ and $U(p)^{-r}$, we get the self-duality pairing $\langle\cdot,\cdot\rangle_r = [\cdot, \tau \circ U(p)^{-r}(\cdot)]$ of $\mathcal{G}_r[p^m]$. Writing $R_s^r : \mathcal{G}_r \hookrightarrow \mathcal{G}_s$ for the inclusion, and $N_r^s = \sum_{j=1}^{p^{s-r}} \gamma_r^j : \mathcal{G}_s \to \mathcal{G}_r$ with $\gamma_r = \gamma^{p^{r-1}}$, we can verify by computation that $\langle R_s^r(x), y\rangle_s = \langle x, N_r^s(y)\rangle_r$ (see [Ohta 1995, §4.1], for instance). From this we get (D) over K_∞.

5. Construction over $\mathbb{Z}_{(p)}[\mu_{p^\infty}]$

Hereafter, for simplicity, we assume that N is cube-free, and we make the construction of $\mathcal{G}$ over $R_\infty := \mathbb{Z}_{(p)}[\mu_{p^\infty}]$. Under this assumption, the ordinary Hecke algebra $h_2^{\mathrm{ord}}(\Gamma_r^s; \mathbb{Z}_p) \subset \mathrm{End}_{\mathbb{Z}_p}(J_s^r[p^\infty]^{\mathrm{ord}})$ generated by Hecke operators $T(n)$ and $U(q)$ is known to be reduced (it has no nontrivial nilradical; see [Hida 2013, Corollary 1.2]). From this fact, $S_2(\Gamma_s^r)$ has a basis of Hecke eigenforms for all Hecke operators $T(n)$ and $U(q)$. If N is not cube-free, we need to consider old-forms $f(dz) \in S_2(\Gamma_1(Np^r))$ for Hecke eigenforms f and for suitable $d \mid N$ (and abelian varieties associated to such forms), which complicates the arguments, though all arguments we give actually go through if we consider cusp forms which are eigenforms for $T(n)$ with n prime to N. Hereafter, if we say f is a Hecke eigenform, we mean that f is an eigenvector of $T(n)$ for all n prime to Np and $U(q)$ for all primes $q \mid Np$. The Hecke eigenforms we consider may not be new-forms of exact level Np^r.

The Tate module $T_r = TJ_r[p^\infty]^{\mathrm{ord}}(\overline{\mathbb{Q}})$ carries Galois representations (constructed by Eichler and Shimura) of Hecke eigenforms satisfying the following properties (see [Hida 2012, §4.2]):

(1) Cusp forms in $S_2(\Gamma_1^0)$ (with $\Gamma_1^0 = \Gamma_1(N) \cap \Gamma_0(p)$).

(2) All cusp forms in $S_2(\Gamma_1(Np^m))$ whose Neben character has p-conductor equal to p^m for $m = 1, 2, \ldots, r$.

By a theorem of Langlands and Carayol (see [Carayol 1986]), the ℓ-adic Galois representation ($\ell \neq p$) associated to such a Hecke eigenform f does not ramify at p over $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}[\mu_{p^r}])$ except for in case (1). In case (1), it is semistable at p. Thus the abelian subvariety A_f attached to f extends to a semiabelian scheme over $\mathbb{Z}_{(p)}[\mu_{p^r}]$ (see [Serre and Tate 1968, §1] and [Bosch et al. 1990, §7.4]). Let ${}^tG_r = \sum_{f \text{ as above}} A_f \subset J_r$. Thus we have an inclusion ${}^tG_r \hookrightarrow J_r$ defined over $\mathbb{Q}$. Let $J_r \cong {}^tJ_r \twoheadrightarrow G_r$ be the dual quotient. Thus, by definition, we have a commutative diagram defined over $\mathbb{Q}$ for all n prime to Np and all primes $q \mid Np$:

$$\begin{array}{ccc} J_r & \longrightarrow & G_r \\ {\scriptstyle T^*(n),U^*(q)}\downarrow & & \downarrow{\scriptstyle T^*(n),U^*(q)} \\ J_r & \longrightarrow & G_r \end{array}$$

where the superscript "$*$" indicates the Rosati involution induced from the polarization of G_r (coming from the canonical polarization of J_r). Since the space spanned by the Hecke eigenforms described in (1) and (2) above is stable under $T(n)$, $T^*(n)$ and $U(q)$, $U^*(q)$, actually we also have the following commutative diagram defined over $\mathbb{Q}$:

$$\begin{array}{ccc} J_r & \longrightarrow & G_r \\ {\scriptstyle T(n),U(q)}\downarrow & & \downarrow{\scriptstyle T(n),U(q)} \\ J_r & \longrightarrow & G_r \end{array}$$

We can also justify this by noting that the Rosati involution induced by the polarization on G_r from the canonical polarization on J_r sends $T^*(n)|_{G_r}$ to $T(n)|_{G_r}$ and $U^*(q)|_{G_r}$ to $U(q)|_{G_r}$ for all n prime to Np and all primes $q \mid Np$.

A Hecke eigenform $f \in S_2(\Gamma_1(Np^m))$, new at p, satisfies (1) or (2) above if and only if $f|U(p) = f|U^*(p) \neq 0$ (see [Miyake 1989, Theorem 4.6.17]). Thus $U(p) \in \mathrm{End}({}^tG_{r/\mathbb{Q}})$ is an isogeny. Recall the quotient field K of R. By [Hida 2013, Proposition 1.1, Corollary 1.2],

(5-1) a sufficiently large power $U(p)^M$ projects $J_{r,K}$ onto ${}^tG_{r,K}$.

In other words, we have the following commutative diagram for general R:

$$
\begin{array}{ccc}
{}^tG_{r,R} & \xrightarrow{\ i\ } & J_{r,R} \\
\downarrow & \swarrow & \downarrow \\
{}^tG_{r,R} & \xrightarrow[i]{} & J_{r,R}
\end{array}
\tag{5-2}
$$

where the vertical and diagonal arrows are given by $U(p)^M$, and ${}^tG_{r,R} \xrightarrow{i} J_{r,R}$ is the Néron extension of the inclusion $i : {}^tG_{r,K} \hookrightarrow J_{r,K}$ (note that the extended i might not be an immersion). We also have the dual $i^* : J_{r,R} \to G_{r,R}$ which is the Néron extension of the projection $i^* : J_{r,K} \to G_{r,K}$.

For any abelian subvariety A of J_r stable under $U(q)$ for all q dividing Np and under $T(n)$ for all n prime to Np, if there exists an abelian subvariety B stable under the same Hecke operators such that $A + B = J_r$ and $A \cap B$ is finite, the abelian subvariety B is uniquely determined by A (the multiplicity-one theorem; see [Gelbart 1975] and [Miyake 1989, §4.6]). The abelian subvariety B is called the *complement* of A in J_r.

By construction, G_r and ${}^tG_r$ extend to semiabelian schemes over $R_r := \mathbb{Z}_{(p)}[\mu_{p^r}]$. The group $\mu = \mu_{p-1} \subset \mathbb{Z}_p^\times$ acts on J_r, ${}^tG_r$ and G_r by the diamond operators. If we define ${}^tG_r^{(0)}$ in ${}^tG_r$ to be the complement of abelian subvariety fixed by μ, then ${}^tG_r^{(0)}$ and its dual quotient $G_r^{(0)}$ extend to abelian schemes over R_r. Anyway, we take the Néron models G_{r,R_r}, ${}^tG_{r,R_r}$, $G_{r,R_r}^{(0)}$ and ${}^tG_{r,R_r}^{(0)}$ over R_r of the abelian varieties G_{r,K_r}, ${}^tG_{r,K_r}$, $G_{r,K_r}^{(0)}$ and ${}^tG_{r,K_r}^{(0)}$, and we take their p-divisible groups. Here the p-divisible group $G_{r,R_r}^{(0)}$ is a Barsotti–Tate group over R_r. The μ-fixed parts $G_{r,R_r}[p^n]^\mu$ and ${}^tG_{r,R_r}[p^n]^\mu$ are at worst quasifinite flat groups schemes.

Theorem 5.1. *Recall $\mathcal{G}_{r,R_r} = J_{r,R_r}[p^\infty]^{\mathrm{ord}}$. We have the two isomorphisms*

$$
{}^tG_{r,R_r}[p^\infty]^{\mathrm{ord}} \xrightarrow[\sim]{\ i\ } \mathcal{G}_{r,R_r} \xrightarrow[\sim]{\ i*\ } G_{r,R_r}[p^\infty]^{\mathrm{ord}}
$$

canonically over $R_r := \mathbb{Z}_{(p)}[\mu_{p^r}]$.

This theorem might appear tautological. However note that a priori $\mathcal{G}_{r,R_r}[p^n]$ is not known even to be a flat group scheme, but we know that ${}^tG_r[p^\infty]^{(0),\mathrm{ord}}$ and $G_r[p^\infty]^{(0),\mathrm{ord}}$ are Barsotti–Tate groups, where the superscript "(0)" indicates the complement of the μ-fixed part. Thus to show that $\mathcal{G}_{r,R_r}^{(0)} = \bigoplus_{0<a<p-1} \mathcal{G}_{r,R_r}(a)$ is a Barsotti–Tate group, it appears that we need to make a difficult analysis of the inclusion $\mathcal{G}_{r,R_r} \subset J_{r,R_r}$ (in the category of fppf abelian sheaves over R_r) to claim that $\mathcal{G}^{(0)}[p^n]$ is represented by a finite-flat group scheme as in the theorem, since $\mathcal{G}_{r,R_r}^{(0)}[p^n]$ is a priori not even known to be represented by a flat group over R_r. However, suppose that we find two ($U(p)$-equivariant) morphisms of group schemes

$$
{}^tG_{r,R_r}[p^n]^{\mathrm{ord}} \xrightarrow{\ \mathcal{L}\ } \mathcal{G}_{r,R_r}[p^n] \quad \text{and} \quad \mathcal{G}_{r,R_r}[p^n] \xrightarrow{\ \mathcal{R}\ } G_{r,R_r}[p^n]^{\mathrm{ord}}
$$

(for each n) such that the composite $\mathcal{R} \circ \mathcal{L}$ is an isomorphism, which implies that $\mathcal{R}$ (resp. $\mathcal{L}$) is an epimorphism (resp. a monomorphism) of fppf abelian sheaves. Note that the category of fppf abelian sheaves over R_r is an abelian category. By (5-1) and (5-2), $\mathrm{Ker}(\mathcal{R})$ is projected down into $\mathrm{Im}(\mathcal{L})$ by $U(p)^M$. Since $U(p)^M$ are automorphisms of the three fppf abelian sheaves, $\mathcal{L}$ and $\mathcal{R}$ must be isomorphisms, showing $\mathcal{G}_r^{(0)}[p^n]$ is a finite flat group over R_r for all n (and hence $\mathcal{G}_r^{(0)}$ is a Barsotti–Tate group over R_r). This point is the nontriviality of the theorem. A geometric analysis in depth of $\mathcal{G}_{r,R_r}$ has been done in [Cais 2012, §5.4], and $\mathcal{G}^{(0)} = \bigoplus_{0<a<p-1} \mathcal{G}(a)$ is shown directly by Cais to be a Λ-BT group, but our short-cut might be worth recording.

To prove the theorem, without making a difficult analysis of the group scheme $\mathcal{G}_{r,R_r}[p^n]$, the following lemma is quite useful:

Lemma 5.2. *Let R be a discrete valuation ring with fraction field K. Let G_K and G'_K be either both Barsotti–Tate groups or both abelian varieties over K. If G_K and G'_K are abelian varieties, let G_R and G'_R be the identity connected component of the Néron models over R of G_K and G'_K. If G_K and G'_K are Barsotti–Tate groups, we assume we have Barsotti–Tate groups G_R and G'_R over R whose generic fibers are isomorphic to G_K and G'_K, respectively.*

(1) *Suppose that we have a surjective morphism $f_K : G_K \to G'_K$ and an endomorphism $g_K : G_K \to G_K$ such that the map* $\mathrm{Ker}(f_K : G_K \to G'_K) \hookrightarrow G_K$ *factors through* $\mathrm{Ker}(g_K : G_K \to G_K) \hookrightarrow G_K$. *Then for the extensions $f : G_R \to G'_R$ and $g : G_R \to G_R$ over R,* $\mathrm{Ker}(f)$ *is a closed subscheme of* $\mathrm{Ker}(g)$ *in the abelian case and is a closed ind-subgroup scheme in the Barsotti–Tate case.*

(2) *Suppose we have an injective morphism $f_K : G'_K \to G_K$ and an endomorphism $g_K : G_K \to G_K$ such that* $\mathrm{Coker}(f_K : G'_K \to G_K)$ *is the surjective image of* $\mathrm{Coker}(g_K : G_K \to G_K)$. *Then, for the extensions $f : G'_R \to G_R$ and $g : G_R \to G_R$ over R,* $\mathrm{Coker}(f)$ *is a quotient group of* $\mathrm{Coker}(g)$.

Here, strictly speaking, a "surjective" morphism between fppf abelian sheaves means an epimorphism in the abelian category of fppf abelian sheaves over K.

Proof. We first prove assertion (1). We note that the category of groups schemes fppf over a base S is a full subcategory of the category of abelian fppf sheaves. Thus we may regard G_K and G'_K as abelian fppf sheaves over K in this proof. Since the category of fppf abelian sheaves is an abelian category (because of the existence of the sheafification functor from presheaves to sheaves under fppf topology described in [Milne 1980, §II.2]), the assumption that the map $\mathrm{Ker}(f_K : G_K \to G'_K) \hookrightarrow G_K$ factors through $\mathrm{Ker}(g_K : G_K \to G_K) \hookrightarrow G_K$ (that is, $\mathrm{Ker}(f_K) \subset \mathrm{Ker}(g_K)$) implies that there exists a morphism $f'_K : G'_K \to G_K$ of fppf abelian sheaves over K such that $f'_K \circ f_K = g_K$. If we have unique extensions $f : G_R \to G'_R$, $g : G_R \to G_R$,

$f' : G'_R \to G_R$ of these morphisms, we have $f' \circ f = g$ by uniqueness. This implies $\mathrm{Ker}(f)$ is a closed subscheme of $\mathrm{Ker}(g)$ in the abelian case and is a closed ind-group scheme in the Barsotti–Tate case.

If G_K and G'_K are abelian schemes, since G_R and G'_R are the connected components of the Néron models of G_K and G'_K, any generic morphism ϕ_K of these schemes extends to a unique morphism over R (see [Bosch et al. 1990, Proposition 7.4.3]).

If G_R and G'_R are Barsotti–Tate groups, the extensions f and f' exist and are unique by [Tate 1967, Theorem 4]. This finishes the proof of (1).

The second assertion is heuristically the dual of the first, with respect to taking dual abelian schemes or Cartier dual Barsotti–Tate groups of G_R and G'_R. We give a direct proof supplied by the referee as the duality of cokernels with kernels may not be valid in this generality. By hypothesis, G'_K is an abelian subvariety (or p-divisible subgroup) of G_K and $g_K : G_K \to G_K$ factors through G'_K. By the Néron functoriality (or Tate's theorem) $g : G_R \to G_R$ factors through $f : G'_R \to G_R$, so $g(G_R) \subset f(G'_R)$ as fppf abelian subsheaves of G_R. This latter inclusion is precisely the meaning of the conclusion of (2). □

Proof of Theorem 5.1. Over $\mathbb{Q}$ we have ${}^t G_{r,\mathbb{Q}}[p^\infty]^{\mathrm{ord}} \subset \mathcal{G}_{r,\mathbb{Q}}$, from the definition. Let $B_\mathbb{Q}$ be the identity connected component of $\mathrm{Ker}(J_{r,\mathbb{Q}} \to G_{r,\mathbb{Q}})$, which is the complement of ${}^t G_{r,\mathbb{Q}}$. By definition, e kills $B_\mathbb{Q}[p^n]$ for all n; thus, it kills the p-primary part $H[p^\infty]$ of the finite group scheme $H = B_\mathbb{Q} \cap {}^t G_{r,\mathbb{Q}}$ over $\mathbb{Q}$. Thus, over $\mathbb{Q}$, we have the identity in the theorem. Since H is finite, H is killed by $M \cdot U(p)^{M'}$ for an integer M prime to p and another integer M' sufficiently large. We apply the first statement of the lemma to the projection $f_K : {}^t G_{r,K_r} \to G_{r,K_r}$ and $g_K = M \cdot U(p)^{M'}$ for $R = R_r$ and $K = K_r$. Thus, by the lemma, we have $\mathrm{Ker}(f) \subset \mathrm{Ker}(M \cdot U(p)^{M'})$; thus, we get a monomorphism ${}^t G_{r,R_r}[p^\infty]^{\mathrm{ord}} \xrightarrow{\iota} G_{r,R_r}[p^\infty]^{\mathrm{ord}}$ of p-abelian fppf sheaves over $R_r = \mathbb{Z}_{(p)}[\mu_{p^r}]$. Thus,

$$
{}^t G_{r,R_r}[p^\infty]^{(0),\mathrm{ord}} \xrightarrow{\iota^{(0)}} G_{r,R_r}[p^\infty]^{(0),\mathrm{ord}}
$$

is a monomorphism of Barsotti–Tate groups of equal corank (here, the corank is the $\mathbb{Z}_p$-corank of the geometric generic fiber). Thus, generically $\iota^{(0)}$ is an isomorphism, which is enough to conclude that $\iota^{(0)}$ is an isomorphism by a result of Tate (see [Tate 1967, Corollary 2 on p. 181]). Since $\iota^{(0)} = \mathcal{R} \circ \mathcal{L}$ for the Néron extension $\mathcal{L} : {}^t G_{r,R_r}[p^\infty]^{(0),\mathrm{ord}} \to \mathcal{G}^{(0)}_{r,R_r}$ of the inclusion ${}^t G_{r,K_r} \to J_{r,K_r}$ and the Néron extension $\mathcal{R} : \mathcal{G}^{(0)}_{r,R_r} \to G_{r,R_r}[p^\infty]^{(0),\mathrm{ord}}$ of the projection $J_{r,K_r} \to G_{r,K_r}$, we conclude that $\mathcal{L}$ is a monomorphism of fppf p-abelian sheaves. By (5-1) and (5-2) combined with the injectivity of $\mathcal{L}$, a high power $U(p)^M$ projects $\mathcal{G}_{r,R_r}[p^n]$ into ${}^t G_{r,R_r}[p^n]^{\mathrm{ord}}$ for all $n > 0$, where we regard $U(p)^M$ as the Néron extension of a projection $U(p)^M : J_{r,K_r} \to {}^t G_{r,K_r}$; compare (5-1). Thus $\mathcal{L}$ is an epimorphism of abelian fppf sheaves, and so also an isomorphism of the Barsotti–Tate group ${}^tG_{r,R_r}[p^\infty]^{(0),\mathrm{ord}}$ onto the fppf abelian sheaf $\mathcal{G}^{(0)}_{r,R_r}$. In other words, the sheaf $\mathcal{G}^{(0)}_{r,R_r}$ is represented by a Barsotti–Tate group ${}^tG_{r,R_r}[p^\infty]^{(0),\mathrm{ord}}$. Now $J_{r,R_r}[p^n]^{(0),\mathrm{ord}}$ is proven to be a finite flat group scheme (without any analysis of the complicated group scheme $J_{r,R_r}[p^n]$), and $\mathcal{G}^{(0),\mathrm{ord}}_{r,R_r}$ is a Barsotti–Tate group over R_r.

In general, for any flat quasifinite group scheme $A_{/R}$, we have a functorial exact sequence

$$0 \longrightarrow FA \longrightarrow A \longrightarrow EA \longrightarrow 0, \tag{5-3}$$

where FA is a finite flat group scheme and EA is étale quasifinite with trivial closed fiber (see [Mazur 1978, Lemma 1.1]). Since ${}^tG_{r,R_r}$ and G_{r,R_r} are semiabelian, their finite p-power torsion points form quasifinite flat group schemes over R_r (see [Bosch et al. 1990, Lemma 7.3.2], for instance). Applying the above exact sequence to ${}^tG_r[p^n]^{\mathrm{ord}}$ and $G_r[p^n]^{\mathrm{ord}}$ for each $n > 0$ and defining

$${}^tG[p^\infty]^{\mathrm{ord,\,BT}} = \varinjlim_n F({}^tG[p^n]^{\mathrm{ord}}) \quad \text{and} \quad G[p^\infty]^{\mathrm{ord,\,BT}} = \varinjlim_n F(G[p^n]^{\mathrm{ord}}),$$

we have the following commutative diagram with exact rows:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & {}^tG_{r,R_r}[p^\infty]^{\mathrm{ord,\,BT}} & \hookrightarrow & {}^tG_{r,R_r}[p^\infty]^{\mathrm{ord}} & \twoheadrightarrow & {}^tE_r^{\mathrm{ord}} & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \iota_{BT}} & & \downarrow{\scriptstyle \iota} & & \downarrow{\scriptstyle \iota_\eta} & & \\ 0 & \longrightarrow & G_{r,R_r}[p^\infty]^{\mathrm{ord,\,BT}} & \hookrightarrow & G_{r,R_r}[p^\infty]^{\mathrm{ord}} & \twoheadrightarrow & E_r^{\mathrm{ord}} & \longrightarrow & 0 \end{array}$$

where the subscript "BT" indicates the maximal Barsotti–Tate subgroups. Here ${}^tE_r^{\mathrm{ord}}[p^n]$ and $E_r^{\mathrm{ord}}[p^n]$ have empty closed fiber, and ${}^tE_r^{\mathrm{ord}}(\overline{\mathbb{Q}})$ and $E_r^{\mathrm{ord}}(\overline{\mathbb{Q}})$ are each isomorphic to $(\mathbb{Q}_p/\mathbb{Z}_p)^m$ for some $m \geq 0$. The morphism ι_η (regarded as ${}^tE_r^{\mathrm{ord}}(\overline{\mathbb{Q}}) \to E_r^{\mathrm{ord}}(\overline{\mathbb{Q}})$) is an isomorphism by the construction over $\mathbb{Q}$ done in Section 4. Since ι_{BT} is a monomorphism, by the same argument as above, ι_{BT} is an isomorphism. This implies that ι is also an isomorphism. Then again $\mathcal{L} : {}^tG_{r,R_r}[p^\infty]^{\mathrm{ord}} \to \mathcal{G}_{r,R_r}$ is an isomorphism by (5-1) and (5-2), which implies $\mathcal{R} : \mathcal{G}_{r,R_r} \cong G_{r,R_r}[p^\infty]^{\mathrm{ord}}$. This finishes the proof. □

Lemma 5.3. *The natural morphism*: ${}^tG_{r,R_s}[p^\infty]^{\mathrm{ord}} \to {}^tG_{s,R_s}[p^\infty]^{\mathrm{ord}}$ *is a closed immersion for* $s > r$.

Proof. We have a morphism of semiabelian schemes ${}^tG_{r,R_s} \xrightarrow{i} {}^tG_{s,R_s}$, whose kernel is killed by $U(p)^M$ for sufficiently large M (by Lemma 5.2 applied to $f = i$ and $g = U(p)^M$). Thus i induces a closed immersion of the Barsotti–Tate part i_{BT} : ${}^tG_{r,R_s}[p^\infty]^{\mathrm{ord,\,BT}} \to {}^tG_{s,R_s}[p^\infty]^{\mathrm{ord,\,BT}}$. Since $E({}^tG_r[p^\infty]^{\mathrm{ord}}) \to E({}^tG_s[p^\infty]^{\mathrm{ord}})$ is a closed immersion, we get the desired result. □

Theorem 5.4. *Over $R_s := \mathbb{Z}_{(p)}[\mu_{p^s}]$, the natural inclusion $\mathcal{G}_{r,R_s}$ into $\mathcal{G}_{s,R_s}$ is a closed immersion whose image is equal to the kernel* $\mathrm{Ker}(\gamma^{p^{r-1}}-1)$ *on $\mathcal{G}_s$ for all $s > r$. In particular, the complement $\mathcal{G}^{(0)}$ of the fixed part of $\mathcal{G}$ by the action of μ is a Λ-BT group over R_∞.*

Proof. The first assertion proves the condition (CT) for the modular Λ-BT group, and hence $\mathcal{G}^{(0)}_{R_\infty}$ is a Λ-BT group over R_∞, as the condition (DV) was already proven in Section 4.

Thus we prove the first assertion. By Lemma 5.2, we have a sequence

$$0 \longrightarrow \mathcal{G}_{r,R_s} \xrightarrow{\ i\ } \mathcal{G}_{s,R_s} \xrightarrow{\gamma^{p^{r-1}}-1} \mathcal{G}_{s,R_s}$$

in which i is a closed immersion by Lemma 5.3. Look at $N_r^s : \mathcal{G}_{s,R_s} \to \mathcal{G}_{s,R_s}$ with $N_r^s = \sum_{\sigma \in \Gamma^{p^{r-1}}/\Gamma^{p^{s-1}}} \sigma$ and the inclusion $i : \mathcal{G}_{r,R_s} \to \mathcal{G}_{s,R_s}$. By applying (2) of Lemma 5.2 to $g_K = N_r^s$ and $f_K = i$ for $K = K_s$ and $R = R_s$, we see $\mathrm{Coker}(i)$ is a surjective image of $\mathrm{Coker}(N_r^s)$. Thus, we have the sequence

$$0 \longrightarrow \mathcal{G}_{r,R_s} \xrightarrow{\ i\ } \mathcal{G}_{s,R_s} \xrightarrow{\ \pi\ } \mathcal{G}_{s-r,R_s} \longrightarrow 0, \tag{5-4}$$

where i is a closed immersion and π is an epimorphism (of abelian fppf sheaves). The generic fiber of the sequence is exact by the result in the previous section. Thus we need to prove the exactness of the sequence (5-4) in the category of abelian fppf sheaves over R_s.

Applying the functor F in (5-3), we get the sequence of the Barsotti–Tate parts:

$$0 \longrightarrow \mathcal{G}^{BT}_{r,R_s} \xrightarrow{\ i^{BT}\ } \mathcal{G}^{BT}_{s,R_s} \xrightarrow{\ \pi^{BT}\ } \mathcal{G}^{BT}_{s-r,R_s} \longrightarrow 0. \tag{5-5}$$

Again π^{BT} is an epimorphism and i^{BT} is a closed immersion. Truncating the sequence to its finite layers, we get the third sequence

$$0 \longrightarrow \mathcal{G}^{BT}_{r,R_s}[p^n] \xrightarrow{\ i_n^{BT}\ } \mathcal{G}^{BT}_{s,R_s}[p^n] \xrightarrow{\ \pi_n^{BT}\ } \mathcal{G}^{BT}_{s-r,R_s}[p^n] \longrightarrow 0 \tag{5-6}$$

with epimorphism π_n^{BT} and closed immersion i_n^{BT} for each $n > 0$. Then $\mathrm{Ker}(\pi_n^{BT})$ is represented by a finite flat group scheme (see [Hida 2012, §1.12.1], for instance). Writing $\mathrm{Ker}(\pi_n^{BT}) = \mathrm{Spec}(A)$, we have $\mathrm{Im}(i_n^{BT}) = \mathrm{Spec}(A/I)$ for an ideal I. Since $\mathrm{rank}_{R_s} A = \mathrm{rank}_{R_s} A/I$ as they have the same generic geometric fiber, we have $I = 0$ and $\mathrm{Ker}(\pi_n^{BT}) = \mathrm{Im}(i_n^{BT})$ for all $n > 0$. In other words, the sequence (5-5) is exact.

By the result in Section 4, we have the exact sequence

$$0 \longrightarrow E(\mathcal{G}_{r,R_s}) \xrightarrow{\ i^{et}\ } E(\mathcal{G}_{s,R_s}) \xrightarrow{\ \pi^{et}\ } E(\mathcal{G}_{s-r,R_s}) \longrightarrow 0.$$

This combined with exactness of (5-5) implies exactness of the sequence (5-4) as desired. □

Remark 5.5. Over a nonnoetherian nondiscrete valuation ring such as R_∞, the distinction between Barsotti–Tate (or crystalline) Galois representations and semistable multiplicative Galois representations is murky. To give an example of this, start with a modular rational elliptic curve E with multiplicative reduction at p of conductor Np (so $p \nmid N$). Then the associated modular form f satisfies $f|U(p) = \epsilon f$ and $f|U^*(p) = \epsilon^{-1} f$ for a root of unity ϵ. Thus f has p-slope 0.

Consider the ordinary universal Galois deformation space S of the Galois representation ρ_E on the p-adic Tate module $T_p E_{/\mathbb{Q}}$. By definition, S is a formal scheme over $\mathbb{Z}_p$. Write $\boldsymbol{h} \subset \mathrm{End}_\Lambda(\mathcal{G})$ for the subalgebra generated over $\Lambda = \mathbb{Z}_p[\![x]\!]$ by all Hecke operators $T(n)$ and $U(q)$. The formal scheme S is often identified with $\mathrm{Spf}(\mathbb{T})$ for a local ring $\mathbb{T}$ of $\boldsymbol{h}$ by an "$R = T$" theorem (see [Wiles 1995] or [Hida 2000b, Theorem 5.29]), so it is flat over $\mathbb{Z}_p$ in such a good case, or even smooth over $\mathbb{Z}_p$ if $\mathbb{T} = \Lambda = \mathbb{Z}_p[\![\Gamma]\!]$. To make our argument easy, suppose that $S = \mathrm{Spf}(\Lambda)$ (so $\mathrm{Spec}(\Lambda)(\mathbb{Z}_p) \cong \Gamma \cong \mathbb{Z}_p$, for which we write $S(\mathbb{Z}_p)$ by abuse of notation). Then the subset $S^{\mathrm{crys}}(\mathbb{Z}_p) \subset S(\mathbb{Z}_p)$ corresponding to crystalline representations is p-adically dense (i.e., f is a p-adic limit of Hecke eigenforms f_k of weight $k > 2$ of level N prime to p). Thus $E[p^n](\overline{\mathbb{Q}})$ is, over $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$, the reduction modulo p^n of a crystalline Galois representation and a multiplicative Galois representation at the same time. Instead of S^{crys}, we can take the subset S^{pBT} of potentially Barsotti–Tate Galois representations in S. Then S^{pBT} is Zariski dense in the scheme $S = \mathrm{Spec}(\Lambda)_{/\mathbb{Z}_p}$. Indeed, identifying $\mathrm{Spec}(\Lambda)$ with

$$\widehat{\mathbb{G}}_{m/\mathbb{Z}_p} = \mathrm{Spec}(\widehat{\mathbb{Z}_p[t, t^{-1}]})$$

(for $\widehat{\mathbb{Z}_p[t, t^{-1}]} = \varprojlim_r \mathbb{Z}_p[t, t^{-1}]/(t^{p^r} - 1)$) and making the identification $\gamma = t$, we have an identification $S^{pBT} = \mu_{p^\infty}(\overline{\mathbb{Q}}_p) - \{1\}$ inside $S(\overline{\mathbb{Q}}_p)$. Thus the Galois module $E[p^n](\overline{\mathbb{Q}}_p)$ can be realized as a generic geometric fiber of a finite flat group scheme G_n defined over a highly wild p-ramified subring R of R_∞. Since the generic fiber does not determine G_n over highly p-ramified ring (see [Raynaud 1974] and [Bosch et al. 1990, §7.5]), we have ambiguity. However, if we can pick G_n inside $\mathcal{G}_{R_\infty}$, it is expected to be unique. Thus $E[p^\infty]$ would be given as a generic fiber of a Barsotti–Tate group over R_∞. In particular, $\mathcal{G}$ is close to a Λ-BT group satisfying the following condition in place of (CT):

(ct) $\mathcal{G}_{R_\infty}[\gamma^{p^{r-1}} - 1]$ (for each $r > 0$) is a Barsotti–Tate group over R_∞.

Thus, decomposing $\mathcal{G}$ as $\mathcal{G}^\mu \oplus \mathcal{G}^{(0)}$, where $\mathcal{G}^\mu$ is the fixed part of $\mathcal{G}$ under $\mu := \mu_{p-1} \subset \mathbb{Z}_p^\times$, control of $\mathcal{G}^\mu$ is not equivalent to having nontrivial (nonflat) cokernel $\mathcal{G}^\mu[\gamma^{p^{r-1}} - 1]/\mathcal{G}_r^\mu$, since $\mathcal{G}_r^\mu(\overline{\mathbb{F}}_p)$ may even be finite. The Barsotti–Tate group

$\mathcal{G}^{\mu}[\gamma^{p^{r-1}}-1]$ over R_∞ does not descend to R_r. We give here two of our three arguments proving (ct) and will give the third in the next section.

Here is our first argument. Take the abelian subvariety $A_{r,\mathbb{Q}} = \sum_f A_f \subset {}^tG_{r,\mathbb{Q}}$, with f running over Hecke eigenforms satisfying condition (2) at the beginning of this section. Thus A_f and A_r have good reduction over R_∞. Then, writing ${}^tG_r = A_r + B_r$ for the complementary abelian subvariety $B_{r,\mathbb{Q}}$, the subvariety B_r is unique and is the image of the Jacobian of X_1^0 in J_r over $\mathbb{Q}$. Then by (DV), one can show that

$$\varinjlim_r A_r^{\mathrm{ord}}[p^\infty](\overline{\mathbb{Q}}) = \varinjlim_r J_r^{\mathrm{ord}}[p^\infty](\overline{\mathbb{Q}}) = \mathcal{G}(\overline{\mathbb{Q}}).$$

Indeed, it is easy to see that $\bigcup_{r>0} \Lambda^*[(\gamma^{p^r}-1)/(\gamma-1)] = \Lambda^*$ for the Pontryagin dual Λ^* of Λ, and hence, identifying $\mathcal{G}(\overline{\mathbb{Q}})$ with $(\Lambda^*)^{2j}$ as Λ-modules, we have

$$\mathcal{G}(\overline{\mathbb{Q}}) \supset \varinjlim_r A_r^{\mathrm{ord}}[p^\infty](\overline{\mathbb{Q}}) \supset \bigcup_{r>0} (\Lambda^*)^{2j}\left[\frac{\gamma^{p^r}-1}{\gamma-1}\right] = (\Lambda^{2j})^* = \mathcal{G}(\overline{\mathbb{Q}}).$$

Then we can go through the argument proving Theorem 5.1, replacing ${}^tG_r$ by A_r to show that $\mathcal{G}_{R_\infty} = \varinjlim_r A_r^{\mathrm{ord}}[p^\infty]_{/R_\infty}$ and get the desired result.

Here is a more direct argument without using A_r. Let

$$\widehat{R} = \widehat{R}_\infty = \bigcup_r \widehat{R}_r = R_\infty \otimes_{\mathbb{Z}} \mathbb{Z}_p \subset \overline{\mathbb{Q}}_p,$$

where $\widehat{R}_r = \varprojlim_n R_r/p^n R_r \cong R_r \otimes_{\mathbb{Z}} \mathbb{Z}_p$ for finite r is the p-adic completion. Define $F(\mathcal{G}_{\widehat{R}}) = \varinjlim_{r,n} F(\mathcal{G}_{r,\widehat{R}}[p^n])$ and $E(\mathcal{G}_{\widehat{R}}) = \varinjlim_{r,n} E(\mathcal{G}_{r,\widehat{R}}[p^n])$ for the functors F, E in (5-3). Since injective limits (in the category of fppf abelian sheaves) are exact, we get an exact sequence of ind-group schemes over $\widehat{R}$:

$$0 \longrightarrow F(\mathcal{G}_{\widehat{R}}) \longrightarrow \mathcal{G}_{\widehat{R}} \longrightarrow E(\mathcal{G}_{\widehat{R}}) \longrightarrow 0.$$

Note that $E(\mathcal{G}_{r,\widehat{R}_r})(\overline{\mathbb{Q}}_p) \cong (\mathbb{Q}_p/\mathbb{Z}_p)^m$ (for m the dimension of multiplicative part of the reduction modulo p of $J^0_{1/\mathbb{Z}_p}$) and that $E(\mathcal{G}_{r,\widehat{R}_r})(\overline{\mathbb{Q}}_p)$ is killed by $x = \gamma - 1$. Thus $E(\mathcal{G}_{\widehat{R}})(\overline{\mathbb{Q}}_p)$ is killed by x (and is still embedded in $(\mathbb{Q}_p/\mathbb{Z}_p)^m$). Note that $\mathcal{G}_{\widehat{R}}(\overline{\mathbb{Q}}_p) \cong (\Lambda^*)^{2j}$ for $2j = \dim_{\mathbb{F}_p} J_1[p](\overline{\mathbb{Q}})$. Since Λ^* is Λ-divisible, it does not have any quotient killed by x (except for $\{0\}$). Thus we get $F(\mathcal{G}_{\widehat{R}}) = \mathcal{G}_{\widehat{R}}$ as we claimed.

There is a third more geometric argument (Proposition 6.3) showing the identity $\mathcal{G}(\overline{\mathbb{F}}_p)[\gamma^{p^{r-1}}-1]/\mathcal{G}_r(\overline{\mathbb{F}}_p) \cong (\mathbb{Q}_p/\mathbb{Z}_p)^m$ (for finite $r > 0$) of the geometric special fibers. Therefore $\mathcal{G}(\overline{\mathbb{F}}_p)$ actually covers the multiplicative part. We give the details of this argument in the following section, after preparing some notation regarding the special fibers of modular curves (over discrete valuation rings).

6. Mod p modular curves

We keep the simplifying assumption that N is cube-free. Hereafter, we write $\mathcal{G}$ and $\mathcal{G}_r$ for the modular Λ-BT group and its r-th layer made out of the Jacobian of X_r. We prove the properties (Od) and (U) for $\mathcal{G}$ over R_∞ now. We consider the following Drinfeld-style moduli problem classifying $(E, \phi'_p, \phi_N)_{/A}$ over $\mathbb{Z}_{(p)}$, where $\phi_N : \mu_N \hookrightarrow E[N]$ is a closed immersion of group schemes over A and ϕ'_p is a pair of isogenies $\pi : E \to E'$ and ${}^t\pi : E' \to E$ of degree p^r together with points $P \in E(A)$ and $P' \in E'(A)$ such that $\mathrm{Ker}(\pi)$ is equal to the relative Cartier divisor $\sum_{j=0}^{p^r-1}[jP] \subset E$ and $\mathrm{Ker}({}^t\pi)$ is equal to the relative Cartier divisor $\sum_{j=0}^{p^r-1}[jP'] \subset E'$. The canonical Cartier duality pairing $\mathrm{Ker}(\pi) \times \mathrm{Ker}({}^t\pi) \to \mu_{p^r}$ gives a point $\zeta_{p^r} = \langle P, P'\rangle$. Thus, this moduli problem is defined over R_r. This problem at p is called the balanced $\Gamma_1(p^r)$ moduli problem in [Katz and Mazur 1985, §3.3]. As shown in Theorem 13.11.4 of the same book, this problem is represented by a regular affine scheme over $\mathbb{Z}_{(p)}[\mu_{p^r}]$ with regular projective compactification X'_r whose generic fiber is X_{r,K_r}. Recall the normalization X_{r,R_r} of $\boldsymbol{P}^1(j)_{/R_r}$ in X_{r,K_r}. Every regular scheme is normal (see [Matsumura 1986, Theorem 19.4], for instance), so $X'_{r,R_r} = X_{r,R_r}$. The special fiber $X_{r,\mathbb{F}_p}$ of X_{r,R_r} has the following description:

$$X_{r,\mathbb{F}_p} = X'_{r,\mathbb{F}_p} = X_{(r,0)} \cup X_{(0,r)} \cup \bigcup_{\substack{a+b=r,\ a>0,\ b>0\\ u\in(\mathbb{Z}/p^{\min(a,b)}\mathbb{Z})^\times}} X_{(a,b,u)},$$

for smooth irreducible projective curves $X_{(a,b,u)}$ intersecting only at supersingular points (see [Katz and Mazur 1985, Theorem 13.11.4]). The curves $X_{(r,0)}$ and $X_{(0,r)}$ are smooth geometrically irreducible (by a theorem of Igusa).

The open curve obtained from $X_{(r,0)}$ by removing supersingular points and cusps represents the moduli problem classifying triples $(E, \mu_{p^r} \hookrightarrow E, \phi_N)$, and the corresponding open curve obtained from $X_{(0,r)}$ classifies $(E, \mathbb{Z}/p^r\mathbb{Z} \hookrightarrow E, \phi_N)$. This curve is called the Igusa curve, and hence we write $I_r = I_{r,\mathbb{F}_p} = X_{(0,r),\mathbb{F}_p}$. We have $I_{r,\mathbb{F}_p} \cong X^{(p^r)}_{(r,0),\mathbb{F}_p}$ (the base change by the p^r-th power Frobenius map) canonically. Since $X_{(r,0)}$ is defined over $\mathbb{F}_p$, we have actually $X_{(r,0),\mathbb{F}_p} \cong I_{r,\mathbb{F}_p}$. All this follows from [Katz and Mazur 1985, Theorem 13.11.4].

We put $Y_r = I_r \cup X_{(r,0)}$ which is the Zariski closure of the image of the disjoint union $I_r \sqcup X_{(r,0)}$ in X_r. This curve Y_r is introduced just above (u) in Section 2. Fix an algebraic closure $\overline{\mathbb{F}}_p$ of $\mathbb{F}_p$. Over $\overline{\mathbb{F}}_p$, the two components of Y_r intersect only at supersingular points (and the crossing is an ordinary double point). On the middle components $X_{(a,b)} = \bigcup_u X_{(a,b,u)}$ with $ab \neq 0$, $\pi : E \to E'$ factors as

$$E \xrightarrow{F^a} E^{(p^a)} \overset{(*)}{\cong} E'^{(p^b)} \xrightarrow{V^b} E'$$

(and the middle isomorphism $(*)$ is determined by the datum $u \in (\mathbb{Z}/p^{\min(a,b)}\mathbb{Z})^\times$ outside the crossing).

As before, let J_r (resp. G_r and ${}^tG_r$) be the identity connected component of the Néron model of $J_{r/\mathbb{Q}}$ (resp. G_r and ${}^tG_r$) over $R_r := \mathbb{Z}_{(p)}[\mu_{p^r}]$. Mazur and Wiles [1984, Chapter 3] have shown the existence of a canonical isogeny $\mathrm{av}(\mathrm{Pic}^0_{Y_r/\mathbb{F}_p}) \to av(G_{r/\mathbb{F}_p})$, where av denotes the abelian variety part. By a theorem of Raynaud [Bosch et al. 1990, Theorem 9.4.5], we have $J_{r,R_r} = \mathrm{Pic}^0_{X_r/R_r}$. Thus, taking the special fiber, we have a surjection $J_{r,\mathbb{F}_p} = \mathrm{Pic}^0_{X_r/\mathbb{F}_p} \twoheadrightarrow \mathrm{Pic}^0_{Y_r/\mathbb{F}_p}$ corresponding to the inclusion $Y_{r,\mathbb{F}_p} = Y'_{r,\mathbb{F}_p} \hookrightarrow X'_{r,\mathbb{F}_p} = X_{r,\mathbb{F}_p}$. Then by Theorem 5.1 combined with [Mazur and Wiles 1984, Proposition on p. 267], we find:

Corollary 6.1. *We have* $\mathcal{G}_{r,\mathbb{F}_p} \cong \mathrm{Pic}^0_{Y_r/\mathbb{F}_p}[p^\infty]^{\mathrm{ord}} \cong G_{r,\mathbb{F}_p}[p^\infty]^{\mathrm{ord}}$.

Proof. Adding the toric part to the isogeny in [Mazur and Wiles 1984], we have an isogeny

$$\mathrm{Pic}^0_{Y_r/\mathbb{F}_p}[p^\infty]^{\mathrm{ord}} \to G_r[p^\infty]^{\mathrm{ord}}_{/\mathbb{F}_p},$$

but the projection: $J_r[p^\infty]^{\mathrm{ord}}_{/\mathbb{F}_p} \cong \mathrm{Pic}^0_{Y_r/\mathbb{F}_p}[p^\infty]^{\mathrm{ord}}$ composed with this isogeny is the special fiber of the isomorphism in Theorem 3.1. □

In [Mazur and Wiles 1984, Section 3.3], it is shown that the $U(p)$ operator on the abelian quotient

$$\mathrm{Pic}^0_{X_{(r,0)}/\mathbb{F}_p} \times \mathrm{Pic}^0_{I_r/\mathbb{F}_p}$$

of $\mathrm{Pic}^0_{Y_r/\mathbb{F}_p}$ has the following matrix shape:

$$\begin{pmatrix} F & * \\ 0 & V\langle p^{(p)}\rangle \end{pmatrix} \text{ on } \mathrm{Pic}^0_{I_r/\mathbb{F}_p} \times \mathrm{Pic}^0_{X_{(r,0)}/\mathbb{F}_p} \tag{6-1}$$

for the p-power relative Frobenius F and its dual V. If $N = 1$, then $U(p) = \left(\begin{smallmatrix} F & 0 \\ 0 & V \end{smallmatrix}\right)$ is semisimple on $\mathrm{Pic}^0_{I_r/\mathbb{F}_p} \times \mathrm{Pic}^0_{X_{(r,0)}/\mathbb{F}_p}$. Here, $\langle p^{(p)}\rangle$ is the diamond operator for $p \in (\mathbb{Z}/N\mathbb{Z})^\times$. This proves the conditions (Od) and (U) for the modular Λ-BT group $\mathcal{G}$. Moreover, writing $j_{r,\mathbb{F}_p} = \mathrm{Pic}^0_{I_r/\mathbb{F}_p}$ (the Jacobian of the r-th layer of the Igusa tower), we confirm that the generic geometric fiber $\mathcal{G}_r(\overline{\mathbb{F}}_p)$ coincides with $j_r[p^\infty](\overline{\mathbb{F}}_p)$ as the Frobenius map F (which equals $U(p)$ on j_r) is an automorphism on the geometric points of j_r and V is topologically nilpotent on $\mathrm{Pic}^0_{X_{(r,0)}/\mathbb{F}_p}[p^\infty](\overline{\mathbb{F}}_p)$.

We prepare some results to show that $\mathcal{G}(\overline{\mathbb{F}}_p)$ is Λ-injective (the third proof of (ct) in Remark 5.5).

Lemma 6.2. *Let $f : X \to Y$ be a finite flat Galois covering with Galois group G of projective smooth connected curves over $\overline{\mathbb{F}}_p$ unramified outside a finite set $S \subset Y(\overline{\mathbb{F}}_p)$. Assume that $G \cong \mathbb{Z}/p^m\mathbb{Z}$ and that every point in S fully ramifies in X, so we have a bijection $f^{-1}(S) \cong S$ induced by f. Then, writing $J_?$ for the Jacobian* *variety for* $? = X, Y$, *the pullback map* $f^*: J_Y(\overline{\mathbb{F}}_p) \hookrightarrow J_X(\overline{\mathbb{F}}_p)$ *is injective, and we have an isomorphism*

$$J_X(\overline{\mathbb{F}}_p)^G/f^*J_Y(\overline{\mathbb{F}}_p) \cong \left\{ D \in \bigoplus_{s \in f^{-1}(S)} \mathbb{Z}/p^m\mathbb{Z}[s] \,\middle|\, \deg(D) = 0 \right\}$$

of finite groups, where $[s]$ *is the divisor on* X *corresponding to the point* s, $J_X(\overline{\mathbb{F}}_p)^G$ *denotes the* G*-invariant subgroup* $H^0(G, J_X(\overline{\mathbb{F}}_p))$, *and* $\deg(\sum_s a_s[s]) = \sum_s a_s$ *for* $a_s \in \mathbb{Z}/p^m\mathbb{Z}$.

We call the quotient group $J_X(\overline{\mathbb{F}}_p)^G/f^*J_Y(\overline{\mathbb{F}}_p)$ the *ambiguous class group* and write it as $\mathrm{Amb}_{X/Y}$.

Proof. Write $\overline{\mathbb{F}}_p(?)$ for the function field of $? = X, Y$. Since $J_Y[n](\overline{\mathbb{F}}_p)$ for a positive integer n is canonically isomorphic to the Galois group of an abelian extension K_n of $\overline{\mathbb{F}}_p(Y)$ unramified everywhere, while X/Y ramifies fully at S, $\overline{\mathbb{F}}_p(X)$ is linearly disjoint from K_n over $\overline{\mathbb{F}}_p(Y)$; so, $f^*: J_Y[n] \to J_X[n]$ is injective. Since $J_Y(\overline{\mathbb{F}}_p) = \bigcup_n J_Y[n]$, we get the injectivity of $f^*: J_Y \to J_X$.

Let $\mathrm{Div}_?$ for $? = X_{\overline{\mathbb{F}}_p}, Y_{\overline{\mathbb{F}}_p}$ be the divisor group of ? and $\mathrm{Div}^0_?$ be the subgroup of degree 0 divisors. Then for the subgroup $P_? = \{\mathrm{div}(g) \mid g \in \overline{\mathbb{F}}_p(?)^\times\}$ of principal divisors, we have $\mathrm{Pic}_{?/\overline{\mathbb{F}}_p}(\overline{\mathbb{F}}_p) = \mathrm{Div}_?/P_?$. Consider the subgroup $R_S = \bigoplus_{s \in f^{-1}(S)} \mathbb{Z}[s] \subset \mathrm{Div}_X$.

Write $D \sim D'$ if the two divisors are linearly equivalent. If $D^\sigma \sim D$ for $D \in \mathrm{Div}_X$ ($\sigma \in G$), writing $D^\sigma - D = \mathrm{div}(g_\sigma)$, we find $g_\sigma^\tau g_\tau / g_{\sigma\tau} \in \overline{\mathbb{F}}_p^\times$. Thus $g \mapsto g_\sigma$ is a 1-cocycle of G having values in $\overline{\mathbb{F}}_p(X)^\times/\overline{\mathbb{F}}_p^\times$. By the long exact sequence attached to the short exact sequence $\overline{\mathbb{F}}_p^\times \hookrightarrow \overline{\mathbb{F}}_p(X)^\times \twoheadrightarrow \overline{\mathbb{F}}_p(X)^\times/\overline{\mathbb{F}}_p^\times$, combined with the fact that $H^2(G, \overline{\mathbb{F}}_p^\times) = 0$ (since $\overline{\mathbb{F}}_p^\times$ is a prime-to-p-torsion module), we conclude from $H^1(G, \overline{\mathbb{F}}_p(X)^\times) = 0$ (Hilbert's theorem 90) that $H^1(G, \overline{\mathbb{F}}_p(X)^\times/\overline{\mathbb{F}}_p^\times) = 0$. Thus $g_\sigma = h - h^\sigma$ for $h \in \overline{\mathbb{F}}_p(X)^\times$, and $D + \mathrm{div}(h) \in \mathrm{Div}_X^G$. This shows that $\mathrm{Pic}_X(\overline{\mathbb{F}}_p)^G$ is the surjective image of Div_X^G.

We have $\mathrm{Div}_X^G = f^*\mathrm{Div}_Y + R_S$ with $R_S \cap f^*\mathrm{Div}_Y = p^m R_S$ as $s \in f^{-1}(S)$ ramifies fully in X/Y. Thus

$$\mathrm{Div}_X^G/f^*\mathrm{Div}_Y \cong R_S/p^m R_S.$$

Suppose $D \in \mathrm{Div}_X^G$ is principal, so $D = \mathrm{div}(g)$ for $g \in \overline{\mathbb{F}}_p(X)$. Then, for $\sigma \in G$, $g^{\sigma-1} = g^\sigma/g$ is a constant in $\overline{\mathbb{F}}_p^\times$. Thus $\sigma \mapsto g^{\sigma-1}$ is a homomorphism of G into $\overline{\mathbb{F}}_p^\times$, which must be trivial as $\overline{\mathbb{F}}_p^\times$ does not have any nontrivial p-subgroup. Thus $g \in f^*(\overline{\mathbb{F}}_p(Y))$. This shows $R_S/p^m R_S$ injects into $\mathrm{Pic}_X(\overline{\mathbb{F}}_p)^G/f^*\mathrm{Pic}_Y(\overline{\mathbb{F}}_p)$, and in fact $R_S/p^m R_S \cong \mathrm{Pic}_X(\overline{\mathbb{F}}_p)^G/f^*\mathrm{Pic}_Y(\overline{\mathbb{F}}_p)$ as $\mathrm{Pic}_X(\overline{\mathbb{F}}_p)^G$ is the surjective image of $\mathrm{Div}_X^G = f^*\mathrm{Div}_Y + R_S$. Since J_X is the degree 0 component of Pic_X, we confirm that

$$J_X(\overline{\mathbb{F}}_p)^G/J_Y(\overline{\mathbb{F}}_p) \cong \{D \in R_S/p^m R_S \mid \deg(D) = 0\}. \qquad \square$$

Proposition 6.3. *The geometric special fiber $\mathcal{G}(\overline{\mathbb{F}}_p)$ is Λ-injective (isomorphic to $(\Lambda^*)^j$), $\mathcal{G}(\overline{\mathbb{F}}_p) \cong (\mathcal{G}/\mathcal{G}^\circ)(\overline{\mathbb{Q}}_p)$ by the reduction map as Λ-modules, and $\mathcal{G}_{\mathbb{F}_p}[\gamma^{p^{r-1}}-1]$ is a Barsotti–Tate group over $\mathbb{F}_p$. Here, $\mathcal{G}^\circ$ is the connected component of $\mathcal{G}$ over $\widehat{R}_\infty = R_\infty \otimes_{\mathbb{Z}_{(p)}} \mathbb{Z}_p$ (whose group of generic geometric points $\mathcal{G}^\circ(\overline{\mathbb{Q}}_p)$ is the kernel of the reduction map).*

Here, j is as in Section 4 and is given by $2j = \dim_{\mathbb{F}_p} J_1[p](\overline{\mathbb{Q}})$. In the following proof, we consider the Igusa tower unramified outside supersingular points:

$$\cdots \longrightarrow I_{r+1,\mathbb{F}_p} \longrightarrow I_{r,\mathbb{F}_p} \longrightarrow \cdots \longrightarrow I_{0,\mathbb{F}_p} := X^0_{1/\mathbb{F}_p}$$

over $\mathbb{F}_p$. Write $j_{r,\mathbb{F}_p}$ for the Jacobian variety of $I_{r,\mathbb{F}_p}$. Mazur and Wiles [1983] studied $\mathcal{G}^{(0)}(\overline{\mathbb{F}}_p) = \varinjlim_r j_r[p^\infty](\overline{\mathbb{F}}_p)^{(0)}$, where the superscript "(0)" indicates the complement of the fixed part of $\mu := \mu_{p-1} \subset \mathbb{Z}_p^\times$. In particular, they showed the control

$$\mathcal{G}^{(0)}(\overline{\mathbb{F}}_p)[\gamma^{p^{r-1}} - 1] = j_r[p^\infty](\overline{\mathbb{F}}_p)^{(0)}$$

for all $r > 0$ (as [ibid., (1) in §2]). The control fails between $\mathcal{G}(\overline{\mathbb{F}}_p)^\mu$ and $j_r[p^\infty](\overline{\mathbb{F}}_p)^\mu$, and the idea for the proof is to compute the failure using Lemma 6.2.

Proof. We only prove the first two assertions, as the last one follows from the second argument in Remark 5.5 after making the base change from R_∞ to $\mathbb{F}_p$. Let S_r be the finite set of supersingular points of I_r. The diamond operator action is equal to the action of $\mathrm{Gal}(I_\infty/X_1^0) = \mathbb{Z}_p^\times$. Since $I_s \to I_r$ ($s > r > 0$) fully ramifies over each point of S_r having the Galois group isomorphic to $(1 + p^r\mathbb{Z}_p)/(1 + p^s\mathbb{Z}_p) \cong \mathbb{Z}/p^{s-r}\mathbb{Z}_p$, the failure of the control of $j_s[p^\infty](\overline{\mathbb{F}}_p)$ can be computed by Lemma 6.2, and we get

$$j_s[p^\infty][\gamma^{p^{r-1}} - 1]/j_r[p^\infty] \cong \mathrm{Amb}_{I_s/I_r} = \{D \in R_{S_s}/p^{s-r}R_{S_s} \mid \deg(D) = 0\}$$

under the notation in the proof of Lemma 6.2. Passing to the injective limit with respect to s, we get, for $m = |S_0| - 1$,

$$\mathcal{G}(\overline{\mathbb{F}}_p)[\gamma^{p^{r-1}} - 1]/\mathcal{G}_r(\overline{\mathbb{F}}_p) = \varinjlim_s j_s[p^\infty][\gamma^{p^{r-1}} - 1]/j_r[p^\infty] = \varinjlim_s \mathrm{Amb}_{I_s/I_r} \cong (\mathbb{Q}_p/\mathbb{Z}_p)^m.$$

As is well known (see [Hida 2012, (4.14)], for instance), the dimension of the multiplicative part of J_1^0 is given by $m = |S_0| - 1$. This shows that $\mathcal{G}(\overline{\mathbb{F}}_p)[\gamma^{p^{r-1}} - 1]$ is a p-divisible module of finite $\mathbb{Z}_p$-corank for all r, and hence $\mathcal{G}(\overline{\mathbb{F}}_p) \cong (\Lambda^*)^j$ (a Λ-injective module of Λ-corank j) by the same argument proving $\mathcal{G}(\overline{\mathbb{Q}}) \cong (\Lambda^*)^{2j}$ in Section 4. The reduction map (over $\widehat{R}_\infty$) induces an injection $(\mathcal{G}/\mathcal{G}^\circ)(\overline{\mathbb{Q}}_p) = (\mathcal{G}^{BT}/\mathcal{G}^\circ)(\overline{\mathbb{Q}}_p) \to \mathcal{G}(\overline{\mathbb{F}}_p)$ (as we have proven $\mathcal{G}^{BT} = F(\mathcal{G}_{\widehat{R}_\infty}) = \mathcal{G}$ in Remark 5.5). Then, comparing the corank, we get an isomorphism $(\mathcal{G}/\mathcal{G}^\circ)(\overline{\mathbb{Q}}_p) \cong \mathcal{G}(\overline{\mathbb{F}}_p)$. □

7. The α-eigenspace in $\mathcal{G}_{R_\infty}$

The modular Λ-BT group $\mathcal{G}$ has coefficients $\Lambda = \mathbb{Z}_p[\![x]\!] = \mathbb{Z}_p[\![\Gamma]\!]$ (for $\Gamma = 1 + p\mathbb{Z}_p \subset \mathbb{Z}_p^\times$). To get a Λ_W-BT group for $\Lambda_W = W[\![x]\!] = W[\![\Gamma]\!]$, we can extend endomorphisms to a valuation ring W bigger than $\mathbb{Z}_p$; that is, we consider an fppf abelian sheaf $\mathcal{G}_{R_\infty} \otimes W$ defined by $A \mapsto \mathcal{G}_{R_\infty} \otimes W(A) := \mathcal{G}_{R_\infty}(A) \otimes_{\mathbb{Z}_p} W$ for A running over fppf extensions of R_∞. Suppose that W is finite flat over $\mathbb{Z}_p$, then forgetting about the action of W, $\mathcal{G}_{R_\infty} \otimes W$ is isomorphic to $\mathcal{G}_{R_\infty}^{\text{rank}\, W}$ and hence is represented by a Λ_W-BT group. Note that $\mathcal{G}_{R_\infty} \otimes W$ is not a base change of $\mathcal{G}$ to W (the operation is just extending endomorphisms from Λ to Λ_W formally). For the Jacobian J_r and a free $\mathbb{Z}$-module L of finite rank, we can form the endomorphism extension $J_r \otimes L$ which send A to $J_r(A) \otimes_{\mathbb{Z}} L$ as an fppf abelian sheaf over R_r (or $\mathbb{Q}$). This fppf abelian sheaf is represented by an abelian variety, again written $J_r \otimes L$, defined over R_r (or $\mathbb{Q}$). We can take L to be the subalgebra $\mathbb{Z}[\mu_{p^r}, \alpha]$ in $\overline{\mathbb{Q}}$ generated by an algebraic integer α and p^r-th roots of unity.

We fix an embedding $i_p : \overline{\mathbb{Q}} \hookrightarrow \overline{\mathbb{Q}}_p$ and often identify $\alpha \in \overline{\mathbb{Q}}$ with $i_p(\alpha) \in \overline{\mathbb{Q}}_p$ without attaching "i_p" (if confusion is unlikely). For an eigenvalue $\alpha \in \overline{\mathbb{Q}}$ of $U(p)$ on $S_2(\Gamma_1(p^r))$, we consider the subfield $\mathbb{Q}_p(\mu_{p^r}, \alpha)$ in $\overline{\mathbb{Q}}_p$ generated by $i_p(\alpha)$ over $\mathbb{Q}_p(\mu_{p^r})$. Note here $\mathbb{Q}(\alpha)$ may not contain μ_{p^r} even if α is realized as a Hecke eigenvalue of a new form in $S_2(\Gamma_0(Np^{r+1}), \chi)$ for χ having p-conductor p^{r+1}. Let W be the p-adic integer ring of $\mathbb{Q}_p(\mu_{p^r}, \alpha)$. We put $\mathfrak{G}_{R_\infty} := \mathcal{G}_{R_\infty} \otimes W$ (the endomorphism extension). We want to know when $\mathfrak{G}_{R_\infty}[U(p) - \alpha]$ is contained in $\mathfrak{G}_{r,R_\infty} = \mathfrak{G}_{R_\infty}[\gamma^{p^{r-1}} - 1] = \mathcal{G}_{r,R_\infty} \otimes W$.

Look into the Hecke algebra $\boldsymbol{h}_W$ over Λ_W defined by

$$\boldsymbol{h}_W = \Lambda_W[\{T(n) \otimes 1\}_{p \nmid n}, U(p) \otimes 1] \subset \mathrm{End}_\Lambda(\mathfrak{G}_{R_\infty}),$$

where $T(n) \otimes 1$ sends $g \otimes w$ to $T(n)(g) \otimes w$ for $g \otimes w \in (\mathcal{G}_{R_\infty} \otimes W)(A)$ with $w \in W$ and $g \in \mathcal{G}_{R_\infty}(A)$. Hereafter, we just write simply $T(n)$ (resp. $U(p)$) for $T(n) \otimes 1$ (resp. $U(p) \otimes 1$). This is the big p-ordinary Hecke algebra over Λ_W, which is free of finite rank over Λ_W. Take a local ring $\mathbb{T}$ of $\boldsymbol{h}_W$ with maximal ideal $\mathfrak{m}$. We give ourselves a Hecke eigenvalue α given by $f|U(p) = \alpha f$ for $f \in S_2(\Gamma_0(p^r), \varepsilon)$ with $\mathbb{T} \cdot f \neq 0$. Regard ε as a character of $\mathbb{Z}_p^\times \supset \Gamma = 1 + p\mathbb{Z}_p$. Since $W \supset \mu_{p^r}(\overline{\mathbb{Q}}_p)$, W contains $\varepsilon(\gamma)$.

For a module or an fppf abelian sheaf M over R_∞ on which $\boldsymbol{h}_W$ acts via endomorphisms, adding the subscript $\mathbb{T}$, we indicate the $\mathbb{T}$-eigenspace. Therefore if M is an fppf abelian sheaf,

$$M_{\mathbb{T}}(A) = \{1_{\mathbb{T}}(x) \in M(A) \mid x \in M(A)\} = \{h(x) \mid x \in M(A), h \in \mathbb{T}\} = \mathbb{T}(M)$$

for the idempotent $1_{\mathbb{T}}$ of $\mathbb{T}$ in $\boldsymbol{h}_W$. Since $\mathbb{T}$ is a direct ring summand of $\boldsymbol{h}_W$, $M_{\mathbb{T}}$ is a direct summand of M as an fppf sheaf. In particular,

$$(7\text{-}1)\qquad \mathfrak{G}_{\mathbb{T}}(A) = \{1_{\mathbb{T}}(x) \in \mathfrak{G}(A) \mid x \in \mathfrak{G}(A)\} = \{h(x) \mid x \in \mathfrak{G}(A), h \in \mathbb{T}\}.$$

Since $\mathfrak{G}_{\mathbb{T}}$ is a direct summand of the Λ_W-BT group $\mathfrak{G}_{R_\infty}$, $\mathfrak{G}_{\mathbb{T}}$ is a Λ_W-BT group over R_∞, and more generally $\mathfrak{G}_{\mathbb{T},r} := \mathfrak{G}_{\mathbb{T}}[\gamma^{p^{r-1}} - 1]$ is a Barsotti–Tate group over R_∞. We write $t = \gamma - \varepsilon(\gamma) \in \Lambda_W$, where $\gamma = 1 + x$ as before is the fixed generator of Γ. Then we expect:

Conjecture 7.1. *Let $f|U(p) = \alpha f$ for a Hecke eigenform $f \in S_2(\Gamma_0(p^r), \varepsilon)$ with $\mathbb{T}(f) \neq 0$, and put $t = \gamma - \varepsilon(\gamma)$. Let*

$$\mathfrak{G}_{\mathbb{T}}[U(p) - \alpha, t^n] = \{x \in \mathfrak{G}_{\mathbb{T}}(\overline{\mathbb{Q}}) : x \mid U(p) = \alpha x \text{ and } t^n x = 0\}.$$

Then there exists a positive integer $s \geq r$ independent of $n < \infty$ such that the p-divisible part $\mathfrak{G}_{\mathbb{T}}[U(p) - \alpha, t^n]^{\mathrm{div}}$ is contained in $\mathfrak{G}_{\mathbb{T},s}(\overline{\mathbb{Q}})$.

The scalar $\alpha \in W$ is regarded as an operator $g \otimes w \mapsto g \otimes \alpha w$ acting on $\mathfrak{G}_{\mathbb{T}}$ under the notation introduced above. Since the Tate module $T\mathfrak{G}_{\mathbb{T},s} \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ (after extending scalars to $\mathbb{Q}_p$) is a multiplicity-free semisimple $\mathbb{T}$-module, this conjecture implies that $T(\mathfrak{G}_{\mathbb{T}}[U(p) - \alpha, t^n]^{\mathrm{div}}) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ is a multiplicity-free semisimple $\mathbb{T}$-module, and moreover, by an isogeny, the Barsotti–Tate group $\mathfrak{G}_{\mathbb{T}}[U(p) - \alpha, t^n]^{BT}$ can be brought into the abelian variety $J_s \otimes_{\mathbb{Z}} \mathbb{Z}[\mu_{p^r}, \alpha]$ (the endomorphism extension). Thus this conjecture is a semisimplicity conjecture for the α-eigenspace of $U(p)$ and conjecturally answers (to some extent) the question (Q5).

In the following section, we relate a weaker version of this conjecture to the nonvanishing problem of a certain $\mathscr{L}$-invariant which was conjectured earlier.

8. The adjoint $\mathscr{L}$-invariant

We use the notation introduced in the previous section. Let $f \in S_2(\Gamma_1(p^r))$ be a Hecke eigenform with $f|U = \alpha f$ for an algebraic integer α with either $|\alpha^\sigma| = \sqrt{p}$ for all $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ or $\alpha = \pm 1$. Take a prime $\mathfrak{p}$ of the integer ring $\overline{\mathbb{Z}}$ of $\overline{\mathbb{Q}}$. We assume that $\alpha \not\equiv 0 \bmod \mathfrak{p}$. Such an eigenform is called a $\mathfrak{p}$-ordinary eigenform. We now relate Conjecture 7.1 to a conjecture of Greenberg on the nonvanishing of an $\mathscr{L}$-invariant. This may be the only heuristic reason supporting the validity of Conjecture 7.1 at this moment. Let $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(W)$ be the $\mathfrak{p}$-adic Galois representation of f.

Suppose $W \supset \mathbb{Z}_p[\mu_{p^r}, \alpha]$ as before. Recall that $\boldsymbol{h}_W := \boldsymbol{h} \otimes_{\mathbb{Z}_p} W$ acts on the endomorphism extension (not the base change) $\mathfrak{G} = \mathcal{G} \otimes_{\mathbb{Z}_p} W$ by $h \otimes w(x \otimes w') = h(x) \otimes ww'$ for $x \in \mathfrak{G}(A) = \mathcal{G}(A) \otimes_{\mathbb{Z}_p} W$. Then $\gamma = 1 + x$ acts on f as $f|\gamma = \varepsilon(\gamma) f$ for a finite order character $\varepsilon : \Gamma \to W^\times$. Let $\mathbb{T}$ be the local ring of $\boldsymbol{h}_W$ acting nontrivially on f, and consider $\mathfrak{G}_{\mathbb{T}}$ defined in (7-1). Thus $\mathfrak{G}_{\mathbb{T}}$ is a Λ-adic Barsotti–Tate group over R_∞. Let $t = \gamma - \varepsilon(\gamma)$. Here is a weaker version of Conjecture 7.1 directly related to the adjoint $\mathscr{L}$-invariant:

Conjecture 8.1. *The p-divisible part* $\mathfrak{G}_\alpha$ *of*

$$\{x \in \mathfrak{G}_{\mathbb{T}}(\overline{\mathbb{Q}}) \mid x|U = \alpha \cdot x \text{ and } t^2 x = 0\}$$

is contained in $\mathfrak{G}_{\mathbb{T}}[\gamma^s - 1](\overline{\mathbb{Q}})$ *for sufficiently large s. In particular, the action of* Γ *on the Tate module* $T\mathfrak{G}_\alpha \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ *is semisimple.*

Let P be the kernel of $\lambda : \mathbb{T} \to W$ given by $f|h = \lambda(h)f$. Then the P-adically completed localization $\mathbb{T}_P$ is canonically isomorphic to $K[\![t]\!]$ for the quotient field K of W. Let $\boldsymbol{a}(p)$ be the image of $U = U(p)$ in $\mathbb{T}_P$. Consider $\boldsymbol{a}'(p) = d\boldsymbol{a}(p)/dt = d\boldsymbol{a}(p)/dx$. Numerically, $\boldsymbol{a}'(p)$ is almost always a unit in Λ, but there are exceptions; for example, if we take $p = 53$ and $\alpha = -1$, $\boldsymbol{a}'(p)$ is a nonunit. Suppose $\boldsymbol{a}'(p) \in \Lambda^\times$. Then $\boldsymbol{a}(p) = \alpha$ can happen on the Λ-adic Tate module $T = T\mathfrak{G}_{\mathbb{T}}$ with multiplicity 1. Thus $T/(\boldsymbol{a}(p) - \alpha)T \cong W^2$, and

$$\mathfrak{G}_\alpha = \mathrm{Ker}(U - \alpha : \mathfrak{G}_{\mathbb{T}} \to \mathfrak{G}_{\mathbb{T}}) \subset J_r[p^\infty](\overline{\mathbb{Q}}) \otimes_{\mathbb{Z}_p} W.$$

By the control theorem of $\boldsymbol{h}$ (see [Hida 2012, Sections 3.1–3.2], for instance), $\boldsymbol{a}(p)(\gamma^k - 1)$ is a $U(p)$-eigenvalue for a weight $k+2$ modular form; so, by the solution of the Ramanujan–Petersson conjecture due to Deligne, $|\boldsymbol{a}(p)(\gamma^k - 1)| = p^{(k+1)/2}$ Thus as a function of x, $\boldsymbol{a}(p)$ assumes infinitely many distinct values. Thus $\boldsymbol{a}(p)$ is transcendental over W. In particular, $\boldsymbol{a}'(p) \neq 0$. Thus, for almost all f, $t^2 \nmid (\boldsymbol{a}(p) - \alpha)$, and the conjecture holds for almost all f. Here, for simplicity, we used Deligne's result to conclude $t^2 \nmid (\boldsymbol{a}(p) - \alpha)$ for most f; there is a more elementary proof of this in [Hida 2011].

We let ρ act by conjugation on the trace 0 subspace of $M_2(W)$, which is called the adjoint square representation $\mathrm{Ad}(\rho)$ of ρ. Since $\mathrm{Ad}(\rho)([p, \mathbb{Q}_p])$ has an eigenvalue 1, the p-adic L-function $L_p^{\mathrm{an}}(s, \mathrm{Ad}(\rho)) = L_p^{\mathrm{an}}(s, \mathrm{Ad}(f))$ has an exceptional zero at $s = 1$ (see [Hida 2011, Section 2]). Following [Mazur et al. 1986], we give an analytic definition of the $\mathscr{L}$-invariant of $L_p^{\mathrm{an}}(s, \mathrm{Ad}(f))$ as

$$L_p^{\mathrm{an}}(1, \mathrm{Ad}(f)) = \mathscr{L}^{\mathrm{an}}(\mathrm{Ad}(f)) \frac{L(1, \mathrm{Ad}(f))}{c_+(\mathrm{Ad}(f))}$$

for a Shimura period $c_+(\mathrm{Ad}(f)(1))$, appropriately normalized. Note here that $L(1, \mathrm{Ad}(f))$ is nonzero. We have a power series $\Phi^{\mathrm{an}}(x) \in W[\![x]\!]$ such that $L_p^{\mathrm{an}}(s, \mathrm{Ad}(f)) = \Phi^{\mathrm{an}}(\gamma^{s-1} - 1)$ regarding $\gamma \in 1 + p\mathbb{Z}_p$. We have $\Phi^{\mathrm{an}}(x) = x\Psi^{\mathrm{an}}(x)$ with $\Psi^{\mathrm{an}}(x) \in W[\![x]\!]$, and $\mathscr{L}^{\mathrm{an}}(\mathrm{Ad}(f))$ is a nonzero constant multiple of $\Psi^{\mathrm{an}}(0)$.

Let $\mathbb{Q}_\infty/\mathbb{Q}$ be the cyclotomic $\mathbb{Z}_p$-extension. The arithmetic p-adic L-function is defined as $L_p(s, \mathrm{Ad}(\rho)) = \Phi(\gamma^{s-1} - 1)$ for the characteristic power series $\Phi(x) \in W[\![x]\!]$ of the adjoint square Selmer group $\mathrm{Sel}_{\mathbb{Q}_\infty}(\mathrm{Ad}(\rho))$ defined by Greenberg (see [Greenberg 1994] and [Hida 2011]), where we identify $\mathrm{Gal}(\mathbb{Q}_\infty/\mathbb{Q})$ with $\Gamma = 1 + p\mathbb{Z}_p$ by the cyclotomic character, and the Iwasawa algebra $W[\![\Gamma]\!]$ with $W[\![x]\!]$ by $\Gamma \ni 1+p \mapsto 1+x$. It is known that $\Phi(x) = x\Psi(x)$ with $\Psi(x) \in W[\![x]\!]$ (see

[Greenberg 1994]). Thus, the arithmetic L-function $L_p(s, \mathrm{Ad}(\rho))$ has a zero at $s = 1$. Greenberg has defined his $\mathscr{L}$-invariant $\mathscr{L}(\mathrm{Ad}(\rho))$ by purely Galois cohomological means and proved that $\Psi(0)$ is a multiple of $\mathscr{L}(\mathrm{Ad}(\rho))$ by a simple constant (up to units in W; see [ibid.]).

The main conjecture in this setting predicts the equality $\Phi(x) = \Phi^{\mathrm{an}}(x)$ up to units in $W[\![x]\!]$ (assuming that ρ is residually absolutely irreducible). The conjecture has been proven in many cases by [Urban 2006]. We assume the following condition:

(H) $\bar{\rho} = (\rho \bmod \mathfrak{m}_W)$ is absolutely irreducible over $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}[\sqrt{p^*}])$ for $p^* = (-1)^{(p-1)/2}p$, or the semisimplification of ρ restricted to $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ is the sum of two distinct characters, or $\alpha = \pm 1$.

Under this circumstance, regarding $\gamma = 1 + p \in 1 + p\mathbb{Z}_p$, it is known that

$$\mathscr{L}(\mathrm{Ad}(\rho)) = -2\log_p(\gamma)\alpha^{-1}\frac{d\boldsymbol{a}(p)}{dx}\Big|_{t=0}. \tag{8-1}$$

This follows from [Greenberg and Stevens 1993] if $\alpha = 1$, because in this case $\mathscr{L}(\mathrm{Ad}(\rho)) = \mathscr{L}(\rho)$. Otherwise, it is proven in [Hida 2004] and [2011]. Though Greenberg made the following conjecture in a more general setting, if we limit ourselves to $\mathrm{Ad}(\rho)$ for the ordinary modular Galois representation ρ, Conjecture 8.1 is equivalent to the following conjecture under (H):

Conjecture 8.2 (R. Greenberg). $\mathscr{L}(\mathrm{Ad}(\rho)) \neq 0$.

References

[Bosch et al. 1990] S. Bosch, W. Lütkebohmert, and M. Raynaud, *Néron models*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) **21**, Springer, Berlin, 1990. MR 91i:14034 Zbl 0705.14001

[Brown 1982] K. S. Brown, *Cohomology of groups*, Graduate Texts in Mathematics **87**, Springer, New York, 1982. MR 83k:20002 Zbl 0584.20036

[Cais 2012] B. Cais, "The Geometry of Hida Families and Λ-adic Hodge theory", preprint, 2012, Available at http://math.arizona.edu/~cais/Papers/Published/hidafinalv3.pdf.

[Carayol 1986] H. Carayol, "Sur les représentations l-adiques associées aux formes modulaires de Hilbert", *Ann. Sci. École Norm. Sup.* (4) **19**:3 (1986), 409–468. MR 89c:11083 Zbl 0616.10025

[Gelbart 1975] S. S. Gelbart, *Automorphic forms on adèle groups*, Annals of Mathematics Studies **83**, Princeton University Press, 1975. MR 52 #280 Zbl 0329.10018

[Greenberg 1994] R. Greenberg, "Trivial zeros of p-adic L-functions", pp. 149–174 in *p-adic monodromy and the Birch and Swinnerton–Dyer conjecture* (Boston, MA, 1991), edited by B. Mazur and G. Stevens, Contemp. Math. **165**, Amer. Math. Soc., Providence, RI, 1994. MR 95h:11063 Zbl 0838.11070

[Greenberg and Stevens 1993] R. Greenberg and G. Stevens, "p-adic L-functions and p-adic periods of modular forms", *Invent. Math.* **111**:2 (1993), 407–447. MR 93m:11054 Zbl 0778.11034

[Hida 1986] H. Hida, "Galois representations into $\mathrm{GL}_2(\mathbf{Z}_p[[X]])$ attached to ordinary cusp forms", *Invent. Math.* **85**:3 (1986), 545–613. MR 87k:11049 Zbl 0612.10021

[Hida 2000a] H. Hida, "Control theorems and applications", Lecture notes, Tata institute of fundamental research, 2000, Available at http://www.math.ucla.edu/∼hida/Control.dvi.gz.

[Hida 2000b] H. Hida, *Modular forms and Galois cohomology*, Cambridge Studies in Advanced Mathematics **69**, Cambridge University Press, 2000. MR 2002b:11071 Zbl 0952.11014

[Hida 2004] H. Hida, "Greenberg's $\mathcal{L}$-invariants of adjoint square Galois representations", *Int. Math. Res. Not.* **2004**:59 (2004), 3177–3189. MR 2005f:11108 Zbl 1081.11039

[Hida 2005] H. Hida, "Λ-adic p-divisible groups, I", Lecture notes, Centre de Recherches Mathématiques, Montréal, 2005, Available at http://www.math.ucla.edu/∼hida/CRM1.pdf.

[Hida 2011] H. Hida, "Constancy of adjoint $\mathcal{L}$-invariant", *J. Number Theory* **131**:7 (2011), 1331–1346. MR 2012k:11071 Zbl 1229.11061

[Hida 2012] H. Hida, *Geometric modular forms and elliptic curves*, 2nd ed., World Scientific, Hackensack, NJ, 2012. MR 2012j:11100 Zbl 1254.11055

[Hida 2013] H. Hida, "Image of Λ-adic Galois representations modulo p", *Invent. Math.* **194**:1 (2013), 1–40. MR 3103254 Zbl 06225180

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

[Matsumura 1986] H. Matsumura, *Commutative ring theory*, Cambridge Studies in Advanced Mathematics **8**, Cambridge University Press, 1986. MR 88h:13001 Zbl 0603.13001

[Mazur 1978] B. Mazur, "Rational isogenies of prime degree", *Invent. Math.* **44**:2 (1978), 129–162. MR 80h:14022 Zbl 0386.14009

[Mazur 1989] B. Mazur, "Deforming Galois representations", pp. 385–437 in *Galois groups over* **Q** (Berkeley, CA, 1987), edited by Y. Ihara et al., Math. Sci. Res. Inst. Publ. **16**, Springer, New York, 1989. MR 90k:11057 Zbl 0714.11076

[Mazur and Wiles 1983] B. Mazur and A. Wiles, "Analogies between function fields and number fields", *Amer. J. Math.* **105**:2 (1983), 507–521. MR 84g:12003 Zbl 0531.12015

[Mazur and Wiles 1984] B. Mazur and A. Wiles, "Class fields of abelian extensions of **Q**", *Invent. Math.* **76**:2 (1984), 179–330. MR 85m:11069 Zbl 0545.12005

[Mazur and Wiles 1986] B. Mazur and A. Wiles, "On p-adic analytic families of Galois representations", *Compositio Math.* **59**:2 (1986), 231–264. MR 88e:11048 Zbl 0654.12008

[Mazur et al. 1986] B. Mazur, J. Tate, and J. Teitelbaum, "On p-adic analogues of the conjectures of Birch and Swinnerton–Dyer", *Invent. Math.* **84**:1 (1986), 1–48. MR 87e:11076 Zbl 0699.14028

[Milne 1980] J. S. Milne, *Étale cohomology*, Princeton Mathematical Series **33**, Princeton University Press, 1980. MR 81j:14002 Zbl 0433.14012

[Miyake 1989] T. Miyake, *Modular forms*, Springer, Berlin, 1989. MR 90m:11062 Zbl 0701.11014

[Ohta 1995] M. Ohta, "On the p-adic Eichler–Shimura isomorphism for Λ-adic cusp forms", *J. Reine Angew. Math.* **463** (1995), 49–98. MR 96e:11061 Zbl 0827.11025

[Raynaud 1974] M. Raynaud, "Schémas en groupes de type $(p, \dots, p)$", *Bull. Soc. Math. France* **102** (1974), 241–280. MR 54 #7488 Zbl 0325.14020

[Serre and Tate 1968] J.-P. Serre and J. Tate, "Good reduction of abelian varieties", *Ann. of Math.* (2) **88** (1968), 492–517. MR 38 #4488 Zbl 0172.46101

[Shimura 1971] G. Shimura, *Introduction to the arithmetic theory of automorphic functions*, Publications of the Mathematical Society of Japan **11**, Princeton University Press, 1971. MR 47 #3318 Zbl 0221.10029

[Tate 1967] J. T. Tate, "p-divisible groups", pp. 158–183 in *Proc. Conf. local fields* (Driebergen, 1966), edited by T. A. Springer, Springer, Berlin, 1967. MR 38 #155 Zbl 0157.27601

[Tilouine 1987] J. Tilouine, "Un sous-groupe p-divisible de la jacobienne de $X_1(Np^r)$ comme module sur l'algèbre de Hecke", *Bull. Soc. Math. France* **115**:3 (1987), 329–360. MR 88m:11043 Zbl 0677.14006

[Urban 2006] E. Urban, "Groupes de Selmer et Fonctions L p-adiques pour les representations modulaires adjointes", preprint, 2006, Available at http://www.math.columbia.edu/~urban/eurp/ADJMC.pdf.

[Wiles 1995] A. Wiles, "Modular elliptic curves and Fermat's last theorem", *Ann. of Math.* (2) **141**:3 (1995), 443–551. MR 96d:11071 Zbl 0823.11029

Received June 23, 2013. Revised October 28, 2013.

HARUZO HIDA
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA, LOS ANGELES
BOX 951555
LOS ANGELES, CA 90095
UNITED STATES

hida@math.ucla.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 268, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.268.313

LE FLOT GÉODÉSIQUE DES QUOTIENTS GÉOMÉTRIQUEMENT FINIS DES GÉOMÉTRIES DE HILBERT

MICKAËL CRAMPON ET LUDOVIC MARQUIS

On étudie le flot géodésique des quotients géométriquement finis Ω/Γ de géométries de Hilbert, en particulier ses propriétés de récurrence.

On prouve, sous une hypothèse géométrique sur les pointes, que le flot géodésique est uniformément hyperbolique. Sans cette hypothèse, on construit un exemple où celui-ci a un exposant de Lyapunov nul.

On fait le lien entre la dynamique du flot géodésique et certaines propriétés du convexe Ω et du groupe Γ. On en déduit des résultats de rigidité, qui étendent ceux de Benoist et Guichard pour les quotients compacts.

Enfin, on s'intéresse au lien entre entropie volumique et exposant critique ; on montre entre autres qu'ils coïncident lorsque le quotient est de volume fini.

We study the geodesic flow of geometrically finite quotients Ω/Γ of Hilbert geometries, in particular its recurrence properties.

We prove that, under a geometric assumption on the cusps, the geodesic flow is uniformly hyperbolic. Without this assumption, we provide an example of a quotient whose geodesic flow has a zero Lyapunov exponent.

We make the link between the dynamics of the geodesic flow and some properties of the convex set Ω and the group Γ. As a consequence, we get various rigidity results which extend previous results of Benoist and Guichard for compact quotients.

Finally, we study the link between volume entropy and critical exponent; for example, we show that they coincide provided the quotient has finite volume.

MSC2010 : 20H10, 22E40, 37D40, 53A20.
Mots-clefs : Hilbert geometry, geodesic flow, geometrically finite, volume entropy.

1. Introduction

Cet article dynamique fait logiquement suite à l'article géométrique [Crampon et Marquis 2012]*, dans lequel nous étudions la notion de finitude géométrique en géométrie de Hilbert. Avec* [Crampon et Marquis 2013]*, ils forment un seul travail que nous avons découpé en trois pour des raisons évidentes de longueur. Concernant la géométrie des variétés géométriquement finies, nous ne rappellerons dans ce texte que les résultats dont nous ferons usage et renvoyons le lecteur à* [Crampon et Marquis 2012] *pour plus d'informations.*

Une géométrie de Hilbert est un espace métrique (Ω, d_Ω) où Ω est un ouvert proprement convexe de l'espace projectif réel $\mathbb{P}^n = \mathbb{P}^n(\mathbb{R})$ et d_Ω est la distance définie sur Ω par

$$d_\Omega(x, y) = \tfrac{1}{2}\big|\ln\big([p : q : x : y]\big)\big|, \quad x, y \in \Omega \text{ distincts;}$$

dans cette formule, les points p et q sont les points d'intersection de la droite (xy) avec le bord $\partial\Omega$ de Ω. Ces géométries ont été introduites par Hilbert comme exemples de géométries dans lesquelles les droites sont des géodésiques. Leur définition imite celle de l'espace hyperbolique dans le modèle projectif de Beltrami, qui correspond à la géométrie de Hilbert définie par un ellipsoïde.

Lorsque l'ouvert convexe Ω est strictement convexe, la géométrie de Hilbert est uniquement géodésique : les droites sont les seules géodésiques. On peut dans ce cas définir le flot géodésique sans recourir à des équations géodésiques, comme on le fait de façon traditionnelle en géométrie riemannienne. Le flot géodésique est ainsi le flot défini sur le fibré homogène $H\Omega = T\Omega \smallsetminus \{0\}/\mathbb{R}^+$ de la façon suivante :

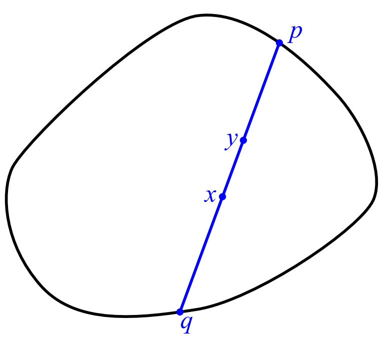


Figure 1. La distance de Hilbert.

si $w = (x, [\xi])$ est un point de $H\Omega$, consistant en un point x de Ω et une direction tangente $[\xi]$, on trouve son image $\varphi^t(w)$ en suivant la droite géodésique partant de x dans la direction $[\xi]$.

Les géométries de Hilbert sont des espaces finslériens : la métrique de Hilbert est engendrée par un champ de normes $F : T\Omega \to \mathbb{R}$ sur Ω, donné par la formule

$$F(x, \xi) = \frac{|\xi|}{2}\left(\frac{1}{|xx^-|} + \frac{1}{|xx^+|}\right), \quad (x, \xi) \in T\Omega,$$

où x^+ et x^- sont les points d'intersection de la droite $\{x + \lambda\xi \mid \lambda \in \mathbb{R}\}$ avec $\partial\Omega$ (voir le paragraphe 2A pour plus de précisions). Lorsque $\partial\Omega$ est de classe $\mathscr{C}^2$ à hessien défini positif, alors on peut définir les géodésiques au moyen d'une équation différentielle, et le flot géodésique est le flot de cette équation.

La géométrie de Hilbert définie par un ouvert strictement convexe à bord $\mathscr{C}^1$ possède *un certain comportement hyperbolique*. Par exemple, dans ce cadre-là, on voit apparaître naturellement, au moyen des horosphères, les variétés stables et instables du flot géodésique. Le flot géodésique est dans ce cas de classe $\mathscr{C}^1$ et l'espace tangent à $H\Omega$ admet une décomposition en

$$TH\Omega = \mathbb{R} \cdot X \oplus E^s \oplus E^u,$$

où X est le générateur du flot, E^s est la distribution stable tangente au feuilletage stable, et E^u est la distribution instable.

Les flots géodésiques des variétés riemanniennes compactes de courbure négative sont les premiers exemples de flots d'Anosov, ou uniformément hyperboliques. Cette propriété d'hyperbolicité ne dépend que des bornes sur la courbure et elle reste donc vraie pour une variété riemannienne non compacte à courbure négative $K < -a^2 < 0$.

Pour une géométrie de Hilbert quelconque, on ne peut espérer obtenir de propriété d'hyperbolicité. En effet, le comportement asymptotique autour d'une géodésique dépend de la régularité du bord du convexe au point extrémal de la géodésique (voir [Crampon 2014] pour une étude détaillée). Par contre, si la géométrie admet un quotient assez petit, on peut s'attendre à des propriétés de récurrence sur le quotient.

C'est le cas lorsqu'il existe un quotient compact : dans [Benoist 2004], Yves Benoist a prouvé que le flot géodésique d'un quotient compact d'une géométrie de Hilbert (avec Ω strictement convexe) était un flot d'Anosov. Notre premier théorème généralise cela au flot géodésique de certaines variétés géométriquement finies.

Les variétés géométriquement finies sont en quelque sorte les variétés non compactes les plus simples. Leur caractéristique essentielle pour nous est que leur cœur convexe se décompose en une partie compacte et un nombre fini de pointes (*cusps*, en anglais). C'est essentiel car le cœur convexe est le support de l'ensemble non errant du flot géodésique ; c'est donc là que se concentre la dynamique. On renvoie au fait 3 ou à l'article [Crampon et Marquis 2012] pour plus de détails.

À chaque fois, on va essayer de comprendre séparément ce qu'il se passe sur la partie compacte puis sur les parties cuspidales. De façon générale, on ne peut rien dire sans faire d'hypothèses sur la géométrie des pointes :

Proposition 1.1 (proposition 8.1). *Il existe une variété géométriquement finie $M = \Omega/\Gamma$ dont le flot géodésique a un exposant de Lyapunov nul. En particulier, le flot géodésique n'est pas uniformément hyperbolique.*

Dans ce texte, nous étudierons donc principalement les variétés géométriquement finies dont les pointes sont « asymptotiquement hyperboliques » : dans une pointe, la métrique de Hilbert est équivalente à une métrique hyperbolique qui a les mêmes géodésiques (non paramétrées) ; voir la définition 4.4. Parmi les variétés géométriquement finies à pointes asymptotiquement hyperboliques, on trouve en particulier les variétés de volume fini, et plus généralement celles dont les sous-groupes paraboliques maximaux sont de rang maximal, c'est-à-dire qu'ils agissent cocompactement sur $\partial\Omega \setminus \{p\}$, où p est le point fixe du groupe parabolique considéré.

Il est fort possible que pour toute variété géométriquement finie $M = \Omega/\Gamma$, il existe un ouvert Ω', Γ-invariant, strictement convexe et à bord $\mathscr{C}^1$, tel que le quotient $M' = \Omega'/\,\Gamma$ est géométriquement fini à pointes asymptotiquement hyperboliques. La raison principale qui nous pousse à penser qu'une telle construction est possible est que les sous-groupes paraboliques d'un tel groupe Γ sont conjugués à des sous-groupes paraboliques de $\mathrm{SO}_{n,1}(\mathbb{R})$.

Pour ces variétés-là, on peut prouver ceci :

Théorème 1.2 (théorème 5.2). *Soient Ω un ouvert strictement convexe et à bord $\mathscr{C}^1$, et $M = \Omega/\Gamma$ une variété géométriquement finie à pointes asymptotiquement hyperboliques. Le flot géodésique de la métrique de Hilbert est uniformément hyperbolique sur son ensemble non errant* $\mathtt{NW}$: *le fibré tangent à HM admet en tout point de* $\mathtt{NW}$ *une décomposition φ^t-invariante*

$$T\,HM = \mathbb{R}\cdot X \oplus E^s \oplus E^u,$$

telle qu'il existe des constantes $\chi, C > 0$ pour lesquelles

$$\|d\varphi^t Z^s\| \leqslant Ce^{-\chi t} \quad \text{et} \quad \|d\varphi^{-t} Z^u\| \leqslant Ce^{-\chi t}, \qquad Z^s \in E^s,\ Z^u \in E^u,\ t \geqslant 0. \tag{1-1}$$

De façon générale, on prouvera aussi les propriétés de récurrence suivantes :

Proposition 1.3 (proposition 6.1). *Soient Ω un ouvert strictement convexe et à bord $\mathscr{C}^1$, et $M = \Omega/\Gamma$ une variété quotient. Le flot géodésique de M est topologiquement mélangeant sur son ensemble non errant.*

Notre deuxième théorème concerne la régularité du bord des ouverts convexes Ω qui admettent un quotient géométriquement fini $M = \Omega/\Gamma$ non compact. Ce résultat est lié au fait que les propriétés hyperboliques des orbites du flot géodésique se lisent directement sur la régularité du bord au niveau de leur point extrémal.

Bien entendu, cela permet de décrire le bord uniquement au niveau de l'ensemble limite Λ_Γ du groupe. Ce n'est pas étonnant puisque celui-ci constitue l'ensemble des points extrémaux des géodésiques récurrentes. De plus, c'est la seule partie du bord qui est *imposée* par le groupe Γ : on peut en effet modifier le bord (presque) à sa guise hors de l'ensemble limite ; c'est d'ailleurs ainsi qu'on obtient l'exemple de la proposition 1.1. Pour un quotient compact ou de volume fini, l'ensemble limite est le bord tout entier et donc le convexe Ω est entièrement déterminé par le groupe Γ.

Théorème 1.4 (corollaire 7.3). *Soient Ω un ouvert strictement convexe et à bord $\mathscr{C}^1$, et $M = \Omega/\Gamma$ une variété géométriquement finie à pointes asymptotiquement hyperboliques. Il existe $\varepsilon > 0$ tel que le bord $\partial\Omega$ du convexe Ω est de classe $\mathscr{C}^{1+\varepsilon}$ en tout point de Λ_Γ.*

Via la caractérisation des quotients de volume fini par leur ensemble limite, on obtient le corollaire suivant.

Corollaire 1.5 (corollaire 7.3). *Soit Ω un ouvert strictement convexe et à bord $\mathscr{C}^1$. Si Ω admet un quotient de volume fini, alors son bord $\partial\Omega$ est de classe $\mathscr{C}^{1+\varepsilon}$ pour un certain $\varepsilon > 0$.*

Lorsque Ω admet un quotient compact, Olivier Guichard a pu déterminer exactement la régularité optimale du bord, c'est-à-dire le plus grand ε tel que le bord $\partial\Omega$ est $\mathscr{C}^{1+\varepsilon}$. Celle-ci est encore une fois déterminée par le groupe Γ, via les valeurs propres de ses éléments hyperboliques. Cela n'est pas étonnant, étant donné que les orbites périodiques sont denses, et que celles-ci sont en bijection avec les classes de conjugaison d'éléments hyperboliques de Γ.

Si on se restreint à l'ensemble limite et l'ensemble non errant, cette observation reste valable pour un quotient quelconque. Nous pouvons ainsi prouver un résultat similaire pour les ouverts convexes qui admettent un quotient géométriquement fini à pointes asymptotiquement hyperboliques. Pour l'énoncé, définissons d'abord

$$\varepsilon(\Lambda_\Gamma) = \sup\{\varepsilon \in [0,1] \mid \text{le bord } \partial\Omega \text{ est } \mathscr{C}^{1+\varepsilon} \text{ en tout point de } \Lambda_\Gamma\}.$$

En suite, pour tout élément hyperbolique $\gamma \in \Gamma$, notons

$$\varepsilon(\gamma) = \sup\{\varepsilon \in [0,1] \mid \text{le bord } \partial\Omega \text{ est } \mathscr{C}^{1+\varepsilon} \text{ au point attractif } x_\gamma^+ \text{ de } \gamma\},$$

et $\varepsilon(\Gamma) = \inf\{\varepsilon(\gamma) \mid \gamma \in \Gamma \text{ hyperbolique}\}$. Ainsi, le bord $\partial\Omega$ est $\mathscr{C}^{1+\varepsilon(\Gamma)}$ en tout point fixe hyperbolique. On obtient alors :

Théorème 1.6 (théorème 7.4). *Soient Ω un ouvert strictement convexe et à bord $\mathscr{C}^1$, et $M = \Omega/\Gamma$ une variété géométriquement finie à pointes asymptotiquement hyperboliques. On a*

$$\varepsilon(\Lambda_\Gamma) = \varepsilon(\Gamma).$$

Notre démonstration de ce théorème est différente de celle de Guichard et repose sur l'extension d'un théorème de Ursula Hamenstädt [1994], qui s'intéresse au meilleur coefficient de contraction d'un flot uniformément hyperbolique.

Pour l'énoncer, il nous faut définir les meilleurs coefficients de contraction du flot sur l'ensemble non errant

$$\chi(\mathtt{NW}) = \sup\{\chi \mid \text{il existe } C > 0 \text{ tel que l'inégalité (1-1) a lieu en tout point de } \mathtt{NW}\},$$

et sur les orbites périodiques

$$\chi(\mathrm{Per}) = \inf\{\chi(w) \mid w \in \mathtt{NW} \text{ périodique}\}.$$

Théorème 1.7 (théorème 7.5). *Soient Ω un ouvert strictement convexe et à bord $\mathscr{C}^1$, et $M = \Omega/\Gamma$ une variété géométriquement finie à pointes asymptotiquement hyperboliques. On a*

$$\chi(\mathrm{Per}) = \chi(\mathtt{NW}).$$

Comme corollaire de ces résultats et du travail précédent [Crampon et Marquis 2012], on obtient un résultat de rigidité :

Théorème 1.8 (corollaire 7.10). *Soit Ω un ouvert strictement convexe et à bord $\mathscr{C}^1$, qui admet une action géométriquement finie d'un groupe Γ contenant un élément parabolique. Si le bord $\partial\Omega$ est de classe $\mathscr{C}^{1+\varepsilon}$ pour tout* $0 < \varepsilon < 1$, *alors Γ est un sous-groupe d'un conjugué de* $\mathrm{SO}_{n,1}(\mathbb{R})$.

Comme cas particulier, on obtient une version volume fini d'un théorème de Benoist [2004] qui concernait les quotients compacts :

Corollaire 1.9 (corollaire 7.11). *Soit Ω un ouvert strictement convexe et à bord $\mathscr{C}^1$ qui admet un quotient de volume fini non compact. Si le bord $\partial\Omega$ est de classe $\mathscr{C}^{1+\varepsilon}$ pour tout* $0 < \varepsilon < 1$, *alors Ω est un ellipsoïde.*

Remarquons que dans l'énoncé de ce théorème, tout comme dans celui du corollaire 1.5, l'une des hypothèses strictement convexe/à bord $\mathscr{C}^1$ est superflue : c'est une conséquence du travail de Daryl Cooper, Darren Long et Stephan Tillmann [Cooper et al. 2011].

À la fin de ce texte, on revient sur la représentation sphérique de $\mathrm{SL}_2(\mathbb{R})$ dans $\mathrm{SL}_5(\mathbb{R})$, que nous avions étudié dans [Crampon et Marquis 2012] car elle permettait de distinguer les deux notions de finitude géométrique que nous y avions introduites. En particulier, on avait vu que l'ensemble des ouverts proprement convexes préservés par cette représentation formait, à action de $\mathrm{SL}_5(\mathbb{R})$ près, une famille croissante

$\{\Omega_r \mid 0 \leqslant r \leqslant \infty\}$. Parmi eux, les convexes Ω_0 et Ω_∞, duaux l'un de l'autre, n'étaient ni strictement convexes ni à bord $\mathscr{C}^1$. Les autres par contre l'étaient. En fait, on peut déterminer précisément leur régularité (voir définition 4.5 pour les notions de régularité $\mathscr{C}^{1+\varepsilon}$ et la β-convexité) :

Proposition 1.10 (proposition 8.2). *Pour* $0 < r < \infty$, *le bord de l'ouvert convexe* Ω_r *est de classe* $\mathscr{C}^{4/3}$ *et* 4*-convexe.*

Nous avons inclus pour finir une première étude de la croissance des groupes discrets dont l'action est géométriquement finie sur Ω. L'objet principal est l'exposant critique δ_Γ du groupe Γ, qui mesure la croissance exponentielle du groupe agissant sur Ω ; à savoir

$$\delta_\Gamma = \limsup_{R\to+\infty} \frac{1}{R} \log \sharp\{g \in \Gamma \mid d_\Omega(x, gx) \leqslant R\}.$$

Lorsque Γ est un groupe cocompact, il est immédiat que l'exposant critique et l'entropie volumique de la géométrie de Hilbert sont égaux. Rappelons que l'entropie volumique de la géométrie de Hilbert (Ω, d_Ω) est le taux de croissance exponentiel des volumes des boules :

$$h_{\mathrm{vol}}(\Omega) = \limsup_{R\to+\infty} \frac{1}{R} \log \mathrm{Vol}_\Omega B(x, R).$$

Lorsque Γ n'est plus cocompact, on a de façon générale $\delta_\Gamma \leqslant h_{\mathrm{vol}}$ mais il n'y a a priori plus de raisons pour que ces deux quantités coïncident, même si Γ est de covolume fini : Françoise Dal'bo, Marc Peigné, Jean-Claude Picaud et Andrea Sambusetti ont construit des exemples de réseaux non uniformes d'espaces de courbure négative pincée où $\delta_\Gamma < h_{\mathrm{vol}}$.

Dans notre cas, le fait que les pointes d'une variété de volume fini soient asymptotiquement hyperboliques entraîne l'égalité :

Théorème 1.11 (théorème 9.2). *Soient* Ω *un ouvert strictement convexe et à bord* $\mathscr{C}^1$, *et* Γ *un sous-groupe discret de* $\mathrm{Aut}(\Omega)$ *de covolume fini. Alors*

$$\delta_\Gamma = h_{\mathrm{vol}}(\Omega).$$

Ce résultat peut même s'étendre au cas des actions géométriquement finies de la manière suivante :

Théorème 1.12 (théorème 9.7). *Soit* Γ *un sous-groupe discret de* $\mathrm{Aut}(\Omega)$ *dont l'action sur* Ω *est géométriquement finie. Alors*

$$\delta_\Gamma = \limsup_{R\to+\infty} \frac{1}{R} \log \mathrm{Vol}_\Omega(B(o, R) \cap C(\Lambda_\Gamma)),$$

où o *est un point quelconque de* Ω.

Plan. Les sections 2 et 3 sont des préliminaires portant respectivement sur les géométries de Hilbert et leur flot géodésique.

La section 4 explique ce qui nous sera utile sur les variétés géométriquement finies, en présentant notamment l'hypothèse d'asymptoticité hyperbolique des pointes.

La section 5 est consacrée à la démonstration du théorème 1.2. Bien que l'idée soit claire et très simple, la démonstration reste malgré tout quelque peu technique.

La section 6 se concentre sur les propriétés de récurrence du flot géodésique d'une variété quelconque $M = \Omega/\Gamma$; en particulier, on y montre la proposition 1.3.

Dans la section 7, on s'intéresse à la régularité du bord de l'ouvert convexe. C'est là qu'on montre le théorème 1.4 et le corollaire 1.5. Une bonne partie de cette section est dédiée au théorème 1.6, via le théorème 1.7 dont la démonstration présente quelques technicités.

La section 8 construit le contre-exemple de la proposition 1.1 et détaille la proposition 1.10. Enfin, dans la section 9, on montre les théorèmes 1.11 et 1.12 qui lient exposant critique et entropie volumique. Là encore, les démonstrations présentent quelques difficultés techniques.

2. Géométries de Hilbert

2A. ***Distance et volume.*** Une *carte affine* A de $\mathbb{P}^n$ est le complémentaire d'un hyperplan projectif. Une carte affine possède une structure naturelle d'espace affine. Un ouvert Ω de $\mathbb{P}^n$ différent de $\mathbb{P}^n$ est *convexe* lorsqu'il est inclus dans une carte affine et qu'il est convexe dans cette carte. Un ouvert convexe Ω de $\mathbb{P}^n$ est dit *proprement convexe* lorsqu'il existe une carte affine contenant son adhérence $\overline{\Omega}$. Autrement dit, un ouvert convexe est proprement convexe lorsqu'il ne contient pas de droite affine. Un ouvert proprement convexe Ω de $\mathbb{P}^n$ est dit *strictement convexe* lorsque son bord $\partial\Omega$ ne contient pas de segment non trivial.

Hilbert a introduit sur un ouvert proprement convexe Ω de $\mathbb{P}^n$ la distance qui porte aujourd'hui son nom. Pour $x \neq y \in \Omega$, on note p, q les points d'intersection de la droite (xy) et du bord $\partial\Omega$ de Ω, de telle façon que x soit entre q et y, et y entre x et p (voir figure 2). On pose

$$d_\Omega(x, y) = \tfrac{1}{2}\ln\big([p : q : x : y]\big) = \tfrac{1}{2}\ln\frac{|qy|\cdot|px|}{|qx|\cdot|py|} \quad \text{et} \quad d_\Omega(x, x) = 0,$$

où :

1. la quantité $[p : q : x : y]$ désigne le birapport des points p, q, x, y ;
2. $|\cdot|$ est une norme euclidienne quelconque sur une carte affine A qui contient l'adhérence $\overline{\Omega}$ de Ω.

Le birapport étant une notion projective, il est clair que d_Ω ne dépend ni du choix de A, ni du choix de la norme euclidienne sur A.

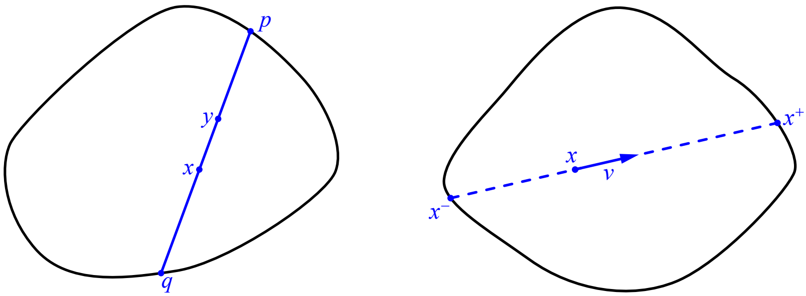


Figure 2. La distance de Hilbert et la norme de Finsler.

Fait 1. *Soit Ω un ouvert proprement convexe de $\mathbb{P}^n$.*

1. *d_Ω est une distance sur Ω.*
2. *(Ω, d_Ω) est un espace métrique complet.*
3. *La topologie induite par d_Ω coïncide avec celle induite par $\mathbb{P}^n$.*
4. *Le groupe* $\mathrm{Aut}(\Omega)$ *des transformations projectives de* $\mathrm{SL}_{n+1}(\mathbb{R})$ *qui préservent Ω est un sous-groupe fermé de* $\mathrm{SL}_{n+1}(\mathbb{R})$ *qui agit par isométries sur (Ω, d_Ω). Il agit donc proprement sur Ω.*

La distance de Hilbert d_Ω est induite par une structure finslérienne sur l'ouvert Ω. On choisit une carte affine A et une métrique euclidienne $|\cdot|$ sur A pour lesquelles Ω apparaît comme un ouvert convexe borné. On identifie le fibré tangent $T\Omega$ de Ω à $\Omega \times A$. Soient $x \in \Omega$ et $v \in A$, on note $x^+ = x^+(x, v)$ (resp. x^-) le point d'intersection de la demi-droite définie par x et v (resp. $-v$) avec $\partial\Omega$ (voir figure 2). On pose

$$F(x, v) = \frac{|v|}{2}\left(\frac{1}{|xx^-|} + \frac{1}{|xx^+|}\right), \tag{2-1}$$

quantité indépendante du choix de A et de $|\cdot|$, puisqu'on ne considère que des rapports de longueurs.

Fait 2. *Soient Ω un ouvert proprement convexe de $\mathbb{P}^n$ et A une carte affine qui contient $\overline{\Omega}$. La distance induite par la métrique finslérienne F est la distance d_Ω. Autrement dit on a les formules suivantes* :

- $\mathscr{F}(x, v) = \frac{d}{dt}\Big|_{t=0} d_\Omega(x, x + tv)$, *pour $v \in A$.*
- $d_\Omega(x, y) = \inf \int_0^1 F(\dot\sigma(t))\, dt$, *où l'infimum est pris sur les chemins σ de classe $\mathscr{C}^1$ tel que $\sigma(0) = x$ et $\sigma(1) = y$.*

Il y a plusieurs manières naturelles de construire un volume pour une géométrie de Finsler, la définition riemannienne acceptant plusieurs généralisations. Nous travaillerons avec le volume de Busemann, noté Vol_Ω.

Pour le construire, on se donne une carte affine A et une métrique euclidienne $|\cdot|$ sur A pour lesquelles Ω apparaît comme un ouvert convexe borné. On note $B_{T_x\Omega}(r) = \{v \in T_x\Omega \mid F(x,v) < r\}$ la boule de rayon $r > 0$ de l'espace tangent à Ω en x, Vol la mesure de Lebesgue sur A associée à $|\cdot|$ et $v_n = \mathrm{Vol}(\{v \in A \mid |v| < 1\})$ le volume de la boule unité euclidienne en dimension n.

Pour tout borélien $\mathscr{A} \subset \Omega \subset A$, on pose :

$$\mathrm{Vol}_\Omega(\mathscr{A}) = \int_{\mathscr{A}} \frac{v_n}{\mathrm{Vol}(B_{T_x\Omega}(1))}\, d\mathrm{Vol}(x)$$

Là encore, la mesure Vol_Ω est indépendante du choix de A et de $|\cdot|$. En particulier, elle est préservée par le groupe $\mathrm{Aut}(\Omega)$.

La proposition suivante permet de comparer deux géométries de Hilbert entre elles.

Proposition 2.1. *Soient Ω_1 et Ω_2 deux ouverts proprement convexes de $\mathbb{P}^n$ tels que $\Omega_1 \subset \Omega_2$.*

- *Les métriques finslériennes F_1 et F_2 de Ω_1 et Ω_2 vérifient : $F_2(w) \leqslant F_1(w)$, $w \in T\Omega_1 \subset T\Omega_2$, l'égalité ayant lieu si et seulement si $x^+_{\Omega_1}(w) = x^+_{\Omega_2}(w)$ et $x^-_{\Omega_1}(w) = x^-_{\Omega_2}(w)$.*
- *Pour tous $x, y \in \Omega_1$, on a $d_{\Omega_2}(x,y) \leqslant d_{\Omega_1}(x,y)$.*
- *Les boules métriques et métriques tangentes vérifient, pour tout $x \in \Omega_1$ et $r > 0$, $B_{\Omega_1}(x,r) \subset B_{\Omega_2}(x,r)$ et $B_{T_x\Omega_1}(r) \subset B_{T_x\Omega_2}(r)$, avec égalité si et seulement si $\Omega_1 = \Omega_2$.*
- *Pour tout borélien $\mathscr{A}$ de Ω_1, on a $\mathrm{Vol}_{\Omega_2}(\mathscr{A}) \leqslant \mathrm{Vol}_{\Omega_1}(\mathscr{A})$.*

2B. ***Fonctions de Busemann et horosphères.*** Nous supposons dans ce paragraphe que l'ouvert proprement convexe Ω de $\mathbb{P}^n$ est strictement convexe et à bord $\mathscr{C}^1$. Dans ce cadre, il est possible de définir les fonctions de Busemann et les horosphères de la même manière qu'en géométrie hyperbolique, et nous ne donnerons pas de détails.

Pour $\xi \in \partial\Omega$ et $x \in \Omega$, notons $c_{x,\xi} : [0, +\infty) \to \Omega$ la géodésique issue de x et d'extrémité ξ, soit $c_{x,\xi}(0) = x$ et $c_{x,\xi}(+\infty) = \xi$. La *fonction de Busemann* basée en $\xi \in \partial\Omega$ $b_\xi(\cdot,\cdot) : \Omega \times \Omega \to \mathbb{R}$ est définie par :

$$b_\xi(x,y) = \lim_{t\to+\infty} d_\Omega(y, c_{x,\xi}(t)) - t = \lim_{z\to\xi} d_\Omega(y,z) - d_\Omega(x,z), \quad x, y \in \Omega.$$

L'existence de ces limites est due aux hypothèses de régularité faites sur Ω. Les fonctions de Busemann sont de classe $\mathscr{C}^1$.

L'horosphère basée en $\xi \in \partial\Omega$ et passant par $x \in \Omega$ est l'ensemble

$$\mathscr{H}_\xi(x) = \{y \in \Omega \mid b_\xi(x,y) = 0\}.$$

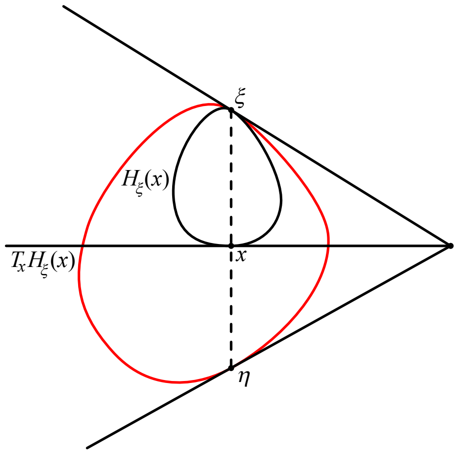


Figure 3. Une horosphère et son espace tangent.

L'horoboule basée en $\xi \in \partial\Omega$ et passant par $x \in \Omega$ est l'ensemble

$$H_\xi(x) = \{y \in \Omega \mid b_\xi(x, y) < 0\}.$$

L'horoboule basée en $\xi \in \partial\Omega$ et passant par $x \in \Omega$ est un ouvert strictement convexe de Ω, dont le bord est l'horosphère correspondante, qui est elle une sous-variété de classe $\mathscr{C}^1$ de Ω.

Dans une carte affine A dans laquelle Ω apparaît comme un ouvert convexe relativement compact, on peut, en identifiant $T\Omega$ avec $\Omega \times A$, construire géométriquement l'espace tangent à $\mathscr{H}_\xi(x)$ en x : c'est le sous-espace affine contenant x et l'intersection $T_\xi\partial\Omega \cap T_\eta\partial\Omega$ des espaces tangents à $\partial\Omega$ en ξ et $\eta = (x\xi) \cap \partial\Omega \smallsetminus \{\xi\}$.

On peut voir que l'horoboule et l'horosphère basées en $\xi \in \partial\Omega$ et passant par $x \in \Omega$ sont les limites des boules et des sphères métriques centrées au point $z \in \Omega$ et passant par x lorsque z tend vers ξ.

2C. ***Dualité.*** À l'ouvert proprement convexe Ω de $\mathbb{P}^n$ est associé l'ouvert proprement convexe dual Ω^* : on considère un des deux cônes $C \subset \mathbb{R}^{n+1}$ au-dessus de Ω, et son dual

$$C^* = \{f \in (\mathbb{R}^{n+1})^* \mid f(x) > 0 \text{ pour chaque } x \in C\}.$$

Le convexe Ω^* est par définition la trace de C^* dans $\mathbb{P}((\mathbb{R}^{n+1})^*)$.

Le bord de $\partial\Omega^*$ est facile à comprendre, car il s'identifie à l'ensemble des hyperplans tangents à Ω. En effet, un hyperplan tangent T_x à $\partial\Omega$ en x est la trace d'un hyperplan H_x de $\mathbb{R}^{n+1}$. L'ensemble des formes linéaires dont le noyau est H_x forme une droite de $(\mathbb{R}^{n+1})^*$, dont la trace x^* dans $\mathbb{P}((\mathbb{R}^{n+1})^*)$ est dans $\partial\Omega^*$. Il n'est pas dur de voir qu'on obtient ainsi tout le bord $\partial\Omega^*$.

Cette remarque permet de voir que le dual d'un ouvert strictement convexe a un bord de classe $\mathscr{C}^1$, et inversement. En particulier, lorsque Ω est strictement convexe et que son bord est de classe $\mathscr{C}^1$, ce qui est le cas que nous étudierons, on obtient une involution continue $x \mapsto x^*$ entre les bords de Ω et Ω^*.

Tout sous-groupe Γ de $\mathrm{SL}_{n+1}(\mathbb{R})$ agit par dualité sur $(\mathbb{R}^{n+1})^*$ et donc sur $\mathbb{P}((\mathbb{R}^{n+1})^*)$, via la formule suivante :

$$(\gamma \cdot f)(x) = f(\gamma^{-1}x), \quad \gamma \in \Gamma,\ x \in \mathbb{R}^{n+1}.$$

Le convexe dual Ω^* est préservé par un élément $\gamma \in \mathrm{SL}_{n+1}(\mathbb{R})$ si et seulement si Ω est préservé par γ. On obtient de cette façon une action de tout sous-groupe Γ de $\mathrm{Aut}(\Omega)$ sur le convexe dual Ω^*. Le sous-groupe discret de $\mathrm{Aut}(\Omega^*)$ ainsi obtenu sera noté Γ^*. Bien entendu, on a $(\Omega^*)^* = \Omega$ et $(\Gamma^*)^* = \Gamma$.

Dans tout ce qui suit, sauf mention explicite, Ω désignera un ouvert proprement convexe, strictement convexe et à bord $\mathscr{C}^1$.

2D. ***Isométries.*** Les isométries d'une géométrie (Ω, d_Ω) avec Ω strictement convexe à bord $\mathscr{C}^1$ ont été classifiées dans [Crampon et Marquis 2012]. Ce sont toutes des transformations projectives qui préservent Ω, et, quitte à considérer leur carré, on les verra donc comme des éléments du groupe linéaire $\mathrm{SL}_{n+1}(\mathbb{R})$, agissant sur $\mathbb{P}^n$. Outre les isométries elliptiques qui sont de torsion et qui ne nous intéresseront pas ici, on trouve les isométries hyperboliques et paraboliques.

— Une *isométrie hyperbolique* γ a exactement deux points fixes $x_\gamma^+, x_\gamma^- \in \partial\Omega$, l'un répulsif et l'autre attractif. Cela veut dire que la suite $(\gamma^n)_{n\in\mathbb{N}}$ converge uniformément sur les compacts de $\overline{\Omega} \smallsetminus \{x_\gamma^-\}$ vers x_γ^+, et la suite $(\gamma^{-n})_{n\in\mathbb{N}}$ converge uniformément sur les compacts de $\overline{\Omega} \smallsetminus \{x_\gamma^+\}$ vers x_γ^-. De plus, les valeurs propres $\lambda_0(\gamma)$ et $\lambda_n(\gamma)$ associées aux points fixes x_γ^+ et x_γ^- sont positives : c'est une conséquence du fait que le rayon spectral de γ est valeur propre de γ ([Benoist 2006, lemme 2.3], par exemple) ; elles sont de multiplicité 1 car, sinon, il y aurait un segment dans le bord de Ω. Finalement, γ agit par translation sur le segment ouvert $]x_\gamma^- x_\gamma^+[$ de Ω, translation de force

$$\tau(\gamma) = \ln \frac{\lambda_0(\gamma)}{\lambda_n(\gamma)}.$$

— Une *isométrie parabolique* γ a exactement un point fixe $p \in \partial\Omega$ et préserve toute horosphère basée en p. De plus, la famille $(\gamma^n)_{n\in\mathbb{Z}}$ converge uniformément sur les compacts de $\overline{\Omega} \smallsetminus \{p\}$ vers p.

On dira qu'un sous-groupe discret $\mathscr{P}$ de $\mathrm{Aut}(\Omega)$, sans torsion, est *parabolique* si tous ses éléments sont paraboliques. Un tel groupe est nilpotent et ses éléments fixent un même point $p \in \partial\Omega$. On dira que le groupe $\mathscr{P}$ est *de rang maximal* si son action sur $\partial\Omega \smallsetminus \{p\}$ est cocompacte.

Si un sous-groupe discret Γ de $\mathrm{Aut}(\Omega)$ est donné, on dira qu'un sous-groupe parabolique de Γ est *maximal* s'il n'est contenu dans aucun autre sous-groupe parabolique.

2E. *Ensemble limite.* Comme en géométrie hyperbolique, on peut définir l'ensemble limite et le domaine de discontinuité d'un sous-groupe discret de $\mathrm{Aut}(\Omega)$ de la façon suivante. On utilise ici de façon essentielle la stricte convexité de Ω.

Définition 2.2. Soit Γ un sous-groupe discret de $\mathrm{Aut}(\Omega)$ et $x \in \Omega$. L'ensemble limite Λ_Γ de Γ est le sous-ensemble de $\partial\Omega$ suivant :

$$\Lambda_\Gamma = \overline{\Gamma \cdot x} \smallsetminus \Gamma \cdot x.$$

Le domaine de discontinuité $\mathcal{O}_\Gamma$ de Γ est le complémentaire de l'ensemble limite de Γ dans $\partial\Omega$.

L'ensemble limite Λ_Γ, s'il n'est pas infini, est vide ou consiste en 1 ou 2 points. On dit que Γ est *non élémentaire* si Λ_Γ est infini. Dans ce dernier cas, l'ensemble limite Λ_Γ est le plus petit fermé Γ-invariant non vide de $\partial\Omega$. En particulier, Λ_Γ est l'adhérence des points fixes des éléments hyperboliques de Γ.

Définition 2.3. Soit Γ un sous-groupe de $\mathrm{SL}_{n+1}(\mathbb{R})$. On dira que Γ est *irréductible* lorsque les seuls sous-espaces vectoriels de $\mathbb{R}^{n+1}$ invariants par Γ sont $\{0\}$ et $\mathbb{R}^{n+1}$. On dira que Γ est *fortement irréductible* si tous ses sous-groupes d'indice fini sont irréductibles, autrement dit, si Γ ne préserve pas une union finie de sous-espaces vectoriels non triviaux.

Lemme 2.4. *Soit* Γ *un sous-groupe discret de* $\mathrm{Aut}(\Omega)$. *Les propositions suivantes sont équivalentes* :

(i) *L'ensemble limite* Λ_Γ *de* Γ *engendre* $\mathbb{P}^n$.

(ii) *Le groupe* Γ *est irréductible.*

(iii) *Le groupe* Γ *est fortement irréductible.*

Démonstration. L'implication (i) $\Rightarrow$ (ii) vient du fait que Λ_Γ est l'adhérence des points fixes des éléments hyperboliques de Γ. Pour les implications (ii) $\Rightarrow$ (i) et (iii) $\Rightarrow$ (i), il suffit de voir que l'espace engendré par Λ_Γ est invariant par Γ.

Montrons pour finir l'implication (i) $\Rightarrow$ (iii). Supposons donc que Λ_Γ engendre $\mathbb{P}^n$. Si G est un sous-groupe d'indice fini de Γ, alors, pour tout élément hyperbolique h de Γ, il existe un entier $n \geqslant 1$ tel que $h^n \in G$. Ainsi, $\Lambda_G = \Lambda_\Gamma$ et donc Λ_G engendre $\mathbb{P}^n$, ce qui équivaut à l'irréductibilité de G. $\square$

3. Le flot géodésique

3A. ***Généralités.*** Le flot géodésique est le principal objet d'étude de ce travail. Nous le définirons sur le fibré tangent homogène, ou en demi-droites, de Ω, qui est le fibré $\pi : H\Omega \to \Omega$, avec $H\Omega = (T\Omega \smallsetminus \{0\})/\mathbb{R}_+$: deux points (x, u) et (y, v) de $T\Omega \smallsetminus \{0\}$ sont identifiés si $x = y$ et $u = \lambda v$ pour un certain réel $\lambda > 0$.

L'image d'un point $w = (x, [\xi]) \in H\Omega$ par le flot géodésique $\varphi^t : H\Omega \to H\Omega$ est le point $\varphi^t(w) = (x_t, [\xi_t])$ obtenu en suivant la géodésique partant de x dans la direction $[\xi]$ pendant le temps t. Il est engendré par le champ de vecteurs X sur $H\Omega$, qui a la même régularité que le bord de Ω. Ainsi, φ^t est au moins de classe $\mathscr{C}^1$.

Nous ferons les calculs de façon intelligente en utilisant l'invariance projective. Une *carte adaptée à un point* $w \in H\Omega$ est une carte affine munie d'une métrique euclidienne telle que :

- la fermeture de Ω est incluse dans la carte ;
- l'intersection des plans tangents à $\partial\Omega$ en x^+ et x^- sont à l'infini de la carte ; autrement dit, ils y sont parallèles ;
- la droite (xx^+) et les plans tangents à $\partial\Omega$ en x^+ et x^- sont orthogonaux.

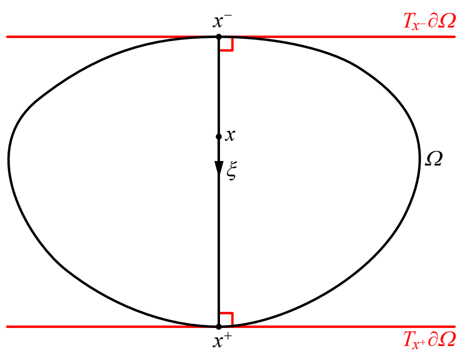


Figure 4. Une carte adaptée en w.

3B. ***Variétés stable et instable.*** On définit les variétés stable $W^s(w)$ et instable $W^u(w)$ de $w = (x, [\xi]) \in H\Omega$ par

$$W^s(w) = \{w' = (y, [yx^+]) \mid y \in \mathscr{H}_{x^+}(x)\},$$
$$W^u(w) = \{w' = (y, [x^-y]) \mid y \in \mathscr{H}_{x^-}(x)\}.$$

Il n'est pas difficile de voir que, comme Ω est strictement convexe à bord $\mathscr{C}^1$, on a

$$W^s(w) = \Big\{w' \in H\Omega \mid \lim_{t\to+\infty} d_\Omega(\pi\varphi^t(w), \pi\varphi^t(w')) = 0\Big\},$$
$$W^u(w) = \Big\{w' \in H\Omega \mid \lim_{t\to-\infty} d_\Omega(\pi\varphi^t(w), \pi\varphi^t(w')) = 0\Big\}.$$

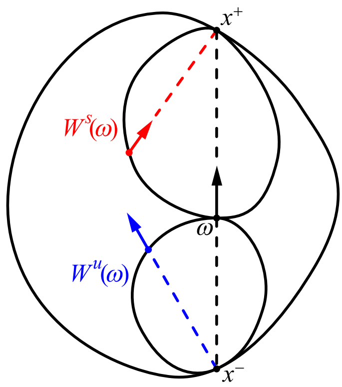


Figure 5. Variétés stable et instable.

Les sous-espaces stable E^s et instable E^u sont les espaces tangents aux variétés stable et instable. On a clairement que $E^s \cap E^u = \{0\}$ et donc la décomposition

$$TH\Omega = \mathbb{R} \cdot X \oplus E^s \oplus E^u,$$

qu'on appellera décomposition d'Anosov.

On peut définir une norme de Finsler $\|\cdot\|$ sur HM de la façon suivante : pour $Z = aX + Z^s + Z^u \in \mathbb{R} \cdot X \oplus E^s \oplus E^u$, on pose

$$\|Z\| = \big(a^2 + F(d\pi\, Z^s) + F(d\pi\, Z^u)\big)^{1/2}. \tag{3-1}$$

Cette métrique est précisément celle qui a été introduite dans [Crampon 2014], et qui apparaît naturellement via une décomposition en sous-fibrés horizontaux et verticaux. Nous n'aurons toutefois pas besoin ici de ces notions.

Remarquons qu'en particulier, si Z est un vecteur tangent stable ou instable, c'est-à-dire $Z = Z^s$ ou $Z = Z^u$, on a $\|Z^s\| = F(d\pi\, Z^s)$ ou $\|Z^u\| = F(d\pi\, Z^u)$.

Rappelons les deux lemmes suivants, dont les démonstrations permettront de fixer certaines notations.

Lemme 3.1. *Soient $w \in H\Omega$ et $Z \in T_w H\Omega$ un vecteur stable* (*respectivement instable*). *L'application $t \mapsto \|d\varphi^t(Z)\|$ est une bijection décroissante* (*respectivement croissante*) *de* $(0, +\infty)$ *dans* $(0, +\infty)$.

Démonstration. Choisissons une carte adaptée au point w. Notons $x = \pi w$ et $x_t = \pi\varphi^t(w)$, $t \in \mathbb{R}$. Supposons que Z est un vecteur stable, tangent à $H\Omega$ en w et notons $z = d\pi(Z)$, $z_t = d\pi\, d\varphi^t(Z)$, $t \in \mathbb{R}$. Rappelons que par définition de la norme, on a $\|d\varphi^t(Z)\| = F(z_t)$. Or,

$$F(z_t) = \frac{|z_t|}{2}\left(\frac{1}{|x_t y_t^+|} + \frac{1}{|x_t y_t^-|}\right),$$

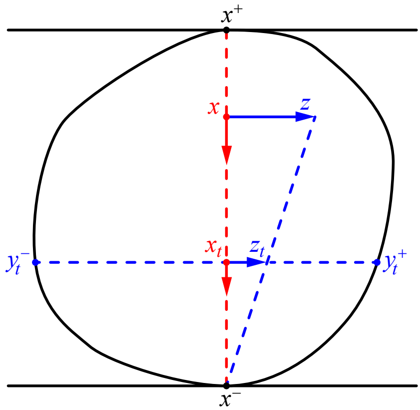


Figure 6. Contraction du flot.

où y_t^+ et y_t^- sont les points d'intersection de $x+\mathbb{R}\cdot z$ avec $\partial\Omega$.

Si on considère l'application

$$h_t : y \in \mathcal{H}_{x^+}(x) \mapsto y_t = \pi\varphi^t(y, [yx^+]) = (yx^+) \cap \mathcal{H}_{x^+}(x_t),$$

on voit que z_t est en fait donné par

$$z_t = dh_t(z) = \frac{|x_t x^+|}{|xx^+|} z.$$

On obtient ainsi

$$F(z_t) = \frac{|z|}{2|xx^+|}\left(\frac{|x_t x^+|}{|x_t y_t^+|} + \frac{|x_t x^+|}{|x_t y_t^-|}\right).$$

Que $t \mapsto \|d\varphi^t(Z)\|$ soit strictement décroissante est alors une conséquence directe de la stricte convexité de Ω. La régularité $\mathcal{C}^1$ au point extrémal $\varphi^{+\infty}(w)$ de l'orbite de w entraîne que $\|d\varphi^t(Z)\|$ tend vers 0 en $+\infty$. La stricte convexité à l'autre point extrémal $\varphi^{-\infty}(w)$ de l'orbite de w implique $\lim_{t\to-\infty} \|d\varphi^t(Z)\| = +\infty$.

Dans le cas où Z est un vecteur instable, on obtient, en gardant les mêmes notations :

$$F(z_t) = \frac{|z|}{2|xx^-|}\left(\frac{|x_t x^-|}{|x_t y_t^+|} + \frac{|x_t x^-|}{|x_t y_t^-|}\right).$$

On peut donc appliquer le même raisonnement. □

Remarquons ici le corollaire suivant, qui dit que la décroissance et la croissance du lemme précédent sont contrôlées.

Corollaire 3.2. *Pour tout vecteur $Z \in TH\Omega$, on a*

$$e^{-2|t|}\|Z\| \leqslant \|d\varphi^t(Z)\| \leqslant e^{2|t|}\|Z\|.$$

Démonstration. Soient $w \in H\Omega$ et $Z^s \in E^s(w)$ un vecteur stable. Posons $z := d\pi\, Z^s$. Soit $Z^u \in E^u(w)$ l'unique vecteur instable tel que $d\pi\, Z^u = z$. On a vu, dans la démonstration du lemme précédent, et avec les mêmes notations que, pour tout $t \in \mathbb{R}$,

$$\|d\varphi^t Z^s\| = \frac{|z|}{2|xx^+|}\left(\frac{|x_t x^+|}{|x_t y_t^+|} + \frac{|x_t x^+|}{|x_t y_t^-|}\right),$$

et

$$\|d\varphi^t Z^u\| = \frac{|z|}{2|xx^-|}\left(\frac{|x_t x^-|}{|x_t y_t^+|} + \frac{|x_t x^-|}{|x_t y_t^-|}\right).$$

Ainsi,

$$\frac{\|d\varphi^t Z^s\|}{\|d\varphi^t Z^u\|} = \frac{|x_t x^+|}{|x_t x^-|}\frac{|xx^+|}{|xx^-|}.$$

L'égalité $d_\Omega(x, x_t) = t$ implique directement que

$$\frac{|x_t x^+|}{|x_t x^-|}\frac{|xx^+|}{|xx^-|} = e^{-2t},$$

et donc

$$\frac{\|d\varphi^t Z^s\|}{\|d\varphi^t Z^u\|} = e^{-2t}. \tag{3-2}$$

Maintenant, le fait que la fonction $t \mapsto \|d\varphi^t Z^s\|$ soit décroissante implique que

$$\limsup_{t\to+\infty} \frac{1}{t} \ln \|d\varphi^t Z^s\| \leqslant 0;$$

de même, comme $t \mapsto \|d\varphi^t Z^u\|$ est croissante,

$$\liminf_{t\to+\infty} \frac{1}{t} \ln \|d\varphi^t Z^u\| \geqslant 0.$$

De l'égalité (3-2), on déduit donc que

$$\liminf_{t\to+\infty} \frac{1}{t} \ln \|d\varphi^t Z^s\| \geqslant -2,$$

et

$$\limsup_{t\to+\infty} \frac{1}{t} \ln \|d\varphi^t Z^u\| \leqslant 2.$$

Ainsi, on a, pour tout $t \geqslant 0$ et $Z^s \in E^s \setminus \{0\}$,

$$e^{-2|t|} \leqslant \frac{\|d\varphi^t (Z^s)\|}{\|Z^s\|} \leqslant 1.$$

De manière similaire, on obtient pour tout $t \geqslant 0$ et $Z^u \in E^u \setminus \{0\}$,

$$1 \leqslant \frac{\|d\varphi^t (Z^u)\|}{\|Z^u\|} \leqslant e^{2|t|}.$$

On obtient le résultat en décomposant un vecteur Z selon $E^s \oplus E^u \oplus \mathbb{R}\cdot X$. □

3C. ***Ensemble non errant.*** Nous voulons par la suite étudier des propriétés de récurrence du flot géodésique. Pour cela, il nous faut regarder l'ensemble des points qui ne partent pas pour toujours à l'infini.

Étant donnée la variété $M=\Omega/\Gamma$, on notera HM le fibré tangent homogène de M, quotient du fibré $H\Omega$ par le groupe Γ. *L'ensemble non errant* du flot géodésique de M est l'ensemble fermé $\mathtt{NW}$ des points $w\in HM$ dont l'orbite passe une infinité de fois dans tout voisinage ouvert de w, dans le passé *et* dans le futur. Cet ensemble est naturellement relié à l'ensemble limite : c'est la projection sur HM de l'ensemble

$$\{w=(x,\xi)\in H\Omega \mid x^+(w), x^-(w)\in\Lambda_\Gamma\}.$$

En particulier, la projection de $\mathtt{NW}$ sur M est incluse dans le cœur convexe de M ; cela nous permettra d'utiliser le fait 3 ci-dessous lorsque nous seront confrontés à des variétés géométriquement finies.

4. Variétés géométriquement finies

4A. ***Décomposition du cœur convexe.*** Les variétés géométriquement finies sont le contexte de cet article. Des définitions équivalentes de la finitude géométrique ont été données dans [Crampon et Marquis 2012]. Rappelons seulement le résultat suivant, essentiel dans le présent travail (voir [Crampon et Marquis 2012, section 8]) :

Fait 3. *Soit $M=\Omega/\Gamma$ une variété géométriquement finie. Le cœur convexe $C(M)$ de M est l'union d'un compact K et d'un nombre fini de pointes*

$$\mathcal{C}_i=(H_i\cap\overline{C(\Lambda_\Gamma)}^{\Omega})/\mathcal{P}_i,\quad 1\leqslant i\leqslant l,$$

où H_i est une horoboule basée en un point $p_i\in\partial\Omega$, et $\mathcal{P}_i$ est le sous-groupe parabolique maximal de Γ fixant p_i, soit $\mathcal{P}_i=\mathrm{Stab}_\Gamma(p_i)$.

4B. ***Groupes paraboliques de rang maximal.*** Les sous-groupes paraboliques maximaux qui apparaissent ici sont conjugués à des sous-groupes paraboliques d'isométries hyperboliques ; c'est un des résultats principaux de [Crampon et Marquis 2012]. Un cas particulier est celui où les sous-groupes paraboliques sont de rang maximal, c'est-à-dire que leur action sur $\partial\Omega\smallsetminus\{p\}$, où p est le point fixe du groupe en question, est cocompacte. Dans ce cas, on a le résultat suivant.

Théorème 4.1 [Crampon et Marquis 2012, section 7]. *Soit $\mathcal{P}$ un sous-groupe parabolique de* $\mathrm{Aut}(\Omega)$, *de rang maximal et de point fixe $p\in\partial\Omega$. Il existe deux ellipsoïdes $\mathcal{P}$-invariants $\mathcal{E}^{\mathrm{int}}$ et $\mathcal{E}^{\mathrm{ext}}$ tels que :*

- $\partial\mathcal{E}^{\mathrm{int}}\cap\partial\mathcal{E}^{\mathrm{ext}}=\partial\mathcal{E}^{\mathrm{int}}\cap\partial\Omega=\partial\mathcal{E}^{\mathrm{ext}}\cap\partial\Omega=\{p\}$;
- $\mathcal{E}^{\mathrm{int}}\subset\Omega\subset\mathcal{E}^{\mathrm{ext}}$;
- $\mathcal{E}^{\mathrm{int}}$ *est une horoboule de* $\mathcal{E}^{\mathrm{ext}}$.

Anticipons un peu. Pour pouvoir dire quelque chose du flot géodésique d'une variété géométriquement finie, il va nous falloir maîtriser ce qui se passe dans les parties qui partent à l'infini, les pointes $\mathscr{C}_i$. De façon générale, cela ne sera pas possible, comme le montre le contre-exemple que nous donnons dans la partie 8.1. Toutefois, lorsque les sous-groupes paraboliques sont de rang maximal, les deux ellipsoïdes du théorème précédent nous donnent deux métriques hyperboliques dans chaque pointe qui contrôlent la métrique de Hilbert.

Corollaire 4.2. *Soit $M = \Omega/\Gamma$ une variété géométriquement finie. Supposons que les sous-groupes paraboliques maximaux de Γ soient tous de rang maximal. Alors, pour toute constante $C > 1$, on peut trouver une décomposition*

$$C(M) = K \sqcup \bigsqcup_{1\leqslant i\leqslant l} \mathscr{C}_i$$

du cœur convexe de M, et, sur chaque $\mathscr{C}_i$, deux métriques hyperboliques $\mathtt{h}_i$ et $\mathtt{h}'_i$, telles que :

- *F, $\mathtt{h}_i$ et $\mathtt{h}'_i$ ont les mêmes géodésiques, à paramétrisation près ;*
- *$C^{-1}\mathtt{h}_i \leqslant \mathtt{h}'_i \leqslant F \leqslant \mathtt{h}_i \leqslant C\mathtt{h}'_i$.*

Démonstration. Soit $p \in \Lambda_\Gamma$ un point parabolique, $\mathscr{P}$ le sous-groupe parabolique maximal de Γ fixant p. Soient $\mathscr{E}^{\text{int}}$ et $\mathscr{E}^{\text{ext}}$ deux ellipsoïdes donnés par le théorème 4.1. Ils définissent deux métriques hyperboliques $\mathtt{h}$ et $\mathtt{h}'$ telles que $\mathtt{h}' \leqslant F \leqslant \mathtt{h}$.

L'adhérence de Zariski $\mathscr{U}$ de $\mathscr{P}$ dans $\mathrm{SL}_{n+1}(\mathbb{R})$ est isomorphe à $\mathbb{R}^{n-1}$. Les ensembles $\partial\mathscr{E}^{\text{int}} \smallsetminus \{p\}$ et $\partial\mathscr{E}^{\text{ext}} \smallsetminus \{p\}$ sont des orbites de $\mathscr{U}$. Par exemple, dans une

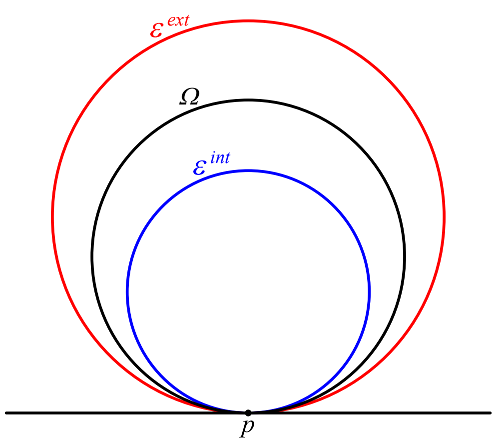


Figure 7. Les ellipsoïdes tangents.

certaine base de $\mathbb{R}^{n+1}$, $\partial\mathscr{E}^{\text{ext}}$ est d'équation

$$x_n x_{n+1} = x_1^2 + \cdots + x_{n-1}^2,$$

et alors $\partial\mathscr{E}^{\text{int}}$ est d'équation

$$-x_n^2 + 2a x_n x_{n+1} = 2a(x_1^2 + \cdots + x_{n-1}^2)$$

pour un certain $a > 0$.

Soient $C > 1$ et x un point de $\mathscr{E}^{\text{int}}$. Pour $t \geqslant 0$, on note $x(t)$ le point du segment $[xp]$ tel que $d_\Omega(x, x(t)) = t$. On considère la fonction

$$f_x : t \mapsto \max_{v \in T_{x(t)}\Omega \setminus \{0\}} \frac{\mathtt{h}(x(t), v)}{\mathtt{h}'(x(t), v)}.$$

Lemme 4.3. *Il existe $T = T(C)$ tel que, pour tout $t \geqslant T$, $1 \leqslant f_x(t) \leqslant C$.*

Démonstration. Pour simplifier les calculs, on peut supposer que les équations de $\partial\mathscr{E}^{\text{ext}}$ et $\partial\mathscr{E}^{\text{int}}$ sont les précédentes et qu'on travaille dans la carte affine $\{x_{n+1} = 1\}$ avec la structure euclidienne induite par celle de $\mathbb{R}^{n+1}$. On peut aussi supposer que le point x a pour coordonnées $(0, \dots, 0, x_n)$ pour un certain $x_n > 0$; le point p est ici l'origine et $T_p\partial\mathscr{E}^{\text{ext}} = \{t_n = 0\}$. La définition de la métrique de Finsler (formule (2-1)) donne immédiatement que, pour $v \in \mathbb{R} \cdot xp \smallsetminus \{0\}$ ou $v \in \{t_n = 0\} \smallsetminus \{0\}$,

$$\lim_{t \to +\infty} \frac{\mathtt{h}(x(t), v)}{\mathtt{h}'(x(t), v)} = 1.$$

Il existe donc $T(C)$ tel que pour $t \geqslant T(C)$ on ait $\dfrac{\mathtt{h}(x(t), v)}{\mathtt{h}'(x(t), v)} \leqslant C$ pour $v \in \mathbb{R} \cdot xp \smallsetminus \{0\}$ ou $v \in \{t_n = 0\} \smallsetminus \{0\}$.

De plus, pour tout $t \geqslant 0$, les sous-espaces $\mathbb{R} \cdot xp$ et $\{t_n = 0\}$ de $T_{x(t)}\Omega$ sont orthogonaux, tant pour $\mathtt{h}(x(t), \cdot)$ que pour $\mathtt{h}'(x(t), \cdot)$. En décomposant le vecteur $v \in T_{x(t)}\Omega \smallsetminus \{0\}$ selon $\mathbb{R} \cdot xp$ et $\{t_n = 0\}$, on voit que

$$\frac{\mathtt{h}(x(t), v)}{\mathtt{h}'(x(t), v)} \leqslant C$$

dès que $t \geqslant T(C)$. □

Comme les métriques $\mathtt{h}$ et $\mathtt{h}'$ sont invariantes par $\mathscr{U}$, on a aussi $1 \leqslant f_{u \cdot x}(t) \leqslant C$, pour tout $u \in \mathscr{U}$ et tout $t \geqslant T$. Il existe donc une horoboule H^{int} de $\mathscr{E}^{\text{int}}$ basée en p telle que, sur H^{int}, $C^{-1}\mathtt{h} \leqslant \mathtt{h}' \leqslant F \leqslant \mathtt{h} \leqslant C\mathtt{h}'$. Comme $\mathscr{P}$ agit de façon cocompacte sur $\partial\Omega \smallsetminus \{p\}$, H^{int} contient une horoboule H de Ω basée en p telle que, sur H :

- F, $\mathtt{h}$ et $\mathtt{h}'$ ont les mêmes géodésiques (les droites), à paramétrisation près ;
- $C^{-1}\mathtt{h} \leqslant \mathtt{h}' \leqslant F \leqslant \mathtt{h} \leqslant C\mathtt{h}'$.

On peut maintenant conclure. Considérons un ensemble de représentants $\{p_i\}_{i=1}^{l}$ des points paraboliques de Λ_Γ. On note $\mathscr{P}_i$ le sous-groupe parabolique maximal de Γ

qui fixe le point p_i. On peut faire la construction précédente pour chaque point p_i. On obtient ainsi une horoboule H_i de Ω basée en p_i et, sur H_i, deux métriques hyperboliques $\mathcal{P}_i$-invariantes $\mathtt{h}_i$ et $\mathtt{h}_i'$, vérifiant les propriétés précédentes. Elles induisent par projection deux métriques hyperboliques sur la pointe $\mathcal{C}_i = H_i/\mathcal{P}_i$, qui satisfont aux conditions de l'énoncé.

On peut supposer que $H_i \subset C(\Lambda_\Gamma)$, $1 \leqslant i \leqslant l$. D'après le fait 3, il est aussi possible de prendre les H_i telles que l'union

$$\bigcup_{\substack{\gamma\in\Gamma \\ 1\leqslant i\leqslant l}} \gamma \cdot H_i$$

soit disjointe. La pointe $\mathcal{C}_i = H_i/\mathcal{P}_i$ s'identifie ainsi à une partie de $C(M)$. L'ensemble $K = C(M) \setminus \bigsqcup_i \mathcal{C}_i$ est nécessairement compact et cela donne la décomposition annoncée. □

4C. *Pointes asymptotiquement hyperboliques.* Nous allons suivre le chemin indiqué par les groupes paraboliques de rang maximal en nous restreignant à ces variétés géométriquement finies dont nous savons contrôler la métrique de Hilbert dans les pointes :

Définition 4.4. On dira qu'une variété $M = \Omega/\Gamma$ géométriquement finie est *à pointes asymptotiquement hyperboliques* s'il existe une décomposition du cœur convexe $C(M) = K \sqcup \bigsqcup_{1\leqslant i\leqslant l} \mathcal{C}_i$ telle que, sur chaque $\mathcal{C}_i$, il existe une métrique hyperbolique $\mathtt{h}_i$, ayant les mêmes géodésiques (non paramétrées) que F et qui soit équivalente à F, c'est-à-dire que, pour un certain $C_i \geqslant 1$,

$$C_i^{-1}\mathtt{h}_i \leqslant F \leqslant C_i \mathtt{h}_i.$$

Si la condition d'être géométriquement fini porte sur le groupe Γ, celle d'hyperbolicité asymptotique des pointes porte sur Ω. Le lemme 4.6 qui suit donne une condition sur le bord de Ω, inspirée par les observations précédentes, pour que les pointes soient asymptotiquement hyperboliques. Pour l'énoncer, nous faisons quelques rappels :

Définition 4.5. Soient $\varepsilon > 0$, $\beta > 1$. On dit qu'une fonction $f : U \subset \mathbb{R}^n \to \mathbb{R}$, définie et de classe $\mathcal{C}^1$ sur un ouvert U, est :

– de classe $\mathcal{C}^{1+\varepsilon}$ si, pour une certaine constante $C > 0$,

$$|f(x) - f(y) - d_x f(y-x)| \leqslant C|x-y|^{1+\varepsilon}, \quad x, y \in U;$$

– β-convexe si, pour une certaine constante $C > 0$,

$$|f(x) - f(y) - d_x f(y-x)| \geqslant C|x-y|^{\beta}, \quad x, y \in U.$$

On dit que f est de classe $\mathscr{C}^{1+\varepsilon}$ ou β-convexe en un point $x \in U$ si on a les inégalités précédentes pour tout y dans un voisinage de x.

Lemme 4.6. *Soit $M = \Omega/\Gamma$ une variété géométriquement finie. Si le bord $\partial\Omega$ est de classe $\mathscr{C}^{1+1}$ et 2-convexe en chaque point parabolique de Λ_Γ, alors la variété M est à pointes asymptotiquement hyperboliques.*

Démonstration. L'hypothèse de régularité de $\partial\Omega$ aux points paraboliques nous permet, pour chaque point parabolique p de Λ_Γ de stabilisateur le groupe $\mathscr{P} = \mathrm{Stab}_\Gamma(p)$, de trouver deux ellipsoïdes $\mathscr{E}^{\mathrm{int}}$ et $\mathscr{E}^{\mathrm{ext}}$ tels que :

– $\partial\mathscr{E}^{\mathrm{int}} \cap \partial\mathscr{E}^{\mathrm{ext}} = \partial\mathscr{E}^{\mathrm{int}} \cap \partial\Omega = \partial\mathscr{E}^{\mathrm{ext}} \cap \partial\Omega = \{p\}$;

– $\mathscr{E}^{\mathrm{int}} \subset \Omega \subset \mathscr{E}^{\mathrm{ext}}$.

Comme les sous-groupes paraboliques de Γ sont conjugués à des sous-groupes de $\mathrm{SO}_{n,1}(\mathbb{R})$, on peut choisir ces deux ellipsoïdes de telle façon qu'ils soient $\mathscr{P}$-invariants. Soit H une horoboule de Ω basée en p, d'adhérence incluse dans $\mathscr{E}^{\mathrm{int}}$. Les ellipsoïdes $\mathscr{E}^{\mathrm{int}}$ et $\mathscr{E}^{\mathrm{ext}}$ définissent sur H deux métriques hyperboliques $\mathscr{P}$-invariantes $\mathtt{h}$ et $\mathtt{h}'$, qui ont les mêmes géodésiques que F, et telles que $\mathtt{h}' \leqslant F \leqslant \mathtt{h}$.

On peut maintenant voir qu'il existe une constante $C \geqslant 1$ telle que $1/C \leqslant \mathtt{h}/\mathtt{h}' \leqslant C$ sur $H \cap C(\Lambda_\Gamma)$. Comme H est $\mathscr{P}$-invariante, il suffit de le montrer sur un domaine fondamental D de P sur $H \cap C(\Lambda_\Gamma)$. Soit $\mathscr{H}$ l'horosphère au bord de H. Comme M est géométriquement finie, l'intersection $\mathscr{D} = \bar{D}^\Omega \cap \mathscr{H}$ est compacte. Il suffit donc de voir que pour tout $x \in \mathscr{D}$, la fonction

$$f_x : t \in [0, +\infty) \mapsto \max_{v \in T_{x(t)}\Omega\setminus\{0\}} \frac{\mathtt{h}(x(t), v)}{\mathtt{h}'(x(t), v)}$$

est bornée dans $(0, +\infty)$, où $x(t)$ est le point du segment $[xp]$ tel que $d_\Omega(x, x(t)) = t$. C'est un petit calcul.

Comme les métriques $\mathtt{h}$ et $\mathtt{h}'$ sont $\mathscr{P}$-invariantes, elles donnent deux métriques hyperboliques sur le quotient $H \cap C(\Lambda_\Gamma)/\mathscr{P}$, qui vérifient les conditions voulues. On conclut alors comme dans la démonstration du corollaire 4.2. □

On doit pouvoir obtenir la même conclusion que celle du corollaire 4.2 sous l'hypothèse que le bord $\partial\Omega$ est deux fois différentiable en chaque point parabolique de Λ_Γ. Toutefois, cette observation plus précise ne nous sera pas utile dans ce texte : nous n'utiliserons que le corollaire 4.2, dans la section 9.

De façon générale, les sous-groupes paraboliques étant conjugués à des sous-groupes de $\mathrm{SO}_{n,1}(\mathbb{R})$, on peut se poser la :

Question 4.7. Soit Γ un sous-groupe discret de $\mathrm{Aut}(\Omega)$ dont l'action est géométriquement finie. Existe t-il un ouvert convexe Ω' sur lequel Γ agit de façon géométriquement finie à pointes asymptotiquement hyperboliques ?

Beaucoup de résultats dynamiques ne dépendent pas du convexe que l'on considère et le résultat précédent permettrait de se ramener à une situation géométrique et dynamique agréable, qui sera notre propos dans cette article. Par exemple, le spectre des longueurs ne dépend que du groupe Γ, les longueurs des géodésiques fermées étant données par les valeurs propres des éléments hyperboliques du groupe.

4D. ***Cas particuliers.*** Parmi les variétés géométriquement finies, on peut distinguer celles qui ont volume fini, et celles dont le cœur convexe est compact.

Dans [Crampon et Marquis 2012], on a pu voir que les quotients Ω/Γ qui ont volume fini sont précisément les variétés géométriquement finies dont l'ensemble limite est le bord $\partial\Omega$ tout entier. En particulier, si Ω/Γ est une variété de volume fini, les sous-groupes paraboliques maximaux de Γ sont de rang maximal. Remarquons que dans tous les cas, une pointe $\mathscr{C}$ d'une variété géométriquement finie a un volume fini (voir [ibid., partie 8]).

Les variétés *convexes cocompactes* sont celles dont le cœur convexe est compact ; autrement dit, le quotient Ω/Γ est géométriquement finie et le groupe Γ ne contient pas d'éléments paraboliques.

5. Hyperbolicité uniforme du flot géodésique

Rappelons d'abord quelques définitions.

Définition 5.1. Soit W une variété munie d'une métrique de Finsler $\|\cdot\|$ continue. Soient $\varphi^t : W \to W$ un flot de classe $\mathscr{C}^1$ engendré par le champ de vecteurs X sur W, et V une partie φ^t-invariante de W. On dit que le flot φ^t est *uniformément hyperbolique* sur V s'il existe une décomposition φ^t-invariante

$$TW = \mathbb{R}\cdot X \oplus E^s \oplus E^u$$

du fibré tangent à W en tout point de V, et des constantes $a, C > 0$ pour lesquelles

$$\|d\varphi^t Z^s\| \leqslant Ce^{-at}\|Z^s\| \quad \text{et} \quad \|d\varphi^{-t} Z^u\| \leqslant Ce^{-at}\|Z^u\|, \qquad Z^s \in E^s,\ Z^u \in E^u,\ t \geqslant 0.$$

Dans le cas où W est une variété compacte et $V = W$, on parle plus souvent de *flot d'Anosov*. Les distributions E^s et E^u s'appellent les distributions stables et instables du flot. Le but de cette partie est de montrer une telle propriété d'hyperbolicité pour notre flot géodésique, restreint à son ensemble non errant.

Dans le cas où la variété M est compacte, l'ensemble non errant est HM tout entier, et Yves Benoist a déjà prouvé que le flot géodésique était d'Anosov. Si la variété M est convexe-cocompacte, c'est-à-dire que son cœur convexe est compact, l'ensemble non errant est lui-même compact, et une démonstration similaire fonctionnerait pour prouver l'uniforme hyperbolicité sur l'ensemble non errant.

Nous étendons ce résultat au cas d'une variété géométriquement finie à pointes asymptotiquement hyperboliques :

Théorème 5.2. *Soit $M = \Omega/\Gamma$ une variété géométriquement finie à pointes asymptotiquement hyperboliques. Le flot géodésique est uniformément hyperbolique sur l'ensemble non errant, de décomposition*

$$THM = \mathbb{R}\cdot X \oplus E^s \oplus E^u.$$

Nous montrerons le théorème en plusieurs temps. Fixons une fois pour toutes une décomposition du cœur convexe $C(M)$ de M en une partie compacte K et une union finie de pointes $\mathscr{C}_i$, $1 \leqslant i \leqslant l$, chacun d'entre eux portant une métrique hyperbolique $\mathtt{h}_i$ telle que :

– F et $\mathtt{h}_i$ ont les mêmes géodésiques, à paramétrisation près ;

– $C^{-1}\mathtt{h}_i \leqslant F \leqslant C\mathtt{h}_i$, pour une certaine constante $C \geqslant 1$.

Pour la partie compacte, on se servira du lemme suivant :

Lemme 5.3. *Soient V une partie compacte de HM et $T > 0$. Il existe un réel $0 < b(V,T) < 1$ tel que, si $\varphi^t(w) \in V$ pour $0 \leqslant t \leqslant T$, alors, pour tout $Z \in E^s(w)$,*

$$\|d\varphi^T(Z)\| \leqslant b(V,T)\|Z\|.$$

Démonstration. C'est une simple conséquence du lemme 3.1. Notons V_T l'ensemble des $w \in V$ tels que $\varphi^t(w) \in V$ pour $0 \leqslant t \leqslant T$ et $E_1 = \{Z \in E^s(w) \mid w \in V_T, \|Z\| = 1\}$. Les ensembles V_T et E_1 sont compacts. La fonction $Z \in E_1 \mapsto \|d\varphi^T(Z)\|$ est continue et atteint donc son maximum pour un certain vecteur Z_M. Le lemme 3.1 nous dit que $\|d\varphi^T(Z_M)\| < 1$, d'où le résultat. □

Pour les pointes, c'est un peu plus délicat. Choisissons une des pointes $\mathscr{C}_i$, et oublions les indices : on note $\mathscr{C}$ la pointe et $\mathtt{h}$ la métrique hyperbolique sur $\mathscr{C}$.

Lemme 5.4. *Pour tout $0 < a < 1$, on peut trouver un temps $T_a = T_a(C) > 0$ tel que, pour tout $w \in H\mathscr{C}$ tel que $\varphi^t(w) \in H\mathscr{C}$ pour $0 \leqslant t \leqslant T_a$ et $Z \in E^s(w)$, on a*

$$\|d\varphi^{T_a} Z\| \leqslant a\|Z\|.$$

Avant de montrer ce dernier lemme, voyons d'abord comment en déduire le théorème :

Démonstration du théorème 5.2. Rappelons la décomposition du cœur convexe en

$$C(M) = K \sqcup \bigsqcup_i \mathscr{C}_i.$$

Choisissons un réel $0 < a < 1$ et un temps $T_a > 0$ comme dans le lemme 5.4, et posons

$$K_a = \bigcup_{-T_a \leqslant t \leqslant T_a} \varphi^t\big(HM|_K\big).$$

Pour tout point w de l'ensemble non errant $\mathtt{NW}$, le morceau d'orbite $\{\varphi^t(w)\}_{0\leqslant t\leqslant T_a}$ est inclus soit dans K_a, soit dans un des $H\mathscr{C}_i$. Les deux lemmes précédents impliquent alors que, pour tout $Z\in E^s|_{\mathtt{NW}}$, on a

$$\|d\varphi^{T_a}Z\|\leqslant A\|Z\|,$$

avec $A=\max(a,b(K_a,T_a))<1$. Ainsi, pour tout $t\geqslant 0$, en posant $N=[t/T_a]$, on a

$$\|d\varphi^t(Z)\|\leqslant A^N\|d\varphi^{t-NT_a}(Z)\|\leqslant \frac{\|d\varphi^{t-NT_a}(Z)\|}{e^{\frac{t-NT_a}{T_a}\ln A}}e^{\frac{t}{T_a}\ln A}\leqslant A^{-1}e^{-\frac{\ln A^{-1}}{T_a}t}\|Z\|.$$

Cela prouve la décroissance uniformément hyperbolique sur la distribution stable. On fait de même pour la distribution instable en considérant φ^{-t}. □

Le reste de cette partie est consacrée à la démonstration du lemme 5.4. Bien entendu, l'idée est de comparer les flots géodésiques des métriques F et $\mathtt{h}$ sur $H\mathscr{C}$, qui satisfont $C^{-1}\mathtt{h}\leqslant F\leqslant C\mathtt{h}$ pour une certaine constante $C>1$. Comme F et $\mathtt{h}$ ont les mêmes géodésiques à paramétrisation près, le flot φ^t est en effet une renormalisation du flot $\varphi^t_{\mathtt{h}}$ de la métrique $\mathtt{h}$: on a

$$\varphi^t(w)=\varphi_{\mathtt{h}}^{\alpha(w,t)}(w)$$

pour un certain $\alpha(w,t)\in\mathbb{R}$. Bien sûr, cette expression ne fait sens que si $\varphi^s(w)$ est dans $H\mathscr{C}$ pour tout $0\leqslant s\leqslant t$. La fonction α est donc définie sur l'ensemble

$$W=\{(w,t)\mid \varphi^s(w)\in H\mathscr{C},\ 0\leqslant s\leqslant t\}\subset H\mathscr{C}\times\mathbb{R}.$$

Soit g la fonction définie sur $H\mathscr{C}$ par $F=g^{-1}\mathtt{h}$. C'est une fonction de classe $\mathscr{C}^1$, qui prend ses valeurs dans l'intervalle $[C^{-1},C]$. Si $X_{\mathtt{h}}$ est le générateur du flot géodésique de $\mathtt{h}$, alors on a $X=gX_{\mathtt{h}}$. On retrouve la fonction α en intégrant g :

$$\alpha(w,t)=\int_0^t g(\varphi^s(w))\,ds;$$

la fonction α est donc de classe $\mathscr{C}^1$ et satisfait

$$C^{-1}t\leqslant\alpha(w,t)\leqslant Ct,\quad t\geqslant 0.$$

L'espace tangent à $H\mathscr{C}$ se décompose de deux façons, selon que l'on considère le flot de F ou de $\mathtt{h}$:

$$TH\mathscr{C}=E^s\oplus E^u\oplus\mathbb{R}\cdot X=E^s_{\mathtt{h}}\oplus E^u_{\mathtt{h}}\oplus\mathbb{R}\cdot X.$$

Sur $H\mathscr{C}$, on dispose des métriques $\|\cdot\|$ et $\|\cdot\|_{\mathtt{h}}$ associées respectivement à F et $\mathtt{h}$ et définies par la formule (3-1) via les décompositions précédentes. La métrique $\|\cdot\|_{\mathtt{h}}$ est bien entendu une métrique riemannienne, qui n'est rien d'autre que la métrique

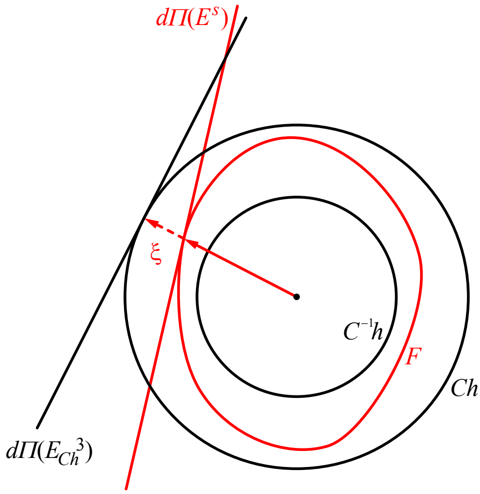


Figure 8. La boule unité de F est coincée entre celles de $C^{-1}\mathtt{h}$ et $C\mathtt{h}$.

de Sasaki, et pour laquelle la décomposition $TH\mathscr{C} = E^s_{\mathtt{h}} \oplus E^u_{\mathtt{h}} \oplus \mathbb{R}\cdot X$ est orthogonale. Rappelons que, si $Z^s_{\mathtt{h}} \in E^s_{\mathtt{h}}$, alors

$$\|d\varphi^t_{\mathtt{h}}(Z^s_{\mathtt{h}})\| = e^{-t}\|Z^s_{\mathtt{h}}\|, \quad t \in \mathbb{R},$$

sous réserve, bien sûr, que Z soit tangent à $H\mathscr{C}$ en un point w tel que $\varphi^s_{\mathtt{h}}(w) \in H\mathscr{C}$ pour $0 \leqslant s \leqslant t$.

Le lemme essentiel est le suivant :

Lemme 5.5. *La distribution E^s est, sur $H\mathscr{C}$, incluse dans un cône d'angle θ pour $\|\cdot\|_{\mathtt{h}}$ autour de la distribution $E^s_{\mathtt{h}}$, avec $0 \leqslant \theta = \theta(C) < \pi/2$.*

Démonstration. Il revient au même de montrer qu'il existe θ tel que, pour tout $w = (x, [\xi])$ dans $H\mathscr{C}$, la projection $d\pi(E^s(w))$ est dans un cône d'angle θ pour $\mathtt{h}$ autour de $d\pi(E^s_{\mathtt{h}}(w))$.

De la démonstration de la proposition 3.6 de [Crampon 2009], on peut tirer que la projection $d\pi(E^s(w))$ coïncide avec l'espace tangent en $\xi \in T_x\mathscr{C}$ à la boule unité tangente de la norme $F(x,\cdot)$; où ξ est le vecteur de norme 1 de $[\xi]$. La même chose est bien sûr valable pour $d\pi(E^s_{\mathtt{h}}(w))$ et la métrique $\mathtt{h}$.

Or, la boule unité tangente de F est coincée entre les boules de rayon C^{-1} et C de $\mathtt{h}$. Cet encadrement et le fait précédent impliquent l'existence de θ. □

Tout vecteur $Z^s \in E^s$ se décompose en

$$Z^s = Z^s_{\mathtt{h}} + Z^u_{\mathtt{h}} + Z^X_{\mathtt{h}} \in E^s_{\mathtt{h}} \oplus E^u_{\mathtt{h}} \oplus \mathbb{R}\cdot X_{\mathtt{h}}.$$

On déduit du lemme précédent que l'angle entre Z^s et sa projection $Z^s_{\mathtt{h}}$ sur $E^s_{\mathtt{h}}$ est toujours inférieur à θ ; de même en ce qui concerne l'angle, pour $\mathtt{h}$, entre $d\pi\, Z^s$ et $d\pi\, Z^s_{\mathtt{h}}$. D'où :

Corollaire 5.6. *Pour tout vecteur $Z^s \in E^s$, on a*

$$\frac{1}{C}\|Z^s_{\mathtt{h}}\|_{\mathtt{h}} \leqslant \|Z^s\| \leqslant \frac{C}{\cos\theta}\|Z^s_{\mathtt{h}}\|_{\mathtt{h}}.$$

Démonstration. On a

$$\|Z^s\| = F(d\pi\, Z^s) \leqslant C\mathtt{h}(d\pi\, Z^s) \leqslant \frac{C}{\cos\theta}\mathtt{h}(d\pi\, Z^s_{\mathtt{h}}) = \frac{C}{\cos\theta}\|Z^s_{\mathtt{h}}\|_{\mathtt{h}},$$

et

$$\|Z^s_{\mathtt{h}}\|_{\mathtt{h}} = \mathtt{h}(d\pi\, Z^s_{\mathtt{h}}) \leqslant \mathtt{h}(d\pi\, Z^s) \leqslant C F(d\pi\, Z^s) = C\|Z^s\|. \qquad \square$$

Démonstration du lemme 5.4. Fixons $(w, t) \in W$ et un vecteur stable $Z^s \in E^s(w)$, qui se décompose en

$$Z^s = Z^s_{\mathtt{h}} + Z^u_{\mathtt{h}} + Z^X_{\mathtt{h}} \in E^s_{\mathtt{h}} \oplus E^u_{\mathtt{h}} \oplus \mathbb{R}\cdot X_{\mathtt{h}}.$$

On a donc

$$d\varphi^t(Z^s) = d\varphi^t(Z^s_{\mathtt{h}}) + d\varphi^t(Z^u_{\mathtt{h}}) + d\varphi^t(Z^X_{\mathtt{h}}). \tag{5-1}$$

D'autre part, considérons les fonctions φ et $\varphi_{\mathtt{h}}$ définies par

$$\varphi(w, t) = \varphi^t(w), \quad \varphi_{\mathtt{h}}(w, t) = \varphi^t_{\mathtt{h}}(w);$$

les fonctions φ et $\varphi_{\mathtt{h}}$ sont définies, respectivement, sur W et sur l'ensemble

$$\{(w, t) \mid \varphi^s_{\mathtt{h}}(w) \in H\mathscr{C},\ 0 \leqslant s \leqslant t\}.$$

On a ainsi $\varphi(w, t) = \varphi_{\mathtt{h}}(w, \alpha(w, t))$, d'où

$$d\varphi^t(Z^s) = \frac{\partial\varphi}{\partial w}(w, t)(Z^s) = \frac{\partial\varphi_{\mathtt{h}}}{\partial t}\frac{\partial\alpha}{\partial w}(w, t)(Z^s) + d\varphi_{\mathtt{h}}^{\alpha(w,t)}(Z^s). \tag{5-2}$$

L'application $\dfrac{\partial\varphi_{\mathtt{h}}}{\partial t}$ a son image dans $\mathbb{R}\cdot X$; le premier terme

$$\frac{\partial\varphi_{\mathtt{h}}}{\partial t}\frac{\partial\alpha}{\partial w}(w, t)(Z^s)$$

de la dernière expression est donc un vecteur de $\mathbb{R}\cdot X$. Comme $d\varphi^t_{\mathtt{h}}$ préserve la décomposition $TH\mathscr{C} = E^s_{\mathtt{h}} \oplus E^u_{\mathtt{h}} \oplus \mathbb{R}\cdot X_{\mathtt{h}}$, on déduit de (5-1) et (5-2) que

$$(d\varphi^t(Z^s))^s_{\mathtt{h}} = d\varphi_{\mathtt{h}}^{\alpha(w,t)}(Z^s_{\mathtt{h}}).$$

On a alors, d'après le corollaire 5.6,

$$\|d\varphi^t Z^s\| \leqslant \frac{C}{\cos\theta}\|d\varphi_{\mathtt{h}}^{\alpha(w,t)}(Z^s_{\mathtt{h}})\|_{\mathtt{h}} = \frac{C}{\cos\theta}e^{-\alpha(w,t)}\|Z^s_{\mathtt{h}}\|_{\mathtt{h}} \leqslant \frac{C^2}{\cos\theta}e^{-t}\|Z^s\|.$$

Mais on peut écrire

$$\frac{C^2}{\cos\theta}e^{-t} = e^{-t\left(1-\frac{1}{t}\ln\frac{C^2}{\cos\theta}\right)}.$$

Aussi, en prenant $T_a = \ln \dfrac{C^2}{a\cos\theta}$, on obtient $\|d\varphi^{T_a} Z^s\| \leqslant a\|Z^s\|$. □

Remarque 5.7. Dans le lemme 5.5, on pourrait voir que plus C est proche 1, plus $\theta(C)$ peut être pris proche de 0. Or, dans la pointe $\mathscr{C}$ choisi pour la constante C, on a

$$\|d\varphi^t Z^s\| \leqslant \frac{C^2}{\cos\theta(C)}e^{-t}\|Z^s\|,$$

pour tout vecteur stable Z^s tangent en un point $w \in H\mathscr{C}$ tel que $\varphi^s(w) \in H\mathscr{C}$, $0 \leqslant s \leqslant t$. En particulier, sous l'hypothèse de rang maximal des sous-groupes paraboliques maximaux, on peut, d'après le corollaire 4.2, choisir la pointe de telle façon que la constante $C^2/\cos\theta(C)$ soit aussi proche de 1 qu'on le souhaite.

6. Propriétés de récurrence

6A. ***Transitivité et mélange topologique.*** Le but de cette partie est d'étudier les propriétés de récurrence du flot géodésique d'une variété quotient $M = \Omega/\Gamma$ *quelconque*. Rappelons qu'un flot φ^t sur un espace topologique X est dit :

– *topologiquement transitif* s'il existe une orbite dense ou, de façon équivalente, si pour tous ouverts U et V de X, il existe $T \in \mathbb{R}$ tel que $\varphi^T(U) \cap V \neq \varnothing$;

– *topologiquement mélangeant* si pour tous ouverts U et V de X, il existe $T \in \mathbb{R}$ tel que, pour tout $t \geqslant T$, $\varphi^t(U) \cap V \neq \varnothing$.

Le résultat principal est le suivant.

Proposition 6.1. *Soit $M = \Omega/\Gamma$, avec Γ non élémentaire. Le flot géodésique est topologiquement mélangeant sur son ensemble non errant.*

Une orbite périodique du flot géodésique sur HM se projette sur une géodésique fermée de M, parcourue dans un sens ou dans l'autre. Or, les géodésiques fermées orientées sont en bijection avec les classes de conjugaison d'éléments hyperboliques de Γ. La géodésique fermée orientée définie par un tel $\gamma \in \Gamma$ est précisément la projection sur M de l'axe orienté $(x_\gamma^- x_\gamma^+)$.

À un élément $\gamma \in \Gamma$ correspond ainsi une unique orbite périodique. Bien sûr, l'orbite périodique associée à γ^{-1} se projette sur la même géodésique fermée que celle associée à γ, mais l'orientation est inversée. Le lemme suivant est immédiat :

Lemme 6.2. *Soient g et h deux éléments hyperboliques de* Aut(Ω) *tel que l'intersection des groupes engendrés par g et h est triviale. Posons $k_n = g^n h^n$, alors*

$$\lim_{n\to+\infty} x_{k_n}^+ = x_g^+ \quad et \quad \lim_{n\to+\infty} x_{k_n}^- = x_h^-.$$

Corollaire 6.3. *Soit* Γ *un sous-groupe de* $\mathrm{Aut}(\Omega)$. *L'ensemble* $\{(x_g^+, x_g^-) \mid g \in \Gamma\}$ *est dense dans* $\Lambda_\Gamma \times \Lambda_\Gamma$. *En particulier, les orbites périodiques de* HM *sont denses dans* $\mathtt{NW}$.

Démonstration. Il suffit de se rappeler que l'action de Γ sur Λ_Γ est minimale, puisque Λ_Γ est le plus petit fermé Γ-invariant de $\partial\Omega$; en particulier, l'ensemble des x_g^+ pour $g \in \Gamma$ est dense dans Λ_Γ. Fixons une métrique riemannienne quelconque sur $\partial\Omega$. Si on prend un couple (x, y) dans $\Lambda_\Gamma \times \Lambda_\Gamma$, il existe, pour tout $\varepsilon > 0$, des éléments g et h hyperboliques de Γ tels que x_g^+ et x_h^- sont ε-proches de, respectivement, x et y. Le lemme précédent affirme alors que pour n assez grand, si $k_n = g^n h^n \in \Gamma$, les points $x_{k_n}^+$ et $x_{k_n}^-$ sont 2ε-proches de, respectivement, x et y. □

Démonstration de la proposition 6.1. Prenons U et V deux ouverts de $\mathtt{NW}$. Les orbites périodiques étant denses dans $\mathtt{NW}$, il existe une orbite périodique passant dans U, et une autre, distincte de la première, passant dans V. Considérons des relevés (xy) et $(x'y')$ dans $H\Omega$ de ces orbites. Ce sont les axes d'éléments γ et γ' distincts de Γ. Le projeté de l'orbite (xy') sur HM est alors une orbite qui rencontre à la fois U et V. Ainsi, il existe $t \geqslant 0$ tel que $\varphi^t(U) \cap V \neq \varnothing$ et le flot est topologiquement transitif.

Comme le flot est topologiquement transitif, le mélange topologique est équivalent au fait que le spectre des longueurs des orbites périodiques engendre un sous-groupe dense de $\mathbb{R}$ (exercice 18.3.4 du livre [Katok et Hasselblatt 1995]). Or, la longueur de l'orbite périodique définie par l'élément hyperbolique $\gamma \in \Gamma$ est exactement $\frac{1}{2}\ln(\lambda_0(\gamma)/\lambda_n(\gamma))$, où $\lambda_0(\gamma)$ et $\lambda_n(\gamma)$ sont le module de, respectivement, sa plus grande et plus petite valeur propre. Si le groupe engendré par les longueurs n'était pas dense dans $\mathbb{R}$, il existerait $l > 0$ tel que pour tout $\gamma \in \Gamma$, il existe $k_\gamma \in \mathbb{N}$ tel que

$$\tag{6-1} \tfrac{1}{2}\ln\frac{\lambda_0(\gamma)}{\lambda_n(\gamma)} = k_\gamma l.$$

Quitte à se restreindre au sous-espace projectif engendré par Λ_Γ, on peut supposer que l'action de Γ est irréductible. D'après le lemme 2.4, elle est même fortement irréductible. La proposition 6.5 nous dit que l'adhérence de Zariski G de Γ est alors un groupe semi-simple. Le théorème 6.4 qui suit implique que la relation (6-1) ne peut être vérifiée pour tout $\gamma \in \Gamma$. □

Le théorème permettant de conclure la démonstration est dû à Yves Benoist [2000b]. Rappelons-en ici un énoncé dans notre contexte particulier.

Soit G un sous-groupe de Lie semi-simple de $\mathrm{SL}_{n+1}(\mathbb{R})$. À tout élément g de G, on associe le vecteur $\ln(g) = (\ln\lambda_0(g), \dots, \ln\lambda_n(g)) \in \mathbb{R}^{n+1}$, où

$$\lambda_0(g) \geqslant \lambda_1(g) \geqslant \cdots \geqslant \lambda_n(g)$$

désignent les modules des valeurs propres de g. Pour un sous-groupe Γ de G, on note $\ln\Gamma$ l'ensemble des $\ln\gamma$ pour $\gamma\in\Gamma$. Le résultat est le suivant :

Théorème 6.4 [ibid.]. *Soient G un sous-groupe de Lie semi-simple de* $\mathrm{SL}_{n+1}(\mathbb{R})$ *et* Γ *un sous-groupe de G. Si* Γ *est Zariski-dense dans G, alors le sous-groupe engendré par* $\ln\Gamma$ *est dense dans le sous-espace vectoriel de* $\mathbb{R}^{n+1}$ *engendré par* $\ln G$.

L'autre résultat que l'on a utilisé était la proposition suivante, due à Yves Benoist. Pour faciliter la lecture de ce texte, nous en donnons une démonstration sous l'hypothèse Ω strictement convexe, qui n'est toutefois pas nécessaire.

Proposition 6.5 [Benoist 2000a, remarque suivant le corollaire 3.2]. *Soit* Γ *un sous-groupe irréductible de* $\mathrm{SL}_{n+1}(\mathbb{R})$ *qui préserve un ouvert* $\Omega\subset\mathbb{P}^n$ *proprement convexe et strictement convexe. La composante connexe G de l'adhérence de Zariski de* Γ *est un groupe de Lie semi-simple.*

Démonstration. D'après le lemme 2.4, Γ est fortement irréductible. Quitte à considérer un sous-groupe d'indice fini, on peut donc supposer que Γ est Zariski-connexe et ainsi que G est d'indice fini dans l'adhérence de Zariski de Γ. Le groupe G est Zariski-fermé à indice fini près et agit de façon irréductible sur $\mathbb{R}^{n+1}$. Il est donc réductif. [1]

En effet, soit N un sous-groupe unipotent et distingué de G. Considérons le sous-espace vectoriel E de $\mathbb{R}^{n+1}$ des points fixes de N. Comme N est distingué dans G, l'espace vectoriel E est préservé par G. Comme l'action de G sur $\mathbb{R}^{n+1}$ est irréductible, on en déduit que E est trivial. Or, le théorème de Kolchin affirme que tout groupe unipotent fixe un vecteur non trivial. On a donc nécessairement $E=\mathbb{R}^{n+1}$ et $N=\{1\}$. Ainsi, G est bien réductif.

Par conséquent, pour montrer que G est semi-simple, il suffit de montrer que son centre est trivial. Soit a un élément du centre de G. Comme Γ est fortement irréductible, d'après le lemme 2.4, il possède au moins un élément hyperbolique γ. Notons ρ^+ son rayon spectral. L'espace propre $\ker(\gamma-\rho^+)$ est donc de dimension 1 (voir le paragraphe 2D) et préservé par a. Par suite, a possède une valeur propre λ réelle. L'espace propre $\ker(a-\lambda)$ est préservé par Γ et doit donc être trivial. On en déduit que a est une homothétie. Comme $a\in\Gamma\subset\mathrm{SL}_{n+1}(\mathbb{R})$, on conclut que $a=1$ ou $a=-1$. □

6B. ***Lemme de fermeture et conséquences.*** Nous rappelons ici un résultat classique pour les flots d'Anosov et une de ses conséquences, qui nous servira dans la partie suivante. Yves Coudene et Barbara Schapira [2010] on donné une démonstration de ces deux résultats dans le cadre de la courbure négative (ou nulle), qui s'adapte sans changement aucun. On pourra aussi consulter [Eberlein 1996] en ce qui concerne le lemme de fermeture :

1. C'est-à-dire que tout sous-groupe unipotent distingué est réduit à l'identité.

Lemme 6.6 [Coudene et Schapira 2010, Appendice ; Eberlein 1996, Proposition 4.5.15]. *Soient $M = \Omega/\Gamma$ une variété quotient, avec Γ non élémentaire, et K une partie compacte de HM. Fixons $\varepsilon > 0$.*

Il existe $\delta > 0$ et $T > 0$ tels que, si $w \in K$ satisfait $d(w, \varphi^t(w)) < \delta$ pour un certain $t > T$, alors il existe un point $w' \in HM$ tel que :

- *w' est périodique de période $t' \in (t - \varepsilon, t + \varepsilon)$;*
- *pour tout $0 < s < \min\{t, t'\}$, $d(\varphi^s(w), \varphi^s(w')) < \varepsilon$.*

En version courte, cela signifie que si un point revient assez proche de sa position d'origine après un temps t, alors il existe une orbite périodique qui suit son orbite pendant le temps t, et ce en restant aussi proche qu'on le veut.

Prenons une variété quotient $M = \Omega/\Gamma$, avec Γ non élémentaire. Notons $\mathcal{M}$ l'ensemble des mesures de probabilité sur $\mathtt{NW}$ invariantes par le flot, qu'on munit de la convergence étroite des mesures : une suite (η_n) de $\mathcal{M}$ converge vers η si, pour toute fonction f continue sur $\mathtt{NW}$, $\int f\, d\eta_n$ converge vers $\int f\, d\eta$. L'ensemble $\mathcal{M}$ est un convexe dont les points extrémaux sont les mesures ergodiques. Parmi les mesures ergodiques, on peut distinguer le sous-ensemble $\mathcal{M}_{\mathrm{Per}}$ constitué des mesures de Lebesgue portées par les orbites périodiques. Le lemme de fermeture entraîne :

Proposition 6.7 [Coudene et Schapira 2010, Corollaire 2.3]. *Soit $M = \Omega/\Gamma$ une variété quotient, avec Γ non élémentaire. L'enveloppe convexe de $\mathcal{M}_{\mathrm{Per}}$ est dense dans $\mathcal{M}$.*

7. Régularité du bord

Définition 7.1. On dira qu'un point $w = (x, [\xi]) \in H\Omega$, ou le rayon géodésique $\{\pi\varphi^t(w)\}_{t\geqslant 0}$ qu'il définit, est *hyperbolique* si, pour tout vecteur stable $Z^s \in E^s(w)$, on a
$$\limsup_{t\to+\infty} \frac{1}{t} \ln \|d\varphi^t Z^s\| < 0.$$

Si w est un point hyperbolique, il existe alors $\chi > 0$ tel que, pour tout vecteur stable $Z^s \in E^s(w)$,
$$\limsup_{t\to+\infty} \frac{1}{t} \ln \|d\varphi^t Z^s\| \leqslant -\chi.$$

Pour tout $w \in H\Omega$, on notera $\chi(w) \geqslant 0$ le plus grand des réels χ qui vérifie l'inégalité précédente pour tout vecteur $Z^s \in E^s(w)$; autrement dit,
$$\chi(w) = \inf_{Z^s\in E^s(w)} \liminf_{t\to+\infty} -\frac{1}{t} \ln \|d\varphi^t Z^s\|.$$

Le point w est donc hyperbolique si et seulement si $\chi(w) > 0$.

En fait, on peut facilement caractériser les points de $H\Omega$ qui sont hyperboliques, et même déterminer $\chi(w)$, selon la régularité du bord $\partial\Omega$ au point extrémal x^+ du rayon défini par w.

Proposition 7.2. *Un point $w \in H\Omega$ est hyperbolique, de coefficient $\chi(w) > 0$, si et seulement si $\partial\Omega$ est de classe $\mathscr{C}^{1+\varepsilon}$ en x^+, pour tout $0 < \varepsilon < (2/\chi(w) - 1)^{-1}$.*

Démonstration. Reprenons les notations du lemme 3.1 : on a choisi une carte adaptée au point $w \in H\Omega$, Z est un vecteur stable tangent à $H\Omega$ au point w, et $z = d\pi(Z)$, $z_t = (d\pi\, d\varphi^t Z)$. Tout se passe dans un plan et on peut donc supposer qu'on est en dimension 2. On a vu durant la démonstration du lemme 3.1 que

$$\|d\varphi^t(Z)\| = F(z_t) = \frac{|z|}{2|xx^+|}\left(\frac{|x_t x^+|}{|x_t y_t^+|} + \frac{|x_t x^+|}{|x_t y_t^-|}\right).$$

Ainsi, le rayon géodésique défini par w est hyperbolique si et seulement si

$$\liminf_{t\to+\infty} \frac{1}{t} \ln\left(\frac{|x_t x^+|}{|x_t y_t^+|} + \frac{|x_t x^+|}{|x_t y_t^-|}\right) < 0;$$

autrement dit si et seulement s'il existe $\chi, C > 0$ tels que pour $t \geqslant 0$,

$$\frac{|x_t x^+|}{|x_t y_t^+|} + \frac{|x_t x^+|}{|x_t y_t^-|} < Ce^{-\chi t},$$

soit

$$\frac{|x_t x^+|}{|x_t y_t^+|} < Ce^{-\chi t} \quad \text{et} \quad \frac{|x_t x^+|}{|x_t y_t^-|} < Ce^{-\chi t}.$$

Mais de l'égalité $d_\Omega(x, x_t) = t = \frac{1}{2}\ln[x^+ : x^- : x : x_t]$, on tire

$$|x_t x^+| = e^{-2t}\frac{|x_t x^-|}{|xx^-|}|xx^+|,$$

et donc il existe une constante $C_0 \geqslant 1$ pour laquelle

$$C_0^{-1}e^{-2t} \leqslant |x_t x^+| \leqslant C_0 e^{-2t}.$$

Ainsi, le rayon géodésique défini par w est hyperbolique si et seulement s'il existe $\chi, D > 0$ tels que pour $t \geqslant 0$,

$$\frac{|x_t x^+|}{|x_t y_t^+|} < D|x_t x^+|^{\chi/2} \quad \text{et} \quad \frac{|x_t x^+|}{|x_t y_t^-|} < D|x_t x^+|^{\chi/2},$$

soit

$$|x_t x^+|^{1-\chi/2} < D|x_t y_t^+| \quad \text{et} \quad |x_t x^+|^{1-\chi/2} < D|x_t y_t^-|.$$

Appelons $f : T_{x^+}\partial\Omega \mapsto \mathbb{R}$ le graphe de $\partial\Omega$, de telle façon que

$$|x_t x^+| = f(|x_t y_t^+|) = f(-|x_t y_t^-|).$$

La condition précédente est alors équivalente à : pour tout s (assez petit),

$$f(s) < Ds^{1/(1-\chi/2)}.$$

Or, dans notre carte adaptée, on a $d_{x^+} f = 0$. L'inégalité précédente veut donc dire que la fonction f est $\mathscr{C}^{1+\varepsilon}$ en x^+ avec $\varepsilon = 1/(2/\chi - 1)$. □

Cette proposition et le théorème 5.2 ont la conséquence suivante :

Corollaire 7.3. *Supposons que Ω admette une action géométriquement finie à pointes asymptotiquement hyperboliques d'un sous-groupe discret Γ de* $\mathrm{Aut}(\Omega)$. *Alors il existe $\varepsilon > 0$ tel que le bord $\partial\Omega$ de Ω est de classe $\mathscr{C}^{1+\varepsilon}$ en tout point de Λ_Γ.*

En particulier, si Ω admet une action de covolume fini, alors le bord $\partial\Omega$ de Ω est de classe $\mathscr{C}^{1+\varepsilon}$ pour un certain $\varepsilon > 0$.

7A. *Régularité optimale du bord.* Soit Γ un sous-groupe discret de $\mathrm{Aut}(\Omega)$. Notons

(7-1) $$\varepsilon(\Lambda_\Gamma) = \sup\{\varepsilon \in [0, 1] \mid \text{le bord } \partial\Omega \text{ est } \mathscr{C}^{1+\varepsilon} \text{ en tout point de } \Lambda_\Gamma\}.$$

Pour tout élément hyperbolique $\gamma \in \Gamma$, notons

$$\varepsilon(\gamma) = \sup\{\varepsilon \in [0, 1] \mid \text{le bord } \partial\Omega \text{ est } \mathscr{C}^{1+\varepsilon} \text{ en } x_\gamma^+\},$$

et $\varepsilon(\Gamma) = \inf\{\varepsilon(\gamma) \mid \gamma \in \Gamma \text{ hyperbolique}\}$. Le bord $\partial\Omega$ est ainsi de classe $\mathscr{C}^{1+\varepsilon(\Gamma)}$ en tout point fixe d'un élément hyperbolique de Γ. Rappelons-nous que l'ensemble des points fixes d'éléments hyperboliques de γ est dense dans Λ_Γ, dès que Γ n'est pas élémentaire ; on pourrait donc s'attendre au théorème qui suit, qui est toutefois faux en général.

Théorème 7.4. *Supposons que Ω admette une action géométriquement finie à pointes asymptotiquement hyperboliques Γ d'un sous-groupe discret Γ de* $\mathrm{Aut}(\Omega)$. *Alors*

$$\varepsilon(\Lambda_\Gamma) = \varepsilon(\Gamma).$$

Dans le cas où le groupe Γ est cocompact, l'ensemble limite Λ_Γ est tout le bord $\partial\Omega$ et ce résultat a déjà été prouvé par Olivier Guichard. Nous allons donner ici une toute autre démonstration. Toutefois, remarquons que la méthode de Guichard est plus précise car elle permet de prouver que le supremum dans la définition (7-1) est en fait un maximum, c'est-à-dire que $\partial\Omega$ est exactement $\mathscr{C}^{1+\varepsilon(\Gamma)}$ (et pas plus, sauf si Ω est un ellipsoïde).

Remarquons que l'hypothèse faite sur les pointes est essentielle. En effet, il est possible de faire en sorte que la régularité en un point parabolique soit aussi mauvaise que l'on veut car lorsque le groupe parabolique n'est pas de rang maximal, il n'impose la régularité du bord au point fixe que dans certaines directions. On consultera la partie 8.1 à ce propos.

Notre démonstration repose sur une approche dynamique et en particulier sur l'extension d'un résultat de Ursula Hamenstädt concernant le "meilleur" coefficient de contraction du flot géodésique [Hamenstädt 1994]. Au vu de la proposition 7.2, cette approche est en fait totalement naturelle. Dans toute la suite, nous supposerons que Γ est sans torsion, ce qui ne change rien d'après le lemme de Selberg, et nous étudierons plus en détail le flot géodésique de $M = \Omega/\Gamma$.

Rappelons que l'ensemble non errant $\mathtt{NW}$ du flot géodésique sur HM est la projection sur HM de l'ensemble

$$\{w = (x, \xi) \in H\Omega \mid x^+, x^- \in \Lambda_\Gamma\}.$$

Notons $\chi(\mathtt{NW})$ la meilleure constante d'hyperbolicité du flot géodésique sur l'ensemble non errant ; autrement dit, $\chi(\mathtt{NW})$ est le supremum des réels $\chi \geqslant 0$ tels qu'il existe $C > 0$ tel que, pour tout $w \in \mathtt{NW}$ et tous $Z^s \in E^s(w)$, $Z^u \in E^u(w)$, on a

$$\|d\varphi^t(Z^s)\| \leqslant Ce^{-\chi t}\|Z^s\|, \quad \|d\varphi^{-t}(Z^u)\| \leqslant Ce^{-\chi t}\|Z^u\|, \quad t \geqslant 0.$$

On a en fait

$$\chi(\mathtt{NW}) = \inf_{w \in \mathtt{NW}} \chi(w).$$

On a déjà vu que les points périodiques formaient un ensemble Per dense dans $\mathtt{NW}$. Notons

$$\chi(\mathrm{Per}) = \inf\{\chi(w) \mid w \in \mathtt{NW} \text{ périodique}\}.$$

Le résultat, inspiré de celui d'Hamenstädt, est le suivant :

Théorème 7.5. *Soit $M = \Omega/\Gamma$ une variété géométriquement finie à pointes asymptotiquement hyperboliques. On a*

$$\chi(\mathtt{NW}) = \chi(\mathrm{Per}).$$

Voyons tout de suite comment ce dernier résultat implique directement le théorème 7.4. Rappelons que l'ensemble des orbites périodiques est en bijection avec les classes de conjugaison d'éléments hyperboliques de Γ : si $\gamma \in \Gamma$ est hyperbolique, la projection de la géodésique orientée $(x_\gamma^- x_\gamma^+)$ sur HM est une orbite périodique du flot géodésique, qu'on note encore γ. Associé à cette orbite γ, on dispose du plus petit coefficient de contraction :

$$\chi(\gamma) := \inf_{Z^s \in E^s(w)} \liminf_{t \to +\infty} -\frac{1}{t} \ln \|d\varphi^t Z^s\| = \chi(w),$$

où w est un point quelconque de l'orbite γ. L'égalité principale est la suivante, qui découle de la proposition 7.2 :

$$\varepsilon(\Gamma) = \frac{1}{2/\chi(\mathrm{Per}) - 1}. \tag{7-2}$$

Il s'avère qu'on peut exprimer $\chi(\gamma)$ en fonction des valeurs propres de γ, comme l'affirme le lemme suivant, montré dans [Benoist 2004] ou dans [Crampon 2009] :

Lemme 7.6. *Soit* $\gamma \in \mathrm{Aut}(\Omega)$ *un élément hyperbolique. Notons*

$$\lambda_0(\gamma) \geqslant \lambda_1(\gamma) \geqslant \cdots \geqslant \lambda_n(\gamma) > 0$$

les modules de ses valeurs propres, comptées avec multiplicité. Alors

$$\chi(\gamma) = 2\left(1 - \frac{\ln\lambda_0(\gamma) - \ln\lambda_{n-1}(\gamma)}{\ln\lambda_0(\gamma) - \ln\lambda_n(\gamma)}\right).$$

Corollaire 7.7. *Pour toute variété quotient* $M = \Omega/\Gamma$, *on a* $\chi(\mathrm{Per}) \leqslant 1$. *De plus, si* $\chi(\mathrm{Per}) = 1$, *alors* Γ *n'est pas Zariski-dense dans* $\mathrm{SL}_{n+1}(\mathbb{R})$.

Démonstration. En gardant les notations du lemme, on voit que si $\gamma \in \mathrm{Aut}(\Omega)$ est hyperbolique, alors

$$\chi(\gamma^{-1}) = 2\left(1 - \frac{\ln\lambda_1(\gamma) - \ln\lambda_n(\gamma)}{\ln\lambda_0(\gamma) - \ln\lambda_n(\gamma)}\right).$$

Ainsi,

$$\chi(\gamma) + \chi(\gamma^{-1}) = 2\left(1 - \frac{\ln\lambda_1(\gamma) - \ln\lambda_{n-1}(\gamma)}{\ln\lambda_0(\gamma) - \ln\lambda_n(\gamma)}\right) \leqslant 2.$$

Cela implique que soit $\chi(\gamma) \leqslant 1$ soit $\chi(\gamma^{-1}) \leqslant 1$, et donc que

$$\chi(\mathrm{Per}) = \inf\{\chi(\gamma) \mid \gamma \in \Gamma \text{ hyperbolique}\} \leqslant 1.$$

Maintenant, supposons que $\chi(\mathrm{Per}) = 1$, c'est-à-dire que $\chi(\gamma) \geqslant 1$ pour tout élément $\gamma \in \Gamma$ hyperbolique. De l'inégalité $\chi(\gamma) + \chi(\gamma^{-1}) \leqslant 2$, on déduit que $\chi(\gamma) = 1$ pour tout élément $\gamma \in \Gamma$ hyperbolique. D'après le lemme précédent, cela veut dire que, pour tout élément $\gamma \in \Gamma$ hyperbolique,

$$2\left(1 - \frac{\ln\lambda_0(\gamma) - \ln\lambda_1(\gamma)}{\ln\lambda_0(\gamma) - \ln\lambda_n(\gamma)}\right) = 1,$$

soit

$$\ln\lambda_0(\gamma) + \ln\lambda_n(\gamma) - 2\ln\lambda_1(\gamma) = 0.$$

En particulier, l'ensemble $\ln(\Gamma)$ n'engendre pas tout l'espace

$$\ln(\mathrm{SL}_{n+1}(\mathbb{R})) = \left\{x = (x_0, \cdots, x_n) \in \mathbb{R}^{n+1} \;\middle|\; \sum_{i=0}^{n} x_i = 0\right\}.$$

D'après le théorème 6.4, cela implique que Γ n'est pas Zariski-dense dans $\mathrm{SL}_{n+1}(\mathbb{R})$. □

Démonstration du théorème 7.4. D'après l'égalité (7-2), on a

$$\varepsilon(\gamma) = \min\Big(1, \frac{1}{2/\chi(\gamma) - 1}\Big),$$

et donc, via le corollaire 7.7, que

$$\varepsilon(\Gamma) = \frac{1}{2/\chi(\mathrm{Per}) - 1}.$$

De même, on a

$$\varepsilon(\Lambda_\Gamma) = \frac{1}{2/\chi(\mathtt{NW}) - 1}.$$

Le théorème 7.5 implique

$$\varepsilon(\Lambda_\Gamma) = \varepsilon(\Gamma). \qquad \square$$

7B. *Conséquences.* La régularité optimale du bord aux points de Λ_Γ se lit donc sur les valeurs propres des éléments de Γ :

Théorème 7.8. *Supposons que Ω admette une action géométriquement finie à pointes asymptotiquement hyperboliques d'un sous-groupe discret Γ de* Aut(Ω). *Alors*

$$\varepsilon(\Lambda_\Gamma) = \inf\left\{ \frac{\ln\lambda_{n-1}(\gamma) - \ln\lambda_n(\gamma)}{\ln\lambda_0(\gamma) - \ln\lambda_{n-1}(\gamma)} \;\Big|\; \gamma \in \Gamma \text{ hyperbolique} \right\}.$$

Démonstration. C'est un simple calcul à partir de l'égalité

$$\varepsilon(\Lambda_\Gamma) = \frac{1}{2/\chi(\mathrm{Per}) - 1}. \qquad \square$$

Les corollaires ci-dessous sont sûrement plus parlants, et sont complémentaires du résultat de rigidité de Benoist concernant la régularité des convexes divisibles [Benoist 2004, Proposition 6.1]. Il repose sur le résultat suivant, cas particulier d'un des théorèmes principaux de [Crampon et Marquis 2012] :

Théorème 7.9. *Soit Γ un sous-groupe discret de* Aut(Ω). *Si Γ agit de façon géométriquement finie sur Ω et contient un élément parabolique, alors son adhérence de Zariski est soit* $\mathrm{SL}_{n+1}(\mathbb{R})$ *tout entier, soit conjuguée à* $\mathrm{SO}_{n,1}(\mathbb{R})$.

Corollaire 7.10. *Supposons que Ω admette une action géométriquement finie d'un sous-groupe discret Γ de* Aut(Ω) *qui contienne un élément parabolique. Si le bord $\partial\Omega$ est de classe $\mathscr{C}^{1+\varepsilon}$ pour tout $0 < \varepsilon < 1$ en tout point de Λ_Γ, alors Γ est conjugué à un sous-groupe de* $\mathrm{SO}_{n,1}(\mathbb{R})$.

Démonstration. Les hypothèses impliquent que $\varepsilon(\Gamma) = 1$, soit $\chi(\mathrm{Per}) = 1$. Le corollaire 7.7 implique que Γ n'est pas Zariski-dense dans $\mathrm{SL}_{n+1}(\mathbb{R})$. D'après le théorème précédent, Γ est Zariski-dense dans un conjugué de $\mathrm{SO}_{n,1}(\mathbb{R})$. $\square$

Un cas particulier est le suivant, où l'on obtient une *vraie* rigidité :

Corollaire 7.11. *Si Ω admet un quotient de volume fini non compact, alors le bord $\partial\Omega$ est de classe $\mathscr{C}^{1+\varepsilon}$ pour tout $0 < \varepsilon < 1$ si et seulement si Ω est un ellipsoïde.*

7C. ***Démonstration du théorème 7.5.*** Cette démonstration, largement inspirée de [Hamenstädt 1994], est assez technique. En voici d'abord le schéma, qui repousse la partie la plus délicate, incluse dans le lemme 7.12, à la suite.

Démonstration. Notons pour simplifier $\chi = \chi(\mathrm{Per})$. Cela veut dire que pour tout $\varepsilon > 0$ et point $w \in \mathtt{NW}$ périodique, il existe une constante $C_\varepsilon(w)$ telle que, pour tout vecteur tangent stable Z en w, on ait

$$\|d\varphi^t Z\| \leqslant C_\varepsilon(w) e^{-(\chi-\varepsilon)t} \|Z\|.$$

Considérons l'ensemble

$$A_{T,\varepsilon} = \left\{ w \in \mathtt{NW} \;\middle|\; \frac{\|d\varphi^T Z\|}{\|Z\|} \leqslant e^{-(\chi-\varepsilon)T} \text{ pour chaque } Z \in E^s(w) \right\}.$$

Un point w n'est pas dans $A_{T,\varepsilon}$ s'il existe un vecteur stable en w qui est contracté par φ^T avec un exposant inférieur à $\chi - \varepsilon$. En particulier, à ε fixé, pour tout point w périodique, il existe un temps $T(w)$ tel que pour tout $t \geqslant T(w)$, $w \in A_{t,\varepsilon}$. On va montrer qu'en fait l'orbite de tout point $w \in \mathtt{NW}$ sous φ^T passe “la plupart du temps” dans $A_{T,\varepsilon}$ si T est assez grand. Pour cela on pose

$$N_{n,T,\varepsilon} = \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{1}_{A_{T,\varepsilon}}(\varphi^{iT} w).$$

$N_{n,T,\varepsilon}$ compte la proportion des n premiers points de l'orbite de w sous φ^T qui sont dans $A_{T,\varepsilon}$. Le résultat principal est le suivant :

Lemme 7.12. *Pour tous $\varepsilon, \delta > 0$, il existe $T = T(\varepsilon, \delta)$ et $N = N(T)$ tels que, pour $n \geqslant N$ et $w \in \mathtt{NW}$,*

$$N_{n,T,\varepsilon}(w) \geqslant 1 - \delta.$$

Le théorème découle aisément de ce lemme. En effet, pour tous $w \in \mathtt{NW}$ et $Z \in E^s(w)$, on a

$$\begin{aligned}
\frac{\|d\varphi^{nT} Z\|}{\|Z\|} &\leqslant \prod_{i=0}^{n-1} \frac{\|d\varphi^{(i+1)T} Z\|}{\|d\varphi^{iT} Z\|} \\
&\leqslant \prod_{\substack{0\leqslant i\leqslant n-1 \\ \varphi^{iT} w \notin A_{t,\varepsilon}}} \frac{\|d\varphi^{(i+1)T} Z\|}{\|d\varphi^{iT} Z\|} \prod_{\substack{0\leqslant i\leqslant n-1 \\ \varphi^{iT} w \in A_{t,\varepsilon}}} \frac{\|d\varphi^{(i+1)T} Z\|}{\|d\varphi^{iT} Z\|} \\
&\leqslant 1 \cdot e^{-(\chi-\varepsilon) N_{n,T,\varepsilon}(w) nT}.
\end{aligned}$$

La majoration par 1 du premier terme découle du lemme 3.1. Si maintenant, à ε et δ fixés, on prend $T \geqslant T(\varepsilon, \delta)$ et $n \geqslant N(T)$, on obtient, pour tout $Z \in E^s$:

$$\frac{\|d\varphi^{nT} Z\|}{\|Z\|} \leqslant e^{-(\chi-\varepsilon)(1-\delta)nT}.$$

En prenant $\delta = \varepsilon/(\chi - \varepsilon)$, cela donne

$$\frac{\|d\varphi^{nT} Z\|}{\|Z\|} \leqslant e^{-(\chi-2\varepsilon)nT}.$$

On en conclut qu'il existe une constante C telle que pour tous $Z \in E^s$ et $t \geqslant 0$,

$$\frac{\|d\varphi^{t} Z\|}{\|Z\|} \leqslant Ce^{-(\chi-2\varepsilon)t}.$$

C'est gagné, puisque ε est arbitrairement petit. □

Nous allons ici nous servir de la proposition 6.7 pour attaquer la partie technique : le lemme 7.12. On définit pour $k \in \mathbb{N}$ et $w \in \mathtt{NW}$,

$$Q_k(w) = \inf\left\{ -\frac{1}{2^k} \ln \frac{\|d\varphi^{2^k} Z\|}{\|Z\|} \,\middle|\, Z \in E^s(w) \right\}.$$

Comme le flot est uniformément hyperbolique, Q_k est $\geqslant 0$ et majoré, indépendamment de k. Posons alors

$$F_{k,\varepsilon} = \max\{0, (\chi - \varepsilon) - Q_k\}.$$

Ainsi, $F_{k,\varepsilon}(w) > 0$ s'il y a un vecteur stable en w qui est contracté avec un exposant inférieur à $\chi - \varepsilon$. Autrement dit, $F_{k,\varepsilon}(w) > 0$ si et seulement si $w \notin A_{2^k,\varepsilon}$.

Les fonctions Q_k et $F_{k,\varepsilon}$ sont toutes deux positives, continues sur $\mathtt{NW}$, et majorées indépendamment de k. Le premier lemme est le suivant :

Lemme 7.13. *Soit $\varepsilon > 0$. Pour k assez grand, la fonction $F_{k,\varepsilon}$ est à support compact sur* $\mathtt{NW}$.

Démonstration. Rappelons-nous que le coeur convexe $C(M)$ de M se décompose en une partie compacte K et un nombre fini de pointes $\mathscr{C}_i$, $1 \leqslant i \leqslant l$, asymptotiquement hyperboliques. On a montré (voir la remarque 5.7) que pour tout point $w \in HM$ tel que $\varphi^s(w) \in H\mathscr{C}_i$, $0 \leqslant s \leqslant t$, on avait

$$\|d\varphi^t Z^s\| \leqslant Me^{-t}\|Z^s\|, \quad Z^s \in E^s(w),$$

pour une certaine constante $M > 0$. Ainsi, pour tout $\delta > 0$, il existe T_δ tel que pour $t \geqslant T_\delta$, on a, pour tout point $w \in HM$ tel que $\varphi^s(w) \in H\mathscr{C}_i$, $0 \leqslant s \leqslant t$,

$$\|d\varphi^t Z^s\| \leqslant e^{-(1-\delta)t}\|Z^s\|, \quad Z^s \in E^s(w).$$

Comme d'après le corollaire 7.7, on a toujours $\chi \leqslant 1$, on peut prendre

$$\delta = 1 - \frac{1+\chi-\varepsilon}{2}$$

et k tel que $2^k \geqslant T_\delta$. On obtient alors que, pour tout point $w \in HM$ tel que $\varphi^s(w) \in H\mathscr{C}_i$, $0 \leqslant s \leqslant 2^k$,

$$\|d\varphi^{2^k} Z^s\| \leqslant e^{-\frac{1+\chi-\varepsilon}{2}2^k} \|Z^s\|, \quad Z^s \in E^s(w).$$

En particulier, $F_{k,\varepsilon} = 0$ sur l'ensemble

$$\{w \in \mathtt{NW} \mid \varphi^s(w) \in H\mathscr{C}_i,\ 0 \leqslant s \leqslant 2^k\},$$

dont le complémentaire dans NW est compact. La fonction $F_{k,\varepsilon}$ est donc à support compact pour k assez grand. □

Rappelons qu'on a noté $\mathcal{M}$ l'ensemble des mesures de probabilités sur NW invariantes par le flot.

Lemme 7.14. *Pour tous* $\varepsilon, \delta > 0$, *il existe* $k(\varepsilon,\delta)$ *tel que pour tout* $k \geqslant k(\varepsilon,\delta)$ *et tout* $\eta \in \mathcal{M}$,

$$\int F_{k,\varepsilon}\, d\eta < \delta.$$

Démonstration. Fixons $\varepsilon > 0$, et choisissons k assez grand pour que la fonction $F_{k,\varepsilon}$ soit à support compact S sur NW. Notons $\mathfrak{M}$ l'espace vectoriel des mesures de Radon signées sur NW, muni de la topologie de la convergence étroite des mesures. Si A est un compact de NW et $m > 0$, l'ensemble des mesures de $\mathfrak{M}$ à support dans A et de masse totale $\leqslant m$ est compact pour cette topologie. En particulier, en notant, pour $\eta \in \mathfrak{M}$, η_S la mesure définie par

$$\eta_S(B) = \eta(S \cap B), \quad B \text{ Borélien de NW},$$

l'ensemble

$$\mathcal{M}(S) = \{\eta_S \mid \eta \in \mathcal{M}\}$$

est un ensemble compact.

On définit une forme linéaire sur $\mathfrak{M}$ par

$$\Psi_k : \eta \in \mathfrak{M} \mapsto \int F_{k,\varepsilon}\, d\eta \in \mathbb{R}.$$

Remarquons tout de suite que $\Psi_k(\eta) = \Psi_k(\eta_S)$ pour toute mesure $\eta \in \mathfrak{M}$. La forme linéaire Ψ_k est positive et continue, et surtout la suite (Ψ_k) est uniformément bornée : pour tout k,

$$\|\Psi_k\| = \sup_{\|\eta\|\leqslant 1} |\Psi_k(\eta)| \leqslant \|F_{k,\varepsilon}\|_\infty < \chi < +\infty.$$

On munit l'espace $\mathfrak{M}'$ des formes linéaires continues sur $\mathfrak{M}$ de la topologie

$*$-faible : une suite (Φ_n) de $\mathfrak{M}'$ converge vers Φ si pour toute mesure $\eta \in \mathfrak{M}$, $(\Phi_n(\eta))$ converge vers $\Phi(\eta)$. Pour cette topologie, les ensembles bornés sont relativement compacts. En particulier, on peut supposer, quitte à extraire une sous-suite, que la suite (Ψ_k) converge vers $\Psi \in \mathfrak{M}'$.

Notons $C(\mathcal{M}_{\mathrm{Per}})$ l'enveloppe convexe de l'ensemble $\mathcal{M}_{\mathrm{Per}}$ des mesures portées par les orbites périodiques. Si $\eta \in C(\mathcal{M}_{\mathrm{Per}})$, on a, par construction, $\Psi(\eta) = 0$. Par densité de $C(\mathcal{M}_{\mathrm{Per}})$ dans $\mathcal{M}$ (proposition 6.7) et continuité de Ψ, on en déduit que $\Psi = 0$ sur $\mathcal{M}$. Maintenant, en écrivant, pour $Z \in E^s$,

$$-\frac{1}{2^{k+1}} \ln \frac{\|d\varphi^{2^{k+1}} Z\|}{\|Z\|} = \frac{1}{2}\left(-\frac{1}{2^k} \ln \frac{\|d\varphi^{2^k} d\varphi^{2^k} Z\|}{\|d\varphi^{2^k} Z\|} - \frac{1}{2^k} \ln \frac{\|d\varphi^{2^k} Z\|}{\|Z\|}\right),$$

on remarque facilement que

$$Q_{k+1} \geqslant \tfrac{1}{2}(Q_k \circ \varphi^{2^k} + Q_k).$$

De là, on obtient par définition de $F_{k,\varepsilon}$ que

$$F_{k+1,\varepsilon} \leqslant \tfrac{1}{2}(F_{k,\varepsilon} \circ \varphi^{2^k} + F_{k,\varepsilon}).$$

Cela entraîne que, si $\eta \in \mathcal{M}$ est une mesure de probabilité invariante,

$$\Psi_{k+1}(\eta) = \int F_{k+1,\varepsilon}\, d\eta \leqslant \tfrac{1}{2}\left(\int F_{k,\varepsilon} \circ \varphi^{2^k}\, d\eta + \int F_{k,\varepsilon}\, d\eta\right) = \int F_{k,\varepsilon}\, d\eta = \Psi_k(\eta).$$

Ainsi, pour tout $\eta \in \mathcal{M}$, la suite $(\Psi_k(\eta))$ est décroissante.

Ainsi, la suite de fonctions $\Psi_k : \mathcal{M} \to \mathbb{R}$ converge en décroissant vers 0. L'ensemble $\mathcal{M}(S)$ étant compact, le théorème de Dini entraîne que la convergence de (Ψ_k) vers 0 est uniforme sur $\mathcal{M}(S)$, donc sur $\mathcal{M}$ puisque $\Psi_k(\eta) = \Psi_k(\eta_S)$ pour tout $\eta \in \mathcal{M}$. Autrement dit, pour tout $\delta > 0$, il existe $k(\varepsilon, \delta)$ tel que pour $k \geqslant k(\varepsilon, \delta)$ et $\eta \in \mathcal{M}$,

$$\Psi_k(\eta) = \int F_{k,\varepsilon}\, d\eta \leqslant \delta. \qquad \square$$

Nous aurons enfin besoin de l'observation élémentaire qui suit :

Lemme 7.15. *Soient $\varepsilon > 0$ et $T > 0$. Si $w \notin A_{T,\varepsilon}$, alors, pour tout $-\varepsilon/8 < \lambda < \varepsilon/8$, $\varphi^{\lambda T}(w) \notin A_{T,\varepsilon/2}$.*

Démonstration. Soit $\varepsilon > 0$. En vertu du corollaire 3.2, pour tout $t \geqslant 0$ et tout $Z \in E^s$,

$$\frac{\|d\varphi^t Z\|}{\|Z\|} \geqslant e^{-2t}.$$

Ainsi, pour tous $T > 0$, $\lambda \in \mathbb{R}$,

$$\frac{\|d\varphi^{(1+\lambda)T} Z\|}{\|Z\|} = \frac{\|d\varphi^{\lambda T} d\varphi^T Z\|}{\|d\varphi^T Z\|} \frac{\|d\varphi^T Z\|}{\|Z\|} \geqslant e^{-2\lambda T} \frac{\|d\varphi^T Z\|}{\|Z\|}.$$

Par conséquent, si $w \notin A_{T,\varepsilon}$, alors, pour tout $\lambda \geqslant 0$ et tout $Z \in E^s(\varphi^t w)$,

$$\begin{aligned}\frac{\|d_{\varphi^t w}\varphi^T Z\|}{\|Z\|} &= \frac{\|d_w\varphi^{T+t} d_{\varphi^t w}\varphi^{-t} Z\|}{\|d_{\varphi^t w}\varphi^{-t} Z\|} \frac{\|d_{\varphi^t w}\varphi^{-t} Z\|}{\|Z\|} \\ &\geqslant e^{-2\lambda T} \frac{\|d_w\varphi^T d_{\varphi^t w}\varphi^{-t} Z\|}{\|d_{\varphi^t w}\varphi^{-t} Z\|} e^{-2\lambda T} \\ &\geqslant e^{-4\lambda T} e^{-(\chi-\varepsilon)T} \geqslant e^{-(\chi-\varepsilon+4\lambda)T}.\end{aligned}$$

Cela implique que, si $w \notin A_{T,\varepsilon}$, alors $\varphi^{\lambda T}(w) \notin A_{T,\varepsilon/2}$ pour tout $-\dfrac{\varepsilon}{8} < \lambda < \dfrac{\varepsilon}{8}$. □

Démonstration du lemme 7.12. Fixons pour la suite $\varepsilon > 0$. Choisissons $\delta > 0$ et $T = 2^k$ avec $k \geqslant k(\varepsilon/4, \delta\varepsilon^2/16)$ donné par le lemme 7.14. On a donc, pour toute mesure de probabilité invariante η,

$$\int F_{k,\varepsilon/4}\, d\eta < \delta \frac{\varepsilon^2}{16}.$$

On procède par l'absurde en supposant qu'il existe un point $w \in \mathtt{NW}$ et une suite $(n_j)_{j\in\mathbb{N}}$ telle que $N_{n_j,\varepsilon,T}(w) \leqslant 1-\delta$. D'après le lemme 7.15, à chaque fois qu'un point $\varphi^{iT}(w)$, $i \in \mathbb{N}$, de l'orbite de w sous φ^T n'est pas dans $A_{T,\varepsilon}$, alors $\varphi^{t+iT}(w)$ n'est pas dans $A_{T,\varepsilon/2}$ pour tout $-\varepsilon T/8 < t < \varepsilon T/8$; et donc $F_{k,\varepsilon/4}(\varphi^t w) \geqslant \varepsilon/4$.

Or, parmi les points $\varphi^{iT}(w)$, $0 \leqslant i \leqslant n_j$ de l'orbite de w sous φ^T, il y en a $N_{n_j,\varepsilon,T}(w)n_j$ qui sont dans $A_{T,\varepsilon}$; cela implique qu'entre les instants 0 et n_jT, l'orbite de w n'est pas dans $A_{T,\varepsilon/2}$ pendant au moins le temps $n_jT\delta\varepsilon/4$. Autrement dit,

$$\begin{aligned}\frac{1}{n_jT}\int_0^{n_jT} F_{k,\varepsilon/4}(\varphi^t w)\, dt &\geqslant \frac{1}{n_jT}\int_0^{n_jT} F_{k,\varepsilon/4}(\varphi^t w)(1-\mathbf{1}_{A_{T,\varepsilon/2}}(\varphi^t w))\, dt \\ &\geqslant \frac{1}{n_jT}\frac{\varepsilon}{4} n_jT\delta\frac{\varepsilon}{4} = \delta\frac{\varepsilon^2}{16}.\end{aligned}$$

On définit la suite de mesures de probabilités $(\eta_j)_{j\in\mathbb{N}}$ par

$$\int f\, d\eta_j = \frac{1}{n_jT}\int_0^{n_jT} f(\varphi^t(w))\, dt, \qquad f \in C(\mathtt{NW}).$$

Toute valeur d'adhérence η de la suite (η_j) vérifie

$$\int F_{k,\varepsilon/4}\, d\eta \geqslant \delta\frac{\varepsilon^2}{16}.$$

Or, une telle mesure η est nécessairement invariante par le flot, et cela contredit le choix de k. □

8. Quelques exemples

8A. ***Un exemple où le flot géodésique n'est pas uniformément hyperbolique.*** On va construire un exemple "dégénéré" où le flot géodésique a un exposant de Lyapunov nul. Pour cela, on considère un certain groupe fuchsien Γ qui contient un parabolique, et on le fait agir de façon canonique sur $\mathbb{H}^3$; on construit alors un nouvel ouvert convexe Γ-invariant dont le bord n'est de classe $\mathscr{C}^{1+\varepsilon}$ en aucun point parabolique, pour tout $\varepsilon > 0$. D'après la proposition 7.2, toute orbite ultimement incluse dans la pointe aura un exposant de Lyapunov nul.

Le groupe Γ ici présenté n'est donc pas irréductible sur $\mathbb{P}^3$ mais il est sans doute possible de le déformer par pliage, tout en préservant les propriétés de régularité que l'on désirait.

Le résultat s'énonce ainsi :

Proposition 8.1. *Il existe un ouvert proprement convexe* Ω *de* $\mathbb{P}^3$, *strictement convexe à bord* $\mathscr{C}^1$, *et un sous-groupe discret* Γ *de* $\mathrm{Aut}(\Omega)$ *tels que* :

– *l'action de* Γ *sur* Ω *soit géométriquement finie mais non convexe-cocompacte* ;
– *le bord* $\partial\Omega$ *de* Ω *n'est pas de classe* $\mathscr{C}^{1+\varepsilon}$ *aux points paraboliques de* Λ_Γ, *pour tout* $\varepsilon > 0$.

En particulier, *le flot géodésique sur la variété quotient* Ω/Γ *a un exposant de Lyapunov nul* ; *il n'est donc pas uniformément hyperbolique.*

Démonstration. Soient Σ le tore à 1 trou et Γ son groupe fondamental ; Γ est un groupe libre à 2 générateurs. On munit Σ d'une structure hyperbolique de volume

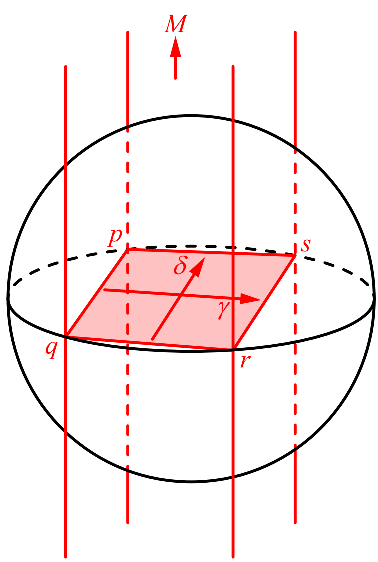


Figure 9. Domaines fondamentaux.

fini de la façon suivante : on se donne un carré idéal P de $\mathbb{H}^2$ et on identifie les côtés opposés de ce carré à l'aide de deux éléments hyperboliques γ et δ. Ainsi le domaine fondamental pour l'action de Γ sur $\mathbb{H}^2$ est le carré idéal en question. Pour simplifier la discussion, on choisit ce carré idéal de telle façon qu'il ait un groupe diédral d'ordre 8 de symétrie.

À présent, on plonge Γ dans $SO_{3,1}(\mathbb{R})$ de façon canonique. Ainsi, Γ agit sur l'espace hyperbolique $\mathbb{H}^3$ de dimension 3. L'ensemble limite de Γ sur $\mathbb{H}^3$ est un cercle, intersection d'un plan projectif Π et de $\partial\mathbb{H}^3$.

Le groupe Γ préserve le plan Π et le point M intersection des hyperplans tangents à $\partial\mathbb{H}^3$ en $\Lambda_\Gamma = \Pi \cap \partial\mathbb{H}^3$. L'ouvert convexe Ω_∞ obtenu en prenant la réunion des deux cônes de bases $\Pi \cap \mathbb{H}^3$ et de sommet M est préservé par Γ et par $SO_{2,1}(\mathbb{R})$. Un domaine fondamental D pour l'action de Γ sur Ω_∞ est la réunion des deux cônes de base P et de sommet M. Nous allons construire une partie convexe D_1 de D telle que la réunion $\bigcup_{\gamma\in\Gamma} \gamma(D_1)$ nous donne l'ouvert convexe Ω désiré.

On note p, q, r, s les sommets de P, γ l'élément qui identifie $[pq]$ avec $[sr]$ et δ celui qui identifie $[rq]$ avec $[sp]$ (voir la figure 10) ; le groupe Γ est engendré par ces deux éléments γ et δ. On appelle $\Pi_{p,q}$ (respectivement $\Pi_{q,s}, \ldots$) le plan contenant p, q (respectivement $q, s, \ldots$) et M.

On commence par s'intéresser à l'intersection de Ω_∞ avec le plan $\Pi_{p,q}$ engendré par p, q et M. On choisit une première courbe $\mathscr{C}_{p,q}$ qui joint p à q puis q à p et qui :

- est incluse dans la face de D contenant $[pq]$, c'est-à-dire $\Pi_{p,q} \cap \bar{D}$;
- est strictement convexe et de classe $\mathscr{C}^1$;
- n'est pas de classe $\mathscr{C}^{1+\varepsilon}$ en p et q pour aucun $\varepsilon > 0$.

On peut remarquer au passage que cette courbe est, par la dernière propriété, incluse dans $\mathbb{H}^3$ au voisinage de p et q. On utilise à présent les symétries du groupe Γ (c'est-à-dire celles de P) pour copier cette courbe et obtenir des courbes $\mathscr{C}_{p,s}$, $\mathscr{C}_{r,q}$ et $\mathscr{C}_{s,r}$ joignant $\{p, s\}$, $\{r, q\}$ et $\{s, r\}$.

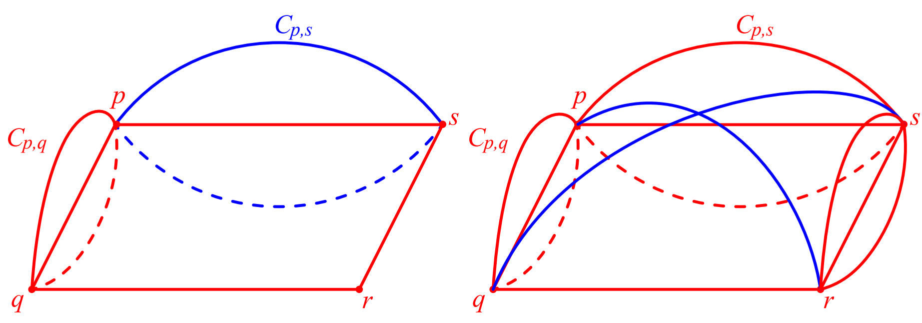


Figure 10. Les courbes et le dôme.

Nous allons “relier” ces courbes pour construire le bord de D_1 (qu’on appellera le dôme). Pour simplifier la discussion, il est bon de remarquer que cet ensemble de courbes admet le même groupe de symétrie que le carré idéal de départ, à savoir un groupe diédral d’ordre 8.

Soit $\gamma^{\mathbb{R}}$ (respectivement $\delta^{\mathbb{R}}$) le groupe à 1-paramètre engendré par γ (respectivement δ). Les orbites d’un point de $\Omega_\infty \smallsetminus \Pi$ sous $\gamma^{\mathbb{R}}$ sont des demi-ellipses d’extrémités γ^- et γ^+ ; de même pour $\delta^{\mathbb{R}}$.

Le domaine fondamental D privé de $\Pi_{p,r} \cup \Pi_{q,s}$ possède 4 composantes connexes D_{δ^-}, D_{γ^-}, D_{δ^+} et D_{γ^+}, naturellement étiquetées par $\delta^-, \gamma^-, \delta^+, \gamma^+$. L’orbite de $\mathscr{C}_{p,q}$ sous $\gamma^{\mathbb{R}}$ est une surface convexe[2] $S_{p,q}$, qui contient $\mathscr{C}_{s,r}$. De même, en considérant l’orbite de $\mathscr{C}_{r,q}$ sous $\delta^{\mathbb{R}}$, on obtient une surface convexe $S_{r,q}$. Soit S la surface obtenue comme la réunion

$$S = (D_{\delta^-} \cap S_{r,q}) \cup (D_{\gamma^-} \cap S_{p,q}) \cup (D_{\delta^+} \cap S_{r,q}) \cup (D_{\gamma^+} \cap S_{p,q}).$$

La surface S possède encore un groupe diédral d’ordre 8 de symétries.

Appelons D_0 l’adhérence dans D de l’enveloppe convexe de la surface S. L’ensemble D_0 est convexe, inclus dans $\mathbb{H}^3$. La réunion $\Omega_0 = \bigcup_{\gamma \in \Gamma} \gamma(D_0)$ est un ouvert proprement convexe. En effet, $\gamma(D_0) \cup D_0$ est encore convexe puisque, par construction, les surfaces $S_{p,q}$ et $\gamma(S_{p,q})$ se recollent pour donner une surface convexe ; bien sûr, le même chose est valable pour δ ; le résultat pour Γ s’en déduit à l’aide d’une récurrence sur la longueur d’un élément pour la métrique des mots de Γ.

L’ouvert proprement convexe Ω_0 est strictement convexe mais son bord n’est pas de classe $\mathscr{C}^1$ a priori. En dehors des courbes $\mathscr{C}_{p,r} = \Pi_{p,r} \cup \partial\Omega_0$ et $\mathscr{C}_{q,s} = \Pi_{q,s} \cup \partial\Omega_0$ et de leurs images par Γ, la surface $\partial\Omega_0$ est de classe $\mathscr{C}^1$. En lissant Ω_0 le long de ces courbes, on obtient qu’il existe un voisinage $\mathscr{V}$ de $\mathscr{C}_{p,r} \cup \mathscr{C}_{q,s}$ dans $\partial\Omega_0$ et un convexe D_1 tel que $\partial D_0 \smallsetminus \mathscr{V} = \partial D_1 \smallsetminus \mathscr{V}$. L’ensemble $\Omega_1 = \bigcup_{\gamma \in \Gamma} \gamma(D_1)$ est alors un ouvert proprement convexe, strictement convexe, à bord $\mathscr{C}^1$ et Γ-invariant, mais son bord n’est pas de classe $\mathscr{C}^{1+\varepsilon}$ aux points paraboliques de Λ_Γ, pour tout $\varepsilon > 0$. □

8B. ***Représentation sphérique de* $\mathbf{SL_2(\mathbb{R})}$ *dans* $\mathbf{SL_5(\mathbb{R})}$.** Dans [Crampon et Marquis 2012], on avait introduit deux notions de finitude géométrique, la finitude géométrique sur Ω et sur $\partial\Omega$. Dans l’article présent, nous n’avons étudié que la première.

Comme exemple d’action géométriquement finie sur $\partial\Omega$ mais pas sur Ω, on avait donné la représentation sphérique de $SL_2(\mathbb{R})$ dans $SL_5(\mathbb{R})$: il s’agit de l’action de $SL_2(\mathbb{R})$ sur l’espace $V_4 = \mathbb{R}_4[X, Y]$ des polynômes homogènes de degré 4 en deux variables, sur lequel $SL_2(\mathbb{R})$ agit par coordonnées. Notons $G < SL_5(\mathbb{R})$ l’image de $SL_2(\mathbb{R})$ par cette représentation.

2. On dit qu’une hypersurface de $\mathbb{P}^n$ est *convexe* si elle est une partie du bord d’un convexe de $\mathbb{P}^n$.

Rappelons ce qui a été vu dans [Crampon et Marquis 2012]. Il existe un point $x \in \mathbb{P}^4$ dont l'orbite sous G est la courbe de Veronese, dont une équation est donnée par :

$$[t:s] \in \mathbb{P}^1 \to [t^4 : t^3 s : t^2 s^2 : t s^3 : s^4] \in \mathbb{P}^4.$$

L'ensemble des ouverts proprement convexes préservés par cette représentation de $\mathrm{SL}_2(\mathbb{R})$ dans $\mathrm{SL}_5(\mathbb{R})$ forme une famille croissante $\{\Omega_r,\ 0 \leqslant r \leqslant \infty\}$. L'ouvert convexe Ω_0 est l'enveloppe convexe de la courbe de Veronese, et Ω_∞ est son dual ; ces deux convexes ne sont ni strictement convexes ni à bord $\mathscr{C}^1$. Les ouverts convexes Ω_r, $0 < r < \infty$ sont les r-voisinages de Ω_0 dans la géométrie de Hilbert $(\Omega_\infty, d_{\Omega_\infty})$. On a vu dans [Crampon et Marquis 2012] que ces convexes étaient strictement convexes et à bord $\mathscr{C}^1$. En fait, on peut voir en procédant comme dans la section 7 que :

Proposition 8.2. *Pour $0 < r < \infty$, le bord de l'ouvert convexe Ω_r est de classe $\mathscr{C}^{4/3}$ et 4-convexe.*

Démonstration. L'ensemble limite Λ_G de l'action de G sur Ω_r est dans tous les cas la courbe de Veronese. Hors de l'ensemble limite, le bord de Ω_r est lisse car l'action de G sur $\partial\Omega_r \smallsetminus \Lambda_G$ est libre et transitive (voir [Crampon et Marquis 2012, section 10]) ; $\partial\Omega_r \smallsetminus \Lambda_G$ s'identifie donc à une orbite de G. De plus, le dual de Ω_r est un certain $\Omega_{r'}$. Le fait que $\partial\Omega_{r'} \smallsetminus \Lambda_G$ soit lisse implique que $\partial\Omega_r \smallsetminus \Lambda_G$ est 2-convexe. Autrement dit, le bord de Ω_r est lisse à hessien défini positif hors de l'ensemble limite.

La courbe de Veronese Λ_G est une courbe algébrique lisse sur laquelle G agit transitivement. La régularité $\partial\Omega_r$ est donc la même en tout point de Λ_G. Or, un point x de Λ_G est un point fixe d'un certain élément hyperbolique $g \in G$. La régularité de $\partial\Omega_r$ en x se lit sur les valeurs propres de g. Il n'est pas dur de voir que si g est l'image par la représentation d'un élément hyperbolique γ de $\mathrm{SL}_2(\mathbb{R})$. Si λ, λ^{-1} sont les valeurs propres de γ, avec $\lambda > 1$, alors celles de g sont $\lambda^4, \lambda^2, 1, \lambda^{-2}, \lambda^{-4}$. D'après la proposition 7.2 et le lemme 7.6, $\partial\Omega_r$ est $\mathscr{C}^{1+\varepsilon}$ en x pour tout $\varepsilon < \varepsilon(g)$ avec

$$\varepsilon(g) = \frac{\ln \lambda^{-2} - \ln \lambda^{-4}}{\ln \lambda^4 - \ln \lambda^{-2}} = \frac{1}{3}.$$

En fait, dans le cas d'un point fixe hyperbolique, on peut être plus précis dans la proposition 7.2 et voir que la valeur $\varepsilon(g)$ est atteinte, autrement dit que le bord est $\mathscr{C}^{1+\varepsilon(g)}$ en x (et pas plus). On obtient donc que $\partial\Omega_r$ est $\mathscr{C}^{4/3}$ et, par dualité, que $\partial\Omega_r$ est 4-convexe. (En effet, le bord $\partial\Omega$ est β-convexe au point x si et seulement si le bord $\partial\Omega^*$ du convexe dual est $\mathscr{C}^{1+\varepsilon}$ au point x^*, avec $1/(1+\varepsilon) + 1/\beta = 1$.) □

9. Entropie volumique et exposant critique

Si Γ est un sous-groupe discret de $\mathrm{Aut}(\Omega)$, on notera, pour $x \in \Omega$ et $R \geqslant 0$,

$$N_\Gamma(x, R) = \sharp\{g \in \Gamma \mid d_\Omega(x, gx) \leqslant R\}$$

le nombre d'éléments g de Γ tels que $gx \in B(x, R)$. L'exposant critique du groupe Γ, défini par

$$\delta_\Gamma = \limsup_{R\to+\infty} \frac{1}{R} \ln N_\Gamma(x, R),$$

mesure le taux de croissance exponentiel du groupe Γ agissant sur Ω ; il est immédiat que la limite précédente ne dépend pas du point x considéré.

L'exposant critique δ_Γ de Γ est nommé ainsi car c'est l'exposant critique des séries de Poincaré de Γ données par

$$g_\Gamma(s, x) = \sum_{\gamma\in\Gamma} e^{-sd_\Omega(x,\gamma x)}, \quad x \in \Omega;$$

cela veut dire que pour $s > \delta_\Gamma$, la série converge, et pour $s < \delta_\Gamma$, elle diverge.

L'entropie volumique d'une géométrie de Hilbert

$$h_{\mathrm{vol}}(\Omega) = \limsup_{R\to+\infty} \frac{1}{R} \ln \mathrm{Vol}_\Omega B(x, R)$$

représente le taux de croissance exponentiel du volume des boules de l'espace métrique (Ω, d_Ω).

9A. ***Groupes de covolume fini.*** Si la géométrie (Ω, d_Ω) admet une action cocompacte d'un sous-groupe discret Γ de $\mathrm{Aut}(\Omega)$, on a évidemment l'égalité

$$\delta_\Gamma = h_{\mathrm{vol}}(\Omega)$$

puisqu'alors l'entropie volumique ne dépend pas de la mesure de volume considérée, pourvu qu'elle soit Γ-invariante ; aussi peut-on prendre la mesure de comptage de l'orbite d'un point x de Ω sous Γ pour retrouver δ_Γ. Si le groupe est "trop petit", cette égalité devient en général fausse, et on a seulement $\delta_\Gamma \leqslant h_{\mathrm{vol}}$. Dans [Dal'Bo et al. 2009], Françoise Dal'bo, Marc Peigné, Jean-Claude Picaud et Andrea Sambusetti ont étudié cette question pour les sous-groupes de covolume fini de variétés de Hadamard, à courbure négative pincée. Ils ont montré le résultat suivant.

Théorème 9.1. – *Soit M une variété riemannienne à courbure strictement négative, de volume fini. Si M est asymptotiquement $\frac{1}{4}$-pincée, alors $h_{\mathrm{vol}} = h_{\mathrm{top}}$.*

– *Pour tout $\varepsilon > 0$, il existe une variété riemannienne de volume fini et de courbure strictement négative $\left(\frac{1}{4} + \varepsilon\right)$-pincée telle que $h_{\mathrm{top}} < h_{\mathrm{vol}}$.*

L'hypothèse de pincement asymptotique concerne la géométrie de la variété à l'infini, c'est-à-dire dans ses pointes. Dans notre contexte, c'est le corollaire 4.2 qui va nous permettre de montrer le prochain résultat :

Théorème 9.2. *Soit Γ un sous-groupe discret de* $\mathrm{Aut}(\Omega)$, *de covolume fini. Alors*

$$\delta_\Gamma = h_{\mathrm{vol}}(\Omega).$$

La démonstration de ce résultat est fort similaire à celle de [Dal'Bo et al. 2009], elle se simplifie par certains aspects et nécessite des arguments un peu différents par d'autres. Elle reste malgré tout un brin technique...

On va commencer par calculer l'exposant critique d'un sous-groupe parabolique de rang maximal.

Lemme 9.3. *Soit Γ un sous-groupe de* $\mathrm{Aut}(\Omega)$ *et* $\mathrm{Aut}(\Omega')$ *avec $\Omega \subset \Omega'$. Appelons $g_{\Gamma,\Omega}(s,x)$ et $g_{\Gamma,\Omega'}(s,x)$ les séries de Poincaré pour l'action de Γ sur Ω et Ω', $\delta_\Gamma(\Omega)$ et $\delta_\Gamma(\Omega')$ leur exposant critique. Alors, pour tout $s > \delta_\Gamma(\Omega')$,*

$$g_{\Gamma,\Omega}(s,x) \leqslant g_{\Gamma,\Omega'}(s,x).$$

En particulier, $\delta_\Gamma(\Omega) \leqslant \delta_\Gamma(\Omega')$.

Démonstration. Pour $x, y \in \Omega$, on a $d_{\Omega'}(x,y) \leqslant d_\Omega(x,y)$. Donc si $x \in \Omega$ et $s > \delta_\Gamma(\Omega')$, on a $g_{\Gamma,\Omega}(s,x) \leqslant g_{\Gamma,\Omega'}(s,x)$. En particulier, la convergence de $g_{\Gamma,\Omega'}(s,x)$ implique celle de $g_{\Gamma,\Omega}(s,x)$, d'où le résultat. □

Lemme 9.4. *L'exposant critique d'un sous-groupe parabolique de rang maximal $\mathscr{P}$ de* $\mathrm{Aut}(\Omega)$ *est $\delta_{\mathscr{P}} = (n-1)/2$ et les séries de Poincaré de $\mathscr{P}$ divergent en $\delta_{\mathscr{P}}$:*

$$\sum_{\gamma \in \mathscr{P}} e^{-\delta_{\mathscr{P}} d_\Omega(x,\gamma x)} = +\infty \quad \textit{pour tout } x \in \Omega.$$

Démonstration. Appelons p le point fixe de $\mathscr{P}$. D'après le théorème 4.1, on peut trouver deux ellipsoïdes $\mathscr{E}^{\mathrm{int}}$ et $\mathscr{E}^{\mathrm{ext}}$ $\mathscr{P}$-invariants tels que

$$\partial\mathscr{E}^{\mathrm{int}} \cap \partial\mathscr{E}^{\mathrm{ext}} = \partial\mathscr{E}^{\mathrm{int}} \cap \partial\Omega = \partial\mathscr{E}^{\mathrm{ext}} \cap \partial\Omega = \{p\} \quad \text{et} \quad \mathscr{E}^{\mathrm{int}} \subset \Omega \subset \mathscr{E}^{\mathrm{ext}}.$$

Il est connu en géométrie hyperbolique que $\delta_{\mathscr{P}}(\mathscr{E}^{\mathrm{int}}) = \delta_{\mathscr{P}}(\mathscr{E}^{\mathrm{ext}}) = (n-1)/2$ et que les séries de Poincaré de $\mathscr{P}$ divergent en l'exposant critique (on pourra consulter, par exemple, [Dal'bo et al. 2000, partie 3], même si les calculs, élémentaires, remontent à Alan Beardon, dans les années 1970). D'après le lemme 9.3, on a de même pour $\mathscr{P}$ agissant sur Ω. □

Nous aurons aussi besoin du lemme suivant :

Lemme 9.5. *Soit $C > 1$ arbitrairement proche de 1 et $\mathscr{P}$ un sous-groupe parabolique maximal de* $\mathrm{Aut}(\Omega)$ *fixant $p \in \partial\Omega$. Alors il existe une horoboule H_C basée en p,* *d'horosphère au bord $\mathcal{H}_C$ et une constante $D > 1$ telles que*

$$\frac{1}{D} N_{\mathcal{P}}\Big(x, \frac{R}{C}\Big) \leqslant \mathrm{Vol}_\Omega(B(x,R) \cap H_C) \leqslant D N_{\mathcal{P}}(x, CR), \quad x \in \mathcal{H}_C,\ R > 0.$$

Démonstration. Dans l'espace hyperbolique, on sait (voir, par exemple, [Dal'Bo et al. 2009, proposition 3.3]) que, pour tout sous-groupe parabolique maximal $\mathcal{P}$, toute horoboule H stable par $\mathcal{P}$, d'horosphère au bord $\mathcal{H}$, il existe un réel $D \geqslant 1$ tel que

$$\text{(9-1)} \quad \frac{1}{D}\mathrm{Vol}_{\mathbb{H}^n}(B(x,R) \cap H_C) \leqslant N_{\mathcal{P}}(x,R) \leqslant D\mathrm{Vol}_{\mathbb{H}^n}(B(x,R) \cap H), \quad x \in \mathcal{H}.$$

Soit donc $\mathcal{P}$ un sous-groupe parabolique maximal de $\mathrm{Aut}(\Omega)$ fixant $p \in \partial\Omega$. Le corollaire 4.2 nous fournit une horoboule H_C basée en p qui porte deux métriques hyperboliques $\mathcal{P}$-invariantes $\mathtt{h}$ and $\mathtt{h}'$ telles que

$$\frac{1}{C}\mathtt{h}' \leqslant \mathtt{h} \leqslant F \leqslant \mathtt{h}' \leqslant C\mathtt{h}.$$

Prenons $x \in \mathcal{H}$. D'après la proposition 2.1, on a, pour tout $R > 0$,

$$B_{\mathtt{h}'}\Big(x, \frac{R}{C}\Big) \subset B_{\mathtt{h}}(x,R) \subset B(x,R) \subset B_{\mathtt{h}'}(x,R) \subset B_{\mathtt{h}}(x, CR),$$

où, par $B_{\mathtt{h}}$ et $B_{\mathtt{h}'}$, on note les boules métriques pour $\mathtt{h}$ et $\mathtt{h}'$. En appelant $\mathrm{Vol}_{\mathtt{h}}$ et $\mathrm{Vol}_{\mathtt{h}'}$ les volumes riemanniens associés à $\mathtt{h}$ et $\mathtt{h}'$, on a, toujours d'après la proposition 2.1,

$$\mathrm{Vol}_{\mathtt{h}'} \leqslant \mathrm{Vol}_\Omega \leqslant \mathrm{Vol}_{\mathtt{h}}.$$

Ainsi,

$$\mathrm{Vol}_{\mathtt{h}'}\Big(B_{\mathtt{h}'}\Big(x, \frac{R}{C}\Big) \cap H\Big) \leqslant \mathrm{Vol}_\Omega(B(x,R) \cap H) \leqslant \mathrm{Vol}_{\mathtt{h}}(B_{\mathtt{h}}(x, CR) \cap H).$$

D'après l'encadrement (9-1), il existe une constante $D > 1$ telle que

$$\frac{1}{D} N^{\mathtt{h}'}_{\mathcal{P}}\Big(x, \frac{R}{C}\Big) \leqslant \mathrm{Vol}_\Omega(B(x,R) \cap H) \leqslant D N^{\mathtt{h}}_{\mathcal{P}}(x, CR),$$

où $N^{\mathtt{h}}_{\mathcal{P}}(x,R)$ est le nombre de points de l'orbite $\mathcal{P} \cdot x$ dans la boule de rayon R pour $\mathtt{h}$; de même pour $\mathtt{h}'$.

(Bien entendu, les horoboules considérées dans l'encadrement (9-1) sont les horoboules hyperboliques et pas celles de F, et il faut donc faire un peu plus attention lorsqu'on dit qu'une telle constante D existe. Si $\mathcal{H}_{\mathtt{h}}$ est l'horosphère pour $\mathtt{h}$ basée en p et passant par x, alors la $\mathtt{h}$-distance maximale entre $\mathcal{H}$ et $\mathcal{H}_{\mathtt{h}}$ est finie, car $\mathcal{P}$ agit de façon cocompacte sur $\mathcal{H} \smallsetminus \{p\}$ et $\mathcal{H}_{\mathtt{h}} \smallsetminus \{p\}$. Donc, pour une certaine constante $D' > 0$, on a, pour tout $R > 0$,

$$\Big|\mathrm{Vol}_{\mathtt{h}}(B_{\mathtt{h}}(x,R) \cap H) - \mathrm{Vol}_{\mathtt{h}}(B_{\mathtt{h}}(x,R) \cap H_{\mathtt{h}})\Big| \leqslant D' N_{\mathcal{P}}(x,R),$$

où $H_{\mathtt{h}}$ est l'horoboule définie par $\mathcal{H}_{\mathtt{h}}$. D'où l'existence de D.)

Pour conclure, il suffit de remarquer que, comme $\mathtt{h} \leqslant F \leqslant \mathtt{h}'$, on a

$$N_{\mathcal{P}}^{\mathtt{h}}(x, R) \leqslant N_{\mathcal{P}}(x, R) \leqslant N_{\mathcal{P}}^{\mathtt{h}'}(x, C). \qquad \square$$

Démonstration du théorème 9.2. On sait déjà que $\delta_\Gamma \leqslant h_{\mathrm{vol}}$, et il faut donc seulement prouver l'inégalité inverse.

Fixons $C > 1$ arbitrairement proche de 1, et choisissons $o \in \Omega$ ainsi qu'un domaine fondamental convexe localement fini pour l'action de Γ sur Ω, qui contienne le point o. Décomposons ce domaine fondamental en

$$C_0 \sqcup \bigsqcup_{1 \leqslant i \leqslant l} C_i,$$

où C_0 est compact et les C_i, $1 \leqslant i \leqslant l$, correspondent aux pointes $\xi_i \in \partial\Omega$: chaque C_i est une partie d'un domaine fondamental pour l'action d'un sous-groupe parabolique maximal $\mathcal{P}_i$ sur une horoboule H_{ξ_i} basée au point ξ_i ; les points ξ_i sont les points de $\partial\Omega$ adhérents au domaine fondamental. On suppose que les H_{ξ_i} sont choisies de telle façon qu'elles satisfassent au lemme 9.5 pour la constante C qu'on a fixée.

La boule $B(o, R)$ de rayon $R \geqslant 0$ peut être décomposée en

$$B(o, R) = (\Gamma \cdot C_0 \cap B(o, R)) \sqcup \bigsqcup_{1 \leqslant i \leqslant l} \big(\Gamma \cdot H_{\xi_i} \cap B(o, R)\big),$$

de telle façon que

$$\mathrm{Vol}_\Omega(B(o, R)) = \mathrm{Vol}_\Omega(\Gamma \cdot C_0 \cap B(o, R)) + \sum_{i=1}^{l} \mathrm{Vol}_\Omega(\Gamma \cdot H_{\xi_i} \cap B(o, R)).$$

Pour le premier terme, on a $\mathrm{Vol}_\Omega(\Gamma \cdot C_0 \cap B(o, R)) \leqslant N_\Gamma(o, R)\mathrm{Vol}_\Omega(C_0)$; c'est donc le second qu'il nous faut étudier.

Pour chaque horoboule $H_{\gamma\xi_i} = \gamma H_{\xi_i}$, appelons $x_{\gamma,i}$ le point d'intersection de la droite $(o\gamma\xi_i)$ avec l'horosphère $\partial H_{\gamma\xi_i} \smallsetminus \{\gamma\xi_i\}$, qui n'est rien d'autre que la projection de o sur $H_{\gamma\xi_i}$. Pour chaque $\gamma \in \Gamma$, notons $\bar{\gamma} \in \Gamma$ un des éléments $g \in \Gamma$, en nombre fini, tels que $x_{\gamma,i} \in g \cdot C_i$; $\bar{\gamma}$ est le “premier élément pour lequel $H_{\gamma\xi_i}$ intersecte $B(o, R)$.” Appelons $\overline{\Gamma}$ l'ensemble de ces éléments $\bar{\gamma}$.

La remarque principale est le lemme ci-dessous, équivalent du fait classique suivant en courbure négative pincée : pour chaque $\theta \in (0, \pi)$, on peut trouver une constante $C(\theta)$ telle que, pour chaque triangle géodésique xyz dont l'angle au point y est au moins θ, le chemin $x \to y \to z$ sur le triangle est une quasi-géodésique entre x et z avec une erreur au plus $C(\theta)$.

Lemme 9.6. *Il existe $r > 0$ tel que, pour chaque $\gamma \in \Gamma$, $1 \leqslant i \leqslant l$ et $z \in H_{\gamma\xi_i}$, le chemin formé des segments $[ox_{\gamma,i}]$ et $[x_{\gamma,i}z]$ est une quasi-géodésique avec une erreur au plus r, c'est-à-dire que*

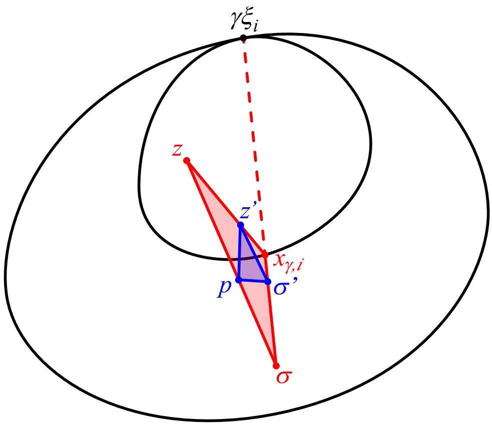
$$d_\Omega(o, z) \geqslant d_\Omega(o, x_{\gamma,i}) + d_\Omega(x_{\gamma,i}, z) - r.$$

Démonstration. Prenons $\gamma \in \Gamma$, $1 \leqslant i \leqslant l$ et $z \in H_{\gamma\xi_i}$. Rappelons que l'espace métrique (Ω, d_Ω) est Gromov-hyperbolique (voir [Crampon et Marquis 2012, section 9]). Aussi existe t-il un réel $\delta \geqslant 0$ pour lequel chaque triangle géodésique est δ-fin. Ainsi, il existe $p \in [oz]$ tel que

$$d_\Omega(p, [x_{\gamma,i}z]) \leqslant \delta, \quad d_\Omega(p, [ox_{\gamma,i}]) \leqslant \delta.$$

On peut donc trouver des points $o' \in [ox_{\gamma,i}]$ et $z' \in [x_{\gamma,i}z]$ de telle façon que

$$d_\Omega(o', p) + d_\Omega(p, z') \leqslant 2\delta.$$



Par l'inégalité triangulaire, la distance entre o' et z' est alors plus petite que 2δ. Puisque $x_{\gamma,i}$ est la projection de o' sur l'horoboule $H_{\gamma\xi_i}$ et que z' est dans l'horoboule $H_{\gamma\xi_i}$, on a $d_\Omega(o', x_{\gamma,i}) \leqslant d_\Omega(o', z') \leqslant 2\delta$ et, par l'inégalité triangulaire, $d_\Omega(x_{\gamma,i}, z') \leqslant 4\delta$. Ainsi, on obtient

$$d_\Omega(o', x_{\gamma,i}) + d_\Omega(x_{\gamma,i}, z') \leqslant 6\delta.$$

Cela amène

$$\begin{aligned} d_\Omega(o, x_{\gamma,i}) + d_\Omega(x_{\gamma,i}, z) &\leqslant d_\Omega(o, o') + d_\Omega(o', x_{\gamma,i}) + d_\Omega(x_{\gamma,i}, z') + d_\Omega(z', z) \\ &\leqslant 6\delta + d_\Omega(o, p) + d_\Omega(p, o') + d_\Omega(z', p) + d_\Omega(p, z) \\ &\leqslant 8\delta + d_\Omega(o, z). \end{aligned}$$ □

Maintenant, si z est un point dans $\gamma \cdot H_{\xi_i} \cap B(o, R)$, pour certains $\gamma \in \Gamma$, $1 \leqslant i \leqslant l$ et $R > 0$, le lemme 9.6 implique que

$$d_\Omega(o, x_{\gamma,i}) + d_\Omega(x_{\gamma,i}, z) \leqslant d_\Omega(o, z) + r \leqslant R + r.$$

Or, il existe $c \geqslant 0$ tel que $d_\Omega(o, x_{\gamma,i}) \geqslant d_\Omega(o, \bar{\gamma} o) - c$: il suffit de prendre pour c la distance maximale entre o et le bord $\partial C_i \cap \partial H_{\xi_i} \smallsetminus \{\xi_i\}$. D'où

$$d_\Omega(x_{\gamma,i}, z) \leqslant R + r - d_\Omega(o, \bar{\gamma} o) + c.$$

Posons $K = r + c$. Pour tous $\gamma \in \Gamma$, $1 \leqslant i \leqslant l$, et $R > 0$, on a ainsi

$$\gamma \cdot H_{\xi_i} \cap B(o, R) \subset \gamma \cdot H_{\xi_i} \cap B(x_{\gamma,i}, R - d_\Omega(o, \bar{\gamma} o) + K).$$

Cela permet d'évaluer le volume $\mathrm{Vol}_\Omega(\Gamma \cdot H_{\xi_i} \cap B(o, R))$. En effet,

$$\begin{aligned}\mathrm{Vol}_\Omega(\Gamma \cdot H_{\xi_i} \cap B(o, R)) &= \sum_{\bar{\gamma} \in \bar{\Gamma}} \mathrm{Vol}_\Omega(\bar{\gamma} \cdot H_{\xi_i} \cap B(o, R)) \\ &\leqslant \sum_{\bar{\gamma} \in \bar{\Gamma}} \mathrm{Vol}_\Omega\big(\bar{\gamma} \cdot H_{\xi_i} \cap B(x_{\gamma,i}, R - d(o, \bar{\gamma} o) + K)\big) \\ &\leqslant \sum_{0 \leqslant k \leqslant [R]} \sum_{\substack{\bar{\gamma} \in \bar{\Gamma} \\ k \leqslant d_\Omega(o, \bar{\gamma} o) \leqslant k+1}} \mathrm{Vol}_\Omega(\bar{\gamma} \cdot H_{\xi_i} \cap B(x_{\gamma,i}, R - k + K)) \\ &\leqslant \sum_{0 \leqslant k \leqslant [R]} N_{\bar{\Gamma}}(o, k, k+1) \mathrm{Vol}_\Omega(H_{\xi_i} \cap B(x_i, R - k + K)),\end{aligned}$$

où $x_i = x_{\mathrm{Id},i}$ et, pour toute partie S de Γ et tout $0 \leqslant r < R$,

$$N_S(o, r, R) = \sharp\{\gamma \in S \mid r \leqslant d_\Omega(o, \gamma o) < R\}.$$

Le lemme 9.5 donne

$$\mathrm{Vol}_\Omega(\Gamma \cdot H_{\xi_i} \cap B(o, R)) \leqslant D \sum_{0 \leqslant k \leqslant [R]} N_{\bar{\Gamma}}(x_i, k, k+1) N_{\mathcal{P}_i}(x_i, C(R - k + K)) \tag{9-2}$$

pour une certaine constante $D > 1$ qui peut être choisie indépendante de i. De plus, comme l'exposant critique de chaque $\mathcal{P}_i$ est $(n-1)/2$, il existe un réel $M \geqslant 1$, indépendant de i mais dépendant de C, tel que

$$\frac{1}{M} e^{(\frac{n-1}{2} - (C-1))R} \leqslant N_{\mathcal{P}_i}(x_i, R)) \leqslant M e^{(\frac{n-1}{2} + (C-1))R}.$$

D'un côté, cela implique que

$$N_{\mathcal{P}_i}(x_i, C(R - k + K)) \leqslant L N_{\mathcal{P}_i}(x_i, C(R - k)),$$

où $L = M^2 e^{(\frac{n-1}{2} - (C-1))CK}$. D'un autre côté, cela nous donne

$$\begin{aligned}N_{\mathcal{P}_i}(x_i, CR) &\leqslant M e^{(\frac{n-1}{2} + (C-1))CR} \\ &= M e^{(\frac{n-1}{2} - (C-1))R} e^{(\frac{n-1}{2} + C + 1)(C-1)R} \\ &\leqslant M^2 e^{(\frac{n-1}{2} + C + 1)(C-1)R} N_{\mathcal{P}_i}(x_i, R).\end{aligned}$$

Avec (9-2), on obtient

$$\begin{aligned}&\mathrm{Vol}_\Omega(\Gamma \cdot H_{\xi_i} \cap B(o, R)) \leqslant \\ &\qquad DLM^2 e^{(\frac{n-1}{2} + C + 1)(C-1)R} \sum_{0 \leqslant k \leqslant [R]} N_{\bar{\Gamma}}(x_i, k, k+1) N_{\mathcal{P}_i}(x_i, R - k).\end{aligned} \tag{9-3}$$

Finalement, remarquons que tout élément $\gamma \in \Gamma$ tel que $d_\Omega(x_i, \gamma x_i) < R$ peut être écrit de façon unique $\gamma = \bar{\gamma}_i p_i$, avec $d_\Omega(x_i, \bar{\gamma}_i x_i) < R$ et $p_i \in \mathcal{P}_i$, de telle façon que

$$d_\Omega(x_i, p_i x_i) + d_\Omega(x_i, \bar{\gamma}_i x_i) \geqslant R.$$

D'où

$$N_\Gamma(x_i, R) \geqslant \sum_{0 \leqslant k \leqslant [R]} N_{\overline{\Gamma}}(x_i, k, k+1) N_{\mathcal{P}_i}(x_i, R-k). \tag{9-4}$$

Les inégalités (9-3) et (9-4) impliquent alors

$$\mathrm{Vol}_\Omega(\Gamma \cdot H_{\xi_i} \cap B(o, R)) \leqslant DLM^2 e^{(\frac{n-1}{2}+C+1)(C-1)R} N_\Gamma(x_i, R);$$

et donc, en mettant tout ensemble

$$\mathrm{Vol}_\Omega(B(o, R)) \leqslant N e^{(\frac{n-1}{2}+C+1)(C-1)R} N_\Gamma(o, R),$$

pour un certain réel $N > 1$. Cela donne

$$h_{\mathrm{vol}} \leqslant \delta_\Gamma + \Big(\frac{n-1}{2} + C + 1\Big)(C-1).$$

Comme C est arbitrairement proche de 1, on obtient $h_{\mathrm{vol}} \leqslant \delta_\Gamma$. □

9B. ***Groupes dont l'action est géométriquement finie sur Ω.*** En fait, par la même démonstration et les résultats de [Crampon et Marquis 2012], on peut obtenir un résultat similaire pour des groupes dont l'action est géométriquement finie sur Ω :

Théorème 9.7. *Soit* Γ *un sous-groupe discret de* $\mathrm{Aut}(\Omega)$ *dont l'action sur* Ω *est géométriquement finie. Alors*

$$\delta_\Gamma = \limsup_{R \to +\infty} \frac{1}{R} \ln \mathrm{Vol}_\Omega(B(o, R) \cap C(\Lambda_\Gamma)),$$

où o *est un point quelconque de* Ω.

Nous avons préféré présenter la démonstration dans le cas du volume fini que nous considérions déjà assez technique pour ne pas la surcharger. Les seuls points à vérifier pour étendre le résultat sont les trois lemmes 9.4, 9.5 et 9.6 ; le reste se lit tel quel en pensant seulement à considérer l'intersection avec $C(\Lambda_\Gamma)$.

Pour le lemme 9.6, il suffit de se souvenir que, d'après [Crampon et Marquis 2012], l'espace $(C(\Lambda_\Gamma), d_\Omega)$ est Gromov-hyperbolique.

Pour les deux autres, il nous faut rappeler quelques éléments de [Crampon et Marquis 2012], dont on conseille de consulter la partie 7.

Tous les groupes paraboliques apparaissant dans une action géométriquement finie d'un groupe Γ sur Ω sont conjugués dans $\mathrm{SL}_{n+1}(\mathbb{R})$ à des sous-groupes paraboliques de $\mathrm{SO}_{n,1}(\mathbb{R})$. En particulier, un sous-groupe parabolique $\mathcal{P}$ de Γ est virtuellement isomorphe à $\mathbb{Z}^d$ pour un certain $1 \leqslant d \leqslant n-1$; d est le rang de $\mathcal{P}$. De plus, si $p \in \partial\Omega$ est le point fixe de $\mathcal{P}$, il existe une coupe Ω_p de dimension $d+1$ de Ω, contenant p dans son adhérence, c'est-à-dire l'intersection de Ω avec un sous-espace projectif de dimension $d+1$, qui est préservée par $\mathcal{P}$; ainsi, $\mathcal{P}$ apparaît comme un sous-groupe parabolique de rang maximal de $\mathrm{Aut}(\Omega_p)$. On obtient la généralisation suivante du lemme 9.4 :

Lemme 9.8. *L'exposant critique d'un sous-groupe parabolique $\mathcal{P}$ de* $\mathrm{Aut}(\Omega)$, *de rang $d \leqslant n-1$, est $\delta_{\mathcal{P}} = d/2$ et les séries de Poincaré de $\mathcal{P}$ divergent en $\delta_{\mathcal{P}}$:*

$$\sum_{\gamma\in\mathcal{P}} e^{-\delta_{\mathcal{P}} d_\Omega(x,\gamma x)} = +\infty \quad \textit{pour tout } x \in \Omega.$$

Démonstration. Comme rien ne dépend pas du point base considéré, on peut le prendre dans le convexe Ω_p pour se ramener au cas original du lemme 9.4. □

Le lemme 9.5 s'étendrait immédiatement sous les conclusions du corollaire 4.2. En général, on peut étendre au moins la majoration, qui est le point que l'on utilise dans la démonstration du théorème :

Lemme 9.9. *Soit $C > 1$ arbitrairement proche de 1 et $\mathcal{P}$ un sous-groupe parabolique maximal de* $\mathrm{Aut}(\Omega)$ *fixant $p \in \partial\Omega$. Il existe une horoboule H_C basée en p d'horosphère au bord $\mathcal{H}_C$ et une constante $D > 1$ telle que*

$$\mathrm{Vol}_\Omega(B(x,R)\cap H\cap C(\Lambda_\Gamma)) \leqslant D N_{\mathcal{P}}(x, CR), \quad x \in \mathcal{H}_C \cap C(\Lambda_\Gamma),\ R \geqslant 1.$$

On aura besoin du lemme 9.12 ci-dessous. Il se déduit du fait suivant :

Lemme 9.10 [Colbois et Vernicos 2006]. *Pour tout $m \geq 1$ et $R > 0$, il existe deux constantes $v_m(R), V_m(R) > 0$ telles que, pour tout ouvert proprement convexe Ω de $\mathbb{P}^m$ et $x \in \Omega$,*

$$v_m(R) \leqslant \mathrm{Vol}_\Omega(B(x,R)) \leqslant V_m(R).$$

Remarque 9.11. Le lemme précédent est contenu dans [Colbois et Vernicos 2006, théorème 12], qui donne en plus des bornes explicites, dont la dépendance en R est exponentielle. Pour l'énoncé présenté ici, on peut donner une démonstration qualitative, basé sur le théorème de compacité de Benzécri [1960] ; ce théorème affirme que l'action de $\mathrm{SL}_{n+1}(\mathbb{R})$ sur l'ensemble des couples (Ω, x), où Ω est un ouvert proprement convexe de $\mathbb{P}^n$ et x un point de Ω, est propre et cocompacte. On peut trouver cette démonstration dans [Crampon et Marquis 2012].

Lemme 9.12. *Soient $r > 0$ et $1 \leqslant d \leqslant n$. Il existe deux constantes $M, m > 0$, dépendant seulement de r, n et d, telles que, pour tout ouvert proprement convexe Ω de $\mathbb{P}^n$, tout sous-espace $\mathbb{P}^d$ de dimension d intersectant Ω et toute partie A compacte* *de l'ouvert proprement convexe $\Omega_d = \mathbb{P}^d \cap \Omega$ de $\mathbb{P}^d$, le volume du r-voisinage $V_r(A)$ de A dans Ω est comparable au volume du r-voisinage de A dans Ω_d :*

$$m \leqslant \frac{\mathrm{Vol}_\Omega(V_r(A))}{\mathrm{Vol}_{\Omega_d}(V_r(A)\cap\Omega_d)} \leqslant M.$$

Démonstration. Considérons un ensemble $\{x_i\}_{1\leqslant i\leqslant N}$ $2r$-séparé maximal de A : deux points x_i, x_j, $i \neq j$, sont à distance au moins $2r$ et il est impossible de rajouter un point à l'ensemble qui satisfasse à cette propriété. En particulier, les boules de rayon r centrées aux points x_i sont disjointes et incluses dans $V_r(A)$, alors que $V_r(A)$ est recouvert par les boules de rayon $4r$ centrées aux points x_i. On a ainsi, en utilisant le lemme 9.10,

$$\begin{aligned} N v_n(r) &\leqslant \sum_{i=1}^N \mathrm{Vol}_\Omega(B(x_i,r)) \leqslant \mathrm{Vol}_\Omega(V_r(A)) \\ &\leqslant \sum_{i=1}^N \mathrm{Vol}_\Omega(B(x_i,4r)) \leqslant N V_n(4r); \end{aligned}$$

de même,

$$N v_d(r) \leqslant \mathrm{Vol}_{\Omega_d}(V_r(A)\cap\Omega_d) \leqslant N V_d(4r).$$

En prenant le quotient, on obtient

$$m := \frac{v_n(r)}{V_d(4r)} \leqslant \frac{\mathrm{Vol}_\Omega(V_r(A))}{\mathrm{Vol}_{\Omega_d}(V_r(A)\cap\Omega_d)} \leqslant \frac{V_n(4r)}{v_d(r)} =: M. \qquad \square$$

Démonstration du lemme 9.9. Notons d le rang de $\mathscr{P}$ et $\Omega_p = \Omega\cap\mathbb{P}^{d+1}$ une coupe de Ω de dimension $d+1$, contenant p dans son adhérence, et préservée par $\mathscr{P}$. Pour toute horoboule H basée en p, l'intersection $C(\Lambda_\Gamma)\cap H$ est dans un d_Ω-voisinage de taille $r = r(H) \geqslant 0$ finie de $\Omega_p \cap C(\Lambda_\Gamma)$. Cela est simplement dû au fait que $\mathscr{P}$ agit de façon cocompacte sur $C(\Lambda_\Gamma)\cap\mathscr{H}$, où $\mathscr{H}$ est l'horosphère au bord de H. De plus, comme le bord de Ω est $\mathscr{C}^1$ en p, on peut, en considérant une horoboule plus petite, prendre r aussi petit que l'on veut.

On fixe l'horoboule H de telle façon que l'intersection $H\cap\Omega_p$, qui est une horoboule de Ω_p basée en p, satisfasse au lemme 9.5 pour Ω_p, avec la constante C. On fixe aussi $r > 0$ tel que $C(\Lambda_\Gamma)\cap H$ est dans un d_Ω-voisinage de taille r de $\Omega_p\cap C(\Lambda_\Gamma)$.

Si x est un point de $\mathscr{H}\cap C(\Lambda_\Gamma)$, il existe un point x' de $\mathscr{H}\cap\Omega_p$ à distance moins que r de x ; on a alors $B(x, R) \subset B(x', R+r)$ et

$$\mathrm{Vol}_\Omega(B(x, R)\cap H\cap C(\Lambda_\Gamma)) \leqslant \mathrm{Vol}_\Omega(B(x', R+r)\cap H\cap C(\Lambda_\Gamma)).$$

Maintenant, si $x' \in \mathcal{H} \cap \Omega_p$, l'ensemble $B(x', R+r) \cap H \cap C(\Lambda_\Gamma)$ est inclus dans le r-voisinage de $B_{\Omega_p}(x', R) \cap H$ dans Ω. On a donc, d'après le lemme 9.12,

$$\mathrm{Vol}_\Omega(B(x', R+r) \cap H \cap C(\Lambda_\Gamma)) \leqslant M \mathrm{Vol}_{\Omega_p}(A_r). \tag{9-5}$$

où A_r est le r-voisinage de $B_{\Omega_p}(x', R) \cap H$ dans Ω_p.

La partie de A_r qui est dans H correspond précisément à l'intersection

$$B_{\Omega_p}(x', R+r) \cap H$$

à laquelle on peut appliquer le lemme 9.5, qui donne :

$$\mathrm{Vol}_{\Omega_p}(B_{\Omega_p}(x', R+r) \cap H) \leqslant D N_{\mathcal{P}}(x, C(R+r)).$$

Le reste de A_r est dans un voisinage de taille $r' = \max\{r, \mathrm{diam}\}$ de l'ensemble fini de points $\mathcal{P} \cdot x \cap B(x, R+r)$, où diam est le diamètre d'un domaine fondamental compact pour l'action de $\mathcal{P}$ sur $\mathcal{H} \cap \Omega_p \cap C(\Lambda_\Gamma)$ (tout cela pour la distance d_{Ω_p}). Le volume de cette partie est donc majoré par

$$N_{\mathcal{P}}(x', R+r) V_d(r'),$$

où $V_d(r')$ est la constante donnée par le lemme 9.10. Au final, on obtient

$$\begin{aligned}\mathrm{Vol}_{\Omega_p}(A_r) &\leqslant N_{\mathcal{P}}(x', R+r) V_d(r') + D N_{\mathcal{P}}(x', C(R+r)) \\ &\leqslant D' N_{\mathcal{P}}(x', C(R+r)),\end{aligned}$$

pour une certaine constante D'.

En regroupant le tout, on arrive à

$$\begin{aligned}\mathrm{Vol}_\Omega(B(x, R) \cap H \cap C(\Lambda_\Gamma)) &\leqslant D' N_{\mathcal{P}}(x', C(R+r)) \\ &\leqslant D' N_{\mathcal{P}}(x, C(R+r)+r).\end{aligned}$$

Cela donne le résultat, la condition $R \geqslant 1$ étant due à la présence du r. □

Le théorème 9.7 soulève la question suivante, sur laquelle nous terminerons ce texte. C'est une question qu'on peut poser de façon très générale mais une réponse dans des cas particuliers serait déjà intéressante.

Question 9.13. Soit Ω un ouvert proprement convexe de $\mathbb{P}^n$ (strictement convexe et à bord $\mathcal{C}^1$). A t-on $\delta_\Gamma = h_{\mathrm{vol}}(C(\Lambda_\Gamma), d_{C(\Lambda_\Gamma)})$ pour tout sous-groupe discret Γ de $\mathrm{Aut}(\Omega)$?

Remerciements

Nous tenons à remercier Françoise Dal'bo pour son intérêt, notamment pour la section 9, ainsi que François Ledrappier qui nous a indiqué l'article [Hamenstädt 1994]. Nous remercions également le rapporteur anonyme pour sa relecture minutieuse et ses remarques pertinentes.

Mickaël Crampon est financé par le programme FONDECYT N° 3120071 de la CONICYT (Chile).

Bibliographie

[Benoist 2000a] Y. Benoist, "Automorphismes des cônes convexes", *Invent. Math.* **141**:1 (2000), 149–193. MR 2001f:22034 Zbl 0957.22008

[Benoist 2000b] Y. Benoist, "Propriétés asymptotiques des groupes linéaires, II", pp. 33–48 dans *Analysis on homogeneous spaces and representation theory of Lie groups* (Okayama, Kyoto, 1997), édité par T. Kobayashi et al., Adv. Stud. Pure Math. **26**, Math. Soc. Japan, Tokyo, 2000. MR 2001e:22005 Zbl 0960.22012

[Benoist 2004] Y. Benoist, "Convexes divisibles, I", pp. 339–374 dans *Algebraic groups and arithmetic* (Mumbai, 2001), édité par S. G. Dani et G. Prasad, Tata Inst. Fund. Res., Mumbai, 2004. MR 2005h:37073 Zbl 1084.37026

[Benoist 2006] Y. Benoist, "Convexes divisibles, IV: Structure du bord en dimension 3", *Invent. Math.* **164**:2 (2006), 249–278. MR 2007g:22007 Zbl 1107.22006

[Benzécri 1960] J.-P. Benzécri, "Sur les variétés localement affines et localement projectives", *Bull. Soc. Math. France* **88** (1960), 229–332. MR 23 #A1325 Zbl 0098.35204

[Colbois et Vernicos 2006] B. Colbois et C. Vernicos, "Bas du spectre et delta-hyperbolicité en géométrie de Hilbert plane", *Bull. Soc. Math. France* **134**:3 (2006), 357–381. MR 2007g:53080 Zbl 1117.53034

[Cooper et al. 2011] D. Cooper, D. Long et S. Tillmann, "On convex projective manifolds and cusps", preprint, 2011. arXiv 1109.0585

[Coudene et Schapira 2010] Y. Coudene et B. Schapira, "Generic measures for hyperbolic flows on non-compact spaces", *Israel J. Math.* **179** (2010), 157–172. MR 2012g:37050 Zbl 1229.53078

[Crampon 2009] M. Crampon, "Entropies of strictly convex projective manifolds", *J. Mod. Dyn.* **3**:4 (2009), 511–547. MR 2011g:37079 Zbl 1189.37034

[Crampon 2014] M. Crampon, "Lyapunov exponents in Hilbert geometry", **34**:2 (2014), 501–533. arXiv 1105.6275

[Crampon et Marquis 2012] M. Crampon et L. Marquis, "Finitude géométrique en géométrie de Hilbert", preprint, 2012. À paraître aux *Annales de l'Institut Fourier*. arXiv 1202.5442

[Crampon et Marquis 2013] M. Crampon et L. Marquis, "Un lemme de Kazhdan–Margulis–Zassenhaus pour les géométries de Hilbert", *Ann. Math. Blaise Pascal* **20**:2 (2013), 363–376. MR 3138033 Zbl 1282.22007

[Dal'bo et al. 2000] F. Dal'bo, J.-P. Otal et M. Peigné, "Séries de Poincaré des groupes géométriquement finis", *Israel J. Math.* **118** (2000), 109–124. MR 2001g:37040

[Dal'Bo et al. 2009] F. Dal'Bo, M. Peigné, J.-C. Picaud et A. Sambusetti, "On the growth of nonuniform lattices in pinched negatively curved manifolds", *J. Reine Angew. Math.* **627** (2009), 31–52. MR 2010c:53049 Zbl 1184.53039

[Eberlein 1996] P. B. Eberlein, *Geometry of nonpositively curved manifolds*, University of Chicago Press, 1996. MR 98h:53002 Zbl 0883.53003

[Hamenstädt 1994] U. Hamenstädt, "Anosov flows which are uniformly expanding at periodic points", *Ergodic Theory Dynam. Systems* **14**:2 (1994), 299–304. MR 95c:58132 Zbl 0805.58049

[Katok et Hasselblatt 1995] A. Katok et B. Hasselblatt, *Introduction to the modern theory of dynamical systems*, Encyclopedia of Mathematics and its Applications **54**, Cambridge University Press, 1995. MR 96c:58055 Zbl 0878.58020

Received March 30, 2012.

MICKAËL CRAMPON
UNIVERSIDAD DE SANTIAGO DE CHILE
DEPARTAMENTO DE MATEMÁTICA Y CIENCIA DE LA COMPUTACIÓN
AVENIDA LAS SOPHORAS 173
ESTACIÓN CENTRAL, SANTIAGO DE CHILE
CHILE

mickael.crampon@usach.cl

LUDOVIC MARQUIS
IRMAR
263 AVENUE DU GÉNÉRAL LECLERC, CS 74205
35042 RENNES CÉDEX
FRANCE

ludovic.marquis@univ-rennes1.fr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 268, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.268.371

NONPLANARITY OF UNIT GRAPHS AND CLASSIFICATION OF THE TOROIDAL ONES

A. K. DAS, H. R. MAIMANI, M. R. POURNAKI AND S. YASSEMI

The unit graph of a ring R with nonzero identity is the graph in which the vertex set is R, and two distinct vertices x and y are adjacent if and only if $x + y$ is a unit in R. In this paper, we derive several necessary conditions for the nonplanarity of the unit graphs of finite commutative rings with nonzero identity, and determine, up to isomorphism, all finite commutative rings with nonzero identity whose unit graphs are toroidal.

1. Introduction

Algebraic combinatorics is an area of mathematics which employs methods of abstract algebra in various combinatorial contexts and vice versa. Associating a graph to an algebraic structure is a research subject in this area and has attracted considerable attention. The research in this subject aims at exposing the relationship between algebra and graph theory and at advancing the application of one to the other. In fact, there are three major problems in this area: (1) characterization of the resulting graphs, (2) characterization of the algebraic structures with isomorphic graphs, and (3) realization of the connections between the algebraic structures and the corresponding graphs. Beck [1988] introduced the idea of a zero-divisor graph of a commutative ring R with nonzero identity. He defined $\Gamma_0(R)$ to be the graph in which the vertex set is R, and two distinct vertices x and y are adjacent if and only if $xy = 0$. He was mostly concerned with coloring of $\Gamma_0(R)$. Beck conjectured that $\chi(R) = \omega(R)$, where $\chi(R)$ and $\omega(R)$ denote, respectively, the chromatic number and the clique number of $\Gamma_0(R)$. Such graphs are called *weakly perfect graphs*. This investigation of coloring of a commutative ring was then continued by Anderson and Naseer [1993]. They gave a counterexample for the above conjecture of Beck. Anderson and Livingston [1999] proposed a different

This research was in part supported by a grant from IPM (H. R. Maimani: No. 91050214, M. R. Pournaki: No. 92130115, and S. Yassemi: No. 92130214), by a grant from INSF (M. R. Pournaki), and by a grant from the Center for International Scientific Studies and Collaboration (CISSC), Iran, and Égide – Programme Gundishapur 2012, Hubert Curien Partnership, France (M. R. Pournaki and S. Yassemi: No. 27462PL).
MSC2010: primary 05C75; secondary 13M05.
Keywords: finite commutative ring, unit graph, genus.

method of associating a zero-divisor graph to a commutative ring R, and according to them this gives a better illustration of the zero-divisor structure of the ring. They defined $\Gamma(R)$ to be the graph in which the vertex set consists of all the nonzero zero-divisors of R, and two distinct vertices x and y are adjacent if and only if $xy = 0$. For a survey and recent results concerning zero-divisor graphs, we refer the reader to [Anderson et al. 2011]. In literature, one can find a number of different types of graphs attached to rings or other algebraic structures. For a survey of recent results concerning graphs attached to rings, we refer the reader to [Maimani et al. 2011a].

The present paper deals with what is known as the unit graph of a ring, a notion that generalizes the idea of Grimaldi [1990] who introduced and studied in detail a graph $G(\mathbb{Z}_n)$ in which the vertex set is the ring $\mathbb{Z}_n$ of integers modulo a positive integer n, and two distinct vertices x and y are adjacent if and only if $x + y$ is a unit in $\mathbb{Z}_n$. In general, given an arbitrary ring R with nonzero identity, its *unit graph* $G(R)$ is defined to be the graph in which the vertex set is R, and two distinct vertices x and y are adjacent if and only if $x + y$ is a unit in R. Some of the properties of this graph have been studied in detail in [Ashrafi et al. 2010; Maimani et al. 2010a; 2010b; 2010c; 2011b]. The graphs in Figure 1 are the unit graphs of the rings indicated.

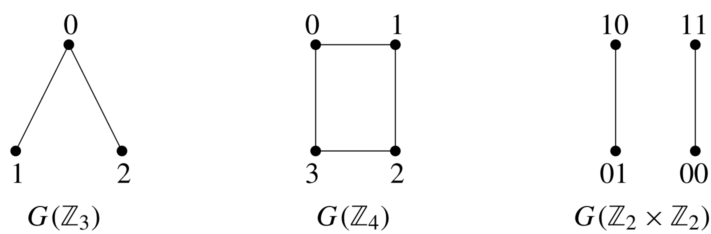


Figure 1. The unit graphs of some specific rings.

It is easy to see that, given any two rings R and S, if $R \cong S$ as rings, then $G(R) \cong G(S)$ as graphs. This point is illustrated in Figure 2 for the unit graphs of two isomorphic rings $\mathbb{Z}_3 \times \mathbb{Z}_2$ and $\mathbb{Z}_6$.

It is also easy to see that if the rings R_1, R_2, S_1 and S_2 are such that $G(R_1) \cong G(R_2)$ and $G(S_1) \cong G(S_2)$, then $G(R_1 \times S_1) \cong G(R_2 \times S_2)$. However, this property does not hold in general for other widely studied graphs associated to rings (for example, the zero-divisor graphs).

In this paper, we derive several necessary conditions for the nonplanarity of the unit graphs of finite commutative rings with nonzero identity; in particular, we show that given any positive integer g, there exists only a finite number of finite commutative rings with nonzero identity whose unit graphs have genus g.

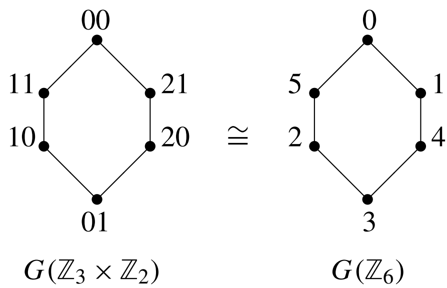


Figure 2. The unit graphs of two isomorphic rings.

Also, in analogy with the results in [Maimani et al. 2012], we determine, up to isomorphism, all finite commutative rings with nonzero identity whose unit graphs are toroidal. It may be recalled here that the *genus* of a graph G, denoted by $\gamma(G)$, is smallest nonnegative integer g such that the graph G can be embedded on the surface obtained by attaching g handles to a sphere. The graphs of genus 0 and 1 are called *planar graphs* and *toroidal graphs* respectively. For unexplained terminology and notations in this paper, we refer the reader to [Chartrand and Oellermann 1993].

2. Some auxiliary results and the related concepts

In this section, we put together certain graph theoretical terminologies and some well-known results which have been used extensively in the forthcoming sections. Note that all graphs considered in this section are finite simple graphs, that is, graphs with finitely many vertices and without loops or multiple edges.

Let x and y be any two vertices in a graph G. Then, x and y are said to be *adjacent* in G if $x \neq y$ and there is an edge $\{x, y\}$ between x and y. A *path* between x and y is a sequence $\{x, x_1\}, \{x_1, x_2\}, \ldots, \{x_n, y\}$ of distinct edges, which is also written as $\{x, x_1, x_2, \ldots, x_n, y\}$, where the vertices $x, x_1, x_2, \ldots, x_n, y$ are all distinct (except, possibly, x and y). A path between x and y is called a *cycle* if $x = y$. The number of edges in a path or a cycle, is called its *length*.

A graph G is said to be *connected* if there is a path between every pair of distinct vertices in G. A *chord* of a cycle in a graph is an edge of the graph which does not lie in the edge set of the cycle but whose endpoints lie in the vertex set of the cycle. A *chordless cycle* of a graph is a cycle without any chord.

A cycle of a graph, embedded on a surface, is called *contractible with respect to the embedding* if it can be contracted continuously on the surface to a point. A cycle of a toroidal graph is said to be *flat* if it is contractible in every torus embedding of the graph. Given a cycle C of a graph G, we write $G - C$ to denote the graph obtained from G by deleting the vertices of C and the edges of the graph incident to the vertices of C.

A graph G is said to be *complete* if there is an edge between every pair of distinct vertices in G. We denote the complete graph with n vertices by K_n. A *bipartite graph* is the one whose vertex set can be partitioned into two disjoint parts in such a way that the two end vertices of every edge lie in different parts. Among the bipartite graphs, the *complete bipartite graph* is the one in which two vertices are adjacent if and only if they lie in different parts. The complete bipartite graph, with parts of size m and n, is denoted by $K_{m,n}$.

A *subdivision of an edge* $\{x, y\}$ in a graph is a path $\{x, x_1, x_2, \dots, x_n, y\}$ obtained by inserting some new vertices $x_1, x_2, \dots, x_n$ into the edge $\{x, y\}$. A *subdivision of a graph* G is the result of some subdivisions of the edges of G. Furthermore, every graph can be considered as a subdivision of itself. A remarkably simple characterization of planar graphs was given by Kuratowski in 1930. Kuratowski's theorem [Chartrand and Oellermann 1993, page 153] says that a graph is planar if and only if it contains no subdivision of $K_{3,3}$ or K_5. As a consequence of Kuratowski's theorem, one has the following result.

Lemma 2.1 [Neufeld and Myrvold 1997, Theorem 2.1]. *If a cycle C of a toroidal graph G is such that $G - C$ is nonplanar, then C is flat in G. Furthermore, if flat C is chordless and $G - C$ is connected, then C is a flat face in any torus embedding of G.*

Given a graph G, we denote its vertex set by $V(G)$ and its edge set by $E(G)$. If G_1 and G_2 are any two graphs, then their *disjoint union*, denoted by $G_1 \sqcup G_2$, is defined to be the graph in which the vertex set is $V(G_1) \sqcup V(G_2)$ and the edge set is $E(G_1) \sqcup E(G_2)$. The following result, which follows from [Battle et al. 1962, Corollary 2], often enables us to reformulate some results which are otherwise true for connected graphs.

Lemma 2.2. *If a graph G is isomorphic to the disjoint union $G_1 \sqcup G_2$ of two graphs G_1 and G_2, then $\gamma(G) = \gamma(G_1) + \gamma(G_2)$.*

If G is a graph and $x \in V(G)$, then the *degree* of x in G is defined as the number of vertices adjacent to x in G, and is denoted by $\deg(x)$. If r is a nonnegative integer such that $\deg(x) = r$ for all $x \in V(G)$, then the graph G is said to be *r-regular*. In general, we write $\delta(G)$ to denote the minimum of the degrees of the vertices of G. In this connection, using Lemma 2.2, one may reformulate [Wickham 2008, Proposition 2.1] as follows.

Lemma 2.3. *If G is a graph (not necessarily connected) having n vertices with $n \geq 3$, then*

$$\delta(G) \leq 6 + \frac{12(\gamma(G) - 1)}{n}.$$

The *girth* of a graph G is the minimum of the lengths of all cycles in G, and is denoted by $\mathrm{gr}(G)$. If G is *acyclic*, that is, if G has no cycles, then we write

$\mathrm{gr}(G) = \infty$. It has been proved in [Archdeacon 1996, Section 2.3] that if G is a connected graph (but not acyclic) having n vertices and m edges, then

$$\gamma(G) \geq \frac{m(k-2)}{2k} - \frac{n}{2} + 1,$$

where $k = \mathrm{gr}(G)$. Therefore, using the facts that in an acyclic graph the total number of edges is less than the total number of vertices, and that the girth of a bipartite graph (which is not acyclic) is at least four, we have, in view of Lemma 2.2, the following result.

Lemma 2.4. *If G is a bipartite graph (not necessarily connected) having n vertices and m edges with $n \geq 3$, then*

$$\gamma(G) \geq \frac{m}{4} - \frac{n}{2} + 1.$$

We conclude the section with two useful results.

Lemma 2.5 [White 1973, Theorem 6–38]. *If $n \geq 3$, then*

$$\gamma(K_n) = \left\lceil \frac{(n-3)(n-4)}{12} \right\rceil.$$

Lemma 2.6 [White 1973, Theorem 6–37]. *If $m, n \geq 2$, then*

$$\gamma(K_{m,n}) = \left\lceil \frac{(m-2)(n-2)}{4} \right\rceil.$$

3. Some necessary conditions for the nonplanarity of unit graphs

In this section, we derive a few necessary conditions for the nonplanarity of the unit graphs of finite commutative rings with nonzero identity. However, we begin with a known result.

Lemma 3.1 [Ashrafi et al. 2010, Proposition 2.4]. *Let R be a finite commutative ring with nonzero identity, and $U(R)$ be the set of all unit elements of R. Let $x \in R$. Then*

$$\deg(x) = \begin{cases} |U(R)| - 1 & \text{if } x \in U(R) \text{ and } 2 \in U(R), \\ |U(R)| & \text{otherwise.} \end{cases}$$

Let us now derive the first necessary condition for the nonplanarity of unit graphs.

Proposition 3.2. *Let R be a finite commutative ring with nonzero identity such that $\gamma(G(R)) = g > 0$. Then either $|R| \leq 12(g-1)$ or $|U(R)| \leq 7$.*

$\lvert U(R)\rvert$	$\lvert Z(R)\rvert$	$\lvert R\rvert$	R
7	1	8	$\mathbb{F}_8$
6	1	7	$\mathbb{Z}_7$
6	3	9	$\mathbb{Z}_9,\ \mathbb{Z}_3[x]/\langle x^2\rangle$
4	1	5	$\mathbb{Z}_5$
4	4	8	$\mathbb{Z}_8,\ \mathbb{Z}_2[x]/\langle x^3\rangle,\ \mathbb{Z}_2[x]/\langle 2x, x^2-2\rangle,$ $\mathbb{Z}_2[x,y]/\langle x,y\rangle^2,\ \mathbb{Z}_4[x]/\langle 2,x\rangle^2$
3	1	4	$\mathbb{F}_4$
2	1	3	$\mathbb{Z}_3$
2	2	4	$\mathbb{Z}_4,\ \mathbb{Z}_2[x]/\langle x^2\rangle$
1	1	2	$\mathbb{Z}_2$

Table 1. Finite commutative local rings with at most 7 units.

Proof. If $|R| > 12(g-1)$, then, using Lemma 2.3, we deduce that

$$\delta(G(R)) \le 6 + \frac{12(g-1)}{|R|} < 6+1 = 7.$$

Since $\delta(G(R))$ is an integer, we have $\delta(G(R)) \le 6$. Hence, it follows from Lemma 3.1 that $|U(R)| \le 7$. □

Let R be a finite commutative ring with nonzero identity. Let $Z(R)$ denote the set of all zero-divisors of R. It is easy to see that $U(R) \sqcup Z(R) = R$ and so $|U(R)| + |Z(R)| = |R|$. The structure theorem for finite commutative rings says that R is isomorphic to a direct product of finite commutative local rings with nonzero identity, and such a product is unique up to the order in which the factors are arranged (see [McDonald 1974]). If R itself is a local ring, then we have the following result which is essentially due to Raghavendran.

Lemma 3.3 [Raghavendran 1969, Theorem 2]. *Let R be a finite commutative local ring with nonzero identity. Then $|R| = p^{nr}$, $|Z(R)| = p^{(n-1)r}$ and $|U(R)| = p^{(n-1)r}(p^r - 1)$ for some prime p and some positive integers n and r.*

Now, Lemma 3.3 together with some well-known results on the structures of small local rings (see, for example, [Corbas and Williams 2000a; 2000b]) enable us to obtain the following result.

Proposition 3.4. *Let R be a finite commutative local ring with nonzero identity such that $|U(R)| \le 7$. Then the possible forms of R are given by Table 1.*

As a consequence of the above result, we derive a necessary condition for the nonplanarity of the unit graphs of finite commutative local rings with nonzero identity.

Corollary 3.5. *Let R be a finite commutative local ring with nonzero identity such that $\gamma(G(R)) = g > 0$. Then $|R| \le \max\{9, 12(g-1)\}$. In particular, the number of finite commutative local rings with nonzero identity such that $\gamma(G(R)) = g > 0$ is finite.*

Proof. If $|R| \le 12(g-1)$, then we are done. Otherwise, by Proposition 3.2, we have $|U(R)| \le 7$. Therefore, by Proposition 3.4, $|R| \le 9$. This completes the proof of the first part. The last part of the corollary is obvious, because, given any positive integer g, the number of rings R with $|R| \le \max\{9, 12(g-1)\}$ is clearly finite. □

In Figure 1, one can see that $G(\mathbb{Z}_2 \times \mathbb{Z}_2)$ is isomorphic to the disjoint union of two copies of $G(\mathbb{Z}_2)$. In fact, it is not difficult to make a more general observation that if S is a finite commutative ring with nonzero identity, then the unit graph $G(\mathbb{Z}_2 \times \mathbb{Z}_2 \times S)$ is isomorphic to the disjoint union of two copies of the unit graph $G(\mathbb{Z}_2 \times S)$. Therefore, in view of Lemma 2.2, we have the following result.

Lemma 3.6. *Let S be a finite commutative ring with nonzero identity. Then we have $\gamma(G(\mathbb{Z}_2 \times \mathbb{Z}_2 \times S)) = 2\gamma(G(\mathbb{Z}_2 \times S))$. In particular, $\gamma(G((\mathbb{Z}_2)^t)) = 0$ for all $t \ge 1$.*

The following result plays an important role in getting rid of all finite commutative rings with nonzero identity whose unit graphs are planar.

Lemma 3.7 [Ashrafi et al. 2010, Theorem 5.14]. *Let R be a finite commutative ring with nonzero identity. Then the unit graph $G(R)$ is planar if and only if R is isomorphic to one of $\mathbb{Z}_5$, $\mathbb{Z}_3 \times \mathbb{Z}_3$ or $(\mathbb{Z}_2)^t \times S$, where $t \ge 0$ and S is one of the rings $\mathbb{Z}_2$, $\mathbb{Z}_3$, $\mathbb{Z}_4$, $\mathbb{F}_4$ and $\left\{\left[\begin{smallmatrix} a & b \\ 0 & a \end{smallmatrix}\right] \mid a, b \in \mathbb{Z}_2\right\} \cong \mathbb{Z}_2[x]/\langle x^2\rangle$.*

We are now in a position to state and prove the main result of this section.

Theorem 3.8. *Let R be a finite commutative ring with nonzero identity such that $\gamma(G(R)) = g > 0$. Then either*

$$|R| \le 12(g-1) \quad \textit{or} \quad R \cong (\mathbb{Z}_2)^t \times S,$$

where $0 \le t \le 1 + \log_2 g$ and S is one of the finite rings given by Table 2.

Proof. Let $|R| > 12(g-1)$. In this case, Proposition 3.2 implies that $|U(R)| \le 7$. Using the structure theorem for finite commutative rings (see the discussion preceding Lemma 3.3) along with Lemma 3.6 and the fact that $g > 0$, we conclude that

$$R \cong (\mathbb{Z}_2)^t \times R_1 \times R_2 \times \cdots \times R_k,$$

where $0 \le t \le 1 + \log_2 g$, $k \ge 1$ and each R_i is a finite commutative local ring with nonzero identity having at least three elements. Now, we have

$$|U(R)| = |U(R_1 \times R_2 \times \cdots \times R_k)| = |U(R_1)| \times |U(R_2)| \times \cdots \times |U(R_k)|.$$

Clearly, $|U(R)| \neq 1$. Since $|U(R)| \leq 7$, we have the following possibilities:

(1) $|U(R)| = p$, where $p = 2, 3, 5$ or 7. In this case, we have $k = 1$ and $|U(R_1)| = p$.

(2) $|U(R)| = 6$. In this case, either we have $k = 1$ and $|U(R_1)| = 6$, or we have $k = 2$, $|U(R_1)| = 2$ and $|U(R_2)| = 3$.

(3) $|U(R)| = 4$. In this case, either we have $k = 1$ and $|U(R_1)| = 4$, or we have $k = 2$, $|U(R_1)| = 2$ and $|U(R_2)| = 2$.

The result now follows from Proposition 3.4 and Lemma 3.7. □

As an immediate corollary, we have the following result.

Corollary 3.9. *Let R be a finite commutative ring with nonzero identity such that $\gamma(G(R)) = g > 0$. Then $|R| \leq 32g$. In particular, given any positive integer g, the number of finite commutative rings with nonzero identity such that $\gamma(G(R)) = g$ is finite.*

Proof. By Theorem 3.8, either $|R| \leq 12(g-1)$ or $R \cong (\mathbb{Z}_2)^t \times S$, where $0 \leq t \leq 1 + \log_2 g$ and S is a ring with $|S| \leq 16$. In the second case, $|R| = 2^t|S| = 2^{t-1}(2|S|) \leq 32g$. Hence, it follows that $|R| \leq \max\{32g, 12(g-1)\} = 32g$. The last part of the corollary is obvious, because, given any positive integer g, the number of rings R with $|R| \leq 32g$ is clearly finite. □

4. Classification of rings with toroidal unit graphs

In this section, we determine, up to isomorphism, all finite commutative rings with nonzero identity whose unit graphs have genus one, that is, whose unit graphs are toroidal graphs.

Let R be a finite commutative ring with nonzero identity such that $\gamma(G(R)) = 1$. Then, by Theorem 3.8, R is isomorphic to either S or $\mathbb{Z}_2 \times S$, where S is one of the finite rings mentioned in Table 2. There are 36 such possibilities for R, among which we single out the ones whose unit graphs have genus 1. For this purpose, the following result is very useful; in fact, in combination with Lemma 2.4, it helps

$\lvert U(R)\rvert$	S
7	$\mathbb{F}_8$
6	$\mathbb{Z}_7$, $\mathbb{Z}_9$, $\mathbb{Z}_3[x]/\langle x^2\rangle$, $\mathbb{Z}_3 \times \mathbb{F}_4$, $\mathbb{Z}_4 \times \mathbb{F}_4$, $\mathbb{Z}_2[x]/\langle x^2\rangle \times \mathbb{F}_4$
4	$\mathbb{Z}_5$ (for $t \neq 0$), $\mathbb{Z}_8$, $\mathbb{Z}_2[x]/\langle x^3\rangle$, $\mathbb{Z}_2[x]/\langle 2x, x^2-2\rangle$, $\mathbb{Z}_2[x,y]/\langle x,y\rangle^2$, $\mathbb{Z}_4[x]/\langle 2,x\rangle^2$, $\mathbb{Z}_3 \times \mathbb{Z}_3$ (for $t \neq 0$), $\mathbb{Z}_3 \times \mathbb{Z}_4$, $\mathbb{Z}_3 \times \mathbb{Z}_2[x]/\langle x^2\rangle$, $\mathbb{Z}_4 \times \mathbb{Z}_4$, $\mathbb{Z}_4 \times \mathbb{Z}_2[x]/\langle x^2\rangle$, $\mathbb{Z}_2[x]/\langle x^2\rangle \times \mathbb{Z}_2[x]/\langle x^2\rangle$

Table 2. Possible seed rings for toroidal unit graphs.

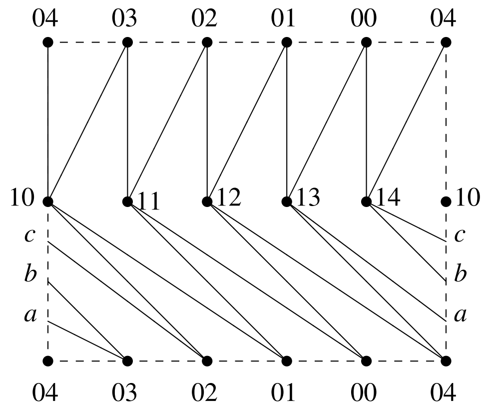


Figure 3. Embedding of the unit graph of $\mathbb{Z}_2 \times \mathbb{Z}_5$ on a torus.

in determining some lower bounds for the genus of the unit graphs of rings of the type $\mathbb{Z}_2 \times S$, where S is a finite commutative ring with nonzero identity.

Lemma 4.1 [Ashrafi et al. 2010, Theorem 3.5]. *Let R be a finite commutative ring with nonzero identity, and $\mathfrak{m}$ be a maximal ideal of R such that $R/\mathfrak{m} \cong \mathbb{Z}_2$. Then the unit graph $G(R)$ is a bipartite graph. Moreover, $G(R)$ is a complete bipartite graph if and only if R is a local ring.*

Let us now start the process of classification by looking at some toroidal unit graphs.

Proposition 4.2. $\gamma(G(\mathbb{Z}_2 \times \mathbb{Z}_5)) = 1$.

Proof. By Lemma 3.7, we have $\gamma(G(\mathbb{Z}_2 \times \mathbb{Z}_5)) \geq 1$. But Figure 3 gives an embedding of the unit graph $G(\mathbb{Z}_2 \times \mathbb{Z}_5)$ on a torus, and so $\gamma(G(\mathbb{Z}_2 \times \mathbb{Z}_5)) = 1$. □

Proposition 4.3. $\gamma(G(\mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_3)) = 1$.

Proof. By Lemma 3.7, we have $\gamma(G(\mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_3)) \geq 1$. But Figure 4 gives an embedding of $G(\mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_3)$ on a torus, and so $\gamma(G(\mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_3)) = 1$. □

Proposition 4.4. $\gamma(G(\mathbb{Z}_3 \times \mathbb{Z}_4)) = \gamma(G(\mathbb{Z}_3 \times \mathbb{Z}_2[x]/\langle x^2 \rangle)) = 1$.

Proof. By Lemma 3.7, we have $\gamma(G(\mathbb{Z}_3 \times \mathbb{Z}_4)) \geq 1$. But Figure 5 gives an embedding of the unit graph $G(\mathbb{Z}_3 \times \mathbb{Z}_4)$ on a torus, and so $\gamma(G(\mathbb{Z}_3 \times \mathbb{Z}_4)) = 1$. On the other hand, since the unit graph $G(\mathbb{Z}_4)$ is isomorphic to the unit graph $G(\mathbb{Z}_2[x]/\langle x^2 \rangle)$, we have

$$G(\mathbb{Z}_3 \times \mathbb{Z}_4) \cong G\left(\mathbb{Z}_3 \times \frac{\mathbb{Z}_2[x]}{\langle x^2 \rangle}\right).$$

This completes the proof. □

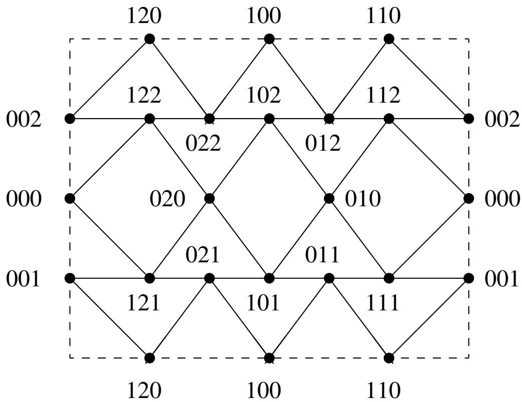


Figure 4. Embedding of the unit graph of $\mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_3$ on a torus.

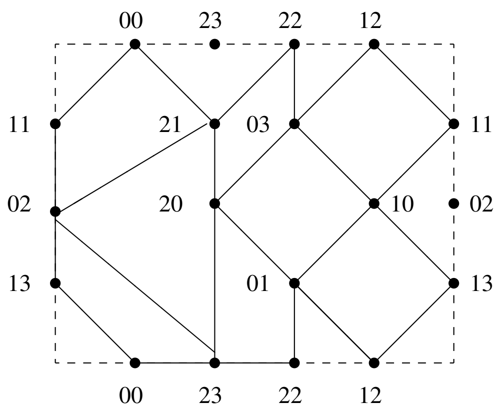


Figure 5. Embedding of the unit graph of $\mathbb{Z}_3 \times \mathbb{Z}_4$ on a torus.

Proposition 4.5. *If S is one of the rings*

$$\mathbb{Z}_7, \quad \mathbb{Z}_8, \quad \frac{\mathbb{Z}_2[x]}{\langle x^3 \rangle}, \quad \frac{\mathbb{Z}_2[x]}{\langle 2x, x^2-2 \rangle}, \quad \frac{\mathbb{Z}_2[x, y]}{\langle x, y \rangle^2} \quad \textit{and} \quad \frac{\mathbb{Z}_4[x]}{\langle 2, x \rangle^2},$$

then $\gamma(G(S)) = 1$.

Proof. Note that the unit graph $G(\mathbb{Z}_7)$ can be regarded as a subgraph of K_7, and so, by Lemmas 2.5 and 3.7, we have $\gamma(G(\mathbb{Z}_7)) = 1$. On the other hand, each of the remaining rings is a local ring with 8 elements of which exactly 4 are zero-divisors, and so it follows from Lemma 4.1 that the associated unit graph of each of these rings is a complete bipartite graph, namely, $K_{4,4}$. The proof is now completed by Lemma 2.6. □

Next we look at some unit graphs which have genus more than 1.

Proposition 4.6. $\gamma(G(\mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_4)) = \gamma(G(\mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_2[x]/\langle x^2 \rangle)) = 2.$

Proof. It is not difficult to see that $G(\mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_4)$ is isomorphic to the disjoint union of two copies of $G(\mathbb{Z}_3 \times \mathbb{Z}_4)$. Therefore, by Lemma 2.2 and Proposition 4.4, we have $\gamma(G(\mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_4)) = 2$. On the other hand, since the unit graph $G(\mathbb{Z}_4)$ is isomorphic to the unit graph $G(\mathbb{Z}_2[x]/\langle x^2\rangle)$, we have

$$G(\mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_4) \cong G\left(\mathbb{Z}_2 \times \mathbb{Z}_3 \times \frac{\mathbb{Z}_2[x]}{\langle x^2\rangle}\right).$$

This completes the proof. □

Proposition 4.7. *If S is one of the rings*

$$\mathbb{Z}_8, \quad \frac{\mathbb{Z}_2[x]}{\langle x^3\rangle}, \quad \frac{\mathbb{Z}_2[x]}{\langle 2x, x^2-2\rangle}, \quad \frac{\mathbb{Z}_2[x,y]}{\langle x,y\rangle^2} \quad \textit{and} \quad \frac{\mathbb{Z}_4[x]}{\langle 2,x\rangle^2},$$

then $\gamma(G(\mathbb{Z}_2 \times S)) = 2$. On the other hand, $\gamma(G(\mathbb{Z}_2 \times \mathbb{Z}_7)) \geq 5$.

Proof. In view of the proof of Proposition 4.5, one may note that, for each of the given choices of S, the unit graph $G(\mathbb{Z}_2 \times S)$ is isomorphic to the disjoint union of two copies of $K_{4,4}$. Hence, the first part follows from Lemmas 2.2 and 2.6.

For the second part, note that the unit graph $G(\mathbb{Z}_2 \times \mathbb{Z}_7)$ is 6-regular with 14 vertices. Also, by Lemma 4.1, it is bipartite, and so it has 42 edges. Therefore, by Lemma 2.4, we have $\gamma(G(\mathbb{Z}_2 \times \mathbb{Z}_7)) \geq 5$. This completes the proof. □

Let us now recall that a subgraph H of a graph G is called a *spanning subgraph* if they have the same sets of vertices. A 1-regular spanning subgraph H of a graph G is called a *perfect matching* of G. Given a graph G with a subgraph H, we write $G \setminus H$ to denote the subgraph of G in which the vertex set is $V(G)$ and the edge set is $E(G) \setminus E(H)$.

Proposition 4.8. $\gamma(G(\mathbb{F}_8)) = 2$ *and* $\gamma(G(\mathbb{Z}_2 \times \mathbb{F}_8)) \geq 7$.

Proof. The unit graph $G(\mathbb{F}_8)$ is isomorphic to K_8, and so, by Lemma 2.5, we have $\gamma(G(\mathbb{F}_8)) = 2$. On the other hand, the unit graph $G(\mathbb{Z}_2 \times \mathbb{F}_8)$ has 16 vertices, and, by Lemma 4.1, it is bipartite. In fact, this graph is isomorphic to the graph $K_{8,8} \setminus M$, where M is a perfect matching of $K_{8,8}$, and so it has 56 edges. Therefore, by Lemma 2.4, we have $\gamma(G(\mathbb{Z}_2 \times \mathbb{F}_8)) \geq 7$. □

Proposition 4.9. $\gamma(G(\mathbb{Z}_3 \times \mathbb{F}_4)) \geq 2$ *and* $\gamma(G(\mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{F}_4)) \geq 7$.

Proof. Note that the unit graph $G := G(\mathbb{Z}_3 \times \mathbb{F}_4)$ is the union of two planar subgraphs plotted in Figure 6 which intersect at exactly four vertices, namely, $00, 01, 0a$ and $0a^2$ (indicated by bigger bullets). It is easy to see that the graph G has 12 vertices and 36 edges. Moreover, it is a 6-regular graph. Also, it is easy to see that it is not a planar graph, as it contains a subdivision of K_5, namely, $\{20, 21, 2a, 2a^2, 0a, 11, 0a^2, 20\}$.

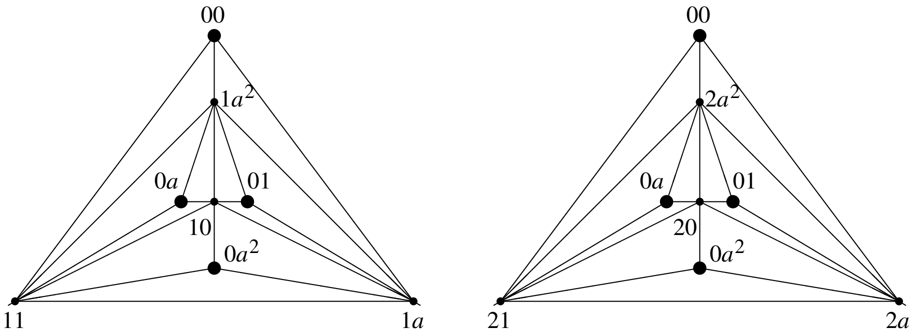


Figure 6. Two planar subgraphs of the unit graph of $\mathbb{Z}_3 \times \mathbb{F}_4$.

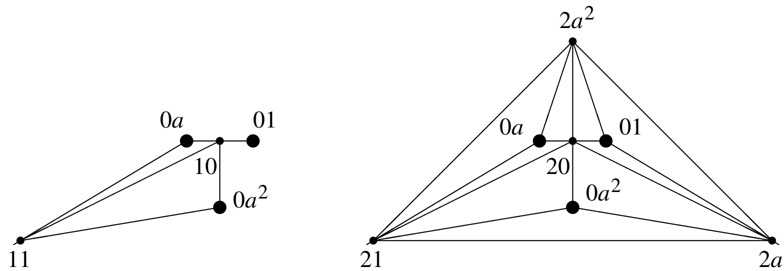


Figure 7. The graph $G(\mathbb{Z}_3 \times \mathbb{F}_4) - \{00, 1a^2, 1a, 00\}$.

Let us now assume that G is toroidal, that is, $\gamma(G) = 1$. Then, by Euler's formula, G has $36 - 12 = 24$ faces. Also, note that G is symmetrical in nature. We use Lemma 2.1 to show that every 3-cycle C in G is a face, and arrive at a contradiction.

First, let us consider a 3-cycle having empty interior. By symmetry, it is enough to take $C = \{00, 1a^2, 1a, 00\}$. Then $G - C$ is given as indicated in Figure 7. Clearly, $G - C$ is nonplanar, as it contains a subdivision of K_5, namely, $\{20, 21, 2a, 2a^2, 0a, 11, 0a^2, 20\}$. Moreover, C is chordless and $G - C$ is connected. Therefore, by Lemma 2.1, C is a face. Note that there are 24 such 3-cycles in G.

Next, we consider the 3-cycle $C = \{1a^2, 10, 1a, 1a^2\}$. Then $G - C$ is given as indicated in Figure 8. Again, it is clear that $G - C$ is nonplanar, as it contains a subdivision of K_5, namely, $\{20, 21, 2a, 2a^2, 0a, 11, 0a^2, 20\}$. Moreover, C is chordless and $G - C$ is connected. Therefore, by Lemma 2.1, C is a face.

Since we have already found 25 faces, our assumption that G is toroidal is wrong. Hence, we conclude that $\gamma(G(\mathbb{Z}_3 \times \mathbb{F}_4)) \geq 2$.

For the second part, note that the unit graph $G(\mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{F}_4)$ is 6-regular with 24 vertices. Also, by Lemma 4.1, it is bipartite, and so it has 72 edges. Therefore, by Lemma 2.4, we have $\gamma(G(\mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{F}_4)) \geq 7$. This completes the proof. □

Arguing in the same manner as above, we also have the following result.

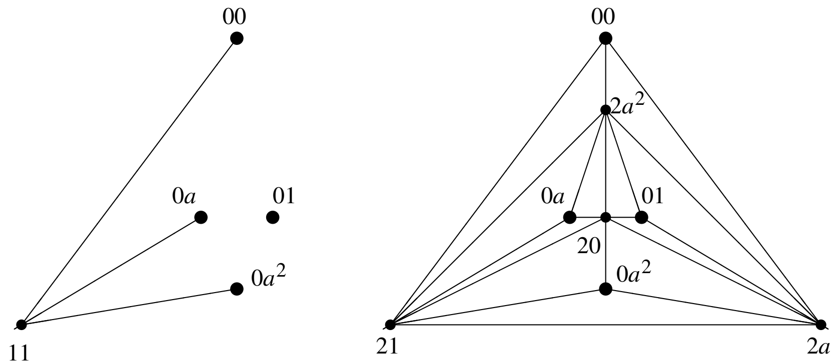


Figure 8. The graph $G(\mathbb{Z}_3 \times \mathbb{F}_4) - \{1a^2, 10, 1a, 1a^2\}$.

Proposition 4.10. *If S is one of the rings $\mathbb{Z}_9$ and $\mathbb{Z}_3[x]/\langle x^2\rangle$, then $\gamma(G(S)) \geq 2$ and $\gamma(G(\mathbb{Z}_2 \times S)) \geq 6$.*

Proof. Note that the unit graphs of the rings $\mathbb{Z}_9$ and $\mathbb{Z}_3[x]/\langle x^2\rangle$ are isomorphic. Therefore, it is enough to prove the result only for $\mathbb{Z}_9$. It is easy to see, from Figure 9, that the unit graph $G(\mathbb{Z}_9)$ is a nonplanar graph with 9 vertices and 24 edges. Moreover, the number of 3-cycles in it is 20. If C is one such 3-cycle, then it is easy to see that C is chordless, and $G(\mathbb{Z}_9) - C$ is connected and nonplanar. In fact, $G(\mathbb{Z}_9) - C$ is either a subdivision of K_5 or has a subgraph isomorphic to $K_{3,3}$, depending on whether the 3-cycle C contains one or none of the vertices 0, 3 and 6. Therefore, if $G(\mathbb{Z}_9)$ is toroidal, then it follows from Lemma 2.1 that every 3-cycle in $G(\mathbb{Z}_9)$ is a face, and so $G(\mathbb{Z}_9)$ has at least 20 faces, whereas from Euler's formula it follows that $G(\mathbb{Z}_9)$ has 15 faces. Hence, we have $\gamma(G(\mathbb{Z}_9)) \geq 2$.

For the second part, note that the unit graph $G(\mathbb{Z}_2 \times \mathbb{Z}_9)$ is 6-regular with 18 vertices. Also, by Lemma 4.1, it is bipartite, and so it has 54 edges. Therefore, by Lemma 2.4, we have $\gamma(G(\mathbb{Z}_2 \times \mathbb{Z}_9)) \geq 6$. This completes the proof. □

Proposition 4.11. *If S is one of the rings*

$$\mathbb{Z}_4 \times \mathbb{Z}_4, \quad \mathbb{Z}_4 \times \frac{\mathbb{Z}_2[x]}{\langle x^2\rangle} \quad \textit{and} \quad \frac{\mathbb{Z}_2[x]}{\langle x^2\rangle} \times \frac{\mathbb{Z}_2[x]}{\langle x^2\rangle},$$

then $\gamma(G(S)) = 2$ and $\gamma(G(\mathbb{Z}_2 \times S)) = 4$.

Proof. Consider the following two subsets of the vertex set of $G(\mathbb{Z}_4 \times \mathbb{Z}_4)$:

$$V_1 = \{(0,0), (0,2), (2,0), (2,2), (1,1), (1,3), (3,1), (3,3)\},$$
$$V_2 = \{(0,1), (0,3), (2,1), (2,3), (1,0), (1,2), (3,0), (3,2)\}.$$

It is easy to see that the two subgraphs $\langle V_1\rangle$ and $\langle V_2\rangle$ are disjoint, and their union is $G(\mathbb{Z}_4 \times \mathbb{Z}_4)$. Moreover, $\langle V_1\rangle \cong \langle V_2\rangle \cong K_{4,4}$. Therefore, it follows from Lemma 2.2 and Lemma 2.6 that $\gamma(G(\mathbb{Z}_4 \times \mathbb{Z}_4)) = 2$. Also, it is easy to see that $G(\mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{Z}_4)$

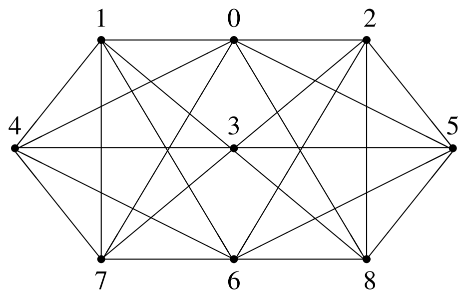


Figure 9. The unit graph of $\mathbb{Z}_9$.

is isomorphic to the disjoint union of two copies of $G(\mathbb{Z}_4 \times \mathbb{Z}_4)$. Therefore, in view of Lemma 2.2, it follows that $\gamma(G(\mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{Z}_4)) = 4$. On the other hand, since the unit graph $G(\mathbb{Z}_4)$ is isomorphic to the unit graph $G(\mathbb{Z}_2[x]/\langle x^2\rangle)$, we have

$$G(\mathbb{Z}_4 \times \mathbb{Z}_4) \cong G\left(\mathbb{Z}_4 \times \frac{\mathbb{Z}_2[x]}{\langle x^2\rangle}\right) \cong G\left(\frac{\mathbb{Z}_2[x]}{\langle x^2\rangle} \times \frac{\mathbb{Z}_2[x]}{\langle x^2\rangle}\right)$$

and

$$G(\mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{Z}_4) \cong G\left(\mathbb{Z}_2 \times \mathbb{Z}_4 \times \frac{\mathbb{Z}_2[x]}{\langle x^2\rangle}\right) \cong G\left(\mathbb{Z}_2 \times \frac{\mathbb{Z}_2[x]}{\langle x^2\rangle} \times \frac{\mathbb{Z}_2[x]}{\langle x^2\rangle}\right).$$

Hence, the result follows. □

Proposition 4.12. *If S is one of the rings $\mathbb{Z}_4 \times \mathbb{F}_4$ and $\mathbb{Z}_2[x]/\langle x^2\rangle \times \mathbb{F}_4$, then $\gamma(G(S)) \geq 2$ and $\gamma(G(\mathbb{Z}_2 \times S)) \geq 9$.*

Proof. Note that $G(\mathbb{Z}_4)$ is a spanning subgraph of $G(\mathbb{F}_4)$. This implies that $G(\mathbb{Z}_4 \times \mathbb{Z}_4)$ is an spanning subgraph of $G(\mathbb{Z}_4 \times \mathbb{F}_4)$. Therefore, by Proposition 4.11, we have $\gamma(G(\mathbb{Z}_4 \times \mathbb{F}_4)) \geq 2$. Also note that the unit graph $G(\mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{F}_4)$ is 6-regular with 32 vertices. Moreover, by Lemma 4.1, it is bipartite, and so it has 96 edges. Therefore, by Lemma 2.4, we have $\gamma(G(\mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{F}_4)) \geq 9$. On the other hand, since the unit graph $G(\mathbb{Z}_4)$ is isomorphic to the unit graph $G(\mathbb{Z}_2[x]/\langle x^2\rangle)$, we have

$$G(\mathbb{Z}_4 \times \mathbb{F}_4) \cong G\left(\frac{\mathbb{Z}_2[x]}{\langle x^2\rangle} \times \mathbb{F}_4\right)$$

and

$$G(\mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{F}_4) \cong G\left(\mathbb{Z}_2 \times \frac{\mathbb{Z}_2[x]}{\langle x^2\rangle} \times \mathbb{F}_4\right).$$

This completes the proof. □

Let us now summarize what we have achieved so far: If S is a finite commutative ring with nonzero identity, then, with $\gamma_S = \gamma(G(S))$ and $\gamma_{2S} = \gamma(G(\mathbb{Z}_2 \times S))$, one has Table 3.

S	γ_S	γ_{2S}
$\mathbb{Z}_5,\ \mathbb{Z}_3\times\mathbb{Z}_3$	0	1
$\mathbb{Z}_3\times\mathbb{Z}_4,\ \mathbb{Z}_3\times\mathbb{Z}_2[x]/\langle x^2\rangle,\ \mathbb{Z}_8,\ \mathbb{Z}_2[x]/\langle x^3\rangle,$ $\mathbb{Z}_2[x]/\langle 2x, x^2-2\rangle,\ \mathbb{Z}_2[x,y]/\langle x,y\rangle^2,\ \mathbb{Z}_4[x]/\langle 2,x\rangle^2$	1	2
$\mathbb{Z}_7$	1	≥ 5
$\mathbb{F}_8$	2	≥ 7
$\mathbb{Z}_3\times\mathbb{F}_4$	≥ 2	≥ 7
$\mathbb{Z}_9,\ \mathbb{Z}_3[x]/\langle x^2\rangle$	≥ 2	≥ 6
$\mathbb{Z}_4\times\mathbb{Z}_4,\ \mathbb{Z}_4\times\mathbb{Z}_2[x]/\langle x^2\rangle,\ \mathbb{Z}_2[x]/\langle x^2\rangle\times\mathbb{Z}_2[x]/\langle x^2\rangle$	2	4
$\mathbb{Z}_4\times\mathbb{F}_4,\ \mathbb{Z}_2[x]/\langle x^2\rangle\times\mathbb{F}_4$	≥ 2	≥ 9

Table 3. Genus of some unit graphs.

Finally, using Table 3 and Theorem 3.8, one easily obtains the following classification theorem.

Theorem 4.13. *Let R be a finite commutative ring with nonzero identity. Then the unit graph $G(R)$ is a toroidal graph if and only if R is isomorphic to one of*

$$\mathbb{Z}_2\times\mathbb{Z}_5,\quad \mathbb{Z}_2\times\mathbb{Z}_3\times\mathbb{Z}_3,\quad \mathbb{Z}_3\times\mathbb{Z}_4,\quad \mathbb{Z}_3\times\frac{\mathbb{Z}_2[x]}{\langle x^2\rangle},\quad \mathbb{Z}_8,$$
$$\frac{\mathbb{Z}_2[x]}{\langle x^3\rangle},\quad \frac{\mathbb{Z}_2[x]}{\langle 2x, x^2-2\rangle},\quad \frac{\mathbb{Z}_2[x,y]}{\langle x,y\rangle^2},\quad \frac{\mathbb{Z}_4[x]}{\langle 2,x\rangle^2}\quad \textit{or}\quad \mathbb{Z}_7.$$

Acknowledgment

The authors are grateful to the anonymous referee for bringing to their notice an inadvertent omission in Theorem 3.8, for making many constructive suggestions, and for supplying some additional references.

References

[Anderson and Livingston 1999] D. F. Anderson and P. S. Livingston, "The zero-divisor graph of a commutative ring", *J. Algebra* **217**:2 (1999), 434–447. MR 2000e:13007 Zbl 0941.05062

[Anderson and Naseer 1993] D. D. Anderson and M. Naseer, "Beck's coloring of a commutative ring", *J. Algebra* **159**:2 (1993), 500–514. MR 94e:13009 Zbl 0798.05067

[Anderson et al. 2011] D. F. Anderson, M. C. Axtell, and J. A. Stickles, "Zero-divisor graphs in commutative rings", pp. 23–45 in *Commutative algebra: Noetherian and non-Noetherian perspectives*, edited by M. Fontana et al., Springer, New York, 2011. MR 2012b:13023 Zbl 1225.13002

[Archdeacon 1996] D. Archdeacon, "Topological graph theory: a survey", *Congr. Numer.* **115** (1996), 5–54. MR 98g:05044 Zbl 0897.05026

[Ashrafi et al. 2010] N. Ashrafi, H. R. Maimani, M. R. Pournaki, and S. Yassemi, "Unit graphs associated with rings", *Comm. Algebra* **38**:8 (2010), 2851–2871. MR 2011j:13010 Zbl 1219.05150

[Battle et al. 1962] J. Battle, F. Harary, Y. Kodama, and J. W. T. Youngs, "Additivity of the genus of a graph", *Bull. Amer. Math. Soc.* **68** (1962), 565–568. MR 27 #5247 Zbl 0142.41501

[Beck 1988] I. Beck, "Coloring of commutative rings", *J. Algebra* **116**:1 (1988), 208–226. MR 89i:13006 Zbl 0654.13001

[Chartrand and Oellermann 1993] G. Chartrand and O. R. Oellermann, *Applied and algorithmic graph theory*, McGraw-Hill, New York, 1993. MR 1211413

[Corbas and Williams 2000a] B. Corbas and G. D. Williams, "Rings of order p^5, I: Nonlocal rings", *J. Algebra* **231**:2 (2000), 677–690. MR 2001h:16020a Zbl 1017.16014

[Corbas and Williams 2000b] B. Corbas and G. D. Williams, "Rings of order p^5, II: Local rings", *J. Algebra* **231**:2 (2000), 691–704. MR 2001h:16020b Zbl 1017.16015

[Grimaldi 1990] R. P. Grimaldi, "Graphs from rings", *Congr. Numer.* **71** (1990), 95–103. MR 90m:05122 Zbl 0747.05091

[Maimani et al. 2010a] H. R. Maimani, M. R. Pournaki, and S. Yassemi, "A class of weakly perfect graphs", *Czechoslovak Math. J.* **60(135)**:4 (2010), 1037–1041. MR 2011h:05104 Zbl 1224.05376

[Maimani et al. 2010b] H. R. Maimani, M. R. Pournaki, and S. Yassemi, "Rings which are generated by their units: a graph theoretical approach", *Elem. Math.* **65**:1 (2010), 17–25. MR 2011d:13039 Zbl 1209.13031

[Maimani et al. 2010c] H. R. Maimani, M. R. Pournaki, and S. Yassemi, "Weakly perfect graphs arising from rings", *Glasg. Math. J.* **52**:3 (2010), 417–425. MR 2011h:05099 Zbl 1243.05085

[Maimani et al. 2011a] H. R. Maimani, M. R. Pournaki, A. Tehranian, and S. Yassemi, "Graphs attached to rings revisited", *Arab. J. Sci. Eng.* **36**:6 (2011), 997–1011. MR 2845527

[Maimani et al. 2011b] H. R. Maimani, M. R. Pournaki, and S. Yassemi, "Necessary and sufficient conditions for unit graphs to be Hamiltonian", *Pacific J. Math.* **249**:2 (2011), 419–429. MR 2012b:05164 Zbl 1214.05074

[Maimani et al. 2012] H. R. Maimani, C. Wickham, and S. Yassemi, "Rings whose total graphs have genus at most one", *Rocky Mountain J. Math.* **42**:5 (2012), 1551–1560. MR 3001816 Zbl 1254.05164

[McDonald 1974] B. R. McDonald, *Finite rings with identity*, Pure and Applied Mathematics **28**, Marcel Dekker, New York, 1974. MR 50 #7245 Zbl 0294.16012

[Neufeld and Myrvold 1997] E. Neufeld and W. Myrvold, "Practical toroidality testing", pp. 574–580 in *Proceedings of the Eighth Annual ACM-SIAM Symposium on Discrete Algorithms* (New Orleans, LA, 1997), edited by M. Saks, ACM, New York, 1997. MR 1447705

[Raghavendran 1969] R. Raghavendran, "Finite associative rings", *Compositio Math.* **21**:2 (1969), 195–229. MR 40 #174 Zbl 0179.33602

[White 1973] A. T. White, *Graphs, groups and surfaces*, North-Holland Mathematics Studies **8**, North-Holland, Amsterdam, 1973. MR 49 #4783 Zbl 0268.05102

[Wickham 2008] C. Wickham, "Classification of rings with genus one zero-divisor graphs", *Comm. Algebra* **36**:2 (2008), 325–345. MR 2009f:13010 Zbl 1137.13015

A. K. Das
Department of Mathematics
North-Eastern Hill University
Permanent Campus
Shillong-793022
Meghalaya
India

akdasnehu@gmail.com

H. R. Maimani
Mathematics Section
Department of Basic Sciences
Shahid Rajaee Teacher Training University
P.O. Box 16785-163

and

School of Mathematics
Institute for Research in Fundamental Sciences (IPM)
P.O. Box 19395-5746
Tehran
Iran

maimani@ipm.ir

M. R. Pournaki
Department of Mathematical Sciences
Sharif University of Technology
P.O. Box 11155-9415

and

School of Mathematics
Institute for Research in Fundamental Sciences (IPM)
P.O. Box 19395-5746
Tehran
Iran

pournaki@ipm.ir

S. Yassemi
School of Mathematics, Statistics and Computer Science
College of Science
University of Tehran

and

School of Mathematics
Institute for Research in Fundamental Sciences (IPM)
P.O. Box 19395-5746
Tehran
Iran

yassemi@ipm.ir

PACIFIC JOURNAL OF MATHEMATICS
Vol. 268, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.268.389

DISCRETE SEMICLASSICAL ORTHOGONAL POLYNOMIALS OF CLASS ONE

DIEGO DOMINICI AND FRANCISCO MARCELLÁN

We study discrete semiclassical orthogonal polynomials of class $s = 1$. By considering particular solutions of the Pearson equation, we obtain five canonical families of such polynomials. We also consider limit relations between these and other families of orthogonal polynomials.

1. Introduction

Discrete orthogonal polynomials with respect to uniform lattices have attracted the interest of researchers from many points of view [Nikiforov et al. 1985]. A first approach comes from the discretization of hypergeometric second-order linear differential equations and thus the classical discrete orthogonal polynomials (Charlier, Krawtchouk, Meixner, Hahn) appear in a natural way. As a consequence of the symmetrization problem for the above second-order difference equations, one can deduce that such polynomials are orthogonal with respect to (discrete) measures. This yields the so-called Pearson equation that the measure satisfies.

In the last twenty years, new families of discrete orthogonal polynomials have been considered in the literature, taking into account the so-called canonical spectral transformations of the orthogonality measure. Under a Uvarov transformation, mass points are added to the discrete measure; sequences of orthogonal polynomials with respect to the new measure have been studied extensively in this case (see [Chihara 1985; Álvarez and Marcellán 1995a; Álvarez et al. 1995], among others). Under a Christoffel transformation, the discrete measure is multiplied by a polynomial; a few results are available in this case [Ronveaux and Salto 2001].

From a structural point of view, some effort has been made to translate to the discrete case the well-known theory of semiclassical orthogonal polynomials (see [Maroni 1991]). In particular, characterizations of such polynomials in terms of structure relations of the first and second kind, as well as discrete holonomic equations (second-order linear difference equations with polynomial coefficients of fixed degree and where the degree of the polynomial appears as a parameter) were

MSC2010: primary 33C47; secondary 34M55, 33E17, 42C05.
Keywords: discrete orthogonal polynomials, Pearson equation, discrete semiclassical polynomials, Laguerre–Freud equations, Painlevé equations.

given in [Marcellán and Salto 1998]. Linear spectral perturbations of semiclassical linear functionals have been studied in the Uvarov case [Godoy et al. 1997].

On the other hand, we must point out that the linear canonical spectral transformations (Christoffel, Uvarov, Geronimus) of classical discrete orthogonal polynomials yield discrete semiclassical orthogonal polynomials. But, as a first step, the problem of classification of discrete semiclassical linear functional of class one remains open. Symmetric discrete semiclassical linear functionals of class one have been described in [Maroni and Mejri 2008]. A classification of D-semiclassical linear functional of class one was given in [Belmehdi 1992] and of those of class two in [Marcellán et al. 2012].

This article provides a constructive method for finding D_w-semiclassical orthogonal polynomials, based on the Pearson equation satisfied by the corresponding linear functional. We will focus our attention on the classification of D_1-semiclassical linear functionals of class $s = 1$. In such a way, new families of linear functionals appear. Notice that an alternative method is based on the Laguerre–Freud equations satisfied by the coefficients of the three-term recurrence relations associated with these orthogonal polynomials. Their complexity increases with the class of the linear functional and the solution is cumbersome. Basic references concerning this approach are [Foupouagnigni et al. 1998] as well as [Maroni and Mejri 2008].

The structure of the article is as follows: Section 2 deals with the basic definitions and the theoretical background we will need in the sequel. In Section 3 we describe the D_1-classical linear functionals as D_1-semiclassical of class $s = 0$. The fact that most of the semiclassical linear functionals of class $s = 1$ are related to the class $s = 0$ will prove to be very useful later on. Indeed, in Section 4, a classification of such semiclassical linear functionals is given. Some of them are not known in the literature, as far as we know. Finally, Section 5 studies limit relations for semiclassical orthogonal polynomials of class $s = 1$.

2. Preliminaries and basic background

Definition 1. Let $\{\mu_n\}_{n\geq 0}$ be a sequence of complex numbers and let $\mathscr{L}$ be a linear complex-valued function defined on the linear space $\mathbb{P}$ of polynomials with complex coefficients by

$$\langle \mathscr{L}, x^n \rangle = \mu_n.$$

Then $\mathscr{L}$ is called the moment functional determined by the moment sequence $\{\mu_n\}_{n\geq 0}$, and μ_n is called the moment of order n.

Given a moment functional $\mathscr{L}$, the formal Stieltjes function of $\mathscr{L}$ is defined by

$$S_{\mathscr{L}}(z) = -\sum_{n=0}^{\infty} \frac{\mu_n}{z^{n+1}}.$$

For any moment functional $\mathcal{L}$ and any polynomial $q(x)$, we define the moment functional $q\mathcal{L}$ by

$$\langle q\mathcal{L}, P\rangle = \langle \mathcal{L}, qP\rangle, \quad P \in \mathbb{P}.$$

Definition 2. Let $\mathcal{L}$ be the linear functional associated with the moment sequence $\{\mu_n\}_{n\geq 0}$ and

$$\Delta_n = \det \begin{bmatrix} \mu_0 & \mu_1 & \cdots & \mu_n \\ \mu_1 & \mu_2 & \cdots & \mu_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_n & \mu_{n+1} & \cdots & \mu_{2n} \end{bmatrix}.$$

We call $\mathcal{L}$ *regular* if $\Delta_n \neq 0$ for all $n \in \mathbb{N}_0 := \{n \in \mathbb{Z} : n \geq 0\}$. We call it *positive definite* if $\Delta_n > 0$ for all $n \in \mathbb{N}_0$.

Definition 3. A sequence of polynomials $\{P_n(x)\}_{n\geq 0}$, with $\deg P_n = n$, is said to be an *orthogonal polynomial sequence* with respect to a regular linear functional $\mathcal{L}$ if there exists a sequence of nonzero real numbers $\{\zeta_n\}_{n\geq 0}$ such that

$$\langle \mathcal{L}, P_k P_n\rangle = \zeta_n \delta_{k,n}, \quad k, n \in \mathbb{N}_0.$$

If $\zeta_n = 1$, then $\{P_n(x)\}_{n\geq 0}$ is said to be an *orthonormal polynomial sequence*. If the linear functional is positive definite, such a sequence is unique under the assumption that each entry has a positive real leading coefficient.

Theorem 4 [Chihara 1978, Theorem 4.4]. *Let $\{b_n\}_{n\geq 0}$ and $\{\gamma_n\}_{n\geq 0}$, with $\gamma_n \neq 0$ for every $n \in \mathbb{N}_0$, be arbitrary sequences of complex numbers and let $\{P_n(x)\}$ be a sequence of monic polynomials defined by the three-term recurrence relation*

$$P_{n+1}(x) = (x - b_n)P_n(x) - \gamma_n P_{n-1}(x), \tag{1}$$

with $P_{-1} = 0$ and $P_0 = 1$. Then, there is a unique linear functional $\mathcal{L}$ such that $\mathcal{L}(1) = \gamma_0$ and

$$\langle \mathcal{L}, P_k(x)P_n(x)\rangle = \gamma_0\gamma_1 \cdots \gamma_n \delta_{k,n}.$$

If the linear functional is positive definite and $\{p_n(x)\}_{n\geq 0}$ is the corresponding orthonormal polynomial sequence, formula (1) becomes

$$a_{n+1}p_{n+1}(x) = (x - b_n)p_n(x) - a_n p_{n-1}(x),$$

where a_n is a real number and $a_n^2 = \gamma_n$.

Definition 5. Let $\mathcal{L}$ be a linear functional and $U^* : \mathbb{P} \to \mathbb{P}$ a linear operator. The linear functional $U\mathcal{L}$ is defined by

$$\langle U\mathcal{L}, P\rangle = -\langle \mathcal{L}, U^* P\rangle, \quad P \in \mathbb{P}.$$

Example 6. If U is the standard derivative operator D, we have $U^* = U = D$.

Definition 7. A regular linear functional $\mathscr{L}$ is called *U-semiclassical* if it satisfies the Pearson equation $U(\phi\mathscr{L}) + \psi\mathscr{L} = 0$ or, equivalently,

$$\langle U(\phi\mathscr{L}) + \psi\mathscr{L}, P\rangle = 0, \quad P \in \mathbb{P},$$

where ϕ, ψ are two polynomials and ϕ is monic. The corresponding orthogonal sequence $\{P_n(x)\}_{n\geq 0}$ is called U-semiclassical.

Semiclassical linear functionals with respect to several choices of operators have been studied in the literature. For example, when $U = D$ (the standard derivative operator), the theory of D-semiclassical linear functionals has been exhaustively studied by P. Maroni and coworkers (see [Maroni 1991] for an excellent survey on this topic).

If $U = D_\omega$, where

$$D_\omega f(x) = \frac{f(x+\omega) - f(x)}{\omega}, \quad \omega \neq 0,$$

a regular linear functional $\mathscr{L}$ is said to be *D_ω-semiclassical* if there exist polynomials ϕ, ψ, where ϕ is monic and $\deg\psi \geq 1$, such that $D_\omega(\phi\mathscr{L}) + \psi\mathscr{L} = 0$.

Notice that

$$D_1 f(x) = f(x+1) - f(x) = \Delta f(x),$$
$$D_{-1} f(x) = f(x) - f(x-1) = \nabla f(x)$$

are the forward and backward difference operators, respectively, and

$$\lim_{\omega\to 0} D_\omega f(x) = Df(x) = f'(x).$$

If $U = D_\omega$, we define $U^* = D_{-\omega}$. With this definition, we have $\Delta^* = \nabla$ and when $\omega \to 0$ we recover the identity $U^* = D = U$.

The concept of the class of a D_ω-semiclassical linear functional plays a central role in giving a constructive theory of such linear functionals.

Definition 8. If $\mathscr{L}$ is a D_ω-semiclassical linear functional, the *class* s of $\mathscr{L}$ is defined by

$$s = \min_{\phi,\psi} \max\{\deg\phi - 2, \deg\psi - 1\},$$

among all polynomials ϕ, ψ such that the Pearson equation holds. Notice that the class s is always nonnegative.

For any complex number c, we introduce the linear map $\theta_c : \mathbb{P} \to \mathbb{P}$, defined by

$$\theta_c(p)(x) = \frac{p(x) - p(c)}{x - c}.$$

Theorem 9 [Maroni 1991]. *A regular linear functional $\mathscr{L}$ satisfying the Pearson equation*

$$D_\omega(\phi\mathscr{L}) + \psi\mathscr{L} = 0$$

is of class s if and only if

$$\prod_{c\in Z(\phi)} \big(|\psi(c-\omega) + (\theta_c\phi)(c-\omega)| + |\langle\mathscr{L}, \theta_{c-\omega}(\psi+\theta_c\phi)\rangle|\big) > 0,$$

where $Z(\phi)$ denotes the set of zeros of the polynomial $\phi(x)$.

When there exists $c \in Z(\phi)$ such that

$$\psi(c-\omega) + (\theta_c\phi)(c-\omega) = \langle\mathscr{L}, \theta_{c-\omega}(\psi+\theta_c\phi)\rangle = 0,$$

the Pearson equation becomes

$$D_\omega[(\theta_c\phi)\mathscr{L}] + [\theta_{c-\omega}(\psi+\theta_c\phi)]\mathscr{L} = 0.$$

Remark 10. When $s = 0$, we obtain the D_ω-classical orthogonal polynomials (see [Abdelkarim and Maroni 1997]). For $\omega = 1$, several characterizations of classical orthogonal polynomials were given in [García et al. 1995]. Indeed, we explain in more detail in the next section the main characteristics of these polynomials and their corresponding linear functionals.

The D_1-semiclassical linear functionals have been studied by F. Marcellán and L. Salto [1998] and they are characterized following the same ideas as in the D case. P. Maroni and M. Mejri [2008] deduced the Laguerre–Freud equations for the coefficients of the three-term recurrence relation of D_w-semiclassical orthogonal polynomials of class $s = 1$. In the symmetric case, when the moments of odd order vanish, they deduced the explicit values of these coefficients, and the integral representations of the corresponding linear functionals are given.

On the other hand, the Pearson equation yields a difference equation for the moments of the linear functional, and, as a consequence, we get a linear difference equation with polynomial coefficients satisfied by the Stieltjes function associated with the linear functional:

Theorem 11. *If $\mathscr{L}$ is a D_ω-semiclassical moment functional, the formal Stieltjes function of $\mathscr{L}$ satisfies the nonhomogeneous first-order linear difference equation*

$$\phi(z)D_\omega S_{\mathscr{L}}(z) = a(z)S_{\mathscr{L}}(z) + b(z),$$

where $a(z)$ and $b(z)$ are polynomials depending on ϕ and ψ, with $\deg a \le s+1$ and $\deg b \le s$.

3. Discrete semiclassical orthogonal polynomials

We consider linear functionals

$$\langle \mathscr{L}, P\rangle = \sum_{x=0}^{\infty} P(x)\rho(x),$$

for some positive weight function $\rho(x)$ supported on a countable subset of the real line. With this choice, the Pearson equation

$$\langle \Delta(\phi\mathscr{L}) + \psi\mathscr{L}, P\rangle = 0, \quad P \in \mathbb{P},$$

yields

$$\Delta(\phi\rho) + \psi\rho = 0. \tag{2}$$

We rewrite this equation as

$$\frac{\rho(x+1)}{\rho(x)} = \frac{\phi(x)-\psi(x)}{\phi(x+1)} = \frac{\lambda(x)}{\phi(x+1)}, \tag{3}$$

with

$$\phi(x) = x(x+\beta_1)(x+\beta_2)\cdots(x+\beta_r),$$

and

$$\lambda(x) = c(x+\alpha_1)(x+\alpha_2)\cdots(x+\alpha_l).$$

Since the Pochhammer symbol $(\alpha)_x$ defined by $(\alpha)_0 = 1$ and

$$(a)_x = a(a+1)\cdots(a+x-1), \quad x \in \mathbb{N}, \tag{4}$$

satisfies the identity

$$\frac{(\alpha)_{x+1}}{(\alpha)_x} = x+\alpha, \quad x \in \mathbb{N}_0,$$

we obtain

$$\rho(x) = \frac{(\alpha_1)_x\cdots(\alpha_l)_x}{(\beta_1+1)_x\cdots(\beta_r+1)_x}\frac{c^x}{x!}. \tag{5}$$

We will denote the orthogonal polynomials associated with $\rho(x)$ by

$$P_n^{(l,r)}(x;\alpha_1,\ldots,\alpha_l,\beta_1,\ldots,\beta_r;c).$$

The moments of the weight function (5) are given by

$$\mu_n = \sum_{x=0}^{\infty} x^n \frac{(\alpha_1)_x\cdots(\alpha_l)_x}{(\beta_1+1)_x\cdots(\beta_r+1)_x}\frac{c^x}{x!}, \quad n = 0,1,\ldots.$$

By [Olver et al. 2010, 16.2], they exist if one of the following conditions holds:

(6) $l \le r$ and $c \in \mathbb{C}$.

(7) $l \ge r+1$, $c \in \mathbb{C}$, and one or more of the top parameters α_i is a nonpositive integer.

(8) $l = r+1$, and $|c| < 1$.

(9) $l = r+1$, $|c| = 1$, and $\operatorname{Re}(\beta_1 + \cdots + \beta_r - \alpha_1 - \cdots - \alpha_l) > 0$.

3.1. ***Discrete classical polynomials.*** Let $s = 0$. We solve the Pearson equation (3) with $\deg \psi = 1$ and $1 \le \deg \phi \le 2$. Three canonical cases appear (see [Nikiforov et al. 1985]), according to the following table, where $\lambda = \psi + \phi$:

$\deg \lambda$	$\deg \phi$	$\deg \psi$
0	1	1
1	1	1
2	2	1

Case 1: If $\deg \lambda = 0$ and $\deg \phi = 1$, we can take

$$\lambda(x) = c, \quad \phi(x) = x, \quad \psi(x) = \phi(x) - \lambda(x) = x - c, \tag{10}$$

and from (5) we obtain

$$\rho(x) = \frac{c^x}{x!}, \quad c > 0, \quad x \in \mathbb{N}_0. \tag{11}$$

The family of orthogonal polynomials associated with the weight function (11) is known as the Charlier polynomials; we denote them by $P_n^{(0,0)}(x;c)$. They have the hypergeometric representation (see [Koekoek et al. 2010, 9.14.1])

$$P_n^{(0,0)}(x;c) = {}_2F_0\left(\begin{matrix} -n, -x \\ - \end{matrix} ; -\frac{1}{c} \right), \tag{12}$$

where the hypergeometric function ${}_pF_q(z)$ is defined by

$${}_pF_q\left(\begin{matrix} a_1, \ldots, a_p \\ b_1, \ldots, b_q \end{matrix} ; z \right) = \sum_{k=0}^{\infty} \frac{(a_1)_k \cdots (a_p)_k}{(b_1)_k \cdots (b_q)_k} \frac{z^k}{k!}. \tag{13}$$

The monic Charlier polynomials $\widehat{P}_n^{(0,0)}(x;c)$ are given by

$$\widehat{P}_n^{(0,0)}(x;c) = (-c)^n P_n^{(0,0)}(x;c). \tag{14}$$

It is usual to denote these polynomials by

$$C_n(x;a) = P_n^{(0,0)}(x;a). \tag{15}$$

Case 2: If $\deg \lambda = 1$ and $\deg \phi = 1$, we can take

$$\lambda(x) = c(x+\alpha), \quad \phi(x) = x, \quad \psi(x) = (1-c)x - c\alpha, \tag{16}$$

and from (5) we have

$$\rho(x) = (\alpha)_x \frac{c^x}{x!}, \quad \alpha > 0, \quad 0 < c < 1, \quad x \in \mathbb{N}_0. \tag{17}$$

From (8), the condition $0 < c < 1$ is needed for the moments to exist. The first moment μ_0 is given by

$$\mu_0 = \sum_{x=0}^{\infty} (\alpha)_x \frac{c^x}{x!} = (1-c)^{-\alpha}. \tag{18}$$

The family of orthogonal polynomials associated with the weight function (17) is known as the Meixner polynomials; we denote them by $P_n^{(1,0)}(x;\alpha;c)$. They have the hypergeometric representation (see [Koekoek et al. 2010, 9.10.1])

$$P_n^{(1,0)}(x;\alpha;c) = {}_2F_1\left(\begin{matrix} -n, -x \\ \alpha \end{matrix}; 1-\frac{1}{c} \right), \tag{19}$$

and the monic Meixner polynomials $\widehat{P}_n^{(1,0)}(x;\alpha;c)$ are given by

$$\widehat{P}_n^{(1,0)}(x;\alpha;c) = (\alpha)_n \left(\frac{c}{c-1} \right)^n P_n^{(1,0)}(x;\alpha;c). \tag{20}$$

It is usual to denote these polynomials by

$$M_n(x;\beta,c) = P_n^{(1,0)}(x;\beta;c).$$

If we want c to be unbounded, we can use (7) and set $\alpha = -N$, with $N \in \mathbb{N}$. For the weight function to be positive we need $c < 0$, and we obtain the Krawtchouk polynomials $P_n^{(1,0)}(x;-N;c)$, with

$$\rho(x) = (-N)_x \frac{c^x}{x!}, \quad c < 0, \quad N \in \mathbb{N}, \quad x \in [0, N], \tag{21}$$

and

$$\phi(x) = x, \quad \psi(x) = (1-c)x + cN. \tag{22}$$

It is usual to denote these polynomials by

$$K_n(x;p,N) = P_n^{(1,0)}\left(x; -N; \frac{p}{p-1} \right).$$

Case 3: If $\deg \lambda = 2$ and $\deg \phi = 2$, we can take

$$\lambda(x) = c(x+\alpha_1)(x+\alpha_2), \quad \phi(x) = x(x+\beta).$$

Thus,

$$\psi(x) = \phi(x) - \lambda(x) = (1-c)x^2 + x(\beta - c\alpha_1 - c\alpha_2) - c\alpha_1\alpha_2,$$

and since $\deg \psi = 1$, we must have $c = 1$. Hence,

$$\phi(x) = x(x+\beta), \quad \psi(x) = x(\beta - \alpha_1 - \alpha_2) - \alpha_1\alpha_2, \tag{23}$$

and

$$\rho(x) = \frac{(\alpha_1)_x(\alpha_2)_x}{(\beta+1)_x}\frac{1}{x!}, \quad x \in \mathbb{N}_0. \tag{24}$$

From (9), we need $\operatorname{Re}(\beta + 1 - \alpha_1 - \alpha_2) > 0$ for the moments to exist. The first moment μ_0 is given by (see [Olver et al. 2010, 15.4.20])

$$\mu_0 = \sum_{x=0}^{\infty} \frac{(\alpha_1)_x(\alpha_2)_x}{(\beta+1)_x}\frac{1}{x!} = \frac{\Gamma(\beta+1)\Gamma(\beta+1-\alpha_1-\alpha_2)}{\Gamma(\beta+1-\alpha_1)\Gamma(\beta+1-\alpha_2)}.$$

Thus, we need $\alpha_1, \alpha_2 > 0$ and $\beta+1 > \alpha_1+\alpha_2$. The family of orthogonal polynomials associated with the weight function (24) is known as the Hahn polynomials; we denote them by $P_n^{(2,1)}(x;\alpha_1,\alpha_2,\beta;1)$. They have the hypergeometric representation [Erdélyi et al. 1953, 10.23.12]

$$P_n^{(2,1)}(x;\alpha_1,\alpha_2,\beta;1) = {}_3F_2\left(\begin{matrix} -n,\ -x,\ n+\alpha_1+\alpha_2-\beta-1 \\ \alpha_1,\ \alpha_2 \end{matrix};\ 1\right). \tag{25}$$

In the literature (see [Koekoek et al. 2010, 9.5.1]), the so-called Hahn polynomials $Q_n(x;\alpha,\gamma,N)$ correspond to the choice $\alpha_1 = \alpha+1,\ \alpha_2 = -N,\ \gamma = -N-\beta-1$, with $N \in \mathbb{N}$.

Another family of Hahn polynomials is

$$h_n(x;\alpha,\beta,N) = P_n^{(2,1)}(x;\beta+1, 1-N, -N-\alpha;1); \tag{26}$$

see page 34 in [Nikiforov et al. 1985]. In fact two different families of Hahn polynomials are considered in that reference; the polynomials involved in the corresponding Pearson equations are related by negating the variable x. Indeed, for the second family we also have a relation

$$\tilde{h}_n(x;\mu,\nu,N) = P_n^{(2,1)}(x;1-N-\nu, 1-N, \mu;1), \tag{27}$$

as well as

$$h_n(x;\alpha,\beta,N) = \tilde{h}_n(x;-N-\alpha,-N-\beta,N). \tag{28}$$

4. Discrete semiclassical polynomials of class one

When $s = 1$, we solve the Pearson equation (2) with $\deg \psi = 2$ and $1 \le \deg \phi \le 3$, and obtain five canonical cases:

$\deg\lambda$	$\deg\phi$	$\deg\psi$
0	2	2
1	2	2
2	1	2
2	2	2
3	3	2

Case 4: If $\deg\lambda = 0$ and $\deg\phi = 2$, we can take

$$\lambda(x) = c, \quad \phi(x) = x(x+\beta), \quad \psi(x) = x^2 + \beta x - c,$$

and from (5) we have

$$\rho(x) = \frac{1}{(\beta+1)_x}\frac{c^x}{x!}, \quad x \in \mathbb{N}_0, \tag{29}$$

where $\beta > -1$ and $c > 0$. The family of orthogonal polynomials associated with the weight function (29) is known as the generalized Charlier polynomials; we denote them by $P_n^{(0,1)}(x;\beta;c)$ and study them in Section 4.1 below.

Case 5: If $\deg\lambda = 1$ and $\deg\phi = 2$, we can take

$$\lambda(x) = c(x+\alpha), \quad \phi(x) = x(x+\beta), \quad \psi(x) = x^2 + (\beta - c)x - c\alpha.$$

From (5), we have

$$\rho(x) = \frac{(\alpha)_x}{(\beta+1)_x}\frac{c^x}{x!}, \quad x \in \mathbb{N}_0, \tag{30}$$

where $\alpha(\beta+1) > 0$ and $c > 0$. The family of orthogonal polynomials associated with the weight function (30) is known as the generalized Meixner polynomials; we denote them by $P_n^{(1,1)}(x;\alpha,\beta;c)$ and study them in Section 4.2.

Case 6: If $\deg\lambda = 2$ and $\deg\phi = 1$, we can take

$$\lambda(x) = c(x+\alpha_1)(x+\alpha_2), \quad \phi(x) = x.$$

From (5), we have

$$\rho(x) = (\alpha_1)_x(\alpha_2)_x\frac{c^x}{x!},$$

and from (7) we need $\alpha_2 = -N$, with $N \in \mathbb{N}$, for the moments to exist. Setting $\alpha_1 = \alpha$, we get

$$\lambda(x) = c(x+\alpha)(x-N), \quad \phi(x) = x, \quad \psi(x) = -cx^2 + x(Nc - c\alpha + 1) + Nc\alpha.$$

The family of orthogonal polynomials associated with the weight function

$$\rho(x) = (\alpha)_x(-N)_x\frac{c^x}{x!}, \quad x \in [0, N], \tag{31}$$

with $c<0$ and $\alpha>0$, will be referred to as the *generalized Krawtchouk polynomials*; we will denote them by $P_n^{(2,0)}(x;\alpha,-N;c)$ and study them in Section 4.3.

Case 7: If $\deg\lambda=2$ and $\deg\phi=2$, we can take

$$\lambda(x)=c(x+\alpha_1)(x+\alpha_2),\quad \phi(x)=x(x+\beta),$$
$$\psi(x)=(1-c)x^2+(\beta-c\alpha_1-c\alpha_2)x-c\alpha_1\alpha_2.$$

From (5), we have

$$\rho(x)=\frac{(\alpha_1)_x(\alpha_2)_x}{(\beta+1)_x}\frac{c^x}{x!},\quad x\in\mathbb{N}_0, \tag{32}$$

and from (8) we need $0<c<1$, with $\alpha_1\alpha_2(\beta+1)>0$. The family of orthogonal polynomials associated with the weight function (32) will be referred to as the *generalized Hahn polynomials of type I*; we will denote them by $P_n^{(2,1)}(x;\alpha_1,\alpha_2;\beta;c)$ and study them in Section 4.4.

Case 8: If $\deg\lambda=3$ and $\deg\phi=3$, we can take

$$\lambda(x)=c(x+\alpha_1)(x+\alpha_2)(x+\alpha_3),\quad \phi(x)=x(x+\beta_1)(x+\beta_2),$$
$$\psi(x)=x(x+\beta_1)(x+\beta_2)-c(x+\alpha_1)(x+\alpha_2)(x+\alpha_3).$$

For $\psi(x)$ to be of second degree we need $c=1$. Thus,

$$\lambda(x)=(x+\alpha_1)(x+\alpha_2)(x+\alpha_3),\quad \phi(x)=x(x+\beta_1)(x+\beta_2),$$
$$\psi(x)=-x^2(\alpha_1+\alpha_2-\beta_1+\alpha_3-\beta_2)-x(\alpha_1\alpha_2+\alpha_1\alpha_3+\alpha_2\alpha_3-\beta_1\beta_2)-\alpha_1\alpha_2\alpha_3,$$

and from (5) we obtain

$$\rho(x)=\frac{(\alpha_1)_x(\alpha_2)_x(\alpha_3)_x}{(\beta_1+1)_x(\beta_2+1)_x}\frac{1}{x!},\quad x\in\mathbb{N}_0. \tag{33}$$

For the moments to exist, (9) gives $\beta_1+\beta_2-\alpha_1-\alpha_2-\alpha_3>0$, while positivity demands that $\alpha_1\alpha_2\alpha_3(\beta_1+1)(\beta_2+1)>0$. The family of orthogonal polynomials associated with the weight function (33) will be referred to as *generalized Hahn polynomials of type II*; we will denote them by $P_n^{(3,2)}(x;\alpha_1,\alpha_2,\alpha_3;\beta_1,\beta_2;1)$ and study them in Section 4.5.

4.1. ***Generalized Charlier polynomials.*** This is Case 4 above, with weight function given by (29) (with $\beta>-1$ and $c>0$). The first moments are

$$\mu_0=c^{-\frac{\beta}{2}}I_\beta(2\sqrt{c})\Gamma(\beta+1),\quad \mu_1=c^{\frac{1-\beta}{2}}I_{\beta+1}(2\sqrt{c})\Gamma(\beta+1),$$

where $I_\nu(z)$ is the modified Bessel function of the first kind [Olver et al. 2010, 10.25.2].

Hounkonnou, Hounga and Ronveaux studied the semiclassical polynomials associated with the weight function

$$\rho_r(x) = \frac{c^x}{(x!)^r}, \quad r = 0, 1, \ldots \tag{34}$$

(see [Hounkonnou et al. 2000]). For $r = 2$, they derived the Laguerre–Freud equations for the recurrence coefficients and a second-order difference equation. Note that from (34) we have

$$\frac{\rho_r(x+1)}{\rho_r(x)} = \frac{c}{(x+1)^r},$$

and from (3) we conclude that

$$\lambda_r(x) = c, \quad \phi_r(x) = x^r, \quad \psi_r(x) = x^r - c,$$

and therefore the orthogonal polynomials associated with $\rho_r(x)$ are of class $r-1$. The case $r = 2$ is a particular example of (29) with $\beta = 0$.

Van Assche and Foupouagnigni [2003] also considered (34) with $r = 2$. Simplifying the Laguerre–Freud equations from [Hounkonnou et al. 2000], they got

$$u_{n+1} + u_{n-1} = \frac{1}{\sqrt{c}} \frac{n u_n}{1 - u_n^2} \quad \text{and} \quad v_n = \sqrt{c}\, u_{n+1} u_n,$$

with $\gamma_n = c(1 - u_n^2)$ and $\beta_n = v_n + n$. They showed that these equations are related to the discrete Painlevé II equation $\mathrm{dP_{II}}$ [Van Assche 2007]

$$x_{n+1} + x_{n-1} = \frac{(an+b)x_n + c}{1 - x_n^2}.$$

They also obtained the asymptotic behavior

$$\lim_{n\to\infty} \gamma_n = c, \quad \lim_{n\to\infty} v_n = 0,$$

and concluded that the asymptotic zero distribution is given by the uniform distribution on [0, 1], as is the case for the usual Charlier polynomials [Kuijlaars and Van Assche 1999].

Smet and Van Assche [2012] studied the orthogonal polynomials associated with the weight function (29). They obtained the Laguerre–Freud equations

$$(a_{n+1}^2 - c)(a_n^2 - c) = c(b_n - n)(b_n - n + \beta), \quad b_n + b_{n-1} = n - 1 - \beta + \frac{cn}{a_n^2}, \tag{35}$$

for the orthonormal polynomials. They showed that these equations are a limiting case of the discrete Painlevé IV equation $\mathrm{dP_{IV}}$ [Van Assche 2007]

$$x_{n+1}x_n = \frac{(y_n - \delta n - E)^2 - A}{y_n^2 - B},$$
$$y_n + y_{n-1} = \frac{\delta n + E - \delta/2 - C}{1 + Dx_n} + \frac{\delta n + E - \delta/2 + C}{1 + x_n/D}.$$

Finally, Filipuk and Van Assche [2013] related the system (35) to the (continuous) fifth Painlevé equation $\mathrm{P_V}$.

4.2. ***Generalized Meixner polynomials.*** This is Case 5 above, and the weight function is given by (30), with $\alpha(\beta+1) > 0$ and $c > 0$. The first moments are

$$\mu_0 = M(\alpha, \beta+1; c), \quad \mu_1 = \frac{\alpha c}{\beta+1} M(\alpha+1, \beta+2; c),$$

where $M(a, b; z)$ is the confluent hypergeometric function [Olver et al. 2010, 13.2.2].

Ronveaux [1986] considered the semiclassical polynomials associated with the weight function

$$\rho_r(x) = \prod_{j=1}^{r} (\alpha_j)_x \frac{c^x}{(x!)^r}, \quad r = 1, 2, \ldots,$$

and in [Ronveaux 2001] he made some conjectures on the asymptotic behavior of the recurrence coefficients.

Smet and Van Assche [2012] studied the orthogonal polynomials associated with the weight function (30). They obtained the Laguerre–Freud equations

$$\begin{aligned} (u_n + v_n)(u_{n+1} + v_n) &= \frac{\alpha-1}{c^2} v_n (v_n - c)\left(v_n - c\frac{\alpha-1-\beta}{\alpha-1}\right), \\ (u_n + v_n)(u_{n+1} + v_{n-1}) &= \frac{u_n}{u_n - \frac{cn}{\alpha-1}} (u_n + c)\left(u_n + c\frac{\alpha-1-\beta}{\alpha-1}\right), \end{aligned} \tag{36}$$

for the orthonormal polynomials, with

$$a_n^2 = cn - (\alpha-1)u_n, \quad b_n = n + \alpha + c - \beta - 1 - \frac{\alpha-1}{c} v_n.$$

They also proved that the system (36) is a limiting case of the asymmetric discrete Painlevé IV equation α-$\mathrm{dP_{IV}}$ [Van Assche 2007].

Filipuk and Van Assche [2011] showed that the system (36) can be obtained from the Bäcklund transformation of the fifth Painlevé equation $\mathrm{P_V}$. The particular case of (30) when $\beta = 0$ was considered by Boelen, Filipuk, and Van Assche [Boelen et al. 2011].

If we set $\alpha = -N$, $N \in \mathbb{N}$, in (30), we obtain

$$\rho(x) = \frac{(-N)_x}{(\beta+1)_x} \frac{c^x}{x!},$$

where we now have $\beta > -1$ and $c < 0$. This case was analyzed in [Boelen et al. 2013].

Singular limits. If we let $\alpha \to 0$ and $\beta \to -1$ in (30), we have $\rho(x) \to \tilde{\rho}(x)$, where $\tilde{\rho}(x)$ is a new weight function satisfying the Pearson equation

$$\Delta[(x-1)x\tilde{\rho}] + [x-(c+1)]x\tilde{\rho} = 0. \tag{37}$$

Assuming $\tilde{\rho}$ satisfies $x\tilde{\rho}(x) = xu(x)$ for some weight function $u(x)$ we get

$$\Delta[(x-1)xu] + [x-(c+1)]xu = 0. \tag{38}$$

Using the product rule

$$\Delta(fg) = f\Delta g + g\Delta f + \Delta f \Delta g \tag{39}$$

in (38), we have

$$xu + (x-1)\Delta(xu) + \Delta(xu) + [x-(c+1)]xu = 0,$$

or

$$x\Delta(xu) + [x-(c+1)+1]xu = 0.$$

Dividing by x, we obtain

$$\Delta(xu) + (x-c)u = 0.$$

Comparing with (10), we see that $u(x)$ is the weight function corresponding to the Charlier polynomials (11), and therefore (37) implies that

$$\tilde{\rho}(x) = \frac{c^x}{x!} + M\delta(x), \tag{40}$$

where $\delta(x)$ is the Dirac delta function.

The orthogonal polynomials $P_n^{(1,1)}(x; 0, -1; c)$ associated with the weight function (40) were first studied by Chihara [1985]. He showed that they satisfy the three-term recurrence relation (1) with

$$b_n = c\frac{n}{n+1}\frac{D_n}{D_{n+1}} + (n+1)\frac{D_{n+1}}{D_n}, \quad \gamma_n = c\frac{n^2}{n+1}\frac{D_n^2}{D_{n-1}D_{n+1}},$$

where

$$D_n = \frac{c^n}{n!}\frac{M}{e^c + MK_{n-1}}, \quad K_n = \sum_{j=0}^{n} \frac{c^j}{j!}, \quad K_{-1} = 0.$$

Note that for D_n to be well-defined for all n, we need $M > -1$, since $K_n \nearrow e^c$.

Bavinck and Koekoek [1995] obtained a difference equation satisfied by these polynomials and Álvarez-Nodarse, García, and Marcellán [Álvarez et al. 1995] found the hypergeometric representation

$$P_n^{(1,1)}(x;0;-1;c) = (-c)^n {}_3F_1\left(\begin{matrix}-n, -x, 1+x/D_n \\ x/D_n\end{matrix}; -\frac{1}{c}\right).$$

Since $\lim_{z\to\infty} \dfrac{(1+z)_x}{(z)_x} = 1$, we see that

$$\lim_{M\to 0} P_n^{(1,1)}(x;0,-1;c) = \widehat{C}_n(x;c),$$

where $\widehat{C}_n(x;c)$ is the monic Charlier polynomial (14).

4.3. ***Generalized Krawtchouk polynomials.*** This is Case 6 above, and the weight function is given by (31), with $c < 0$, $N \in \mathbb{N}$, and $\alpha > 0$. The first moments are

$$\mu_0 = C_N\left(-\alpha; -\frac{1}{c}\right), \quad \mu_1 = -c\alpha N C_{N-1}\left(-\alpha-1; -\frac{1}{c}\right),$$

where $C_n(x;a)$ is the Charlier polynomial (15).

To our knowledge, these polynomials have not appeared before in the literature.

4.4. ***Generalized Hahn polynomials of type I.*** This is Case 7 above, and the weight function is given by (32), with $0 < c < 1$ and $\alpha_1\alpha_2(\beta+1) > 0$. The first moments are

$$\mu_0 = {}_2F_1\left(\begin{matrix}\alpha_1, \alpha_2 \\ \beta+1\end{matrix}; c\right), \quad \mu_1 = c\frac{\alpha_1\alpha_2}{\beta+1}\, {}_2F_1\left(\begin{matrix}\alpha_1+1, \alpha_2+1 \\ \beta+2\end{matrix}; c\right),$$

where ${}_2F_1\left(\begin{matrix}a, b \\ c\end{matrix}; z\right)$ is the hypergeometric function.

Singular limits. (a) If we let $\alpha_2 \to 0, \beta \to -1$ and $\alpha_1 = \alpha$ in (32), we have $\rho(x) \to \tilde{\rho}(x)$, where $\tilde{\rho}(x)$ is a new weight function satisfying the Pearson equation

$$\Delta[(x-1)x\tilde{\rho}] + [(1-c)x - (1+c\alpha)]x\tilde{\rho} = 0. \tag{41}$$

Assuming that $\tilde{\rho}(x)$ satisfies $x\tilde{\rho}(x) = xu(x)$ for some weight function $u(x)$, we get

$$\Delta[(x-1)xu] + [(1-c)x - (1+c\alpha)]xu = 0. \tag{42}$$

Using the product rule (39) in (42), we have

$$xu + (x-1)\Delta(xu) + \Delta(xu) + [(1-c)x - (1+c\alpha)]xu = 0,$$

or

$$x\Delta(xu) + [(1-c)x - (1+c\alpha) + 1]xu = 0.$$

Dividing by x, we obtain $\Delta(xu)+[(1-c)x-c\alpha]u=0$. Comparing with (16), we see that $u(x)$ is the weight function corresponding to the Meixner polynomials (17), and therefore (37) implies that

$$\tilde{\rho}(x)=(\alpha)_x\frac{c^x}{x!}+M\delta(x). \tag{43}$$

The orthogonal polynomials associated with the weight function (43) were first studied by Chihara [1985]. He showed that they satisfy the three-term recurrence relation (1) with

$$b_n=\frac{c(\alpha+n)}{c-1}\frac{n}{n+1}\frac{B_n}{B_{n+1}}+\frac{n+1}{c-1}\frac{B_{n+1}}{B_n},\quad \gamma_n=\frac{c}{(c-1)^2}\frac{n^2(\alpha+n)}{n+1}\frac{B_n^2}{B_{n-1}B_{n+1}},$$

where

$$B_n=\frac{c^n(\alpha)_n}{(1-c)n!}\frac{M}{(1-c)^{-\alpha}+MK_{n-1}},\quad K_n=\sum_{j=0}^{n}(\alpha)_j\frac{c^j}{j!},\quad K_{-1}=0.$$

For B_n to be well-defined for all n, we need $M>-1$, since $K_n\nearrow(1-c)^{-\alpha}$.

In [Brezinski et al. 1991], Richard Askey proposed the problem of finding a second-order difference equation satisfied by these polynomials. The problem was solved in [Bavinck and van Haeringen 1994]; in [Álvarez et al. 1995] the hypergeometric representation

$$P_n^{(2,1)}(x;\alpha,0,-1;c)=(\alpha)_n\left(\frac{c}{c-1}\right)^n{}_3F_2\left(\begin{matrix}-n,\ -x,\ 1+x/B_n\\ \alpha,\ x/B_n\end{matrix};\ 1-\frac{1}{c}\right)$$

was given. In this case,

$$\lim_{M\to 0}P_n^{(2,1)}(x;\alpha,0,-1;c)=\widehat{M}_n(x;\alpha,c),$$

where $\widehat{M}_n(x;\alpha,c)$ is the monic Meixner polynomial (20).

(b) If $\alpha_1=-N$, $N\in\mathbb{N}$, we can remove the restriction that $0<c<1$ and take any $c<0$, with $\alpha_2\notin[-N,0]$, $\beta\notin[-N-1,-1]$, and $\alpha_2(\beta+1)>0$. If we let $\alpha_2\to-(N-1)$ and $\beta\to-N$, we have $\rho(x)\to\tilde{\rho}(x)$, where $\tilde{\rho}(x)$ is a new weight function satisfying the Pearson equation

$$\begin{aligned}&\psi(x)=(1-c)x^2+(\beta-c\alpha_1-c\alpha_2)x-c\alpha_1\alpha_2,\\ &\Delta[x(x-N)\tilde{\rho}]+[(1-c)x+c(N-1)](x-N)\tilde{\rho}=0.\end{aligned}\tag{44}$$

Assuming that $\tilde{\rho}(x)$ satisfies $(x-N)\tilde{\rho}(x)=(x-N)u(x)$ for some weight function $u(x)$, we get

$$\Delta[x(x-N)u]+[(1-c)x+c(N-1)](x-N)u=0. \tag{45}$$

Using the product rule (39) in (45), we have

$$xu + (x - N + 1)\Delta(xu) + [(1-c)x + c(N-1)](x-N)u = 0,$$

or

$$(x - N + 1)\Delta(xu) + (x - N + 1)(x + Nc - cx)u = 0.$$

Dividing by $x - N + 1$, we obtain $\Delta(xu) + [(1-c)x + cN]u = 0$. Comparing with (22), we see that $u(x)$ is the weight function corresponding to the Krawtchouk polynomials (21), and therefore (44) implies that

$$\tilde{\rho}(x) = (-N)_x \frac{c^x}{x!} + M\delta(x - N).$$

4.5. ***Generalized Hahn polynomials of type II.*** This is Case 8, and the weight function is (33), with $\beta_1 + \beta_2 - \alpha_1 - \alpha_2 - \alpha_3 > 0$ and $\alpha_1\alpha_2\alpha_3(\beta_1+1)(\beta_2+1) > 0$. The first moments are

$$\mu_0 = {}_3F_2\left(\begin{matrix}\alpha_1, \alpha_2, \alpha_3\\ \beta_1{+}1, \beta_2{+}1\end{matrix}; 1\right),$$

$$\mu_1 = \frac{\alpha_1\alpha_2\alpha_3}{(\beta_1+1)(\beta_2+1)}\, {}_3F_2\left(\begin{matrix}\alpha_1{+}1, \alpha_2{+}1, \alpha_3{+}1\\ \beta_1{+}2, \beta_2{+}2\end{matrix}; 1\right),$$

where ${}_3F_2\left(\begin{matrix}a_1, a_2, a_3\\ b_1, b_2\end{matrix}; z\right)$ is the hypergeometric function.

To our knowledge, these polynomials have not appeared before in the literature.

Singular limits. (a) If we let $\alpha_3 \to 0$, $\beta_2 \to -1$, $\beta_1 = \beta$ in (33), we have $\rho(x) \to \tilde{\rho}(x)$, where $\tilde{\rho}(x)$ is a new weight function satisfying the Pearson equation

$$\Delta[(x-1)(x+\beta)x\tilde{\rho}] + [(\beta - 1 - \alpha_1 - \alpha_2)x - \alpha_1\alpha_2 - \beta]x\tilde{\rho} = 0. \tag{46}$$

Assuming that $\tilde{\rho}(x)$ satisfies $x\tilde{\rho}(x) = xu(x)$ for some weight function $u(x)$, we get

$$\Delta[(x-1)(x+\beta)xu] + [(\beta - 1 - \alpha_1 - \alpha_2)x - \alpha_1\alpha_2 - \beta]xu = 0. \tag{47}$$

Using the product rule (39) in (47), we have

$$(x+\beta)xu + x\Delta[(x+\beta)xu] + [(\beta - 1 - \alpha_1 - \alpha_2)x - \alpha_1\alpha_2 - \beta]xu = 0,$$

or

$$x\Delta[(x+\beta)xu] + [(\beta - \alpha_1 - \alpha_2)x - \alpha_1\alpha_2]xu = 0.$$

Dividing by x, we obtain

$$\Delta[(x+\beta)xu] + [(\beta - \alpha_1 - \alpha_2)x - \alpha_1\alpha_2]u = 0.$$

Comparing with (23), we see that $u(x)$ is the weight function corresponding to the Hahn polynomials (24); therefore (46) implies that

$$\tilde{\rho}(x) = \frac{(\alpha_1)_x(\alpha_2)_x}{(\beta+1)_x}\frac{1}{x!} + M\delta(x). \tag{48}$$

(b) Similarly, if we let $\alpha_3 = -N, \alpha_2 \to -(N-1), \beta_2 \to -N$, $\alpha_1 = \alpha$, $\beta_1 = \beta$, $\alpha(\beta+1) < 0$ in (33), we have $\rho(x) \to \tilde{\rho}(x)$, where $\tilde{\rho}(x)$ is a new weight function satisfying the Pearson equation

$$\Delta[x(x+\beta)(x-N)\tilde{\rho}] + [(\beta-\alpha+N-1)x + \alpha(N-1)](x-N)\tilde{\rho} = 0. \tag{49}$$

Assuming that $\tilde{\rho}(x)$ satisfies $(x-N)\tilde{\rho}(x) = (x-N)u(x)$ for some weight function $u(x)$, we get

$$\Delta[x(x+\beta)(x-N)u] + [(\beta-\alpha+N-1)x + \alpha(N-1)](x-N)u = 0. \tag{50}$$

Using the product rule (39) in (50), we have

$$(x+\beta)xu + (x-N+1)\Delta[(x+\beta)xu] \\ + [(\beta-\alpha+N-1)x + \alpha(N-1)](x-N)u = 0,$$

or

$$(x-N+1)\Delta[(x+\beta)xu] + (x-N+1)[(\beta-\alpha+N)x + \alpha N]u = 0.$$

Dividing by $x-N+1$, we obtain

$$\Delta[(x+\beta)xu] + [(\beta-\alpha+N)x + \alpha N]u = 0.$$

Comparing with (23), we see that $u(x)$ is the weight function corresponding to the truncated Hahn polynomials (26), and therefore (49) implies that

$$\tilde{\rho}(x) = \frac{(\alpha)_x(-N)_x}{(\beta+1)_x}\frac{1}{x!} + M\delta(x-N). \tag{51}$$

The orthogonal polynomials associated with the weight functions (48) and (51) were first studied in [Álvarez and Marcellán 1995b].

5. Limit relations between polynomials

From the identities (see [Koekoek et al. 2010])

$$\lim_{\lambda\to\infty}\frac{(\lambda\alpha)_x}{\lambda^x} = \alpha^x \quad \text{and} \quad \lim_{\lambda\to\infty}\frac{(\lambda\alpha)_x}{(\lambda\beta)_x} = \left(\frac{\alpha}{\beta}\right)^x,$$

we have the following limit relations:

1. generalized Hahn polynomials of type II to generalized Hahn polynomials of type I
$$\lim_{\alpha\to\infty} P_n^{(3,2)}(x;\alpha_1,\alpha_2,\alpha,\beta,\alpha/c;1) = P_n^{(2,1)}(x;\alpha_1,\alpha_2,\beta;c),$$
2. generalized Hahn polynomials of type I to generalized Krawtchouk polynomials
$$\lim_{\beta\to\infty} P_n^{(2,1)}(x;\alpha,-N,\beta;c\beta) = P_n^{(2,0)}(x;\alpha,-N;c),$$
3. generalized Hahn polynomials of type I to generalized Meixner polynomials
$$\lim_{\alpha_2\to\infty} P_n^{(2,1)}(x;\alpha,\alpha_2,\beta;c/\alpha_2) = P_n^{(1,1)}(x;\alpha,\beta;c),$$
4. generalized Meixner polynomials to generalized Charlier polynomials
$$\lim_{\alpha\to\infty} P_n^{(1,1)}(x;\alpha,\beta;c/\alpha) = P_n^{(0,1)}(x;\beta;c),$$
5. generalized Meixner polynomials to Meixner polynomials
$$\lim_{\beta\to\infty} P_n^{(1,1)}(x;\alpha,\beta;c\beta) = M_n(x;\alpha;c),$$
6. generalized Charlier polynomials to Charlier polynomials
$$\lim_{\beta\to\infty} P_n^{(0,1)}(x;\beta;c\beta) = C_n(x;c).$$

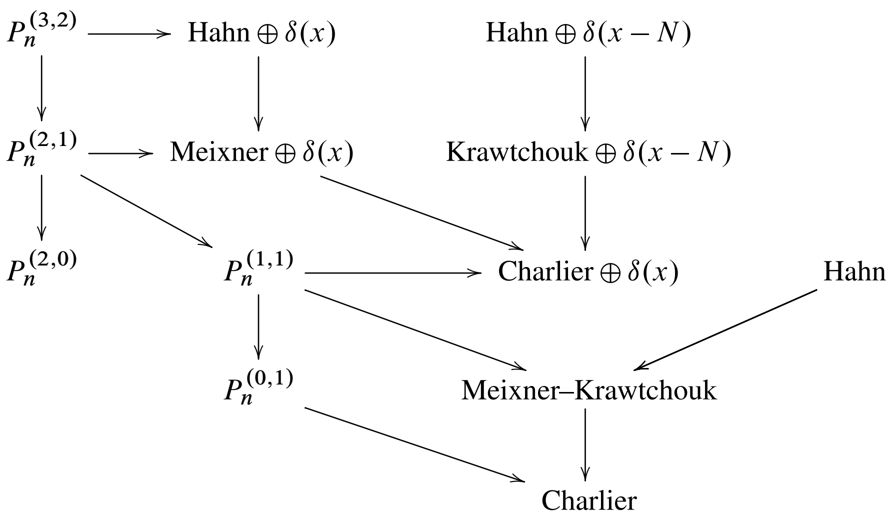
We also have the following singular limits, where "$\oplus\,\delta(x-x_0)$" denotes the addition of a delta function to the measure of orthogonality at the point x_0:

1. generalized Meixner polynomials to Charlier-Dirac polynomials
$$\lim_{\substack{\alpha\to 0\\ \beta\to -1}} P_n^{(1,1)}(x;\alpha,\beta;c) = C_n(x;c)\oplus\delta(x),$$
2. generalized Hahn polynomials of type I to truncated Hahn polynomials
$$\lim_{\substack{\alpha_2\to -N\\ c\to 1}} P_n^{(2,1)}(x;\alpha,\alpha_2,\beta;c) = Q_n(x;\alpha,\beta,N),$$
3. generalized Hahn polynomials of type I to Meixner–Dirac polynomials
$$\lim_{\substack{\alpha_2\to 0\\ \beta\to -1}} P_n^{(2,1)}(x;\alpha,\alpha_2,\beta;c) = M_n(x;\alpha;c)\oplus\delta(x),$$
4. generalized Hahn polynomials of type I to Krawtchouk–Dirac polynomials
$$\lim_{\substack{\alpha_2\to -N+1\\ \beta\to -N}} P_n^{(2,1)}(x;-N,\alpha_2,\beta;c) = K_n(x;-N;c)\oplus\delta(x-N),$$
5. generalized Hahn polynomials of type II to Hahn–Dirac polynomials
$$\lim_{\substack{\alpha_2\to 0\\ \beta_2\to -1}} P_n^{(3,2)}(x;\alpha,\alpha_2,-N,\beta,\beta_2;1) = Q_n(x;\alpha,\beta,N)\oplus\delta(x),$$

6. generalized Hahn polynomials of type II to Hahn–Dirac polynomials

$$\lim_{\substack{\alpha_2 \to -N+1 \\ \beta_2 \to -N}} P_n^{(3,2)}(x; \alpha, \alpha_2, -N, \beta, \beta_2; 1) = Q_n(x; \alpha, \beta, N) \oplus \delta(x - N).$$

We can summarize these results in the following scheme:



6. Concluding remarks

We have described the discrete semiclassical orthogonal polynomials of class $s = 1$ using the different choices for the polynomials in the canonical Pearson equation that the corresponding linear functional satisfies. We have focused our attention to the case where the linear functional has a representation in terms of a discrete positive measure supported on a countable subset of the real line. Some new families of orthogonal polynomials appear, as well as some families of orthogonal polynomials (generalized Charlier, generalized Krawtchouk, and generalized Meixner) which have attracted the interest of researchers in the last years, since the coefficients of their three-term recurrence relations are related to discrete and continuous Painlevé equations. We have also studied limit relations between these families of orthogonal polynomials, having in mind an analogue of the Askey tableau for classical orthogonal polynomials. It would be very interesting to find the equations satisfied by the coefficients of the three-term recurrence relations for the above new sequences of semiclassical orthogonal polynomials. Furthermore, an analysis of the class $s = 2$ will also be welcome in order to get a complete classification of such a class as well as to check if new families of orthogonal polynomials appear as in the case of the D-semiclassical orthogonal polynomials pointed out in [Marcellán et al. 2012].

Acknowledgements

We thank the referees for remarks and suggestions that have helped improve the presentation of this article. Marcellán's work has been supported by Dirección General de Investigación Científica y Técnica, Ministerio de Economía y Competitividad of Spain, grant MTM2012-36732-C03-01.

References

[Abdelkarim and Maroni 1997] F. Abdelkarim and P. Maroni, "The D_ω-classical orthogonal polynomials", *Results Math.* **32**:1-2 (1997), 1–28. MR 99b:33009 Zbl 0889.42020

[Álvarez and Marcellán 1995a] R. Álvarez-Nodarse and F. Marcellán, "Difference equation for modifications of Meixner polynomials", *J. Math. Anal. Appl.* **194**:1 (1995), 250–258. MR 96k:33009 Zbl 0834.39005

[Álvarez and Marcellán 1995b] R. Álvarez-Nodarse and F. Marcellán, "The modification of classical Hahn polynomials of a discrete variable", *Integral Transform. Spec. Funct.* **3**:4 (1995), 243–262. MR 97j:33010 Zbl 0849.33007

[Álvarez et al. 1995] R. Álvarez-Nodarse, A. G. García, and F. Marcellán, "On the properties for modifications of classical orthogonal polynomials of discrete variables", *J. Comput. Appl. Math.* **65**:1-3 (1995), 3–18. MR 98b:33018 Zbl 0865.42023

[Bavinck and Koekoek 1995] H. Bavinck and R. Koekoek, "On a difference equation for generalizations of Charlier polynomials", *J. Approx. Theory* **81**:2 (1995), 195–206. MR 96h:33002 Zbl 0865.33006

[Bavinck and van Haeringen 1994] H. Bavinck and H. van Haeringen, "Difference equations for generalized Meixner polynomials", *J. Math. Anal. Appl.* **184**:3 (1994), 453–463. MR 95h:33004 Zbl 0824.33005

[Belmehdi 1992] S. Belmehdi, "On semi-classical linear functionals of class $s = 1$: classification and integral representations", *Indag. Math.* (*N.S.*) **3**:3 (1992), 253–275. MR 94e:33038 Zbl 0783.33003

[Boelen et al. 2011] L. Boelen, G. Filipuk, and W. Van Assche, "Recurrence coefficients of generalized Meixner polynomials and Painlevé equations", *J. Phys. A* **44**:3 (2011), Article ID #035202. MR 2011j:33036 Zbl 1217.34135

[Boelen et al. 2013] L. Boelen, G. Filipuk, C. Smet, W. Van Assche, and L. Zhang, "The generalized Krawtchouk polynomials and the fifth Painlevé equation", *J. Differ. Equations Appl.* **19**:9 (2013), 1437–1451. MR 3173496 Zbl 1279.33013

[Brezinski et al. 1991] C. Brezinski, L. Gori, and A. Ronveaux (editors), *Orthogonal polynomials and their applications* (Erice, 1990), IMACS Ann. Comput. Appl. Math. **9**, J. C. Baltzer, Basel, 1991. MR 94m:42003 Zbl 0812.00027

[Chihara 1978] T. S. Chihara, *An introduction to orthogonal polynomials*, Mathematics and its Applications **13**, Gordon and Breach, New York, 1978. MR 58 #1979 Zbl 0389.33008

[Chihara 1985] T. S. Chihara, "Orthogonal polynomials and measures with end point masses", *Rocky Mountain J. Math.* **15**:3 (1985), 705–719. MR 87a:42037 Zbl 0586.33007

[Erdélyi et al. 1953] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher transcendental functions*, vol. II, McGraw-Hill, New York, 1953. MR 15,419i Zbl 0052.29502

[Filipuk and Van Assche 2011] G. Filipuk and W. Van Assche, "Recurrence coefficients of a new generalization of the Meixner polynomials", *SIGMA Symmetry Integrability Geom. Methods Appl.* **7** (2011), Paper 068. MR 2861208 Zbl 1250.33015

[Filipuk and Van Assche 2013] G. Filipuk and W. Van Assche, "Recurrence coefficients of generalized Charlier polynomials and the fifth Painlevé equation", *Proc. Amer. Math. Soc.* **141**:2 (2013), 551–562. MR 2996960 Zbl 1269.34095

[Foupouagnigni et al. 1998] M. Foupouagnigni, M. N. Hounkonnou, and A. Ronveaux, "Laguerre–Freud equations for the recurrence coefficients of D_ω semi-classical orthogonal polynomials of class one", *J. Comput. Appl. Math.* **99**:1-2 (1998), 143–154. MR 2000e:33008 Zbl 0931.33009

[García et al. 1995] A. G. García, F. Marcellán, and L. Salto, "A distributional study of discrete classical orthogonal polynomials", *J. Comput. Appl. Math.* **57**:1-2 (1995), 147–162. MR 96m:33004 Zbl 0853.33009

[Godoy et al. 1997] E. Godoy, F. Marcellán, L. Salto, and A. Zarzo, "Perturbations of discrete semiclassical functionals by Dirac masses", *Integral Transform. Spec. Funct.* **5**:1-2 (1997), 19–46. MR 98f:33015 Zbl 0877.33002

[Hounkonnou et al. 2000] M. N. Hounkonnou, C. Hounga, and A. Ronveaux, "Discrete semi-classical orthogonal polynomials: generalized Charlier", *J. Comput. Appl. Math.* **114**:2 (2000), 361–366. MR 2001m:33010 Zbl 0971.42015

[Koekoek et al. 2010] R. Koekoek, P. A. Lesky, and R. F. Swarttouw, *Hypergeometric orthogonal polynomials and their q-analogues*, Springer, Berlin, 2010. MR 2011e:33029 Zbl 1200.33012

[Kuijlaars and Van Assche 1999] A. B. J. Kuijlaars and W. Van Assche, "Extremal polynomials on discrete sets", *Proc. London Math. Soc.* (3) **79**:1 (1999), 191–221. MR 2000g:42033 Zbl 1033.42024

[Marcellán and Salto 1998] F. Marcellán and L. Salto, "Discrete semi-classical orthogonal polynomials", *J. Differ. Equations Appl.* **4**:5 (1998), 463–496. MR 2000e:33010 Zbl 0916.33006

[Marcellán et al. 2012] F. Marcellán, M. Sghaier, and M. Zaatra, "On semiclassical linear functionals of class $s = 2$: classification and integral representations", *J. Differ. Equations Appl.* **18**:6 (2012), 973–1000. MR 2929632 Zbl 1250.42081

[Maroni 1991] P. Maroni, "Une théorie algébrique des polynômes orthogonaux: application aux polynômes orthogonaux semi-classiques", pp. 95–130 in *Orthogonal polynomials and their applications* (Erice, 1990), edited by C. Brezinski et al., IMACS Ann. Comput. Appl. Math. **9**, Baltzer, Basel, 1991. MR 95i:42018 Zbl 0944.33500

[Maroni and Mejri 2008] P. Maroni and M. Mejri, "The symmetric D_ω-semi-classical orthogonal polynomials of class one", *Numer. Algorithms* **49**:1-4 (2008), 251–282. MR 2010a:42100 Zbl 1181.42028

[Nikiforov et al. 1985] A. F. Nikiforov, S. K. Suslov, and V. B. Uvarov, *Классические ортогональные полиномы дискретной переменной*, Nauka, Moscow, 1985. Translated as *Classical orthogonal polynomials of a discrete variable*, Springer, Berlin, 1991. MR 92m:33019 Zbl 0576.33001

[Olver et al. 2010] F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark (editors), *NIST handbook of mathematical functions*, Cambridge University Press, 2010. MR 2012a:33001 Zbl 1198.00002

[Ronveaux 1986] A. Ronveaux, "Discrete semi-classical orthogonal polynomials: generalized Meixner", *J. Approx. Theory* **46**:4 (1986), 403–407. MR 87f:33030 Zbl 0591.42015

[Ronveaux 2001] A. Ronveaux, "Asymptotics for recurrence coefficients in the generalized Meixner case", *J. Comput. Appl. Math.* **133**:1–2 (2001), 695–696.

[Ronveaux and Salto 2001] A. Ronveaux and L. Salto, "Discrete orthogonal polynomials: polynomial modification of a classical functional", *J. Differ. Equations Appl.* **7**:3 (2001), 323–344. MR 2004d:33010 Zbl 0991.33002

[Smet and Van Assche 2012] C. Smet and W. Van Assche, "Orthogonal polynomials on a bi-lattice", *Constr. Approx.* **36**:2 (2012), 215–242. MR 2957309 Zbl 1262.34103

[Van Assche 2007] W. Van Assche, "Discrete Painlevé equations for recurrence coefficients of orthogonal polynomials", pp. 687–725 in *Difference equations, special functions and orthogonal polynomials* (Munich, 2005), edited by S. Elaydi et al., World Scientific, Hackensack, NJ, 2007. MR 2009k:42055 Zbl 1128.39013

[Van Assche and Foupouagnigni 2003] W. Van Assche and M. Foupouagnigni, "Analysis of non-linear recurrence relations for the recurrence coefficients of generalized Charlier polynomials", *J. Nonlinear Math. Phys.* **10**:Suppl. 2 (2003), 231–237. MR 2005f:33020

Received November 8, 2012. Revised January 16, 2014.

DIEGO DOMINICI
DEPARTMENT OF MATHEMATICS
STATE UNIVERSITY OF NEW YORK AT NEW PALTZ
1 HAWK DRIVE
NEW PALTZ, NY 12561-2443
UNITED STATES

dominicd@newpaltz.edu

FRANCISCO MARCELLÁN
DEPARTAMENTO DE MATEMÁTICAS
UNIVERSIDAD CARLOS III DE MADRID
AVENIDA DE LA UNIVERSIDAD 30
28911 LEGANÉS, MADRID
SPAIN

pacomarc@ing.uc3m.es

PACIFIC JOURNAL OF MATHEMATICS
Vol. 268, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.268.413

A NOTE ON CONFORMAL RICCI FLOW

PENG LU, JIE QING AND YU ZHENG

In this note we study the conformal Ricci flow that Arthur Fischer introduced in 2004. We use DeTurck's trick to rewrite the conformal Ricci flow as a strong parabolic-elliptic partial differential equation. Then we prove short-time existence for the conformal Ricci flow on compact manifolds as well as on asymptotically flat manifolds. We show that the Yamabe constant is monotonically increasing along conformal Ricci flow on compact manifolds. We also show that the conformal Ricci flow is the gradient flow for the ADM mass on asymptotically flat manifolds.

1. Introduction

Suppose that M^m is a smooth m-dimensional manifold and that g_0 is a Riemannian metric on M of constant scalar curvature s_0. All manifolds in this note are assumed to have no boundary. The conformal Ricci flow on M is defined as

$$\begin{cases} \frac{\partial}{\partial t} g + 2\left(\mathrm{Ric} - \frac{s_0}{m} g\right) = -2pg & \text{in } M \times (0, T), \\ s_{g(t)} = s_0 & \text{in } M \times [0, T) \end{cases} \tag{1-1}$$

for a family of metrics $g(t)$ with initial condition $g(0) = g_0$ and a family of functions $p = p(t)$ on $M \times [0, T)$, where $s_{g(t)}$ is the scalar curvature of the evolving metric $g(t)$. The conformal Ricci flow (1-1) was introduced by Arthur Fischer [2004] as a modified Ricci flow that preserves the constant scalar curvature of the evolving metrics. It is so named because of the role that conformal geometry plays in maintaining constant scalar curvature. It was shown in [Fischer 2004] that on compact manifolds the conformal Ricci flow is equivalent to

$$\begin{cases} \frac{\partial}{\partial t} g + 2\left(\mathrm{Ric} - \frac{s_0}{m} g\right) = -2pg & \text{in } M \times (0, T), \\ (m-1)\Delta p + s_0 p = -\left|\mathrm{Ric} - \frac{s_0}{m} g\right|^2 & \text{in } M \times [0, T), \end{cases} \tag{1-2}$$

with the initial condition $g(0) = g_0$. Based on the fact that the conformal Ricci flow (1-2) is of parabolic-elliptic nature, analogous to Navier–Stokes equations, the

Jie Qing's research was partially supported by NSF grant number DMS-1005295 and CNSF grant number 10728103. Yu Zheng's research was partially surpported by CNSF grant number 11271132.
MSC2010: primary 53C25; secondary 58J05.
Keywords: conformal Ricci flow, short-time existence, asymptotically flat manifolds, ADM mass.

function p was named the conformal pressure function in [Fischer 2004]. Using the theory of dynamical systems on infinite-dimensional manifolds, Fischer [2004] proved that conformal Ricci flow exists, at least for a short time, on compact manifolds with scalar curvature $s_0 < 0$. He also observed that the Yamabe constant monotonically increases along conformal Ricci flow on compact manifolds of negative Yamabe type. Therefore one can hope that the conformal Ricci flow does a good job of constructing Einstein metrics, considering the behavior of the Hilbert–Einstein action on the space of Riemannian metrics.

In this paper we adopt DeTurck's trick [1983; 2003] to eliminate the degeneracy of (1-2) from the symmetry of diffeomorphisms and consider

$$\begin{cases} \dfrac{\partial}{\partial t} g + 2\Big(\mathrm{Ric} - \dfrac{s_0}{m} g\Big) = -2pg + \mathscr{L}_W g, \\ (m-1)\Delta p + s_0 p = -\Big|\mathrm{Ric} - \dfrac{s_0}{m} g\Big|^2. \end{cases} \tag{1-3}$$

This is DeTurck's conformal Ricci flow for an appropriately chosen vector field W (cf. Equation (3-5)) with an initial metric $g(0) = g_0$ of constant scalar curvature s_0. Equation (1-3) is a strong parabolic-elliptic partial differential equation. We use the contractive mapping theorem to prove an isomorphism property for the linearized DeTurck conformal Ricci flow and then we use the implicit function theorem to prove short-time existence for the DeTurck conformal Ricci flow. From this we obtain short-time existence for the conformal Ricci flow based on the discussion in Section 3.1.

Theorem 1.1. *Let $(M^m,\ g_0)$ be a compact Riemannian manifold of constant scalar curvature s_0 with no boundary. Suppose that the elliptic operator $(m-1)\Delta_{g_0} + s_0$ is invertible. Then there exists a small positive number T such that the conformal Ricci flow $g(t)$ with the initial metric g_0 exists for $t \in [0, T]$.*

This extends the existence result in [Fischer 2004] to include some compact manifolds with scalar curvature $s_0 > 0$. For parabolic Hölder spaces and the theory of linear and nonlinear parabolic equations used in the proof, we take references mostly from [Lunardi 1995]. We also extend the monotonicity of the Yamabe constant in [Fischer 2004] as follows:

Theorem 1.2. *Let (M^m, g_0) be a compact Riemannian manifold and let $g(t)$, $t \in [0, T)$, be the solution of conformal Ricci flow with $g(0) = g_0$. Suppose that g_0 is the only Yamabe metric in the conformal class $[g_0]$ with scalar curvature $s_{g_0} = s_0$ and that $(m-1)\Delta_{g_0} + s_0$ is invertible. Then there is $T_0 \in (0, T]$ such that each metric $g(t)$, $t \in [0, T_0)$, is a Yamabe metric and the Yamabe constant $Y[g(t)]$ is strictly increasing for $t \in [0, T_0)$ unless g_0 is an Einstein metric.*

This theorem indicates that the conformal Ricci flow is somehow a better family of constant scalar curvature metrics than those obtained in [Koiso 1979].

On asymptotically flat manifolds we use weighted Hölder spaces defined in [Lee and Parker 1987], and define weighted parabolic Hölder spaces based on the similar ones in [Lunardi 1995; Oliynyk and Woolgar 2007].

Theorem 1.3. *Let* $(M^m,\ g_0)$ *be a scalar flat and asymptotically flat manifold with* $g_0 - g_e \in C^{4,\alpha}_{-\tau}$, *where* $\alpha \in (0, 1)$, $\tau \in (0, m-2)$, *and* g_e *is the standard Euclidean metric. Then there exists a small positive number* T *such that the conformal Ricci flow* $g(t)$ *from the initial metric* g_0 *exists for* $t \in [0, T]$ *and* $g(t) - g_e \in C^{1,2+\alpha}_{-\tau}([0, T] \times M)$.

It is easily seen that (1-1) and (1-2) are equivalent on asymptotically flat manifolds because of the uniqueness of bounded solutions to linear parabolic equations on such manifolds. The scalar flat assumption in Theorem 1.3 is less stringent than it looks. Thanks to Schoen and Yau [1979, Lemma 3.3 and Corollary 3.1] we know that one can always conformally deform an asymptotically flat metric with nonnegative scalar curvature into a scalar flat and asymptotically flat metric.

Conformal Ricci flow is the gradient flow for the ADM mass on asymptotically flat manifolds (see Definition 4.1) in the following sense:

Theorem 1.4. *Let* $g(t)$ *be the conformal Ricci flow obtained in Theorem 1.3 for* $\tau \in (\frac{m-2}{2}, m-2)$. *Then*

$$\frac{d}{dt}\boldsymbol{m}(g(t)) = -2 \int_M |\mathrm{Ric}_{g(t)}|^2 \, d\mathrm{vol}_{g(t)} \, .$$

In particular, the ADM mass $\boldsymbol{m}(g(t))$ *is strictly decreasing under conformal Ricci flow, except when* g_0 *is a Euclidean metric.*

As a quick application of Theorem 1.4, one can easily show the rigidity part of the celebrated positive mass theorem of Schoen and Yau [1979]. The monotonicity of the ADM mass along conformal Ricci flow is sharply in contrast to the invariance of the ADM mass along Ricci flow on asymptotically flat manifolds [Dai and Ma 2007; Oliynyk and Woolgar 2007].

The organization of the paper is as follows: In Section 2 we introduce the conformal Ricci flow and establish the monotonicity of the Yamabe constant on compact manifolds. In Section 3 we prove short-time existence of the conformal Ricci flow, both on compact manifolds and on asymptotically flat manifolds. In Section 4 we recall the definition of the ADM mass and show that conformal Ricci flow on asymptotically flat manifolds is the gradient flow for the ADM mass.

2. Conformal Ricci flow

In this section we first introduce the conformal Ricci flow and then calculate evolution equations for curvatures along conformal Ricci flow. We then discuss the monotonicity of Yamabe quotients and Yamabe constants along conformal Ricci flow.

2.1. ***Conformal Ricci flow.*** Suppose that M^m is a smooth m-dimensional manifold and that g_0 is a Riemannian metric on M with constant scalar curvature s_0. In [Fischer 2004], the conformal Ricci flow on M is defined by (1-1) for a family of metrics $g(t)$ with initial condition $g(0) = g_0$ and a family of functions $p = p(t)$ on $M \times [0, T)$.

As shown in [Fischer 2004], the normalization condition $s_{g(t)} = s_0$ in (1-1) may be replaced by an elliptic equation and one can rewrite (1-1) as (1-2). The equivalence between (1-1) and (1-2) was proved in [Fischer 2004, Proposition 3.2 and 3.4] when M is a compact manifold. Based on the evolution equation for scalar curvature, it is easily seen that (1-1) always implies (1-2). Equation (1-2) implies (1-1) when the solution to the linear heat equation is unique, which is true both in the compact and asymptotically flat cases that we consider in this paper.

One important issue for geometric PDEs is the scaling property. It is easy to see that for any constant $\lambda > 0$, if g and p solve the conformal Ricci flow (1-2), then

$$g_\lambda(\,\cdot\,, t)) = \lambda^{-2} g(\,\cdot\,, \lambda^2 t)) \quad \text{and} \quad p_\lambda(\,\cdot\,, t) = \lambda^2 p(\,\cdot\,, \lambda^2 t) \tag{2-1}$$

also solve the conformal Ricci flow.

2.2. ***Curvature evolution equations under conformal Ricci flow.*** To understand conformal Ricci flow one often needs to calculate how curvatures behave along it. The calculations are straightforward. Consider a general geometric flow

$$\frac{\partial}{\partial t} g = -2T. \tag{2-2}$$

We recall that the evolution equations for curvatures are (see [Chow et al. 2006; Besse 1987])

$$\begin{aligned}
\frac{\partial}{\partial t} s &= 2\Delta\Theta - 2\nabla^i\nabla^j T_{ij} + 2R^{ij} T_{ij},\\
\frac{\partial}{\partial t} R_{ij} &= \Delta T_{ij} - \nabla_i\nabla^k T_{kj} - \nabla_j \nabla^k T_{ki} + \nabla_i\nabla_j\Theta + 2R_{ikjl}T^{kl} - R_{ik}T^k{}_j - R_{jk}T^k{}_i,\\
\frac{\partial}{\partial t} R_{ikjl} &= \nabla_i\nabla_j T_{kl} - \nabla_i\nabla_l T_{kj} - \nabla_k\nabla_j T_{il} + \nabla_k\nabla_l T_{ij} - R_{ikjm}T^m{}_l - R_{ikml}T^m{}_j,
\end{aligned}$$

where $\Theta := g^{ij} T_{ij}$. For the conformal Ricci flow, where

$$T = \mathrm{Ric} - \frac{s_0}{m} g + pg \quad \text{and} \quad \Theta = s - s_0 + mp,$$

we calculate the evolution equations for curvatures under the first equation in (1-2) and get

$$\frac{\partial}{\partial t}s = \Delta s + \frac{2s_0}{m}(s - s_0) + 2p(s - s_0) + 2(m-1)\Delta p + 2s_0 p + 2\Big|\mathrm{Ric} - \frac{s_0}{m}g\Big|^2,$$
$$\frac{\partial}{\partial t}R_{ij} = \Delta R_{ij} + 2R_{ikjl}R^{kl} - 2R_{ik}R^{k}{}_{j} + (m-2)\nabla_i\nabla_j p + \Delta p g_{ij},$$
$$\frac{\partial}{\partial t}\mathrm{Rm} = \Delta\mathrm{Rm} + \mathrm{Rm} * \mathrm{Rm} + \mathrm{Ric} * \mathrm{Rm} + \frac{2s_0}{m}\mathrm{Rm} - 2p\,\mathrm{Rm} + \widetilde{T}(\nabla^2 p),$$

where the operator $*$ stands for contractions of tensors, Rm is the Riemann curvature tensor, and

$$\widetilde{T}(\nabla^2 p)_{ikjl} = g_{kl}\nabla_i\nabla_j p - g_{kj}\nabla_i\nabla_l p - g_{il}\nabla_k\nabla_j p + g_{ij}\nabla_k\nabla_l p.$$

2.3. ***Yamabe constants under conformal Ricci flow.*** On compact manifolds along conformal Ricci flow, we may calculate that

$$\Theta = mp, \qquad \frac{\partial}{\partial t}\,d\mathrm{vol}_{g(t)} = -mp\,d\mathrm{vol}_{g(t)},$$
$$\frac{d}{dt}\,\mathrm{vol}(M) = -m\int_M p\,d\mathrm{vol}_g = \frac{m}{s_0}\int_M\Big|\mathrm{Ric} - \frac{s_0}{m}g\Big|^2\,d\mathrm{vol}_g\,. \tag{2-3}$$

Given a compact Riemannian manifold (M^m, h), the Yamabe quotient is defined as

$$Q[h] := \frac{\int_M s_h\,d\mathrm{vol}_h}{\mathrm{vol}_h(M)^{(m-2)/m}}$$

and the Yamabe constant is defined as

$$Y[h] = \inf_{h\in[h]} Q[h].$$

A Riemannian metric h is said to be a Yamabe metric if and only if

$$Q[h] = Y[h].$$

Thus from (2-3) we have this:

Proposition 2.1. *Suppose that $g(t)$, $t \in [0, T)$, is a solution to the conformal Ricci flow (1-2) on a compact manifold with scalar curvature $s_{g_0} \neq 0$. Then the Yamabe quotient $Q[g(t)]$ is strictly increasing unless g_0 is an Einstein metric.*

Next we give a proof of Theorem 1.2 concerning the evolution of Yamabe constants along conformal Ricci flow. As observed in [Wang and Zheng 2011; Chang and Lu 2007; Anderson 2005; Koiso 1979], the Yamabe constant could in general behave rather irregularly among manifolds of positive Yamabe type.

Proof of Theorem 1.2. We prove the result by contradiction. Assume there is a sequence $t_i \to 0^+$ such that $g(t_i)$ are not Yamabe metrics. Let $\tilde{g}_i$ be a Yamabe metric in the conformal class $[g(t_i)]$ of the same volume as $g(t_i)$. By the compactness of the space of Yamabe metrics of fixed volume, $\tilde{g}_i$ converges to a Yamabe metric $g_\infty \in [g_0]$ (taking a subsequence if necessary). By the assumption that g_0 is the only Yamabe metric in $[g_0]$, we have $g_\infty = g_0$. That is to say that both $g(t_i)$ and $\tilde{g}_i$ converge to g_0. Since $(m-1)\Delta_{g_0} + s_0$ is assumed to be invertible, we can apply Koiso's decomposition theorem [1979, Corollary 2.9] in the set of metrics of the same volume as g_0 to conclude the following: In some small neighborhood of g_0, each metric can be written uniquely as the product of a metric of constant scalar curvature near g_0 and a function. Since $\tilde{g}_i$ and $g(t_i)$ are in the same conformal class, have the same volume and both have constant scalar curvature, we get a contradiction. □

Note that similar arguments have been used in [Wang and Zheng 2011; Chang and Lu 2007; Anderson 2005; Koiso 1979].

3. Short-time existence of conformal Ricci flow

In this section we prove the short-time existence of the conformal Ricci flow, i.e., Theorem 1.1 and 1.3. The first step is to combine the two equations in the conformal Ricci flow into one evolution equation with one nonlocal term. More precisely, (1-2) can be written as

$$\frac{\partial}{\partial t} g + 2\Big(\mathrm{Ric} - \frac{s_0}{m} g\Big) = -2\mathcal{P}(g) g \quad \text{on } M, \tag{3-1}$$

where

$$\mathcal{P}(g) = ((m-1)\Delta + s_0)^{-1} \Big|\mathrm{Ric} - \frac{s_0}{m} g\Big|^2,$$

provided that $(m-1)\Delta_{g(t)} + s_0$ is invertible for all $t \in [0, T]$. The strategy to prove the short-time existence for the conformal Ricci flow is similar to the one used in [DeTurck 1983; 2003] to prove the short-time existence for the Ricci flow. We will first prove the short-time existence for DeTurck conformal Ricci flow written as

$$\frac{\partial}{\partial t} g + 2\Big(\mathrm{Ric} - \frac{s_0}{m} g\Big) = -2\mathcal{P}(g) g + \mathcal{L}_W g \quad \text{on } M. \tag{3-2}$$

To prove the short-time existence for (3-2) we calculate the linearization of the DeTurck conformal Ricci flow and apply an implicit function theorem.

3.1. ***DeTurck's trick.*** As a system of differential equations, the conformal Ricci flow is of parabolic-elliptic nature, similar to the Navier–Stokes equations. The significant difference between the conformal Ricci flow and the Navier–Stokes equations is that the conformal Ricci flow is a geometric flow. Hence we need to find ways to eliminate the degeneracy of the conformal Ricci flow arising from the symmetries of diffeomorphisms.

In this subsection we will follow the idea of the improved version [DeTurck 2003] of the approach to short-time existence for the Ricci flow in [DeTurck 1983] to get rid of the degeneracy of diffeomorphisms for conformal Ricci flow. Before introducing DeTurck's trick, we first recall the following operator G. Let g be a Riemannian metric on M^m. The operator G on symmetric 2-tensor B is defined as

$$G(B) = B - \tfrac{1}{2}\mathrm{Tr}_g(B)g. \tag{3-3}$$

Also recall the divergence operator δ

$$(\delta B)_i := \nabla^j B_{ij} : \Gamma(S^2(M)) \to \Gamma(T^*M)$$

and its adjoint operator δ^*

$$(\delta^*\omega)_{ij} := -\tfrac{1}{2}(\omega_{i,j} + \omega_{j,i}) : \Gamma(T^*M) \to \Gamma(S^2(M)).$$

Note that if X is the dual vector field of ω, then $\delta^*\omega = -\mathscr{L}_X g$, where $\mathscr{L}_X$ denotes the Lie derivative in the X direction.

According to DeTurck's improved version [2003] of his approach to the short-time existence of Ricci flow [DeTurck 1983], we consider the following gauge-fixed conformal Ricci flow on M:

$$\begin{cases} \dfrac{\partial}{\partial t}g + 2\Big(\mathrm{Ric} - \dfrac{s_0}{m}g\Big) = -2pg + 2(\delta^*(\tilde{g}^{-1}\delta G(\tilde{g}))), \\ \qquad (m-1)\Delta p + s_0 p = -\Big|\mathrm{Ric} - \dfrac{s_0}{m}g\Big|^2 \end{cases} \tag{3-4}$$

for a family of metrics $g(t)$ with $g(0) = g_0$ and a family of functions $p(t)$ on $M \times [0, T)$, where $\tilde{g}$ is any fixed metric on M.

Suppose that $g(t)$, $t \in [0, T)$, solves (3-4). Then we consider the time-dependent vector field W

$$W^k := g^{ij}\big(\Gamma^k_{ij}[g] - \Gamma^k_{ij}[\tilde{g}]\big), \tag{3-5}$$

where Γ^k_{ij} are the Christoffel symbols of the corresponding metric. It turns out that (see [Hamilton 1995, Section 6; Shi 1989])

$$2\big(\delta^*(\tilde{g}^{-1}\delta G(\tilde{g}))\big) = \mathscr{L}_W g.$$

Hence we can rewrite (3-4) as

$$\begin{cases} \dfrac{\partial}{\partial t}g + 2\Big(\mathrm{Ric} - \dfrac{s_0}{m}g\Big) = -2pg + \mathscr{L}_W g, \\ \qquad (m-1)\Delta p + s_0 p = -\Big|\mathrm{Ric} - \dfrac{s_0}{m}g\Big|^2. \end{cases} \tag{3-6}$$

The conformal Ricci flow (1-2) and the DeTurck conformal Ricci flow (3-6) are related to each other by coordinate changes in the following sense. Suppose that $\hat{g}(t)$ solves the DeTurck conformal Ricci flow (3-6) and W is given as in (3-5). We consider the one-parameter family of diffeomorphisms φ_t generated by W on M, defined by

$$\frac{\partial}{\partial t}\varphi_t(x) = -W(\varphi_t(x), t), \quad \varphi_0(x) = x, \tag{3-7}$$

for some time period $[0, T)$.

Lemma 3.1. *Let $(\hat{g}(t), \hat{p}(t))$, $t \in [0, T)$, be a solution to the DeTurck conformal Ricci flow (3-6) on the manifold M^m with the initial metric g_0. Assume that the solution $\varphi_t(x)$ to (3-7) exists for $t \in [0, T)$. Let*

$$g(t) := \varphi_t^* \hat{g}(t) \quad \text{and} \quad p(t) := \hat{p}(\varphi_t(x), t).$$

Then $(g(t), p(t))$, $t \in [0, T)$, is a solution to the conformal Ricci flow (1-2) on the manifold M with $g(0) = g_0$.

Proof. By using (3-6) we simply compute that (cf. [Chow et al. 2006, Section 2.6])

$$\begin{aligned}\frac{\partial}{\partial t} g(t) &= \varphi_t^*\Big(\frac{\partial}{\partial t}\hat{g}(t)\Big) + \frac{\partial}{\partial s}\Big|_{s=0}\big(\varphi_{t+s}^*\hat{g}(t)\big)\\ &= -2\varphi_t^*\Big(\mathrm{Ric}_{\hat{g}} - \frac{s_0}{m}\hat{g} + \hat{p}\hat{g}\Big) + \varphi_t^*\big(\mathscr{L}_W \hat{g}\big) - \mathscr{L}_{(\varphi_t^{-1})_* W}(\varphi_t^*\hat{g})\\ &= -2\Big(\mathrm{Ric}_g - \frac{s_0}{m} g + pg\Big).\end{aligned}$$

The second equation for p in (1-2) is readily seen to hold, since the scalar curvature under both flows is kept constant as s_0. □

This lemma is particularly important to us because it enables us to prove the short-time existence of the conformal Ricci flow by proving the short-time existence of the DeTurck conformal Ricci flow. The later will be shown to be a system of parabolic-elliptic equations (see Lemma 3.3).

On the other hand, suppose that $(g(t), p(t))$, $t \in [0, T)$, solves the conformal Ricci flow (1-2) on M with initial metric g_0. Let $\tilde{g}$ be any fixed metric on M. We then consider the harmonic map flow

$$\frac{\partial}{\partial t}\varphi_t = \Delta_{g(t),\tilde{g}}\varphi_t, \quad \varphi_0 = \mathrm{Id} \tag{3-8}$$

for $\varphi_t : M \to M$, where the nonlinear Laplacian in local coordinates is

$$\big(\Delta_{g_1,g_2} f\big)^\gamma = \Delta_{g_1} f^\gamma + \Gamma_{\alpha\beta}^\gamma[g_2]\frac{\partial f^\alpha}{\partial x^i}\frac{\partial f^\beta}{\partial x^j} g_1^{ij}.$$

The following lemma is useful for deriving the uniqueness of the conformal Ricci flow from the uniqueness of the DeTurck conformal Ricci flow. From it readily follows that the uniqueness of the conformal Ricci flow with a given initial metric holds at least on compact manifolds.

Lemma 3.2. *Let* $(g(t), p(t))$, $t \in [0, T)$, *be a solution to the conformal Ricci flow* (1-2) *on the manifold* M^m *with the initial metric* g_0. *Assume that the solution* $\varphi_t : M \to M$ *to the harmonic map flow* (3-8) *exists for* $t \in [0, T)$. *Let*

$$\hat{g}(t) := (\varphi_t^{-1})^* g(t) \quad \textit{and} \quad \hat{p}(x, t) := p(\varphi_t^{-1}(x), t).$$

Then $(\hat{g}(t), \hat{p}(t))$, $t \in [0, T)$, *is a solution to the DeTurck conformal Ricci flow* (3-6) *on the manifold* M *with the initial metric* g_0.

Proof. This follows from a calculation like the one in the proof of Lemma 3.1, after identifying the vector field W with $\Delta_{g(t),\tilde{g}}\varphi_t$ (see [Chow et al. 2006, p. 117]). □

3.2. ***Linearization of DeTurck conformal Ricci flow.*** In this subsection we compute the linearization of the DeTurck conformal Ricci flow (3-2). To do so we set

$$g_\lambda(t) = g(t) + \lambda h(t)$$

for a family of symmetric 2-tensors $h(t)$ and for $\lambda \in (-\epsilon, \epsilon)$. We rewrite the DeTurck conformal Ricci flow as

$$\mathcal{M}(g(t)) = \frac{\partial}{\partial t} g + 2\left(\mathrm{Ric} - \frac{s_0}{m} g\right) + 2\mathcal{P}(g)g - \mathcal{L}_W g =: \frac{\partial}{\partial t} g - \mathcal{F}(g(t)) = 0 \tag{3-9}$$

and calculate

$$\frac{d}{d\lambda}\bigg|_{\lambda=0} \mathcal{M}(g_\lambda).$$

To compute the linearization of $\mathcal{P}$ we first calculate (cf. [Chow et al. 2006, Equation (S.5), p. 547])

$$\frac{d}{d\lambda}\bigg|_{\lambda=0} \Delta_{g_\lambda}\mathcal{P}(g_\lambda) = -h_{ij}\nabla^i\nabla^j\mathcal{P} - \tfrac{1}{2}(2\nabla^i h_{ij} - \nabla_j h^i{}_i)\nabla^j\mathcal{P} + \Delta\mathcal{P}',$$

where $\mathcal{P}' = \frac{d}{d\lambda}\Big|_{\lambda=o} \mathcal{P}(g_\lambda)$. Next we may calculate

$$\frac{d}{d\lambda}\bigg|_{\lambda=0} \left|\mathrm{Ric}_{g_\lambda} - \frac{s_0}{m} g_\lambda\right|^2$$

using the linearization of Ricci curvature

$$\begin{aligned}\frac{d}{d\lambda}\bigg|_{\lambda=0} 2R_{ij}[g_\lambda] = -\Delta h_{ij} - 2R_{ikjl}h_{kl} + R_{ik}h^k{}_j + R_{jk}h^k{}_i \\ - \nabla_i\nabla_j h^k{}_k + \nabla_i\nabla^k h_{kj} + \nabla_j\nabla^k h_{ki}\end{aligned} \tag{3-10}$$

(see [Chow et al. 2006, (2.31)]). In summary we have

$$(m-1)\Delta\mathcal{P}' + s_0\mathcal{P}' - P_1^{ijkl}\nabla_i\nabla_j h_{kl} - P_2^{ijk}\nabla_i h_{jk} + P_3^{ij} h_{ij} = 0,$$

that is,

$$\mathcal{P}' = ((m-1)\Delta + s_0)^{-1}(P_1 * \nabla^2 h + P_2 * \nabla h + P_3 * h), \tag{3-11}$$

where P_1, P_2, P_3 are tensors that depend on curvature of $g(t)$ and up to second-order spatial derivatives of p.

In the calculation of the linearization of $\mathcal{M}$ the crucial step is to calculate

$$\frac{d}{d\lambda}\Big|_{\lambda=0} \mathcal{L}_{W_\lambda} g_\lambda = -\frac{d}{d\lambda}\Big|_{\lambda=0} \delta^*[g_\lambda]\omega_\lambda,$$

where $(\omega_\lambda)_i := (g_\lambda)_{ik} W_\lambda^k$ and $W_\lambda = -\tilde{g}^{-1}\delta[g_\lambda]G[g_\lambda](\tilde{g})$. In fact, the key point of DeTurck's trick is to collect the second-order covariant derivatives of h in the above and realize that they cancel out the second line in (3-10). To see that we first collect terms involving the first-order covariant derivatives of h in

$$\frac{d}{d\lambda}\Big|_{\lambda=0} (\omega_\lambda)_i = \nabla^k h_{ki} - \tfrac{1}{2}\nabla_i h^k{}_k + \text{ other terms.}$$

Then we collect the second-order covariant derivatives of h in

$$\frac{d}{d\lambda}\Big|_{\lambda=0} ((\delta_{g_\lambda})^*\omega_\lambda)_{ij} = -\nabla_i \nabla^k h_{ki} - \nabla_j \nabla^k h_{kj} + \nabla_i \nabla_j h^k{}_k + \text{other terms.} \tag{3-12}$$

Therefore

$$\frac{d}{d\lambda}\Big|_{\lambda=0} \mathcal{M}(g_\lambda) = \frac{\partial}{\partial t} h - \Delta h + 2\mathcal{P}' g + M_1^{ijk}\nabla_i h_{jk} + M_2^{ij} h_{ij}, \tag{3-13}$$

where M_1 depends only $g(t)$ and $\mathcal{P}(g)$ and M_2 depends on the curvature of $g(t)$ and $\mathcal{P}(g)$. To summarize, we have this:

Lemma 3.3. *Suppose that $g(t)$, $t \in [0, T]$, is a family of metrics such that the elliptic operator $(m-1)\Delta_{g(t)} + s_0$ is invertible for all $t \in [0, T]$. Then the linearization of the DeTurck conformal Ricci flow equations (3-2) at the metrics $g(t)$ in the directions of the symmetric 2-tensors $h(t)$ is*

$$D\mathcal{M}(g)(h) = \frac{\partial}{\partial t} h - \Delta h + 2\mathcal{P}' g + M_1 * \nabla h + M_2 * h, \tag{3-14}$$

where

$$\mathcal{P}' = ((m-1)\Delta + s_0)^{-1}(P_1 * \nabla^2 h + P_2 * \nabla h + P_3 * h).$$

Here P_1, P_2, P_3 are tensors depending on curvature of $g(t)$ and up to the second-order derivatives in spatial variables of $\mathcal{P}(g)$, and M_1, M_2 are tensors depending on the curvature of $g(t)$ and the function $\mathcal{P}(g)$.

3.3. *Short-time existence on closed manifolds.* Let us first solve the conformal Ricci flow on a compact manifold M^m. There are many books that are good references for linear and nonlinear systems of parabolic equations. We will mostly use the book [Lunardi 1995, §5.1], in particular its Theorem 5.1.21, for existence and standard estimates. We adopt the definitions of parabolic Hölder spaces from [Lunardi 1995, pp. 175–177]. We use the same notations for parabolic Hölder spaces of functions and of tensor fields when there is no confusion. To define norms for tensor fields we may use the initial metric and local coordinate charts.

3.3.1. *Preliminaries.* Since we deal with systems of parabolic-elliptic equations, we need to consider elliptic estimates with a time parameter. There is an advantage to using only the supremum norm in the time variable, as indicated by the following lemma.

Lemma 3.4. *Let $g(t)$, $t \in [0, T]$, be a family of smooth Riemannian metrics on a compact manifold M^m. Suppose the operator $(m-1)\Delta_{g(t)} + s_0$ is invertible for $t \in [0, T]$. Then the equation*

$$(m-1)\Delta_{g(t)} p(t) + s_0 p(t) = \gamma \tag{3-15}$$

has a unique solution $p \in C^{0,2+\alpha}$ for each $\gamma \in C^{0,\alpha}$. Moreover, p satisfies the estimate

$$\|p\|_{C^{0,2+\alpha}} \le C\|\gamma\|_{C^{0,\alpha}} \tag{3-16}$$

for some constant C independent of γ.

Proof. In the light of the standard Schauder estimates for elliptic PDEs, we only need to verify that $p(t)$ is continuous in the time variable, which is a consequence of the classical Bernstein estimates. □

The following interpolatory inclusion will be useful in the proof of the short-time existence (cf. [Lunardi 1995, Lemma 5.1.1]).

Lemma 3.5. *There is a constant C independent of T such that for any $t_1, t_2 \in [0, T]$ we have*

$$\|h(t_1, \cdot) - h(t_2, \cdot)\|_{C^{k-2,\alpha}} \le C \cdot |t_1 - t_2| \cdot \|h\|_{C^{1,k+\alpha}}$$

for all $h \in C^{1,k+\alpha}([0, T] \times M)$.

3.3.2. *On linearized DeTurck conformal Ricci flow.* We first solve the linearized DeTurck conformal Ricci flow

$$\begin{cases} D\mathcal{M}(g)(h) = \dfrac{\partial}{\partial t}h - \Delta h + 2\mathcal{P}'g + M_1 * \nabla h + M_2 * h = \gamma, \\ \quad h(0, \cdot) = 0 \end{cases} \tag{3-17}$$

for appropriately given metrics $g(t)$ for each $\gamma \in C^{0,\alpha}$.

Proposition 3.6. *Suppose that $g(t)$, $t \in [0, T]$, is a family of metrics such that the elliptic operator $(m-1)\Delta_{g(t)} + s_0$ is invertible for all $t \in [0, T]$. Then, for $\gamma \in C^{0,\alpha}$, the initial value problem for (3-17) has a unique solution $h \in C^{1,2+\alpha}$. Moreover,*

$$\|h\|_{C^{1,2+\alpha}([0,T]\times M)} \le C\|\gamma\|_{C^{0,\alpha}([0,T]\times M)}. \tag{3-18}$$

Proof. To use a contractive mapping-type argument we consider the Banach space

$$E_1([0, T^*]) = \{h \in C^{0,2+\alpha} : h(0, \cdot) = 0\}.$$

Given $\tilde{h} \in E_1([0, T^*])$, based on [Lunardi 1995, Theorem 5.1.21], we first solve a usual system of linear parabolic equations

$$\begin{cases} \dfrac{\partial}{\partial t}h - \Delta h + M_1 * \nabla h + M_2 * h = \tilde{\gamma}, \\ h(0, \cdot) = 0, \end{cases} \tag{3-19}$$

where $\tilde{\gamma} = \gamma - 2\mathcal{P}'(\tilde{h})g \in C^{0,\alpha}$ and $\mathcal{P}'(\tilde{h})$ is defined by (3-11). We remark that it takes some work to extend Theorem 5.1.21 in [Lunardi 1995] to be applicable to our context, but there are no significant issues in doing so. Hence we may define a map

$$\Psi : E_1([0, T^*]) \to E_1([0, T^*]), \quad \Psi(\tilde{h}) = h.$$

Note that if we set

$$v = \Psi(\tilde{h}_1) - \Psi(\tilde{h}_2),$$

then v satisfies

$$\begin{cases} \dfrac{\partial}{\partial t}v - \Delta v + M_1 * \nabla v + M_2 * v = 2(\mathcal{P}'(\tilde{h}_2) - \mathcal{P}'(\tilde{h}_1))g, \\ v(0, \cdot) = 0. \end{cases}$$

Since

$$\|(\mathcal{P}'(\tilde{h}_2) - \mathcal{P}'(\tilde{h}_1))g\|_{C^{0,2+\alpha}} \le C\|\tilde{h}_1 - \tilde{h}_2\|_{C^{0,2+\alpha}}$$

holds by (3-11) and Lemma 3.4, we obtain again from the estimates based on Theorem 5.1.21 in [Lunardi 1995] that

$$\|v\|_{C^{1,4+\alpha}} \le C\|\tilde{h}_1 - \tilde{h}_2\|_{C^{0,2+\alpha}}.$$

In the light of Lemma 3.5, we thus have

$$\|v(t_1) - v(t_2)\|_{C^{2,\alpha}} \le C \cdot |t_1 - t_2| \cdot \|\tilde{h}_1 - \tilde{h}_2\|_{C^{0,2+\alpha}}.$$

In particular,

$$\|\Psi(\tilde{h}_1) - \Psi(\tilde{h}_2)\|_{C^{0,2+\alpha}} \le CT^*\|\tilde{h}_1 - \tilde{h}_2\|_{C^{0,2+\alpha}}.$$

To apply the contractive mapping theorem we observe that

$$\|\Psi(\tilde{h})\|_{C^{0,2+\alpha}} \leq \|\Psi(0)\|_{C^{0,2+\alpha}} + CT^*\|\tilde{h}\|_{C^{0,2+\alpha}},$$

where

$$\|\Psi(0)\|_{C^{1,2+\alpha}} \leq C_0\|\gamma\|_{C^{0,\alpha}}$$

for some constant C_0, from the estimates based on Theorem 5.1.21 in [Lunardi 1995]. Thus

$$\Psi : B_R = \{h \in E_1([0, T^*]) : \|h\|_{C^{0,2+\alpha}} \leq R\} \to B_R,$$

for $R = 2C_0\|\gamma\|_{C^{0,\alpha}}$, is a contractive mapping when T^* is appropriately small. Then, by the uniqueness of the solution of the linear parabolic equation (3-17), one may extend the solution of (3-17) to $[0, T]$ by steps in time of length T^*. The estimate (3-18) follows from the estimates based on Theorem 5.1.21 in [Lunardi 1995]. □

To summarize, we have established that

$$D\mathcal{M}(g) : C^{1,2+\alpha}([0, T] \times M) \cap \{h(0, \cdot) = 0\} \to C^{0,\alpha}([0, T] \times M)$$

is an isomorphism, provided that $g(t)$ satisfies the assumptions in Proposition 3.6.

3.3.3. *Implicit function theorem argument.* Next we solve the DeTurck conformal Ricci flow and then the conformal Ricci flow. Our approach is to use an implicit function theorem. Let us start with the following general implicit function theorem.

Lemma 3.7. *Let X and Y be Banach spaces and let*

$$\mathcal{H} : X \to Y$$

be a C^1 map. Suppose that for a point $x_0 \in X$ there are positive numbers δ and C such that

$$\|(D\mathcal{H}(x))^{-1}\| \leq C \qquad \text{for all } x \in B_\delta(x_0),$$

$$\|D\mathcal{H}(x_1) - D\mathcal{H}(x_2)\| \leq \frac{1}{2C} \quad \text{for all } x_1, x_2 \in B_\delta(x_0).$$

Then, if

$$\|\mathcal{H}(x_0)\| \leq \frac{\delta}{2C},$$

there is $x \in B_\delta(x_0)$ such that

$$\mathcal{H}(x) = 0.$$

To apply the above implicit function theorem to the map

$$\mathcal{M} : C^{1,2+\alpha}([0, T] \times M) \cap \{g(0) = g_0\} \to C^{0,\alpha}([0, T] \times M)$$

to solve the DeTurck conformal Ricci flow we need to show that $\mathcal{M}$ is continuously differentiable. In fact we have the following lemma.

Lemma 3.8. *Let M^m be a compact manifold and let $g(t) \in C^{1,2+\alpha}([0,T]\times M)$ be a family of metrics such that the elliptic operator $(m-1)\Delta_{g(t)} + s_0$ is invertible for all $t \in [0,T]$. Then there is a $\delta_0 > 0$ such that*

$$\|D\mathcal{M}(g_1) - D\mathcal{M}(g_2)\|_{L(C^{1,2+\alpha},C^{0,\alpha})} \le C\|g_1 - g_2\|_{C^{1,2+\alpha}}$$

for $\|g_i - g\|_{C^{1,2+\alpha}} \le \delta_0$ in $C^{1,2+\alpha}([0,T]\times M)$ and $i = 1, 2$.

Proof. We calculate, for any $h \in C^{1,2+\alpha} \bigcap \{h(0,\cdot)=0\}$, that

$$\begin{aligned}(D\mathcal{M}(g_1) - D\mathcal{M}(g_2))h = (\Delta_{g_2} - \Delta_{g_1})h &+ 2\mathcal{P}'[g_1](g_1 - g_2)\\ &+ M_1[g_1] * (\nabla_{g_1} h - \nabla_{g_2} h) + (M_1[g_1] - M_1[g_2]) * \nabla_{g_2} h\\ &+ (M_2[g_1] - M_2[g_2]) * h + 2(\mathcal{P}'[g_1] - \mathcal{P}'[g_2])g_2.\end{aligned}$$

It is easily seen that

$$\begin{aligned}\|M_1[g_1] * (\nabla_{g_1} h - \nabla_{g_2} h) + (M_1[g_1] - M_1[g_2]) * \nabla_{g_2} h\|_{C^{0,\alpha}}&\\ \le C\|g_1 - g_2\|_{C^{1,2+\alpha}}\|h\|_{C^{1,2+\alpha}},&\\ \|(M_2[g_1] - M_2[g_2])h\|_{C^{0,\alpha}} \le C\|g_1 - g_2\|_{C^{1,2+\alpha}}\|h\|_{C^{1,2+\alpha}},&\\ \|\Delta_{g_1} h - \Delta_{g_2} h\|_{C^{0,\alpha}} \le C\|g_1 - g_2\|_{C^{1,2+\alpha}}\|h\|_{C^{1,2+\alpha}}.&\end{aligned}$$

It is also easy to see that

$$\|\mathcal{P}'[g_1](g_1 - g_2)\|_{C^{0,\alpha}} \le C\|g_1 - g_2\|_{C^{1,2+\alpha}}\|h\|_{C^{1,2+\alpha}}$$

under the assumption that $\|g_i - g\|_{C^{1,2+\alpha}} \le \delta_0$, $i = 1, 2$, from the definition of $\mathcal{P}'$ in (3-11).

For the last remaining term we write

$$\begin{aligned}((m-1)\Delta + s_0)(\mathcal{P}'[g_2] - \mathcal{P}'[g_1]) = (m-1)(\Delta_{g_1} - \Delta_{g_2})\mathcal{P}'[g_1]&\\ + (P_1[g_2] - P_1[g_1]) * \nabla^2_{g_2} h + P_1[g_1] * (\nabla^2_{g_2} - \nabla^2_{g_1})h&\\ + (P_2[g_2] - P_2[g_1]) * \nabla_{g_2} h + P_2[g_1] * (\nabla_{g_2} - \nabla_{g_1})h&\\ + (P_3[g_2] - P_3[g_1]) * h&\end{aligned}$$

and apply Lemma 3.4. Then

$$\|\mathcal{P}'[g_1] - \mathcal{P}'[g_2]\|_{C^{0,\alpha}} \le C\|g_1 - g_2\|_{C^{1,2+\alpha}}\|h\|_{C^{1,2+\alpha}},$$

which implies that

$$\|(\mathcal{P}'[g_1] - \mathcal{P}'[g_2])g\|_{C^{0,\alpha}} \le C\|g_1 - g_2\|_{C^{1,2+\alpha}}\|h\|_{C^{1,2+\alpha}}.$$

Thus the proof is complete. □

To apply Lemma 3.7, we consider the initial approximate solution

$$\bar{g}(t) = g_0 + t\mathcal{F}(g_0), \tag{3-20}$$

where $\mathcal{F}$ was introduced in (3-9). We then calculate that

$$\begin{aligned} \mathcal{M}(\bar{g}) &= -\mathcal{F}(g_0 + t\mathcal{F}(g_0)) + \mathcal{F}(g_0) \\ &= -t\int_0^1 D\mathcal{F}(g_0 + \theta t\mathcal{F}(g_0))d\theta \cdot \mathcal{F}(g_0). \end{aligned} \tag{3-21}$$

Now we are ready to state and prove the short-time existence theorem for the conformal Ricci flow (a precise form of Theorem 1.1).

Theorem 3.9. *Let M^m be a compact manifold with no boundary. Suppose that $g_0 \in C^{4,\alpha}$ is a Riemannian metric on M such that the scalar curvature $s_{g_0} = s_0$ is constant and that the elliptic operator $(m-1)\Delta_{g_0} + s_0$ is invertible. Then there exists a small positive number T such that the conformal Ricci flow $g(t)$ exists in $C^{1,2+\alpha}$ from the initial metric g_0 for $t \in [0, T]$.*

Proof. First we notice that Proposition 3.6 holds for the family of metrics $\bar{g}(t) = g_0 + t\mathcal{F}(g_0)$ in $C^{1,2+\alpha}$, for some appropriately small T such that the elliptic operator $(m-1)\Delta_{\bar{g}} + s_0$ is invertible for all $t \in [0, T]$. Therefore there is a constant C and a small number δ_0 such that

$$\|(D\mathcal{M}(g))^{-1}\| \leq C \quad \text{and} \quad \|D\mathcal{M}(g_1) - D\mathcal{M}(g_2)\| \leq \frac{1}{2C}$$

for all $g, g_1, g_2, \in B(\delta_0)$, where $B(\delta_0) = \{g \in C^{1,2+\alpha} : \|g - \bar{g}\|_{C^{1,2+\alpha}} \leq \delta_0\}$, according to Lemma 3.8. Next, after choosing an even smaller T if necessary, we observe from (3-21) that

$$\|\mathcal{M}(\bar{g})\|_{C^{0,\alpha}} \leq \frac{\delta_0}{2C}.$$

Hence Lemma 3.7 implies that DeTurck conformal Ricci flow $\hat{g}(t)$ exists in $C^{1,2+\alpha}$ with the initial metric g_0. Therefore, applying Lemma 3.1, we obtain the short-time existence for conformal Ricci flow from the initial metric g_0, since (3-7) is always solvable for short time. □

3.4. ***Short-time existence on asymptotically flat manifolds.*** In this subsection we establish the short-time existence of the conformal Ricci flow on asymptotically flat manifolds. The idea of the proof is the same as of the proof in last subsection. We remark here that the short-time existence of the Ricci flow on asymptotically flat manifolds has been established independently in [Dai and Ma 2007; Oliynyk and Woolgar 2007]. The approach in [Dai and Ma 2007] is to use the short-time existence result in [Shi 1989] and the maximum principle to show that the Ricci flow in fact remains asymptotically flat when starting from an asymptotically flat metric, while the approach in [Oliynyk and Woolgar 2007] is to establish short-time existence of the Ricci flow based on weighted function spaces. Our approach is similar to the one in [Oliynyk and Woolgar 2007], since neither short-time existence on noncompact manifolds nor a maximum principle are available for the conformal Ricci flow.

3.4.1. *Analysis on asymptotically flat manifolds.* We first briefly introduce asymptotically flat manifolds according to [Lee and Parker 1987] and then construct appropriate parabolic Hölder spaces on them.

Definition 3.10 [Lee and Parker 1987, Definition 6.3]. A Riemannian manifold M^m with a C^2-metric g is called asymptotically flat of order $\tau > 0$ if there exists a decomposition $M = M_0 \cup M_\infty$ (with M_0 compact) and a diffeomorphism $\Psi : M_\infty \to \mathbb{R}^n \setminus B_R(\vec{0})$ for some $R > 0$, satisfying

$$g(z) = g_e(z) + O(\rho^{-\tau}), \quad \partial_k g(z) = O(\rho^{-\tau-1}), \quad \partial_k \partial_l g(z) = O(\rho^{-\tau-2}),$$

where g_e is the standard Euclidean metric and $\rho = \rho(z) = |z| \to \infty$ in the coordinates $z = (z^1, \cdots, z^m)$ induced on M_∞ by the diffeomorphism Ψ.

We give the definition of weighted Hölder spaces $C^{k,\alpha}_\beta$ from [Lee and Parker 1987, p. 75]. Again we will use the same notations for weighted Hölder spaces of functions and of tensor fields if there is no confusion. We use local coordinate charts and a given metric whenever it is necessary for the definition of Hölder spaces of tensor fields on asymptotically flat manifolds.

Fix a number $T > 0$. Analogous to [Lunardi 1995, pp. 175–177], we define a parabolic weighted Hölder space

$$C^{0,k+\alpha}_\beta := \Big\{ h \in C(M \times [0,T]) : h(t) \in C^{k,\alpha}_\beta \text{ and } \max_{t\in[0,T]} \|h(t)\|_{C^{k,\alpha}_\beta} < \infty \Big\}$$

with the norm

$$\|h\|_{C^{0,k+\alpha}_\beta} := \max_{t\in[0,T]} \|h(t)\|_{C^{k,\alpha}_\beta}.$$

Similarly we define the space

$$C^{1,k+\alpha}_\beta := \big\{ h \in C^{0,k+\alpha}_\beta \text{ and } \partial_t h \in C^{0,k-2+\alpha}_{\beta-2} \big\}$$

with the norm

$$\|h\|_{C^{1,k+\alpha}_\beta} := \max_{t\in[0,T]} \|h(t)\|_{C^{k,\alpha}_\beta} + \max_{t\in[0,T]} \|\partial_t h(t)\|_{C^{k-2,\alpha}_{\beta-2}}.$$

We now recall the elliptic theory for weighted Hölder spaces, for example, from [Lee and Parker 1987, Theorem 9.2] in our context.

Lemma 3.11. *Let $(M^m, g(t))$, for $t \in [0, T]$, be a family of asymptotically flat manifolds with $g(t) - g_e \in C^{0,2+\alpha}_{-\tau}$ for $\tau > 0$. Then*

$$\Delta_{g(t)} : C^{0,2+\alpha}_{\beta} \to C^{0,\alpha}_{\beta-2}$$

is an isomorphism for $\beta \in (2-m, 0)$, that is, there is C such that

$$\|u\|_{C^{0,2+\alpha}_{\beta}} \le C\|\Delta_{g(t)}u\|_{C^{0,\alpha}_{\beta-2}}.$$

Analogous to Lemma 3.5 we have a simple interpolatory inclusion.

Lemma 3.12. *There is a constant C independent of T such that for any $t_1, t_2 \in [0, T]$, we have*

$$\|h(\cdot, t_1) - h(\cdot, t_2)\|_{C^{k-2,\alpha}_{\beta-2}} \le C \cdot |t_1 - t_2| \cdot \|h\|_{C^{1,k+\alpha}_{\beta}}$$

for all $h \in C^{1,k+\alpha}_{\beta}(M \times [0, T])$.

3.4.2. *Short-time existence on asymptotically flat manifolds.* Here we assume that the initial metric g_0 on M^m is asymptotically flat and scalar flat. Thanks to [Schoen and Yau 1979, Lemma 3.3 and Corollary 3.1], we know that one can always conformally deform an asymptotically flat metric with nonnegative scalar curvature into a scalar flat asymptotically flat metric. We will use the strategy of Section 3.3 to prove the short-time existence of conformal Ricci flow on asymptotically flat manifolds.

First with changes of notations we are able to prove the isomorphism analogous to Proposition 3.6. An extension of [Lunardi 1995, Theorem 5.1.21] to the weighted parabolic Hölder spaces on asymptotically flat manifolds may be proven by the standard argument through interior estimates and scaling invariance of the interior estimates (cf. [Oliynyk and Woolgar 2007; Bartnik 1986; Lee and Parker 1987]). The key is to realize that one may move in and out the weight for local estimates.

Proposition 3.13. *Suppose that $g(t)$, $t \in [0, T_0]$, is a family of asymptotically flat metrics with $g(t) - g_e \in C^{0,2+\alpha}_{-\tau}$ for $\tau \in (0, m-2)$. Then there is a $T_* \in (0, T_0]$ such that, for any $T \le T_*$ and $\gamma \in C^{0,\alpha}_{-\tau-2}$, the initial value problem for (3-17) has a unique solution $h \in C^{1,2+\alpha}_{-\tau}$. Moreover,*

$$\|h\|_{C^{1,2+\alpha}_{-\tau}(M\times[0,T])} \le C\|\gamma\|_{C^{0,\alpha}_{-\tau-2}(M\times[0,T])}.$$

This is to say that, for $\tau \in (0, m-2)$,

$$D\mathcal{M}(g) : C^{1,2+\alpha}_{-\tau}(M \times [0, T]) \cap \{h(0, \cdot) = 0\} \to C^{0,\alpha}_{-\tau-2}(M \times [0, T])$$

is an isomorphism, provided that $g(t)$ and T satisfy the assumptions in the above Proposition 3.13. The restriction on the order τ of weight is solely used in solving elliptic equations on weighted spaces in Lemma 3.11.

To obtain a short-time existence of the DeTurck conformal Ricci flow we again apply an implicit function theorem (Lemma 3.7) to the map

$$\mathcal{M}: \{g(t) : g(t) - g_e \in C^{1,2+\alpha}_{-\tau}(M \times [0, T]) \text{ and } g(0) = g_0\} \to C^{0,\alpha}_{-\tau-2}(M \times [0, T])$$

for any $\tau \in (0, m-2)$ and T given from Proposition 3.13. Finally, we arrive at the short-time existence of the conformal Ricci flow.

Theorem 3.14. *Let (M^m, g_0) be scalar flat and asymptotically flat with $g_0 - g_e \in C^{4,\alpha}_{-\tau}$ and $\tau \in (0, m-2)$. Then there exists a small positive number T such that the conformal Ricci flow $g(t)$ from the initial metric g_0 exists for $t \in [0, T]$ and $g(t) - g_e \in C^{1,2+\alpha}_{-\tau}(M \times [0, T])$.*

Proof. As in Section 3.3.3 we first verify that

$$\|D\mathcal{M}(g_1) - D\mathcal{M}(g_2)\|_{L(C^{1,2+\alpha}_{-\tau}, C^{0,\alpha}_{-\tau-2})} \le C\|g_1 - g_2\|_{C^{1,2+\alpha}_{-\tau}}.$$

The proof goes like the one for Lemma 3.8 with only changes of notation. We then construct

$$\bar{g}(t) = g_0 + t\mathcal{F}(g_0) \in g_e + C^{1,2+\alpha}_{-\tau}$$

as in Section 3.3.3 for $g_0 - g_e \in C^{4,\alpha}_{-\tau}$. Another issue one needs to take care of is solving (3-7) to construct the conformal Ricci flow from the DeTurck conformal Ricci flow. But, since $W \in C^{0,1+\alpha}_{-\tau-1}$, it is easy to solve (3-7) on the whole manifold M for some short time. The rest of the proof goes like the one in Section 3.3.3 for Theorem 3.9 with little changes except in notation. Notice that the equivalence between (1-1) and (1-2) holds because of the uniqueness of the bounded solution of a linear parabolic equation on an asymptotically flat manifold. □

4. ADM mass under conformal Ricci flow

Asymptotically flat manifolds are used in general relativity to describe isolated gravitational systems. The fundamental geometric invariant of an asymptotically flat manifold is called the mass of the gravitational system. The so-called ADM mass of an asymptotically flat manifold was first defined in [Arnowitt et al. 1960].

In general relativity the world is modeled by a 4-dimensional spacetime X^4 with a Lorentzian metric g. The physical law that describes the gravity induced by matter in the spacetime is the famous Einstein equation

$$\mathrm{Ric}_g - \tfrac{1}{2}s_g g = T,$$

where T is the energy-momentum-stress tensor that is supposed to reflect the nature and state of matter in the spacetime. A time slice of a space-time that represents an isolated gravitational system is an asymptotically flat 3-manifold M^3.

One of the most important solutions of Einstein equations is the Schwarzschild spacetime, which represents the gravitational system of a static point particle of mass $\boldsymbol{m}$ and whose time slice is an asymptotically flat metric

$$g_{\mathrm{Sch}} = g_e + \frac{\boldsymbol{m}}{\rho} g_e + O(\rho^{-2})$$

on the punctured $\mathbb{R}^3$. The crucial test to validate the notion of mass in relativity is whether its predictions reduce to those of Newtonian gravity under the circumstances where Newtonian theory is known to be valid; when gravity is weak, motions are much slower than the speed of the light, and material stresses are much smaller than the mass-energy density (cf. [Wald 1984, 4.4]).

We now follow [Lee and Parker 1987, Definition 8.2] to introduce ADM mass for asymptotically flat manifolds.

Definition 4.1. Given an asymptotically flat Riemannian manifold (M^m, g) with asymptotic coordinates z, we define the ADM mass by (if the limit exists)

$$\boldsymbol{m}(g) = \lim_{R\to\infty} \omega_{m-1}^{-1} \int_{\mathbb{S}_R} (\partial_i g_{ij} - \partial_j g_{ii}) n^j \, d\sigma,$$

where ω_{m-1} is the volume of the unit sphere $\mathbb{S}^{m-1}$, $\vec{n} = (n^1, \dots, n^m)$ is the outward unit normal vector of the sphere $\mathbb{S}_R = \{z \in \mathbb{R}^m, |z| = R\}$ and $d\sigma$ is the area element of $\mathbb{S}_R$.

Recall from [Lee and Parker 1987] that

$$\mathcal{M}_\tau := \left\{ g = g_e + h : h \in C^{1,\alpha}_{-\tau} \text{ and } \partial_j \partial_i h_{ij} - \partial_j \partial_j h_{ii} \in L^1(M, d\mathrm{vol}_{g_e}) \right\}.$$

After Definition 4.1 one wonders if the ADM mass is indeed a geometric invariant for the asymptotically flat metric. This was confirmed by the following result:

Lemma 4.2 [Arnowitt et al. 1960; Bartnik 1986]. *Suppose that g is an asymptotically flat metric in $\mathcal{M}_\tau$ on M^m for $\tau > \frac{m-2}{2}$. Then the ADM mass $\boldsymbol{m}(g)$ is independent of the choice of asymptotic coordinates at infinity.*

Another important fact about the ADM mass is the following, observed in [Lee and Parker 1987, (8.11); see also Lemma 9.4].

Lemma 4.3. *Let $g(t)$ be a smooth family of asymptotically flat metrics in $\mathcal{M}_\tau$ on M^m, for $\tau > \frac{m-2}{2}$. Then the mass $\boldsymbol{m}(g(t))$ is differentiable and*

$$\frac{d}{dt}\left(-\int_M s_{g(t)} \, d\mathrm{vol}_{g(t)} + \omega_{m-1} \boldsymbol{m}(g(t)) \right) = \int_M G[g(t)] \cdot \varphi(t) \, d\mathrm{vol}_{g(t)},$$

where $G[g(t)] = \mathrm{Ric}_{g(t)} - \frac{1}{2} s_{g(t)} g(t)$ is the Einstein tensor and $\varphi(t) = \partial_t g(t)$.

Theorem 3.14 and Lemma 4.3 now entail the following theorem.

Theorem 4.4. *Let g_0 be a scalar flat and asymptotically flat metric on M^m such that $g_0 - g_e \in C^{4,\alpha}_{-\tau}$ for $\tau \in (\frac{m-2}{2}, m-2)$. Then the conformal Ricci flow $g(t)$ starting with $g(0) = g_0$ exists for some short time and*

$$g(t) \in \mathcal{M}_\tau \quad \text{and} \quad g(t) - g_e \in C^{1,2+\alpha}_{-\tau}.$$

Moreover,

$$\frac{d}{dt}\boldsymbol{m}(g(t)) = -2 \int_M |\mathrm{Ric}_{g(t)}|^2 \, d\mathrm{vol}_{g(t)} .$$

In particular, the ADM mass is strictly decreasing under conformal Ricci flow, except when g_0 is the Euclidean metric.

Proof. To verify that the conformal Ricci flow $g(t)$ stays in $\mathcal{M}_\tau$ we only need to verify that

$$\partial_j \partial_i g_{ij}(t) - \partial_j \partial_j g_{ii}(t) \in L^1(M,\, d\mathrm{vol}_{g_e}).$$

Recall that [Lee and Parker 1987, (9.2)]

$$s = \partial_j \partial_i g_{ij} - \partial_j \partial_j g_{ii} + O(\rho^{-2\tau-2}),$$

which implies that

$$\partial_j \partial_i g_{ij} - \partial_j \partial_j g_{ii} = O(\rho^{-2\tau-2}) \in L^1(M,\, d\mathrm{vol}_{g_e})$$

for $\tau \in (\frac{m-2}{2}, m-2)$. It is easily seen that the ADM mass is strictly decreasing except when g_0 is Ricci flat. Then, using [Bando et al. 1989, Theorem 1.5] and [Lee and Parker 1987, Proposition 10.2], one concludes that g_0 is the standard Euclidean metric. Therefore the proof is complete. □

A quick application of the above Theorem 4.4 is a simple and direct proof of the rigidity part of Schoen and Yau's positive mass theorem.

Corollary 4.5 [Schoen and Yau 1979]. *Suppose that $(M^m,\, g)$ is asymptotically flat manifold with nonnegative scalar curvature and that $g - g_e \in C^{4,\alpha}_{-\tau}$ for $\tau > \frac{m-2}{2}$. If the ADM mass $\boldsymbol{m}(g)$ is zero, then (M, g) is isometric to the standard Euclidean space $\mathbb{R}^m$.*

Proof. First we know that g has to be scalar flat. Otherwise one can conformally deform the metric to a scalar flat one and decrease the ADM mass to be negative, which is impossible due to the first part of the positive mass theorem of Schoen and Yau. Next we invoke Theorem 4.4 and come to the same contradiction if g is not flat. □

Acknowledgment

We would like to mention that Lars Andersson and Dan Lee have communicated to us that they are also aware of some version of Theorem 1.4. We are grateful that Dan Lee brought a related paper [Haslhofer 2011] to our attention.

Part of this work was done while the authors were visiting Beijing International Center for Mathematical Research and East China Normal University. The authors are grateful to them for the hospitality.

References

[Anderson 2005] M. T. Anderson, "On uniqueness and differentiability in the space of Yamabe metrics", *Commun. Contemp. Math.* **7**:3 (2005), 299–310. MR 2006d:53031 Zbl 1082.58013

[Arnowitt et al. 1960] R. Arnowitt, S. Deser, and C. W. Misner, "Canonical variables for general relativity", *Phys. Rev.* (2) **117** (1960), 1595–1602. MR 22 #4505b Zbl 0091.21203

[Bando et al. 1989] S. Bando, A. Kasue, and H. Nakajima, "On a construction of coordinates at infinity on manifolds with fast curvature decay and maximal volume growth", *Invent. Math.* **97**:2 (1989), 313–349. MR 90c:53098 Zbl 0682.53045

[Bartnik 1986] R. Bartnik, "The mass of an asymptotically flat manifold", *Comm. Pure Appl. Math.* **39**:5 (1986), 661–693. MR 88b:58144 Zbl 0598.53045

[Besse 1987] A. L. Besse, *Einstein manifolds*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) **10**, Springer, Berlin, 1987. MR 88f:53087 Zbl 0613.53001

[Chang and Lu 2007] S.-C. Chang and P. Lu, "Evolution of Yamabe constant under Ricci flow", *Ann. Global Anal. Geom.* **31**:2 (2007), 147–153. MR 2008e:53121 Zbl 1112.53048

[Chow et al. 2006] B. Chow, P. Lu, and L. Ni, *Hamilton's Ricci flow*, Grad. Studies Math. **77**, Amer. Math. Soc., Providence, RI, 2006. MR 2008a:53068 Zbl 1118.53001

[Dai and Ma 2007] X. Dai and L. Ma, "Mass under the Ricci flow", *Comm. Math. Phys.* **274**:1 (2007), 65–80. MR 2008h:53121 Zbl 1127.53056

[DeTurck 1983] D. M. DeTurck, "Deforming metrics in the direction of their Ricci tensors", *J. Differential Geom.* **18**:1 (1983), 157–162. MR 85j:53050 Zbl 0517.53044

[DeTurck 2003] D. M. DeTurck, "Deforming metrics in the direction of their Ricci tensors: Improved version", pp. 163–165 in *Collected papers on Ricci flow*, edited by H.-D. Cao et al., Series in Geometry and Topology **37**, International Press, Somerville, MA, 2003.

[Fischer 2004] A. E. Fischer, "An introduction to conformal Ricci flow", *Classical Quantum Gravity* **21**:3 (2004), 171–218. MR 2005b:53109 Zbl 1050.53029 arXiv math/0312519

[Hamilton 1995] R. S. Hamilton, "The formation of singularities in the Ricci flow", pp. 7–136 in *Proceedings of the Conference on Geometry and Topology* (Cambridge, MA, 1993), edited by C.-C. Hsiung and S.-T. Yau, Surveys in Differential Geometry **2**, International Press, Cambridge, MA, 1995. MR 97e:53075 Zbl 0867.53030

[Haslhofer 2011] R. Haslhofer, "A mass-decreasing flow in dimension three", preprint, 2011. arXiv 1107.3220

[Koiso 1979] N. Koiso, "A decomposition of the space $\mathcal{M}$ of Riemannian metrics on a manifold", *Osaka J. Math.* **16**:2 (1979), 423–429. MR 80g:53034 Zbl 0416.58007

[Lee and Parker 1987] J. M. Lee and T. H. Parker, "The Yamabe problem", *Bull. Amer. Math. Soc.* (*N.S.*) **17**:1 (1987), 37–91. MR 88f:53001 Zbl 0633.53062

[Lunardi 1995] A. Lunardi, *Analytic semigroups and optimal regularity in parabolic problems*, Progress Nonlinear Diff. Eq. and Appl. **16**, Birkhäuser, Basel, 1995. MR 96e:47039 Zbl 0816.35001

[Oliynyk and Woolgar 2007] T. A. Oliynyk and E. Woolgar, "Rotationally symmetric Ricci flow on asymptotically flat manifolds", *Comm. Anal. Geom.* **15**:3 (2007), 535–568. MR 2009b:53110 Zbl 1138.53057

[Schoen and Yau 1979] R. Schoen and S. T. Yau, "On the proof of the positive mass conjecture in general relativity", *Comm. Math. Phys.* **65**:1 (1979), 45–76. MR 80j:83024 Zbl 0405.53045

[Shi 1989] W.-X. Shi, "Ricci deformation of the metric on complete noncompact Riemannian manifolds", *J. Differential Geom.* **30**:2 (1989), 303–394. MR 90f:53080 Zbl 0686.53037

[Wald 1984] R. M. Wald, *General relativity*, University of Chicago Press, Chicago, IL, 1984. MR 86a:83001 Zbl 0549.53001

[Wang and Zheng 2011] E.-M. Wang and Y. Zheng, "Regularity of the first eigenvalue of the *p*-Laplacian and Yamabe invariant along geometric flows", *Pacific J. Math.* **254**:1 (2011), 239–255. MR 2900670 Zbl 06029190

Received April 20, 2012.

PENG LU
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF OREGON
304 DEADY HALL
EUGENE, OR 97403
UNITED STATES

penglu@uoregon.edu
http://pages.uoregon.edu/penglu/

JIE QING
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA, SANTA CRUZ
MCHENRY 4178
SANTA CRUZ, CA 95064
UNITED STATES

qing@ucsc.edu
http://www.math.ucsc.edu/faculty/qing.html

YU ZHENG
DEPARTMENT OF MATHEMATICS
SHANGHAI KEY LABORATORY OF PMMP
EAST CHINA NORMAL UNIVERSITY
DONGCHUAN RD 500
SHANGHAI 200241
CHINA

zhyu@math.ecnu.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 268, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.268.435

ON REPRESENTATIONS OF $\mathrm{GL}_{2n}(F)$ WITH A SYMPLECTIC PERIOD

ARNAB MITRA

Given a nonarchimedean local field F, we classify the irreducible admissible representations of $\mathrm{GL}_4(F)$ and $\mathrm{GL}_6(F)$ that bear a nontrivial linear form invariant under the groups $\mathrm{Sp}_2(F)$ and $\mathrm{Sp}_3(F)$, respectively. We propose a few conjectures for the case of $\mathrm{GL}_{2n}(F)$, $n > 3$.

1. Introduction

Let $G = \mathrm{GL}_{2n}(F)$ for F a nonarchimedean local field of characteristic 0 and let H be a symplectic subgroup of G of rank n. A representation π of G is said to have a symplectic period (or to be H-distinguished) if $\mathrm{Hom}_H(\pi|_H, \mathbb{C}) \neq 0$. give a complete list of irreducible admissible representations of $\mathrm{GL}_4(F)$ and $\mathrm{GL}_6(F)$ having a symplectic period. We also make a few conjectural statements for $\mathrm{GL}_{2n}(F)$ at the end.

The motivation for this problem comes from the work of Klyachko [1983] in the case of finite fields. He found a set of representations generalizing the Gelfand–Graev model, after which Heumos and Rallis [1990] studied the analogous notion in the p-adic case. They also proved multiplicity-one theorems in the symplectic case.

Continuing this line of investigation, Offen and Sayag [2007a; 2007b; 2008] proved the uniqueness property of the Klyachko models and multiplicity-one results for irreducible admissible representations. They also showed the existence of the Klyachko model for unitary representations. To state the results precisely we need to introduce notation.

Let δ be a square integrable representation of $\mathrm{GL}_r(F)$. Denote by $U(\delta, m)$ the unique irreducible quotient of the representation,

$$\nu^{(m-1)/2}\delta \times \nu^{(m-3)/2}\delta \times \cdots \times \nu^{-(m-1)/2}\delta.$$

Proposition 1.1 [Offen and Sayag 2007a]. *For $i = 1, \ldots, t$, let δ_i be square-integrable representations of $\mathrm{GL}_{r_i}(F)$ and m_i be positive integers. Let χ_i be a*

MSC2010: primary 22E50; secondary 11F70.
Keywords: admissible, symplectic period.

character of $\mathrm{GL}_{2m_ir_i}(F)$. *Then the representation*

$$\chi_1 U(\delta_1, 2m_1) \times \cdots \times \chi_t U(\delta_t, 2m_t)$$

has a symplectic period.

Further define

$$\mathcal{B} = \{U(\delta, 2m),\ \nu^{\alpha} U(\delta, 2m) \times \nu^{-\alpha} U(\delta, 2m)\},$$

where δ varies over the discrete series representations and $\alpha \in \mathbb{R}$ such that $|\alpha| < \frac{1}{2}$.

Theorem 1.2 [Offen and Sayag 2007b]. *Let* $\pi = \tau_1 \times \cdots \times \tau_r$ *such that* $\tau_i \in \mathcal{B}$. *Then* π *has a symplectic period. Conversely, if* π *is an irreducible unitary representation with a symplectic period, there exist* $\tau_1, \ldots, \tau_r \in \mathcal{B}$ *such that* $\pi = \tau_1 \times \cdots \times \tau_r$.

A natural question now is to classify all irreducible admissible representations that admit a symplectic model. For $\mathrm{GL}_4(F)$ and $\mathrm{GL}_6(F)$ we have:

Theorem 1.3. *Using the notation introduced for Proposition 1.1, an irreducible admissible representation of* $\mathrm{GL}_4(F)$ *with a symplectic period is a product of factors* $\chi_i U(\delta_i, 2n_i)$, *where the* χ_i *are (not necessarily unitary) characters of* $F^{\times}$.

Theorem 1.4. *Using the notation introduced for Proposition 1.1, an irreducible admissible representation of* $\mathrm{GL}_6(F)$ *with a symplectic period is either a product of* $\chi_i U(\delta_i, 2n_i)$ *(the* χ_i *are not necessarily unitary), or is a twist of* $Z([1, \nu], [\nu, \nu^4])$ *or its dual.*

A few words about the proofs. It is a consequence of the uniqueness of the Klyachko models that irreducible cuspidal representations (which are generic) cannot have a symplectic period. Since any nonsupercuspidal irreducible representation is a quotient of a representation of the form $\mathrm{Ind}_{P_{k,2n-k}}^{\mathrm{GL}_{2n}}(\rho \otimes \tau)$, $\rho \in \mathrm{Irr}(\mathrm{GL}_k(F))$, $\tau \in \mathrm{Irr}(\mathrm{GL}_{2n-k}(F))$ it is enough to study the problem for representations of these types. For $\mathrm{GL}_4(F)$ and $\mathrm{GL}_6(F)$, this reduces the problem to the analysis of representations of the type $\pi_1 \times \pi_2$ and $\pi_1 \times \pi_2 \times \pi_3$, where each π_i is an irreducible representation of $\mathrm{GL}_2(F)$. For the $\mathrm{GL}_4(F)$ case, using Mackey theory we obtain an exhaustive list of (not necessarily irreducible) representations. Then we study every possible quotient to obtain a complete list of irreducible $\mathrm{Sp}_2(F)$-distinguished representations of $\mathrm{GL}_4(F)$. In the $\mathrm{GL}_6(F)$ case, we first reduce the problem to the case when none of the π_i are cuspidal. Next we reduce it to the case when at most one of the π_i is an irreducible principal series. Then we do a case-by-case analysis (for each π_i to be one of the three types of irreducible representations of $\mathrm{GL}_2(F)$ — a character, an irreducible principal series or a twist of the Steinberg representation, with at most one being an irreducible principal series), analyzing all possible subquotients for symplectic periods. A common way of showing that an irreducible subquotient is not H-distinguished, especially in the $\mathrm{GL}_6(F)$ case, is to express it as a quotient of a representation, which is then shown not to have a symplectic period using Mackey theory.

A word on the organization of the paper. Section 2 notation and preliminary notions used in the paper. Orbit structures and Mackey theory are covered in detail in Section 3. We analyze the representations of the form $\pi_1 \times \pi_2$ and obtain the theorem for $\mathrm{GL}_4(F)$ in Section 4. In Section 5, we analyze the representations of the form $\pi_1 \times \pi_2 \times \pi_3$, collecting all the irreducible $\mathrm{Sp}_3(F)$-distinguished subquotients. Using this analysis we obtain the theorem for $\mathrm{GL}_6(F)$. In Section 6 we make a few conjectures for the general case based on the available examples.

2. Notation and preliminaries

Notation. Throughout the paper, F will denote a nonarchimedean local field of characteristic 0.

Following the notation of [Bernstein and Zelevinsky 1976], we denote the set of all smooth representations of an l-group G by $\mathrm{Alg}(G)$ and the subset of all irreducible admissible representations by $\mathrm{Irr}(G)$. If $\pi \in \mathrm{Alg}(G)$, we denote by $\tilde{\pi}$, its contragredient.

Any character of $\mathrm{GL}_n(F)$ can be thought of as a character of $F^\times$ via the determinant map. Given a character χ of $F^\times$ and a smooth representation π of $\mathrm{GL}_n(F)$ we will denote the twist of π by χ simply by $\chi\pi$, $\chi\pi(g) := \chi(\det(g))\pi(g)$. Unless otherwise mentioned, St_n and 1_n will be used to denote the Steinberg and the trivial character of $\mathrm{GL}_n(F)$. The norm character $\nu(g) := |\det g|$ will be denoted by ν.

Let $P_{n_1,\dots,n_r}$ be the group of block upper triangular matrices corresponding to the tuple $(n_1, \dots, n_r)$. Let $N_{n_1,\dots,n_r}$ denote its unipotent radical. Let $\delta_{P_{n_1,\dots,n_r}}$ denote the modular function of the group $P_{n_1,\dots,n_r}$. Since a parabolic normalizes its unipotent radical, this defines a character of $P_{n_1,\dots,n_r}$ (the module of the automorphism $n \to pnp^{-1}$ of $N_{n_1,\dots,n_r}$ for $p \in P_{n_1,\dots,n_r}$). Call this character $\delta_{N_{n_1,\dots,n_r}}$. Then we have $\delta_{N_{n_1,\dots,n_r}} = \delta_{P_{n_1,\dots,n_r}}$. For an element $p \in P_{n_1,\dots,n_r}$, with its Levi part equal to $\mathrm{diag}(g_1, \dots, g_r)$, we have

$$\delta_{P_{n_1,\dots,n_r}}(p) = |\det g_1|^{n_2+\cdots+n_r} |\det g_2|^{-n_1+n_3+\cdots+n_r} \cdots |\det g_r|^{-n_1\cdots-n_{r-1}}. \tag{2-1}$$

The induced representation of $(\sigma, H, W) \in \mathrm{Alg}(H)$ to G is the following space of locally constant functions

$$\mathrm{Ind}_H^G \sigma = \{f : G \to W \mid f(hg) = \delta_H^{1/2} \delta_G^{-1/2} \sigma(h) f(g) \text{ for all } h \in H,\ g \in G\},$$

where δ_G and δ_H are the modular functions of G and H respectively. G acts on the space by right action. Compact induction from H to G is denoted by $\mathrm{ind}_H^G \sigma$ and is the subspace of $\mathrm{Ind}_H^G \sigma$ consisting of functions compactly supported mod H. Occasionally we will use nonnormalized induction (see Remark 2.22

of [Bernstein and Zelevinsky 1976] for the definition), although unless otherwise mentioned induction is always normalized. Given representations $\rho_i \in \mathrm{Irr}(\mathrm{GL}_{n_i}(F))$ $(i = 1, \dots, r)$, extend $\rho_1 \otimes \cdots \otimes \rho_r$ to $P_{n_1,\dots,n_r}$ so that it is trivial on $N_{n_1,\dots,n_r}$. We denote by $\rho_1 \times \cdots \times \rho_r$ the representation $\mathrm{Ind}_{P_{n_1,\dots,n_r}}^{\mathrm{GL}_n}(\rho_1 \otimes \cdots \otimes \rho_r)$.

The Jacquet functor with respect to a unipotent subgroup N is denoted by r_N and is always normalized.

If $\pi \in \mathrm{Irr}(\mathrm{GL}_n(F))$, then there exists a partition of n and a multiset of cuspidal representations $\{\rho_1, \dots, \rho_r\}$ corresponding to it such that π can be embedded in $\rho_1 \times \cdots \times \rho_r$. This multiset is uniquely determined by π and called its *cuspidal support*. For the purposes of this paper, for a smooth representation of finite length define it to be the union (as a set) of all the supports of its irreducible subquotients.

Preliminaries on segments. We briefly recall the notation and the basic definition of segments as introduced in [Zelevinsky 1980]. Given a cuspidal representation ρ of $\mathrm{GL}_m(F)$, a *segment* is a set of the form $\{\rho, \rho\nu, \dots, \rho\nu^{k-1}\}$, with $k > 0$; we also write it as $[\rho, \rho\nu^{k-1}]$. Given a segment $\Delta = [\rho, \rho\nu^{k-1}]$, the unique irreducible submodule and the unique irreducible quotient of $\rho \times \cdots \times \rho\nu^{k-1}$ are denoted by $Z(\Delta)$ and $Q(\Delta)$ respectively.

For $\Delta_1 = [\rho_1, \nu^{k_1-1}\rho_1]$ and $\Delta_2 = [\rho_2, \nu^{k_2-1}\rho_2]$, we say that Δ_1 and Δ_2 are *linked* if $\Delta_1 \not\subseteq \Delta_2$, $\Delta_2 \not\subseteq \Delta_1$ and $\Delta_1 \cup \Delta_2$ is also a segment. If Δ_1 and Δ_2 are linked and $\Delta_1 \cap \Delta_2 = \phi$, then we say that Δ_1 and Δ_2 are juxtaposed. If Δ_1 and Δ_2 are linked and $\rho_2 = \nu^k \rho_1$, where $k > 0$, we say that Δ_1 *precedes* Δ_2. Given a multiset $\mathfrak{a} = \{\Delta_1, \dots, \Delta_r\}$ of segments, let

$$\pi(\mathfrak{a}) := Z(\Delta_1) \times \cdots \times Z(\Delta_r).$$

If Δ_i does not precede Δ_j for any $i < j$, $\pi(\mathfrak{a})$ is known to have a unique irreducible submodule, which will be denoted by $Z(\Delta_1, \dots, \Delta_r)$. By Theorem 6.1 of [Zelevinsky 1980], this submodule is independent of the ordering of the segments as long as the "does not precede" condition is satisfied. Hence we simply denote it by $Z(\mathfrak{a})$. In this situation, a similar statement holds for quotients as well and the unique irreducible quotient of $Q(\Delta_1) \times \cdots \times Q(\Delta_r)$ is denoted by $Q(\mathfrak{a})$. For example, the trivial character 1_n of $\mathrm{GL}_n(F)$ is $Z([\nu^{-(n-1)/2}, \nu^{(n-1)/2}])$, while St_n is $Q([\nu^{-(n-1)/2}, \nu^{(n-1)/2}])$.

We say a multiset $\mathfrak{a} = \{\Delta_1, \dots, \Delta_r\}$ is on the cuspidal line of ρ, where ρ is a cuspidal representation of some $\mathrm{GL}_n(F)$, if $\Delta_i \subset \{\nu^k \rho\}_{k\in\mathbb{Z}}$ for all i.

Preliminaries on $\mathrm{GL}_n(F)$ and symplectic periods. We now collect a few basic results on $\mathrm{GL}_n(F)$ and symplectic periods needed in the sequel. The following result is used to calculate explicitly the quotients and the submodules in quite a few cases in the proofs of the main theorems.

Theorem 2.1 [Zelevinsky 1980]. *Let Δ_1 and Δ_2 be segments. If Δ_1 and Δ_2 are linked, put $\Delta_3 = \Delta_1 \cup \Delta_2$ and $\Delta_4 = \Delta_1 \cap \Delta_2$. The representation $\pi = Z(\Delta_1)\times Z(\Delta_2)$ is irreducible if and only if Δ_1 and Δ_2 are not linked. If Δ_1 and Δ_2 are linked then π has length 2. If Δ_2 precedes Δ_1 then π has a unique irreducible submodule $Z(\Delta_1, \Delta_2)$ and a unique irreducible quotient $Z(\Delta_3) \times Z(\Delta_4)$. If Δ_1 precedes Δ_2 then π has a unique irreducible submodule $Z(\Delta_3) \times Z(\Delta_4)$ and a unique irreducible quotient $Z(\Delta_1, \Delta_2)$.*

Using the Zelevinsky involution and Rodier's theorem that $Q(\Delta_1, \Delta_2)$ is taken to $Z(\Delta_1, \Delta_2)$ we have a quotient version of this lemma.

Theorem 2.2. *Let Δ_1 and Δ_2 be segments. If Δ_1 and Δ_2 are linked, put $\Delta_3 = \Delta_1 \cup \Delta_2$ and $\Delta_4 = \Delta_1 \cap \Delta_2$. The representation $\pi = Q(\Delta_1)\times Q(\Delta_2)$ is irreducible if and only if Δ_1 and Δ_2 are not linked. If Δ_1 and Δ_2 are linked then π has length 2. If Δ_2 precedes Δ_1 then π has the unique irreducible submodule $Q(\Delta_3) \times Q(\Delta_4)$. If Δ_1 precedes Δ_2 then π has the unique irreducible quotient $Q(\Delta_3) \times Q(\Delta_4)$.*

Lemma 2.3 [Casselman 1995]. *Let $\eta = \rho_1 \times \cdots \times \rho_m$, where $\rho_i \in \mathrm{Irr}(\mathrm{GL}_i(F))$. Define $\breve{\eta} = \tilde{\rho}_m \times \cdots \times \tilde{\rho}_1$. Then π is an irreducible quotient of η if and only if $\tilde{\pi}$ is an irreducible quotient of $\breve{\eta}$.*

Let $\mathrm{Ext}^1_G(\cdot, \mathbb{C})$ be the derived group of the $\mathrm{Hom}_G(\cdot, \mathbb{C})$ functor (for details, see [Prasad 1990; 1993]).

Lemma 2.4. *Let $H = \mathrm{Sp}_n(F)$. Then $\mathrm{Ext}^1_H(\mathbb{C}, \mathbb{C})$ is trivial.*

Proof. An element of $\mathrm{Ext}^1_H(\mathbb{C}, \mathbb{C})$ corresponds to an exact sequence

$$0 \to \mathbb{C} \xrightarrow{i} V \xrightarrow{j} \mathbb{C} \to 0$$

of H-modules, or equivalently a homomorphism from H to the group of upper triangular unipotent subgroup of $\mathrm{GL}_2(\mathbb{C})$. Since H has no abelian quotients, there are no such nontrivial maps and we have the lemma. □

Theorem 2.5 [Offen and Sayag 2008]. *Let $\pi \in \mathrm{Irr}(\mathrm{GL}_n(F))$. If π embeds in a Klyachko model, it does so in a unique Klyachko model and with multiplicity at most one.*

3. Orbit structures and Mackey theory

Let X be a subspace of a symplectic space $(V, \langle\ ,\ \rangle)$ of dimension $2n$. Let

$$X^{\perp} = \{y \in V \mid \langle y, x\rangle = 0 \text{ for all } x \in X\}.$$

Define $\mathrm{Rad}\ X = X \cap X^{\perp}$. Note that $X/\mathrm{Rad}\ X$ inherits the symplectic structure of V, becomes a nondegenerate symplectic space and hence has even dimension.

The next lemma is a variant of the classical theorem of Witt for quadratic forms.

Lemma 3.1 (Witt). (a) *Let X_1, X_2 be subspaces of V of same dimension. Then there exists a symplectic automorphism ϕ of V, taking X_1 to X_2 if and only if* $\dim \operatorname{Rad} X_1 = \dim \operatorname{Rad} X_2$.

(b) *Let X_1, X_2 be subspaces of V and $\phi : X_1 \to X_2$ be a symplectic isomorphism. Then ϕ extends to a symplectic automorphism of V.*

It follows from this lemma that if X is a k-dimensional subspace of V, and P_X is the parabolic subgroup of $\mathrm{GL}(V)$ consisting of automorphisms of V leaving X invariant, then $\mathrm{Sp}(V)\backslash\mathrm{GL}(V)/P_X$ is in bijective correspondence with integers i, $0 \le i \le \dim X$, such that $\dim X - i$ is even. To get a set of representatives for these double cosets, let

$$\{e_1, e_2, \dots, e_n, f_1, f_2, \dots, f_n\}$$

be the standard symplectic basis of V; i.e., $\langle e_i, f_j\rangle = \delta_{ij}$. Define

$$\begin{aligned}
Y_r &:= \langle e_1, \dots, e_r\rangle,\\
Y_r^{\vee} &:= \langle f_1, \dots, f_r\rangle,\\
S_{k,r} &:= \langle e_{r+1}, \dots, e_{(k+r)/2}, f_{r+1}, \dots, f_{(k+r)/2}\rangle,\\
T_{k,r} &:= \langle e_{\frac{k+r}{2}+1}, \dots, e_n, f_{\frac{k+r}{2}+1}, \dots, f_n\rangle,\\
X_{k,r} &:= Y_r + S_{k,r}.
\end{aligned}$$

Note that $\mathrm{GL}(V)/P_X$ is the set of all k-dimensional subspaces of V on which $\mathrm{Sp}(V)$ acts in a natural way. Therefore $\mathrm{Sp}(V)\backslash\mathrm{GL}(V)/P_X$ is represented by a certain set of k-dimensional subspaces of V, which can be taken to be the spaces $X_{k,r}$ with $0 \le r \le k$ such that $k-r$ is even.

Since $\dim X = \dim X_{k,r}$, there exists an automorphism $g \in \mathrm{GL}(V)$ taking X to $X_{k,r}$. This automorphism gives an isomorphism from P_X to $P_{X_{k,r}}$. Using this isomorphism a representation of P_X can be considered to be a representation of $P_{X_{k,r}}$. By Mackey theory, the restriction of the representation $\mathrm{Ind}_{P_X}^{\mathrm{GL}(V)}(\sigma)$ to $\mathrm{Sp}(V)$ is obtained by gluing the representations:

$$\mathrm{ind}_{(\mathrm{Sp}(V)\cap P_{X_{k,r}})}^{\mathrm{Sp}(V)}(\delta_{P_X}^{1/2}\sigma|_{(\mathrm{Sp}(V)\cap P_{X_{k,r}})}),$$

where the induction is nonnormalized. The isomorphism of P_X with $P_{X_{k,r}}$ takes the unipotent radical of P_X to the unipotent radical of $P_{X_{k,r}}$ and hence the representation of $P_{X_{k,r}}$ so obtained is of the same kind that appears in parabolic induction. This is a special case for maximal parabolics of Proposition 3 of [Offen 2006].

For an isotropic subspace Y of V, the subgroup Q_Y of $\mathrm{Sp}(V)$ stabilizing Y is a parabolic subgroup of $\mathrm{Sp}(V)$, with Levi decomposition

$$Q_Y = \big(\mathrm{GL}(Y) \times \mathrm{Sp}(Y^{\perp}/Y)\big) \ltimes U,$$

where U is the subgroup of $\mathrm{Sp}(V)$ preserving $Y \subset Y^{\perp}$ and acting trivially on Y, $Y^{\perp}/Y$ and $V/Y^{\perp}$.

We fix a symplectic basis of V and identify the group of linear transformations with the corresponding group of matrices, although we emphasize that the following proposition and its corollary are independent of the choice of the basis.

Proposition 3.2. *The subgroup $H_{k,r}$ of $\mathrm{Sp}(V)$ stabilizing the subspace $X_{k,r}$ of V is*

$$H_{k,r} = \big(\mathrm{GL}(Y_r) \times \mathrm{Sp}(S_{k,r}) \times \mathrm{Sp}(T_{k,r})\big).U_{k,r},$$

where $U_{k,r}$ is the unipotent group inside $\mathrm{Sp}(V)$ consisting of automorphisms of V of the form

$$\begin{pmatrix} I_r & A & B & C \\ 0 & I_{k-r} & 0 & A' \\ 0 & 0 & I_{2n-(k+r)} & B' \\ 0 & 0 & 0 & I_r \end{pmatrix},$$

where $A \in \mathrm{Hom}(S_{k,r}, Y_r)$, $B \in \mathrm{Hom}(T_{k,r}, Y_r)$, the matrix $C \in \mathrm{Hom}(Y_r^{\vee}, Y_r)$ is symmetric, and $A' \in \mathrm{Hom}(Y_r^{\vee}, S_{k,r})$, $B' \in \mathrm{Hom}(Y_r^{\vee}, T_{k,r})$ are adjoint to A, B.

Proof. Note that $H_{k,r}$ is nothing but the symplectic automorphisms of V preserving the flag $0 \subset Y_r = X_{k,r} \cap X_{k,r}^{\perp} \subset X_{k,r} \subset X_{k,r} + X_{k,r}^{\perp} = X_{k,r} + T_{k,r} = Y_r^{\perp} \subset V$. Hence $H_{k,r}$ acts on the successive quotients of this filtration, giving rise to a surjective homomorphism to $\mathrm{GL}(Y_r) \times \mathrm{Sp}(S_{k,r}) \times \mathrm{Sp}(T_{k,r})$ with kernel $U_{k,r}$ consisting of the subgroup of $\mathrm{Sp}(V)$ preserving the flag and acting trivially on successive quotients. Clearly $U_{k,r}$ acts trivially on the isotropic subspace Y_r, on $Y_r^{\perp}$ and on $Y_r^{\perp}/Y_r = S_{k,r} + T_{k,r}$. The well-known knowledge of the structure of the parabolic in $\mathrm{Sp}(V)$ defined by Y_r proves the assertion of the proposition. □

Corollary 3.3. (1) *The modular character $\delta_{k,r}$ of the group $H_{k,r}$ is*

$$\delta_{k,r}(\mathrm{diag}(g, h_1, h_2, {}^t g^{-1})) = |\det g|^{r+a+b+1},$$

where $r = \dim Y_r$, $a = \dim S_{k,r} = k - r$, $b = \dim T_{k,r} = 2n - (k + r)$, and $g \in \mathrm{GL}(Y_r)$.

(2) *By (2-1) we have $\delta_P(\mathrm{diag}(g, h_1, h_2, {}^t g^{-1})) = |\det g|^{2r+a+b}$, where we set $P = P_{(r+a,b+r)}$. Thus*

$$\frac{\delta_P^{1/2}}{\delta_{k,r}}\big(\mathrm{diag}(g, h_1, h_2, {}^t g^{-1})\big) = |\det g|^{-1-(a+b)/2} = |\det g|^{-(n-r+1)}. \qquad □$$

Define M to be the group $\mathrm{GL}(Y_r) \times \mathrm{Sp}(S_{k,r}) \times \mathrm{Sp}(T_{k,r})$ and identify it with

$$\mathrm{GL}_r(F) \times \mathrm{Sp}_{(k-r)/2}(F) \times \mathrm{Sp}_{(2n-k-r)/2}(F)$$

via the fixed basis. Call H the group $\mathrm{Sp}_n(F)$ defined with respect to this symplectic basis. Further let $N = N_1 \times N_2$, where N_1 and N_2 are the unipotent subgroups of $\mathrm{GL}_k(F)$ and $\mathrm{GL}_{2n-k}(F)$ corresponding to the partitions $(r, k-r)$ and $(2n-k-r, r)$, respectively. Let $\sigma_1 \in \mathrm{Irr}(\mathrm{GL}_k(F))$ and $\sigma_2 \in \mathrm{Irr}(\mathrm{GL}_{2n-k}(F))$. Call σ the representation of $P = P_{(k,2n-k)}$ obtained by extending $\sigma_1 \otimes \sigma_2$ to P in the usual way.

By Frobenius reciprocity and Corollary 3.3, we get

$$\mathrm{Hom}_H\big(\mathrm{ind}_{H_{k,r}}^H(\delta_P^{1/2}\sigma|_{H_{k,r}}), \mathbb{C}\big) = \mathrm{Hom}_{MU_{k,r}}\big(\nu^{-(n-r+1)}\sigma_1 \otimes \sigma_2, \mathbb{C}\big).$$

Clearly,

$$\mathrm{Hom}_{M.U_{k,r}}\big(\nu^{-(n-r+1)}\sigma_1 \otimes \sigma_2, \mathbb{C}\big) = \mathrm{Hom}_{MN}\big(\nu^{-(n-r+1)}\sigma_1 \otimes \sigma_2, \mathbb{C}\big).$$

Since the normalized Jacquet functor is left adjoint to normalized induction by Proposition 1.9(b) of [Bernstein and Zelevinsky 1977], we obtain

$$\begin{aligned}\mathrm{Hom}_{MN}\big(\nu^{-(n-r+1)}\sigma_1 \otimes \sigma_2, \mathbb{C}\big)\\ &= \mathrm{Hom}_M\big(r_N(\nu^{-(n-r+1)}\sigma_1 \otimes \sigma_2), \delta_N^{-1/2}\big)\\ &= \mathrm{Hom}_M\big(\nu^{-(n-r+1)}\delta_{N_1}^{1/2} r_{N_1}(\sigma_1) \otimes \delta_{N_2}^{1/2} r_{N_2}(\sigma_2), \mathbb{C}\big).\end{aligned}$$

Now let A and B have determinant 1. By (2-1), we have

$$\delta_{N_1}\begin{pmatrix} g & * \\ 0 & A \end{pmatrix} = |\det g|^{(k-r)}, \delta_{N_2}\begin{pmatrix} B & * \\ 0 & {}^tg^{-1} \end{pmatrix} = |\det g|^{2n-(k+r)}.$$

Define α to be the character of M such that $\alpha(\mathrm{diag}(g, h_1, h_2, {}^tg^{-1})) = \nu^{-1}(g)$. Plugging in the value of the delta functions we get

$$\mathrm{Hom}_H\big(\mathrm{ind}_{H_{k,r}}^H(\delta_P^{1/2}\sigma|_{H_{k,r}}), \mathbb{C}\big) = \mathrm{Hom}_M\big(\alpha(r_{N_1}(\sigma_1) \otimes r_{N_2}(\sigma_2)), \mathbb{C}\big). \tag{3-1}$$

From this we have the following lemma for $\mathrm{GL}_{2n}(F)$.

Lemma 3.4. *Let $\pi_i = Z(\Delta_1^i, \dots, \Delta_{k_i}^i) \in \mathrm{Irr}(\mathrm{GL}_{n_i}(F))$ for $i = 1, \dots, s$ be such that the following conditions are satisfied*:

(1) *For $i \neq j$, the segments $\Delta_{m_i}^i$ and $\Delta_{m_j}^j$ are disjoint and not linked, for all $m_i = 1, \dots, k_i$ and all $m_j = 1, \dots, k_j$.*

(2) *$\sum_{i=1}^{s} n_i$ is even and $\pi := \pi_1 \times \cdots \times \pi_s$ has a symplectic period.*

Then each n_i is even and every π_i has a symplectic period.

Proof. Condition (1) forces π to be irreducible (by Proposition 8.5 of [Zelevinsky 1980]). Thus it is enough to prove the lemma for $s = 2$.

Let $\pi_1 \in \mathrm{Irr}(\mathrm{GL}_{n_1}(F))$ and $\pi_2 \in \mathrm{Irr}(\mathrm{GL}_{n_2}(F))$. Now, since $r_{N_1}(\pi_1)$ lies in $\mathrm{Alg}(\mathrm{GL}_r(F) \times \mathrm{GL}_{n_1-r}(F))$ and the functor r_{N_1} takes finite length representations into ones of finite length ([ibid.], Proposition 1.4), up to semisimplification it is of the form $\sum_{i=1}^{t_1} \pi_{1i} \otimes \tau_{1i}$ for some $t_1 > 0$, where $\pi_{1i} \in \mathrm{Irr}(\mathrm{GL}_r(F))$ and $\tau_{1i} \in \mathrm{Irr}(\mathrm{GL}_{n_1-r}(F))$ for all $i = 1, \dots, t_1$. Similarly, up to semisimplification, $r_{N_2}(\pi_2)$ is equal to $\sum_{j=1}^{t_2} \tau_{2j} \otimes \pi_{2j}$, where $\tau_{2j} \in \mathrm{Irr}(\mathrm{GL}_{n_2-r}(F))$ and $\pi_{2j} \in \mathrm{Irr}(\mathrm{GL}_r(F))$.

We claim that for any $\theta \in \mathrm{Irr}(\mathrm{GL}_m(F))$, the cuspidal support (page 438) of $r_N(\theta)$ is always a subset (as a set) of the cuspidal support of θ. Assume $\theta = Z(\Delta_1, \dots, \Delta_l)$. The claim follows from the geometrical lemma (Lemma 2.12 of [Bernstein and Zelevinsky 1977]) applied to $r_N(Z(\Delta_1) \times \cdots \times Z(\Delta_l))$, along with the observation that $r_N(\theta)$ is a submodule of it.

Together with condition (1) of the lemma, this claim implies the vanishing of $\mathrm{Hom}_{\mathrm{GL}_r}(\nu^{-1}\pi_{1i} \otimes \tilde{\pi_{2j}}, \mathbb{C})$ for every pair i, j. By (3-1) and the realization of contragredient representations due to Gelfand and Kazhdan (cf. Theorem 7.3 of [Bernstein and Zelevinsky 1976]), this implies

$$\mathrm{Hom}_H\big(\mathrm{ind}_{H_{n_1,r}}^H(\delta_{P_{n_1,n_2}}^{1/2}(\pi_1 \otimes \pi_2)|_{H_{n_1,r}}), \mathbb{C}\big) = 0$$

unless $r = 0$. This along with condition (2) forces n_1, n_2 to be even and π_1, π_2 both to have symplectic periods. □

Lemma 3.5. *Let Δ_1 and Δ_2 be segments of even lengths such that their intersection is of odd length. Then the representation $\theta = Z(\Delta_1, \Delta_2)$ has a symplectic period.*

Proof. If possible, let $\mathrm{Hom}_H(\theta, \mathbb{C}) = 0$. Define the segments $\Delta_3 = \Delta_1 \cup \Delta_2$ and $\Delta_4 = \Delta_1 \cap \Delta_2$. Without loss of generality assume Δ_1 precedes Δ_2. By Theorem 2.1, θ sits inside the following exact sequence of $\mathrm{GL}_{2n}(F)$ modules:

$$0 \to \theta \to Z(\Delta_2) \times Z(\Delta_1) \to Z(\Delta_3) \times Z(\Delta_4) \to 0.$$

Observe that Δ_3 and Δ_4 are segments of odd length. So, $Z(\Delta_3) \times Z(\Delta_4)$ has a mixed Klyachko model by Theorem 3.7 of [Offen and Sayag 2007b] and hence by Theorem 2.5, it is not H-distinguished. Since $\mathrm{Hom}_H\big(Z(\Delta_2) \times Z(\Delta_1), \mathbb{C}\big) = 0$ if $\mathrm{Hom}_H\big(Z(\Delta_3) \times Z(\Delta_4), \mathbb{C}\big) = 0$ and $\mathrm{Hom}_H(\theta, \mathbb{C}) = 0$, we obtain a contradiction with Proposition 1.1. □

Lemma 3.6. *If Δ_1 and Δ_2 are juxtaposed segments of even lengths in the cuspidal line of 1_1 (the trivial representation of $\mathrm{GL}_1(F)$), the representation $\theta = Z(\Delta_1, \Delta_2)$ does not have a symplectic period.*

Proof. Define $\Delta_3 = \Delta_1 \cup \Delta_2$ and let $2n$ be its length. In fact, twisting it by an appropriate power of ν, without loss of generality we can take Δ_3 to be $[\nu^{-\frac{2n-1}{2}}, \nu^{\frac{2n-1}{2}}]$ and hence $Z(\Delta_3) = 1$. Let

$$\Delta_1 = [\nu^{-\frac{2n-1}{2}}, \nu^{\frac{a}{2}}] \quad \text{and} \quad \Delta_2 = [\nu^{\frac{b}{2}}, \nu^{\frac{2n-1}{2}}].$$

Let $k = \frac{2n-1}{2} - \frac{b}{2} + 1$, the length of Δ_2. Now assume $k \leq n$.

Let $\mu_1 = Z(\Delta_2)$ and $\mu_2 = Z(\Delta_1)$. Let us first calculate $\mathrm{Hom}_H(\mu_1 \times \mu_2, \mathbb{C})$. By (3-1), for $r \neq 0$, $\mathrm{Hom}_H\big(\mathrm{ind}_{H_{k,r}}^H(\delta_P^{1/2}\mu_1 \otimes \mu_2|_{H_{k,r}}), \mathbb{C}\big)$ is isomorphic to

$$\mathrm{Hom}_{\mathrm{GL}_r(F)\times \mathrm{Sp}_{\frac{k-r}{2}}(F)\times \mathrm{Sp}_{n-\frac{k+r}{2}}(F)}\big(\nu^{n-\frac{2k-r}{2}-1} \otimes \nu^{n-1-\frac{k-r}{2}} \otimes \nu^{-\frac{k+r}{2}} \otimes \nu^{n-k-\frac{r}{2}}, \mathbb{C}\big),$$

where $\mathrm{GL}_r(F)$ acts on the last term via the contragredient. Now, consider

$$\mathrm{Hom}_{\mathrm{GL}_r(F)}\big(\nu^{n-\frac{2k-r}{2}-1} \otimes \nu^{-(n-k-\frac{r}{2})}, \mathbb{C}\big).$$

This is nonzero only if $n - \frac{2k-r}{2} - 1 = n - k - \frac{r}{2}$, which is impossible since k is even by the hypothesis of the lemma. Thus

$$\mathrm{Hom}_H\big(\mathrm{ind}_{H_{k,r}}^H(\delta_P^{1/2}\mu_1 \otimes \mu_2|_{H_{k,r}}), \mathbb{C}\big) = 0 \quad \text{if } r \neq 0.$$

On the other hand, if $r = 0$ we have

$$\mathrm{Hom}_H\big(\mathrm{ind}_{H_{k,0}}^H(\delta_P^{1/2}\mu_1 \otimes \mu_2|_{H_{k,0}}), \mathbb{C}\big) = \mathrm{Hom}_{\mathrm{Sp}_{\frac{k}{2}}(F)}\big(\mu_1, \mathbb{C}\big) \otimes \mathrm{Hom}_{\mathrm{Sp}_{n-\frac{k}{2}}(F)}(\mu_2, \mathbb{C}) = \mathbb{C}.$$

Hence $\mathrm{Hom}_H(\mu_1 \times \mu_2, \mathbb{C})$ is at most one-dimensional. Now, we have the following exact sequence of $\mathrm{GL}_{2n}(F)$ modules (and hence of $\mathrm{Sp}_n(F)$ modules):

$$0 \to Z(\Delta_1, \Delta_2) \xrightarrow{i} Z(\Delta_2) \times Z(\Delta_1) \xrightarrow{j} \mathbb{C} \to 0.$$

Applying the functor $\mathrm{Hom}_{\mathrm{Sp}_n(F)}\big(.., \mathbb{C}\big)$ to it we obtain the long exact sequence

$$0 \to \mathrm{Hom}_{\mathrm{Sp}_n(F)}\big(\mathbb{C}, \mathbb{C}\big) \xrightarrow{j^*} \mathrm{Hom}_{\mathrm{Sp}_n(F)}\big(Z(\Delta_2) \times Z(\Delta_1), \mathbb{C}\big) \xrightarrow{i^*} \mathrm{Hom}_{\mathrm{Sp}_n(F)}\big(Z(\Delta_1, \Delta_2), \mathbb{C}\big) \xrightarrow{r^*} \mathrm{Ext}^1_{\mathrm{Sp}_n(F)}\big(\mathbb{C}, \mathbb{C}\big) \to \cdots$$

Observing that $\mathrm{Ext}^1_{\mathrm{Sp}_n(F)}\big(\mathbb{C}, \mathbb{C}\big) = 0$ (see Lemma 2.4) we get the following short exact sequence:

$$0 \to \mathbb{C} \xrightarrow{j^*} \mathrm{Hom}_{\mathrm{Sp}_n(F)}\big(Z(\Delta_2) \times Z(\Delta_1), \mathbb{C}\big) \xrightarrow{i^*} \mathrm{Hom}_{\mathrm{Sp}_n(F)}\big(Z(\Delta_1, \Delta_2), \mathbb{C}\big) \to 0.$$

Since j^* is injective, $\mathrm{Im}(j^*) = \mathbb{C}$. By exactness, $\mathrm{Ker}(i^*) = \mathbb{C}$ as well. Since $\mathrm{Hom}_{\mathrm{Sp}_n(F)}\big(Z(\Delta_2) \times Z(\Delta_1), \mathbb{C}\big)$ was shown to be at most one-dimensional, it is equal to $\mathrm{Ker}(i^*)$. Thus $\mathrm{Im}(i^*) = 0$. But again by exactness, i^* is surjective, thus implying that

$$\mathrm{Hom}_{\mathrm{Sp}_n(F)}\big(Z(\Delta_1, \Delta_2), \mathbb{C}\big) = 0.$$

Thus we have the lemma if $k \leq n$. Since an irreducible representation has a symplectic period if and only if its contragredient has so, we have the lemma in the case $k > n$. □

4. Analysis in the $\mathrm{GL}_4(F)$ case: proof of Theorem 1.3

In this section we prove Theorem 1.3. We begin with the following lemma.

Lemma 4.1. *Let θ be an irreducible representation of $\mathrm{GL}_4(F)$ with a symplectic period. Then there exists $\pi_i \in \mathrm{Irr}(\mathrm{GL}_2(F))$, $i = 1, 2$ such that θ appears as a quotient of $\pi_1 \times \pi_2$.*

Proof. If θ is a supercuspidal representation of $\mathrm{GL}_4(F)$, it is generic and hence by Theorem 2.5 it doesn't have a symplectic period. Thus θ appears as a quotient of either $\chi_1 \times \theta_3$, $\theta_3 \times \chi_1$ or $\pi_1 \times \pi_2$ (where $\chi_1 \in \mathrm{Irr}(\mathrm{GL}_1(F))$, $\theta_3 \in \mathrm{Irr}(\mathrm{GL}_3(F))$ and $\pi_1, \pi_2 \in \mathrm{Irr}(\mathrm{GL}_2(F))$). In the last case we have nothing left to prove. If θ is a quotient of $\theta_3 \times \chi_1$, by Lemma 2.3, $\tilde{\theta}$ is a quotient of $\tilde{\chi}_1 \times \tilde{\theta}_3$. Since an irreducible representation has a symplectic period if and only if its contragredient does, by applying Lemma 2.3 again we are reduced to the first case. So assume θ is a quotient of $\chi_1 \times \theta_3$. Now if θ_3 is cuspidal, $\chi_1 \times \theta_3$ is irreducible and generic. Hence by the disjointness of the symplectic and Whittaker models it cannot have a symplectic period. Thus assume θ_3 isn't cuspidal.

Then θ_3 is a quotient of one of the representations of the form $\chi_1' \times \delta_2$, $\delta_2 \times \chi_1'$ or $\chi_1' \times \chi_1'' \times \chi_1'''$, where $\chi_1', \chi_1'', \chi_1'''$ are characters of $\mathrm{GL}_1(F)$ and δ_2 is a supercuspidal of $\mathrm{GL}_2(F)$.

In the first case, $\chi_1 \times \theta_3$ is a quotient of $\chi_1 \times \chi_1' \times \delta_2$. If $\chi_1 \times \chi_1'$ is irreducible, the lemma is proved. If not, $\chi_1 \times \chi_1' \times \delta_2$ is glued from $Z(\chi_1 \times \chi_1') \times \delta_2$ and $Q(\chi_1 \times \chi_1') \times \delta_2$, where $Z(\chi_1 \times \chi_1')$ and $Q(\chi_1 \times \chi_1')$ are respectively the unique irreducible submodule and unique irreducible quotient of $\chi_1 \times \chi_1'$. Thus any irreducible quotient of $\chi_1 \times \theta_3$ has to be a quotient of one of the two.

In the second case, since $\delta_2 \times \chi_1'$ is irreducible, $\chi_1 \times \delta_2 \times \chi_1' \cong \chi_1 \times \chi_1' \times \delta_2$. Thus we are back to the first case.

In the third case, if both $\chi_1 \times \chi_1'$ and $\chi_1'' \times \chi_1'''$ are irreducible we are done. In case at least one of them is reducible, we get the lemma by breaking $\chi_1 \times \chi_1' \times \chi_1'' \times \chi_1'''$, as in the first case, into subquotients of the required form. □

By this lemma, it is enough to consider representations of the form $\pi_1 \times \pi_2$, where π_1 and π_2 are irreducible representations of $\mathrm{GL}_2(F)$. If $\pi = \pi_1 \times \pi_2$ has an H-distinguished quotient, then π itself is H-distinguished. By Mackey theory we get that $(\pi_1 \times \pi_2)|_{\mathrm{Sp}_2(F)}$ is glued from the two subquotients

$$\mathrm{ind}_{H_{2,0}}^{H}(\delta_{P_{2,2}}^{1/2}\pi_1 \otimes \pi_2|_{H_{2,0}}) \quad \text{and} \quad \mathrm{ind}_{H_{2,2}}^{H}(\delta_{P_{2,2}}^{1/2}\pi_1 \otimes \pi_2|_{H_{2,2}}).$$

Analyzing the two subquotients (using (3-1)), it is easy to see that the necessary conditions for π to have a symplectic period are that either π_1, π_2 are characters of $\mathrm{GL}_2(F)$ or $\pi_2 \cong \nu^{-1}\pi_1$. Any irreducible representation of $\mathrm{GL}_2(F)$ is either a supercuspidal, a character, an irreducible principal series or a twist of the Steinberg representation. Thus any irreducible $\mathrm{Sp}_4(F)$-distinguished representation occurs as a quotient of one of the representations listed in the next proposition.

Proposition 4.2. *Let θ be an irreducible admissible representation of* $\mathrm{GL}_4(F)$ *with a symplectic period. Then θ occurs as a quotient of one of the following representations π of* $\mathrm{GL}_4(F)$:

(1) $\pi = \chi_2 \times \chi_2'$, *where χ_2, χ_2' are characters of* $\mathrm{GL}_2(F)$.

(2) $\pi = \sigma_2 \times \nu^{-1}\sigma_2$, *where σ_2 is a supercuspidal of* $\mathrm{GL}_2(F)$.

(3) $\pi = \chi_1 \times \chi_1' \times \nu^{-1}\chi_1 \times \nu^{-1}\chi_1'$, *where χ_1, χ_1' are characters of $F^\times$ and $\chi_1 \times \chi_1'$ is an irreducible principal series.*

(4) $\pi = Q([\chi_1\nu^{-1/2}, \chi_1\nu^{1/2}]) \times Q([\chi_1\nu^{-3/2}, \chi_1\nu^{-1/2}])$, *where χ_1 is a character of $F^\times$.* □

Now we come to the theorem in the $\mathrm{GL}_4(F)$ case. We state and prove an equivalent version of Theorem 1.3 in terms of the Zelevinsky classification.

Theorem 4.3. *This is the complete list of irreducible admissible representations θ of* $\mathrm{GL}_4(F)$ *with a symplectic period*:

(1) $\theta = Z([\sigma_2, \nu\sigma_2])$, *where σ_2 is a cuspidal representation of* $\mathrm{GL}_2(F)$.

(2) $\theta = Z(\Delta_1, \Delta_2)$, *where $\Delta_1 = [\chi_1\nu^{-1/2}, \chi_1\nu^{1/2}]$ and $\Delta_2 = [\chi_1\nu^{-3/2}, \chi_1\nu^{-1/2}]$ (χ_1 is a character of $F^\times$).*

(3) $\theta =$ *a character of* $\mathrm{GL}_4(F)$.

(4) $\theta = \chi_2 \times \chi_2'$, *where χ_2, χ_2' are characters of* $\mathrm{GL}_2(F)$.

Proof. The strategy of the proof is to consider each representation in the list of Proposition 4.2 and to check, for all irreducible quotients of each one, whether they have a symplectic period.

Case I: $\pi = \chi_2 \times \chi_2'$. If $\chi_2 \times \chi_2'$ is irreducible, $\theta = \pi$ has a symplectic period by Proposition 1.1. So assume otherwise. Let $\chi_2 = Z([\chi_1\nu^{-1/2}, \chi_1\nu^{1/2}])$ and $\chi_2' = Z([\chi_1'\nu^{-1/2}, \chi_1'\nu^{1/2}])$. There are four subcases:

(1) $\chi_1 = \chi_1' \nu$. In this case, $\pi = Z([\chi_1'\nu^{1/2}, \chi_1'\nu^{3/2}]) \times Z([\chi_1'\nu^{-1/2}, \chi_1'\nu^{1/2}])$. By Theorem 2.1, it has a unique irreducible quotient, which is

$$\theta = Z([\chi_1'\nu^{-1/2}, \chi_1'\nu^{3/2}]) \times \chi_1'\nu^{1/2}.$$

By Theorem 3.7 of [Offen and Sayag 2007b], it has a mixed Klyachko model. Hence, by Theorem 2.5, it doesn't have a symplectic period.

(2) $\chi_1 = \chi_1'\nu^{-1}$. Here, $\pi = Z([\chi_1'\nu^{-3/2}, \chi_1'\nu^{-1/2}]) \times Z([\chi_1'\nu^{-1/2}, \chi_1'\nu^{1/2}])$. By Theorem 2.1, this has a unique irreducible quotient

$$\theta = \chi_1' Z([\nu^{-3/2}, \nu^{-1/2}], [\nu^{-1/2}, \nu^{1/2}]),$$

which has a symplectic period by Lemma 3.5. Note that θ is a twist of $U(\mathrm{St}_2, 2)$ and the fact that it has a symplectic period also follows from Proposition 1.1.

(3) $\chi_1 = \chi_1'\nu^2$. In this case, $\pi = Z([\chi_1'\nu^{3/2}, \chi_1'\nu^{5/2}]) \times Z([\chi_1'\nu^{-1/2}, \chi_1'\nu^{1/2}])$. This has a unique irreducible quotient $\theta = Z([\chi_1'\nu^{-1/2}, \chi_1'\nu^{5/2}])$. Thus θ is the character $\chi_1\nu$ of $\mathrm{GL}_4(F)$ and has a symplectic period.

(4) $\chi_1 = \chi_1'\nu^{-2}$. Here, $\pi = Z([\chi_1'\nu^{-5/2}, \chi_1'\nu^{-3/2}]) \times Z([\chi_1'\nu^{-1/2}, \chi_1'\nu^{1/2}])$, which, by Theorem 2.1, has a unique irreducible quotient

$$\theta = Z([\chi_1'\nu^{-5/2}, \chi_1'\nu^{-3/2}], [\chi_1'\nu^{-1/2}, \chi_1'\nu^{1/2}]).$$

By Lemma 3.6, it doesn't have a symplectic period.

Case II: $\pi = \sigma_2 \times \nu^{-1}\sigma_2$. In this case, π has a unique irreducible quotient $U(\nu^{-1/2}\sigma_2, 2) \cong Z([\nu^{-1}\sigma_2, \sigma_2])$. By Proposition 1.1 it has a symplectic period.

Case III: $\pi = \chi_1 \times \chi_1' \times \nu^{-1}\chi_1 \times \nu^{-1}\chi_1'$, where $\chi_1 \times \chi_1'$ is irreducible. There are two further subcases:

(1) $\chi_1' \times \chi_1\nu^{-1}$ is irreducible. This again can be broken down into two subcases.

(1a) $\chi_1' \neq \chi_1\nu^2$. In this case, $\pi \cong \chi_1 \times \nu^{-1}\chi_1 \times \chi_1' \times \nu^{-1}\chi_1'$. The "does not precede" condition (page 438) is satisfied and so π has a unique irreducible quotient. Clearly, π has $Z([\nu^{-1}\chi_1, \chi_1]) \times Z([\nu^{-1}\chi_1', \chi_1'])$ as a quotient. If it's irreducible, it has a symplectic period by Proposition 1.1 and has already been accounted for in case I. So assume the contrary. In that case the segments are linked. But the assumption that $\chi_1' \times \chi_1\nu^{-1}$ is irreducible, together with $\chi_1' \neq \chi_1\nu^2$, forces a contradiction. Hence irreducibility of $Z([\nu^{-1}\chi_1, \chi_1]) \times Z([\nu^{-1}\chi_1', \chi_1'])$ is the only possibility.

(1b) $\chi_1' = \chi_1\nu^2$. In this case,

$$\pi \cong \chi_1'\nu^{-2} \times \chi_1' \times \chi_1'\nu^{-3} \times \chi_1'\nu^{-1} \cong \chi_1'\nu^{-2} \times \chi_1'\nu^{-3} \times \chi_1' \times \chi_1'\nu^{-1}.$$

This representation has $\tau = Z([\chi_1'\nu^{-3}, \chi_1'\nu^{-2}]) \times Z([\chi_1'\nu^{-1}, \chi_1'])$ as a quotient. Since the cuspidal support of π is multiplicity free it has a unique irreducible quotient (by Proposition 2.10 of [Zelevinsky 1980]) and so any H-distinguished irreducible quotient of π is a quotient of τ. Thus they have already been accounted for in case I.

(2) $\chi_1' \times \chi_1 \nu^{-1}$ is reducible. This happens if and only if $\chi_1 = \chi_1'$ or $\chi_1 = \chi_1' \nu^2$. Again, we will deal with the cases separately.

(2a) $\chi_1 = \chi_1'$. The representation is of the form $\chi_1 \times \chi_1 \times \chi_1 \nu^{-1} \times \chi_1 \nu^{-1}$. Since it satisfies the "does not precede" condition (page 438) it has a unique irreducible quotient. It can be easily seen that $\theta = Z([\chi_1 \nu^{-1}, \chi_1]) \times Z([\chi_1 \nu^{-1}, \chi_1])$ is an irreducible quotient of this representation (and so is the unique one). θ has a symplectic period and has already been accounted for in case I.

(2b) $\chi_1 = \chi_1' \nu^2$. In this case, $\pi \cong \chi_1' \nu^2 \times \chi_1' \times \chi_1' \nu \times \chi_1' \nu^{-1} \cong \chi_1' \times \chi_1' \nu^{-1} \times \chi_1' \nu^2 \times \chi_1' \nu$. By an argument similar to the one used in (1b) above, we conclude that this representation has already been accounted for in case I.

Case IV: $\pi = Q([\chi_1 \nu^{-1/2}, \chi_1 \nu^{1/2}]) \times Q([\chi_1 \nu^{-3/2}, \chi_1 \nu^{-1/2}])$. By Theorem 2.2, π has a unique irreducible quotient $\chi_1 Q([\nu^{-3/2}, \nu^{-1/2}], [\nu^{-1/2}, \nu^{1/2}])$. As seen in case I(2), it is a twist of $U(\mathrm{St}_2, 2)$ and has a symplectic period (by Proposition 1.1). □

5. Analysis in the $\mathrm{GL}_6(F)$ case

In this section we obtain the theorem for $\mathrm{GL}_6(F)$. The following lemma reduces the analysis to representations of the form $\pi_1 \times \pi_2 \times \pi_3$, where the π_i are irreducible representations of $\mathrm{GL}_2(F)$.

Lemma 5.1. *Let θ be an irreducible representation of $\mathrm{GL}_6(F)$ with a symplectic period. Then either θ is of the form $Z([\sigma_3, \nu\sigma_3])$, where σ_3 is a supercuspidal representation of $\mathrm{GL}_3(F)$ or it occurs as a subquotient of a representation of the form $\pi_1 \times \pi_2 \times \pi_3$, where $\pi_1 \in \mathrm{Irr}(\mathrm{GL}_2(F))$ for $i = 1, 2, 3$.*

Proof. Since supercuspidal representations are generic, they don't have a symplectic period. Thus θ appears as a subquotient of $\tau_1 \times \tau_2$, where τ_1 and τ_2 are irreducible representations of $\mathrm{GL}_k(F)$ and $\mathrm{GL}_{6-k}(F)$ respectively. By interchanging τ_1 and τ_2 if necessary, we can assume $k \le 3$.

Case 1: $k = 1$. If τ_2 is a cuspidal representation of $\mathrm{GL}_5(F)$, since τ_1 is a character, $\tau_1 \times \tau_2$ is irreducible and generic. Thus τ_2 occurs as a subquotient of a representation induced from a maximal parabolic of $\mathrm{GL}_5(F)$. So θ is either a subquotient of $\tau_1 \times \chi \times \tau$ ($\chi \in \mathrm{Irr}(\mathrm{GL}_1(F))$, $\tau \in \mathrm{Irr}(\mathrm{GL}_4(F))$) or $\tau_1 \times \tau' \times \tau''$ ($\tau' \in \mathrm{Irr}(\mathrm{GL}_2(F))$, $\tau'' \in \mathrm{Irr}(\mathrm{GL}_3(F))$). Thus θ is either a subquotient of $\theta_1 \times \tau$ (where $\theta_1 \in \mathrm{Irr}(\mathrm{GL}_2(F))$)

or of $\theta_2 \times \tau''$ (where $\theta_2 \in \mathrm{Irr}(\mathrm{GL}_3(F))$) thus reducing the lemma to the next two cases.

Case 2: $k = 2$. If τ_2 is a cuspidal representation of $\mathrm{GL}_4(F)$, $\tau_1 \times \tau_2$ is irreducible and doesn't have a symplectic period, by Lemma 3.4. Thus, as earlier, τ_2 occurs as a subquotient of a representation induced from a maximal parabolic of $\mathrm{GL}_4(F)$. So θ is either a subquotient of $\tau_1 \times \chi \times \tau$ ($\chi \in \mathrm{Irr}(\mathrm{GL}_1(F)),\ \tau \in \mathrm{Irr}(\mathrm{GL}_3(F))$) or $\tau_1 \times \tau' \times \tau''$ ($\tau' \in \mathrm{Irr}(\mathrm{GL}_2(F)),\ \tau'' \in \mathrm{Irr}(\mathrm{GL}_2(F))$). In the first scenario θ occurs as a subquotient of $\theta_1 \times \theta_2$ (where $\theta_1,\ \theta_2 \in \mathrm{Irr}(\mathrm{GL}_3(F))$), reducing the lemma to the next case, while in the second we have the lemma.

Case 3: $k = 3$. We will first show that if either of τ_1, τ_2 (say τ_1) is cuspidal then θ is of the form $Z([\sigma_3, \nu\sigma_3])$. Choose $\tau_2' \in \mathrm{Irr}(\mathrm{GL}_3(F))$ such that θ is a quotient of either $\tau_1 \times \tau_2'$ or $\tau_2' \times \tau_1$.

Assume the former. Then $\tau_1 \times \tau_2'$ also has a nontrivial $\mathrm{Sp}_3(F)$-invariant linear form. Now, $\tau_1 \times \tau_2'|_{\mathrm{Sp}_3(F)}$ is glued from

$$\mathrm{ind}_{H_{3,3}}^{H}(\delta_{P_{3,3}}^{1/2}\tau_1 \otimes \tau_2'|_{H_{3,3}}) \quad \text{and} \quad \mathrm{ind}_{H_{3,1}}^{H}(\delta_{P_{3,3}}^{1/2}\tau_1 \otimes \tau_2'|_{H_{3,1}}).$$

Since τ_1 is cuspidal, by (3-1), $\mathrm{Hom}_H(\mathrm{ind}_{H_{3,1}}^{H}(\delta_{P_{3,3}}^{1/2}\tau_1 \otimes \tau_2'|_{H_{3,1}}), \mathbb{C}) = 0$. Hence,

$$\mathrm{Hom}_H(\mathrm{ind}_{H_{3,3}}^{H}(\delta_{P_{3,3}}^{1/2}\tau_1 \otimes \tau_2'|_{H_{3,3}}), \mathbb{C}) \neq 0,$$

which is true if and only if $\tau_2' = \nu^{-1}\tau_1$, again by (3-1) and a theorem of Gelfand and Kazhdan (see [Bernstein and Zelevinsky 1976, Theorem 7.3]). Thus θ equals $Z([\nu^{-1}\tau_1, \tau_1])$.

If instead θ is a quotient of $\tau_2' \times \tau_1$, replacing θ by $\tilde{\theta}$ gives us the desired result.

Thus assume now that none of the two are cuspidal. Then $\exists\ \chi_i, \theta_i'$ $(i = 1, 2)$ such that τ_i is a subquotient of $\chi_i \times \theta_i'$ (where $\chi_i \in \mathrm{Irr}(\mathrm{GL}_1(F)),\ \theta_i' \in \mathrm{Irr}(\mathrm{GL}_2(F))$). Thus θ is a subquotient of $\chi_1 \times \chi_2 \times \theta_1' \times \theta_2'$ and hence the lemma is proved. □

Next we prove a hereditary property for $\mathrm{GL}_6(F)$ using the classification theorem for $\mathrm{GL}_4(F)$.

Proposition 5.2. *Let $\pi_1 \in \mathrm{Irr}(\mathrm{GL}_2(F))$ and $\pi_2 \in \mathrm{Irr}(\mathrm{GL}_4(F))$ be two irreducible representations with symplectic periods. Then $\pi_1 \times \pi_2$ has a symplectic period. Similarly, if π_1, π_2, π_3 are irreducible representations of $\mathrm{GL}_2(F)$, with a symplectic period, then $\pi_1 \times \pi_2 \times \pi_3$ has a symplectic period.*

Proof. Any irreducible representation π of $\mathrm{GL}_2(F)$ having a symplectic period is a character, while by Theorem 1.3 any such representation of $\mathrm{GL}_4(F)$ is either a character, an irreducible product of two characters of $\mathrm{GL}_2(F)$ or a representation of the form $U(\delta, 2)$. The proposition now follows from Proposition 1.1. □

The following lemma is a consequence of Lemma 3.4 and the fact that cuspidal representations are generic (and hence not symplectic).

Lemma 5.3. *Let π_1, π_2, π_3 be irreducible admissible representations of* $\mathrm{GL}_2(F)$. *If one or more of the π_i are cuspidal and θ is an* $\mathrm{Sp}_3(F)$*-distinguished subquotient of $\pi = \pi_1 \times \pi_2 \times \pi_3$ then it is of the form $\chi_2 \times Z([\sigma_2, \nu\sigma_2])$, where χ_2 and σ_2 are a character and a supercuspidal of* $\mathrm{GL}_2(F)$ *respectively.*

Proof. Without loss of generality let π_3 be a supercuspidal. Call it σ_2. Now there can be three cases depending on π_1 and π_2.

Case 1: None of π_1 and π_2 are cuspidal. In this case σ_2 is not in the cuspidal support of $\pi_1 \times \pi_2$ and hence any irreducible subquotient of π is of the form $\sigma_2 \times J$, where J is an irreducible subquotient of $\pi_1 \times \pi_2$. By Lemma 3.4, it doesn't have a symplectic period.

Case 2: Both π_1 and π_2 are cuspidal. In this case π is of the form $\sigma_2 \times \sigma_2' \times \sigma_2''$. If none of the pairs are linked or there is exactly one linked pair among the 3, then again by Lemma 3.4, π doesn't have an $\mathrm{Sp}_3(F)$-distinguished irreducible subquotient. So π has to be either of the form $\sigma_2 \times \nu\sigma_2 \times \nu\sigma_2$, $\sigma_2 \times \sigma_2 \times \nu\sigma_2$ or $\sigma_2 \times \nu\sigma_2 \times \nu^2\sigma_2$ (up to a permutation of the π_i).

If $\pi = \sigma_2 \times \nu\sigma_2 \times \nu^2\sigma_2$ (or a permutation), then it has 4 irreducible subquotients. Of these, $Q([\sigma_2, \nu^2\sigma_2])$ is generic and $Z([\sigma_2, \nu^2\sigma_2])$ doesn't have a symplectic period (by Theorem 1.2). Now consider the subquotient $Z([\sigma_2], [\nu\sigma_2, \nu^2\sigma_2])$. It is the unique irreducible quotient of the representation $\mu_1 \times \mu_2$, where $\mu_1 = \sigma_2$ and $\mu_2 = Z([\nu\sigma_2, \nu^2\sigma_2])$. Now, using (3-1), it can be easily checked that $\mathrm{Hom}_H\big(\mathrm{ind}_{H_{2,0}}^{H}(\delta_{P_{2,4}}^{1/2}\mu_1 \otimes \mu_2|_{H_{2,0}}), \mathbb{C}\big)$ and $\mathrm{Hom}_H\big(\mathrm{ind}_{H_{2,2}}^{H}(\delta_{P_{2,4}}^{1/2}\mu_1 \otimes \mu_2|_{H_{2,2}}), \mathbb{C}\big)$ are both 0, thus implying $\mathrm{Hom}_H(\mu_1 \times \mu_2, \mathbb{C}) = 0$. So, $Z([\sigma_2], [\nu\sigma_2, \nu^2\sigma_2])$ doesn't have a symplectic period and by taking contragredients we conclude that neither does $Z([\nu^2\sigma_2], [\sigma_2, \nu\sigma_2])$. Thus π doesn't have any irreducible subquotient carrying a symplectic period.

If $\pi = \sigma_2 \times \nu\sigma_2 \times \nu\sigma_2$ (or a permutation), it is glued from the irreducible representations $\nu\sigma_2 \times Z([\sigma_2, \nu\sigma_2])$ and $\nu\sigma_2 \times Q([\sigma_2, \nu\sigma_2])$. As in the above paragraph, taking $\mu_1 = \nu\sigma_2$ and $\mu_2 = Z([\sigma_2, \nu\sigma_2])$ and using (3-1), it can be easily checked that $\nu\sigma_2 \times Z([\sigma_2, \nu\sigma_2])$ doesn't have a symplectic period. The representation $\nu\sigma_2 \times Q([\sigma_2, \nu\sigma_2])$ is generic and hence doesn't have a symplectic period, by Theorem 2.5. Similarly $\sigma_2 \times \sigma_2 \times \nu\sigma_2$ (or any of its permutations) cannot have an $\mathrm{Sp}_3(F)$-distinguished subquotient either.

Case 3: Exactly one of π_1 and π_2 is cuspidal. Up to a permutation, π then is a representation of the form $\sigma_2 \times \sigma_2' \times \theta'$, where θ' is an irreducible representation of $\mathrm{GL}_2(F)$, which isn't supercuspidal. If σ_2' and σ_2 are linked and θ' is a character then π has an $\mathrm{Sp}_3(F)$-distinguished subquotient of the required form. Otherwise, again by Lemma 3.4, it doesn't have one. □

Thus it reduces the analysis to the cases where each π_i is either a character, an irreducible principal series or a twist of the Steinberg. Note that (up to a permutation of the π_i) there are 10 possible cases. Next we show that if at least two of the π_i are irreducible principal series representations, we need not consider those cases. This reduces the analysis to the remaining 7 cases.

Lemma 5.4. *Let π_1, π_2, π_3 be irreducible admissible representations of $\mathrm{GL}_2(F)$ such that none of them are cuspidal. If two or more of the π_i are irreducible principal series representations and θ is an $\mathrm{Sp}_3(F)$-distinguished subquotient of $\pi = \pi_1 \times \pi_2 \times \pi_3$ then it also appears as a subquotient of $\pi' = \pi_1' \times \pi_2' \times \pi_3'$, where at most one of the π_i' is a principal series representation.*

Proof. If θ is as above, it is a subquotient of a representation of the form $\tau = \chi_1 \times \cdots \times \chi_6$, where each χ_i is a character of $\mathrm{GL}_1(F)$. It is easy to see that Lemma 3.4 implies that unless all the χ_i are in the same cuspidal line, θ is an irreducible product of a character of $\mathrm{GL}_2(F)$ and an irreducible $\mathrm{Sp}_2(F)$-distinguished representation. We count them in the case when all the three π_i are characters. So without loss of generality we can assume the χ_i to be integral powers of the character ν of $\mathrm{GL}_1(F)$. Say a character is linked to another if they are linked as one-element segments (page 438): explicitly, ν^a and ν^b are linked if and only if $a - b = \pm 1$. If no two of the characters appearing in τ are linked, τ is irreducible and generic and so θ cannot be its subquotient. So we can assume $\tau \cong 1 \times \nu \times \nu^a \times \nu^b \times \nu^c \times \nu^d$.

Now, assume that there is a character among $\nu^a, \dots, \nu^d$ (say ν^a) that is not linked to any of the other characters. Collecting all the ν^a together, we see that $\theta = \nu^a \times \cdots \times \nu^a \times J$ for some irreducible representation J such that $\nu^a \times \cdots \times \nu^a$ and J satisfy the hypothesis of Lemma 3.4. So τ cannot have an $\mathrm{Sp}_3(F)$-distinguished subquotient. Thus we further assume that all the characters among $\nu^a, \dots, \nu^d$ are linked to some other character.

Note that if there exists a partition of the characters of τ such that at least two different blocks of the partition consist of linked pairs, τ is glued from subquotients of the form $\tau_1 \times \tau_2 \times \nu^{n_1} \times \nu^{n_2}$, where τ_i is either a character or a twist of the Steinberg. Thus θ can also be obtained in the cases when two of the π_i are characters, two of them are twists of the Steinberg or one of the π_i is a character and another one is a twist of the Steinberg.

Thus if we show that, under the hypothesis that any two of the characters of τ are linked and such a partition of them doesn't exist, τ cannot have an irreducible H-distinguished subquotient we are done. Lemma 5.5 precisely does that. □

Lemma 5.5. *Call the characters ν^a and ν^b linked if and only if $a - b = \pm 1$. Let $\tau \cong 1 \times \nu \times \nu^a \times \nu^b \times \nu^c \times \nu^d$ be such that every character of it is linked to some* *other character. Assume that there doesn't exist any partition of the characters with at least two blocks consisting of linked pairs. Then τ cannot have an irreducible* $\mathrm{Sp}_3(F)$*-distinguished subquotient.*

Proof. If possible, let θ be an $\mathrm{Sp}_3(F)$-distinguished subquotient of τ. The hypothesis of the lemma implies that the cuspidal support of τ can have at most ν^{-1} or ν^2 along with 1 and ν. Moreover, ν^{-1} and ν^2 cannot both be there simultaneously and in case the support only consists of 1 and ν, τ is one of the representations $1\times 1\times 1\times 1\times 1\times \nu$ or $1\times \nu\times \nu\times \nu\times \nu\times \nu$ (up to a permutation of the characters).

If τ has ν^{-1} in the cuspidal support, 1 can be there only with multiplicity one and so the only possible forms for τ, up to a permutation of the characters, are these:

$$\nu^{-1}\times 1\times\nu\times\nu\times\nu\times\nu, \qquad \nu^{-1}\times\nu^{-1}\times 1\times\nu\times\nu\times\nu,$$
$$\nu^{-1}\times\nu^{-1}\times\nu^{-1}\times 1\times\nu\times\nu, \quad \nu^{-1}\times\nu^{-1}\times\nu^{-1}\times\nu^{-1}\times 1\times\nu.$$

Consider the last representation first. There exists a permutation of factors such that θ is a quotient of the representation obtained by taking the product in that order. An easy calculation, using arguments similar to those of case 1(a) below (where all three π_i are characters), shows that no permutation gives a product which is H-distinguished. Thus θ cannot have a symplectic period which is a contradiction. So τ cannot be $\nu^{-1}\times\nu^{-1}\times\nu^{-1}\times\nu^{-1}\times 1\times\nu$ (or any permutation of the characters). Similarly one checks that τ cannot be $\nu^{-1}\times\nu^{-1}\times\nu^{-1}\times 1\times\nu\times\nu$ (or any permutation). Since the other two representations are contragredients of the above two representations, they cannot have any H-distinguished subquotients either. Thus τ cannot be any permutation of one of them either and we conclude that τ cannot have ν^{-1} in its cuspidal support.

Observe that the possible values of τ if its cuspidal support has ν^2 instead of ν^{-1} can all be obtained by appropriately twisting the contragredients of the ones obtained in the ν^{-1} case. So τ cannot be one of them either and hence cannot have ν^2 in its cuspidal support.

Thus τ can only have 1s and νs in its cuspidal support. If $\tau = 1\times 1\times 1\times 1\times 1\times\nu$ (or any permutation of the characters), it is glued from $Z([1,\nu])\times 1\times 1\times 1\times 1$ and $Q([1,\nu])\times 1\times 1\times 1\times 1$. For the first one take $\mu_1 = Z([1,\nu])$, $\mu_2 = 1\times 1\times 1\times 1$ and use (3-1), as in case 1(a) below, to conclude that it doesn't have a symplectic period. The second one is generic and hence also cannot have symplectic period. Thus again τ cannot have θ as a subquotient and so it cannot be a permutation of $1\times 1\times 1\times 1\times 1\times\nu$. Taking contragredients we conclude that it cannot be a permutation of $1\times\nu\times\nu\times\nu\times\nu\times\nu$ either. This shows that even 1 and ν cannot be in the cuspidal support of τ. This is a contradiction to our initial assumption that τ has an $\mathrm{Sp}_3(F)$-distinguished subquotient. □

Thus we need to analyze only the remaining seven cases.

Case 1: π_1, π_2, π_3 are all characters. The representations in this case are of the form

$$\pi = Z([\chi_1, \chi_1\nu]) \times Z([\chi_1', \chi_1'\nu]) \times Z([\chi_1'', \chi_1''\nu]).$$

If there are no links among the three segments, the representation is an irreducible product of characters of $\mathrm{GL}_2(F)$ and is symplectic by Proposition 1.1. Assume now that there is exactly one link. Without loss of generality we can assume that $[\chi_1', \chi_1'\nu]$ and $[\chi_1'', \chi_1''\nu]$ are linked, and that $[\chi_1, \chi_1\nu]$ is not linked to either. Clearly then, $\chi_1 \neq \chi_1', \chi_1''$. So, $[\chi_1, \chi_1\nu]$ is disjoint and not linked to either $[\chi_1', \chi_1'\nu]$ or $[\chi_1'', \chi_1''\nu]$. Observe that if a segment Δ_1 is not linked to Δ_2 and Δ_3 (where Δ_2 and Δ_3 are linked), it is not linked to $\Delta_2 \cup \Delta_3$ or $\Delta_2 \cap \Delta_3$ either. So by Theorem 7.1 and Proposition 8.5 of [Zelevinsky 1980], each irreducible subquotient of π is of the form $Z([\chi_1, \chi_1\nu]) \times \theta'$, where θ' is an irreducible subquotient of $Z([\chi_1', \chi_1'\nu]) \times Z([\chi_1'', \chi_1''\nu])$. Moreover if the subquotient is H-distinguished, observe then that $Z([\chi_1, \chi_1\nu])$ and θ' satisfy the hypothesis of Lemma 3.4. Thus by Lemma 3.4 any irreducible $\mathrm{Sp}_3(F)$-distinguished subquotient is an irreducible product of H-distinguished representations of $\mathrm{GL}_2(F)$ and $\mathrm{GL}_4(F)$.

Hence we look at the cases where there are at least two links among the segments. Without loss of generality we can assume χ_1 to be trivial. Following are the eight possible cases:

(a) $Z([1, \nu]) \times Z([\nu, \nu^2]) \times Z([\nu^3, \nu^4])$

(b) $Z([1, \nu]) \times Z([\nu, \nu^2]) \times Z([\nu^2, \nu^3])$

(c) $Z([1, \nu]) \times Z([\nu^2, \nu^3]) \times Z([\nu^4, \nu^5])$

(d) $Z([1, \nu]) \times Z([\nu, \nu^2]) \times Z([\nu, \nu^2])$

(e) $Z([1, \nu]) \times Z([\nu^2, \nu^3]) \times Z([\nu^2, \nu^3])$

(f) $Z([1, \nu]) \times Z([\nu^2, \nu^3]) \times Z([\nu^3, \nu^4])$

(g) $Z([1, \nu]) \times Z([1, \nu]) \times Z([\nu, \nu^2])$

(h) $Z([1, \nu]) \times Z([1, \nu]) \times Z([\nu^2, \nu^3])$

In each case we will evaluate all possible irreducible subquotients using [Zelevinsky 1980, Theorem 7.1], to determine whether it has a symplectic period.

(a) $\pi = Z([1, \nu]) \times Z([\nu, \nu^2]) \times Z([\nu^3, \nu^4])$. In this case, all irreducible subquotients of π are $Z([1, \nu], [\nu, \nu^2], [\nu^3, \nu^4])$, $Z([1, \nu^2], [\nu], [\nu^3, \nu^4])$, $Z([1, \nu^4], [\nu])$ and $Z([1, \nu], [\nu, \nu^4])$. We now analyze each of these representations.

- $\theta = Z([1, \nu], [\nu, \nu^2], [\nu^3, \nu^4])$ is the only irreducible submodule of $Z([\nu^3, \nu^4]) \times Z([\nu, \nu^2]) \times Z([1, \nu])$. Using Lemma 2.3 and taking contragredients we get that θ is the unique irreducible quotient of $\pi = Z([1, \nu]) \times Z([\nu, \nu^2]) \times Z([\nu^3, \nu^4])$. Since $Z([\nu, \nu^2], [\nu^3, \nu^4])$ is a quotient of $Z([\nu, \nu^2]) \times Z([\nu^3, \nu^4])$, θ is also the unique irreducible quotient of $Z([1,\nu])\times Z([\nu,\nu^2],[\nu^3,\nu^4])$.

Let $\mu_1 = Z([1,\nu])$ and $\mu_2 = Z([\nu,\nu^2],[\nu^3,\nu^4])$. Now, $\mu_1\times\mu_2$ is glued from

$$\operatorname{ind}_{H_{2,0}}^{H}(\delta_{P_{2,4}}^{1/2}\mu_1\otimes\mu_2|_{H_{2,0}}) \quad \text{and} \quad \operatorname{ind}_{H_{2,2}}^{H}(\delta_{P_{2,4}}^{1/2}\mu_1\otimes\mu_2|_{H_{2,2}})$$

(see Section 3). Since μ_2 doesn't have a symplectic period (by Lemma 3.6),

$$\operatorname{Hom}_H\big(\operatorname{ind}_{H_{2,0}}^{H}(\delta_{P_{2,4}}^{1/2}\mu_1\otimes\mu_2|_{H_{2,0}}),\mathbb{C}\big) = \operatorname{Hom}_{\mathrm{Sp}_1}(\mu_1,\mathbb{C})\otimes\operatorname{Hom}_{\mathrm{Sp}_2}(\mu_2,\mathbb{C})$$

is zero, by (3-1). On the other hand, $\operatorname{Hom}_H\big(\operatorname{ind}_{H_{2,2}}^{H}(\delta_{P_{2,4}}^{1/2}\mu_1\otimes\mu_2|_{H_{2,2}}),\mathbb{C}\big)$ equals

$$\operatorname{Hom}_{\mathrm{GL}_2\times\mathrm{Sp}_1}\big(\nu^{-1}\mu_1\otimes r_{(2,2),(4)}(\mu_2),\mathbb{C}\big). \tag{5-1}$$

An application of the geometrical lemma ([Bernstein and Zelevinsky 1977, Lemma 2.12]) shows that $r_{(2,2),(4)}(Z([\nu,\nu^2])\times Z([\nu^3,\nu^4]))$ is glued from the irreducible representations $Z([\nu,\nu^2])\otimes Z([\nu^3,\nu^4])$, $Z([\nu^3,\nu^4])\otimes Z([\nu,\nu^2])$ and $(\nu\times\nu^3)\otimes(\nu^2\times\nu^4)$. Jacquet functor being an exact functor, $r_{(2,2),(4)}(\mu_2)$ is glued from one or more of these terms. It can be checked that replacing $r_{(2,2),(4)}(\mu_2)$ by each of these three representations makes the group (5-1) trivial.

Thus, we get that $Z([1,\nu])\times Z([\nu,\nu^2],[\nu^3,\nu^4])$ doesn't have a symplectic period. If θ had a symplectic period, this would have given a nontrivial $\mathrm{Sp}_3(F)$-invariant linear functional of $Z([1,\nu])\times Z([\nu,\nu^2],[\nu^3,\nu^4])$ (by composing the one for θ with the quotient map), a contradiction. Hence θ doesn't have a symplectic period.

- If $\theta = Z([1,\nu^2],[\nu],[\nu^3,\nu^4])$, then θ is the unique irreducible submodule of $Z([\nu^3,\nu^4])\times\nu\times Z([1,\nu^2])$. Using Lemma 2.3 and taking contragredients, θ is the unique irreducible quotient of $Z([1,\nu^2])\times\nu\times Z([\nu^3,\nu^4])$. Applying Lemma 2.3 again, $\tilde{\theta}$ is the unique irreducible quotient of $Z([\nu^{-4},\nu^{-3}])\times\nu^{-1}\times Z([\nu^{-2},1])$. By taking $\mu_1 = Z([\nu^{-4},\nu^{-3}])$ and $\mu_2 = \nu^{-1}\times Z([\nu^{-2},1])$ and a similar calculation as above shows that $Z([\nu^{-4},\nu^{-3}])\times\nu^{-1}\times Z([\nu^{-2},1])$, and hence $\tilde{\theta}$, doesn't have a symplectic period. Thus even θ is not $\mathrm{Sp}_3(F)$-distinguished.
- If $\theta = Z([1,\nu^4],[\nu])\cong Z([1,\nu^4])\times\nu$ (by Proposition 8.5 of [Zelevinsky 1980]), by Theorem 2.1, it is an irreducible quotient of $Z([\nu^3,\nu^4])\times Z([1,\nu^2])\times\nu$. Now doing a similar calculation as in the first case by taking $\mu_1 = Z([\nu^3,\nu^4])$ and $\mu_2 = Z([1,\nu^2])\times\nu$, we get that $Z([\nu^3,\nu^4])\times Z([1,\nu^2])\times\nu$, and hence θ, doesn't have a symplectic period.
- If $\theta = Z([1,\nu],[\nu,\nu^4])$, it has a symplectic period by Lemma 3.5.

This concludes case (a).

(b) $\pi = Z([1,\nu]) \times Z([\nu,\nu^2]) \times Z([\nu^2,\nu^3])$. Here the irreducible subquotients of π are $Z([1,\nu],[\nu,\nu^2],[\nu^2,\nu^3])$, $Z([1,\nu^2],[\nu],[\nu^2,\nu^3])$, $Z([1,\nu],[\nu,\nu^3],[\nu^2])$, $Z([1,\nu^3],[\nu],[\nu^2])$, $Z([1,\nu^2],[\nu,\nu^3])$, and $Z([1,\nu^3],[\nu,\nu^2])$. We now analyze each of these representations.

- If $\theta = Z([1,\nu],[\nu,\nu^2],[\nu^2,\nu^3])$, we have $\theta = Q([1,\nu^2],[\nu,\nu^3])$ by Theorem A.10(iii) of [Tadić 1986]. Twisting θ by an appropriate power of ν makes it a unitary representation, which then turns out to be $\mathrm{Sp}_3(F)$-distinguished by Theorem 1.2.
- If $\theta = Z([1,\nu^2],[\nu],[\nu^2,\nu^3])$, then θ is the unique irreducible submodule of $Z([\nu^2,\nu^3]) \times \nu \times Z([1,\nu^2])$. Using Lemma 2.3 and taking contragredients we get that θ is the unique irreducible quotient of $\pi = Z([1,\nu^2]) \times \nu \times Z([\nu^2,\nu^3])$. Using Lemma 2.3 again, we get that $\tilde{\theta}$ is the unique irreducible quotient of $Z([\nu^{-3},\nu^{-2}]) \times \nu^{-1} \times Z([\nu^{-2},1])$. By taking $\mu_1 = Z([\nu^{-3},\nu^{-2}])$ and $\mu_2 = \nu^{-1} \times Z([\nu^{-2},1])$, and doing a similar calculation as in (a), we get that $Z([\nu^{-3},\nu^{-2}]) \times \nu^{-1} \times Z([\nu^{-2},1])$ and hence $\tilde{\theta}$, doesn't have a symplectic period. Thus even θ is not $\mathrm{Sp}_3(F)$-distinguished.
- If $\theta = Z([1,\nu],[\nu,\nu^3],[\nu^2])$, it can be obtained by twisting the contragredient of $Z([1,\nu^2],[\nu],[\nu^2,\nu^3])$, which, as showed in the last paragraph, doesn't have a symplectic period.
- If $\theta = Z([1,\nu^3],[\nu],[\nu^2])$, it is the unique irreducible submodule of
$$\nu^2 \times \nu \times Z([1,\nu^3]) \cong Z([1,\nu^3]) \times \nu^2 \times \nu.$$
Thus it is the unique irreducible submodule of $Z([1,\nu^3]) \times Q([\nu,\nu^2])$. Using Lemma 2.3 and taking contragredients we get that θ is the unique irreducible quotient of $Q([\nu,\nu^2]) \times Z([1,\nu^3])$. Now doing a similar calculation as in (a) by taking $\mu_1 = Q([\nu,\nu^2])$ and $\mu_2 = Z([1,\nu^3])$, we get that $Q([\nu,\nu^2]) \times Z([1,\nu^3])$, and hence θ, doesn't have a symplectic period.
- If $\theta = Z([1,\nu^2],[\nu,\nu^3])$, by Theorem A.10(iii) of [Tadić 1986],
$$\theta = Q([1,\nu],[\nu,\nu^2],[\nu^2,\nu^3]).$$
Twisting θ by an appropriate power of ν makes it a unitary representation. By Theorem 1.2 it doesn't have a symplectic period.
- If $\theta = Z([1,\nu^3],[\nu,\nu^2]) \cong Z([1,\nu^3]) \times Z([\nu,\nu^2])$, it has a symplectic period by Proposition 5.2.

This concludes case (b).

(c) $\pi = Z([1,\nu]) \times Z([\nu^2,\nu^3]) \times Z([\nu^4,\nu^5])$. Here the irreducible subquotients of π are $Z([1,\nu],[\nu^2,\nu^3],[\nu^4,\nu^5])$, $Z([1,\nu^3],[\nu^4,\nu^5])$, $Z([1,\nu],[\nu^2,\nu^5])$ and $Z([1,\nu^5])$. Of these, by Lemma 3.6, $Z([1,\nu^3],[\nu^4,\nu^5])$ and $Z([1,\nu],[\nu^2,\nu^5])$

do not have a symplectic period. $Z([1,\nu^5])$ being a character clearly has a symplectic period. We analyze the remaining representation.

- If $\theta = Z([1,\nu],[\nu^2,\nu^3],[\nu^4,\nu^5])$, by definition, it is the unique irreducible submodule of $Z([\nu^4,\nu^5]) \times Z([\nu^2,\nu^3]) \times Z([1,\nu])$. Thus we get that it is the unique irreducible quotient of $\pi = Z([1,\nu]) \times Z([\nu^2,\nu^3]) \times Z([\nu^4,\nu^5])$ (using Lemma 2.3 and taking contragredients). Since $Z([\nu^2,\nu^3],[\nu^4,\nu^5])$ is a quotient of $Z([\nu^2,\nu^3]) \times Z([\nu^4,\nu^5])$, we get θ is a quotient of $Z([1,\nu]) \times Z([\nu^2,\nu^3],[\nu^4,\nu^5])$. By taking $\mu_1 = Z([1,\nu])$ and $\mu_2 = Z([\nu^2,\nu^3],[\nu^4,\nu^5])$, and doing a calculation as in (a), we get that $Z([1,\nu]) \times Z([\nu^2,\nu^3],[\nu^4,\nu^5])$ and hence θ, doesn't have a symplectic period.

This concludes case (c).

(d) $\pi = Z([1,\nu]) \times Z([\nu,\nu^2]) \times Z([\nu,\nu^2])$. In this case, all irreducible subquotients of π are $Z([1,\nu],[\nu,\nu^2],[\nu,\nu^2])$ and $Z([1,\nu^2],[\nu],[\nu,\nu^2])$. We now analyze both of these representations.

- If $\theta = Z([1,\nu^2],[\nu],[\nu,\nu^2]) \cong Z([1,\nu^2]) \times Z([\nu,\nu^2]) \times \nu$, by Theorem 3.7 of [Offen and Sayag 2007b], it has a mixed Klyachko model. Hence by Theorem 2.5, it is not $\mathrm{Sp}_3(F)$-distinguished.
- If $\theta = Z([1,\nu],[\nu,\nu^2],[\nu,\nu^2])$, it is the unique irreducible submodule of $Z([\nu,\nu^2]) \times Z([\nu,\nu^2]) \times Z([1,\nu])$. Thus it is the unique irreducible submodule of $Z([\nu,\nu^2]) \times Z([1,\nu],[\nu,\nu^2]) \cong Z([\nu,\nu^2]) \times Q([1,\nu],[\nu,\nu^2])$ (by Example 11.4 in [Zelevinsky 1980]). By Theorem 1 of [Badulescu et al. 2012], this representation is irreducible and so $\theta \cong Z([\nu,\nu^2]) \times Z([1,\nu],[\nu,\nu^2])$ and has a symplectic period by Proposition 5.2.

This concludes case (d).

(e) $\pi = Z([1,\nu]) \times Z([\nu^2,\nu^3]) \times Z([\nu^2,\nu^3])$. In this case, all irreducible subquotients of π are $Z([1,\nu],[\nu^2,\nu^3],[\nu^2,\nu^3])$ and $Z([1,\nu^3],[\nu^2,\nu^3])$. We now analyze both of these representations.

- If $\theta = Z([1,\nu^3],[\nu^2,\nu^3]) \cong Z([1,\nu^3]) \times Z([\nu^2,\nu^3])$, it has a symplectic period by Proposition 5.2.
- If $\theta = Z([1,\nu],[\nu^2,\nu^3],[\nu^2,\nu^3])$, it is the unique irreducible submodule of $Z([\nu^2,\nu^3]) \times Z([\nu^2,\nu^3]) \times Z([1,\nu])$. Now, $Z([\nu^2,\nu^3]) \times Z([\nu^2,\nu^3]) \times Z([1,\nu])$ is glued from $Z([\nu^2,\nu^3]) \times Z([1,\nu],[\nu^2,\nu^3])$ and $Z([\nu^2,\nu^3]) \times Z([1,\nu^3])$. Using Theorem 1 of [Badulescu et al. 2012], we get the irreducibility of

$$Z([\nu^2,\nu^3]) \times Z([1,\nu],[\nu^2,\nu^3]) \cong Z([\nu^2,\nu^3]) \times Q([1],[\nu,\nu^2],[\nu^3]).$$

Thus $\theta \cong Z([\nu^2,\nu^3]) \times Z([1,\nu],[\nu^2,\nu^3])$, implying that $\tilde{\theta} \cong Z([\nu^{-3},\nu^{-2}]) \times Z([\nu^{-1},1],[\nu^{-3},\nu^{-2}])$. A calculation as in case (a), taking $\mu_1 = Z([\nu^{-3},\nu^{-2}])$

and $\mu_2 = Z([\nu^{-1}, 1], [\nu^{-3}, \nu^{-2}])$, shows that $\tilde{\theta} = \mu_1 \times \mu_2$ doesn't have a symplectic period. Thus even θ is not $\mathrm{Sp}_3(F)$-distinguished.

This concludes case (e).

The remaining cases, (f), (g), and (h), are dealt with by duality: all irreducible subquotients of π are twists of the contragredients of those obtained in cases (a), (d), and (e), respectively. Hence the only subquotients with a symplectic period are up to a twist, duals of the ones already obtained previously.

Case 2: π_1, π_2, π_3 are all twists of Steinberg. The representations that we are looking at in this case are of the form

$$\pi = Q([\chi_1, \chi_1\nu]) \times Q([\chi_1', \chi_1'\nu]) \times Q([\chi_1'', \chi_1''\nu]).$$

The following result will be used repeatedly in the analysis of this case.

Lemma 5.6. *Let $\pi = Q([\nu^a, \nu^{a+1}]) \times Q([\nu^b, \nu^{b+1}]) \times Q([\nu^c, \nu^{c+1}])$. Then π has a symplectic period only if $a = b = c + 1$.*

Proof. Let $\mu_1 = Q([\nu^a, \nu^{a+1}])$ and $\mu_2 = Q([\nu^b, \nu^{b+1}]) \times Q([\nu^c, \nu^{c+1}])$. Since μ_1 doesn't have a symplectic period (by Theorem 2.5), the group

$$\mathrm{Hom}_H\big(\mathrm{ind}_{H_{2,0}}^H(\delta_{P_{2,4}}^{1/2}\mu_1 \otimes \mu_2|_{H_{2,0}}), \mathbb{C}\big) = \mathrm{Hom}_{\mathrm{Sp}_1(F)}\big(\mu_1, 1\big) \otimes \mathrm{Hom}_{\mathrm{Sp}_2(F)}(\mu_2, \mathbb{C})$$

is zero, by (3-1). Thus the other term, $\mathrm{Hom}_H\big(\mathrm{ind}_{H_{2,2}}^H(\delta_{P_{2,4}}^{1/2}\mu_1 \otimes \mu_2|_{H_{2,2}}), \mathbb{C}\big)$, has to be nonzero. Now, $r_{(2,2),(4)}(Q([\nu^b, \nu^{b+1}]) \times Q([\nu^c, \nu^{c+1}])$ is glued from $Q([\nu^b, \nu^{b+1}]) \otimes Q([\nu^c, \nu^{c+1}])$, $Q([\nu^c, \nu^{c+1}]) \otimes Q([\nu^b, \nu^{b+1}])$ and

$$(\nu^{b+1} \times \nu^{c+1}) \otimes (\nu^b \times \nu^c),$$

by Lemma 2.12 of [Bernstein and Zelevinsky 1977]. It can be checked easily that replacing $r_{(2,2),(4)}(\mu_2)$ by the first two of the three representations makes this Hom space 0. Thus,

$$\mathrm{Hom}_{\mathrm{GL}_2(F)\times\mathrm{Sp}_1(F)}\big(\nu^{-1}Q([\nu^a, \nu^{a+1}]) \otimes (\nu^{b+1} \times \nu^{c+1}) \otimes (\nu^b \times \nu^c), \mathbb{C}\big) \neq 0.$$

Solving the equations for this to be nonzero gives the lemma. □

By similar arguments using Lemma 3.4 as in case 1 it can be easily concluded that if there is at most one link among the three segments then π doesn't have an H-distinguished subquotient. Thus we look at the case where there are at least two links among the segments. Since twisting by a character doesn't matter to us, without loss of generality we can assume χ_1 to be trivial. As before we have eight possible cases:

(a) $Q([1, \nu]) \times Q([\nu, \nu^2]) \times Q([\nu^3, \nu^4])$

(b) $Q([1, \nu]) \times Q([\nu, \nu^2]) \times Q([\nu^2, \nu^3])$

(c) $Q([1,\nu]) \times Q([\nu^2,\nu^3]) \times Q([\nu^4,\nu^5])$
(d) $Q([1,\nu]) \times Q([\nu,\nu^2]) \times Q([\nu,\nu^2])$
(e) $Q([1,\nu]) \times Q([\nu^2,\nu^3]) \times Q([\nu^2,\nu^3])$
(f) $Q([1,\nu]) \times Q([\nu^2,\nu^3]) \times Q([\nu^3,\nu^4])$
(g) $Q([1,\nu]) \times Q([1,\nu]) \times Q([\nu,\nu^2])$
(h) $Q([1,\nu]) \times Q([1,\nu]) \times Q([\nu^2,\nu^3])$

(a) $\pi = Q([1,\nu]) \times Q([\nu,\nu^2]) \times Q([\nu^3,\nu^4])$. Here, all irreducible subquotients of π are $Q([1,\nu],[\nu,\nu^2],[\nu^3,\nu^4])$, $Q([1,\nu^2],[\nu],[\nu^3,\nu^4])$, $Q([1,\nu^4],[\nu])$ and $Q([1,\nu],[\nu,\nu^4])$. We now analyze each of these representations.

- If $\theta = Q([1,\nu],[\nu,\nu^2],[\nu^3,\nu^4])$, it is the unique irreducible quotient of
$$Q([\nu^3,\nu^4]) \times Q([\nu,\nu^2]) \times Q([1,\nu]),$$
which doesn't have a symplectic period by Lemma 5.6. Hence θ doesn't have one.

- If $\theta = Q([1,\nu^2],[\nu],[\nu^3,\nu^4])$, it is the unique irreducible quotient of
$$Q([\nu^3,\nu^4]) \times \nu \times Q([1,\nu^2]).$$
Thus it is an irreducible quotient of $Q([\nu^3,\nu^4]) \times Q([1,\nu]) \times Q([\nu,\nu^2])$ (by Theorem 2.2), which doesn't have a symplectic period by Lemma 5.6. Hence θ doesn't have one.

- If $\theta = Q([1,\nu^4],[\nu]) \cong Q([1,\nu^4]) \times \nu$ (by Proposition 8.5 of [Zelevinsky 1980]), it is generic and hence doesn't have a symplectic period (by Theorem 2.5).

- If $\theta = Q([1,\nu],[\nu,\nu^4])$, it is a quotient of $Q([\nu,\nu^2]) \times Q([\nu^3,\nu^4]) \times Q([1,\nu])$, which doesn't have a symplectic period by Lemma 5.6. Hence it doesn't have one too.

This concludes case (a).

(b) $\pi = Q([1,\nu]) \times Q([\nu,\nu^2]) \times Q([\nu^2,\nu^3])$. Here the irreducible subquotients of π are $Q([1,\nu],[\nu,\nu^2],[\nu^2,\nu^3])$, $Q([1,\nu^2],[\nu],[\nu^2,\nu^3])$, $Q([1,\nu],[\nu,\nu^3],[\nu^2])$, $Q([1,\nu^3],[\nu],[\nu^2])$, $Q([1,\nu^2],[\nu,\nu^3])$ and $Q([1,\nu^3],[\nu,\nu^2])$.

- If $\theta = Q([1,\nu],[\nu,\nu^2],[\nu^2,\nu^3])$, twisting θ by an appropriate power of ν makes it a unitary representation. By Theorem 1.2 it doesn't have a symplectic period.

- If $\theta = Q([1,\nu^2],[\nu],[\nu^2,\nu^3])$, it is the unique irreducible quotient of
$$Q([\nu^2,\nu^3]) \times \nu \times Q([1,\nu^2]).$$
This itself is a quotient of $Q([\nu^2,\nu^3]) \times Q([1,\nu]) \times Q([\nu,\nu^2])$, which doesn't have a symplectic period by Lemma 5.6. Hence θ doesn't have one.

- If $\theta = Q([1,\nu],[\nu,\nu^3],[\nu^2])$, it can be obtained by twisting the contragredient of $Q([1,\nu^2],[\nu],[\nu^2,\nu^3])$, which as showed in the last paragraph, doesn't have a symplectic period.
- If $\theta = Q([1,\nu^3],[\nu],[\nu^2])$, it is the unique irreducible quotient of $\nu^2 \times \nu \times Q([1,\nu^3])$ and hence of $Z([\nu,\nu^2]) \times Q([1,\nu^3])$. Now doing a similar calculation as in case 1(a) by taking $\mu_1 = Z([\nu,\nu^2])$ and $\mu_2 = Q([1,\nu^3])$ we get that $Z([\nu,\nu^2]) \times Q([1,\nu^3])$, and hence θ, doesn't have a symplectic period.
- If $\theta = Q([1,\nu^2],[\nu,\nu^3])$, it has a symplectic period by Theorem 1.2.
- If $\theta = Q([1,\nu^3],[\nu,\nu^2]) \cong Q([1,\nu^3]) \times Q([\nu,\nu^2])$, (by Proposition 8.5 of [Zelevinsky 1980]), it is generic and hence doesn't have a symplectic period (by Theorem 2.5).

This concludes case (b).

(c) $\pi = Q([1,\nu]) \times Q([\nu^2,\nu^3]) \times Q([\nu^4,\nu^5])$. Here the irreducible subquotients of π are $Q([1,\nu],[\nu^2,\nu^3],[\nu^4,\nu^5])$, $Q([1,\nu^3],[\nu^4,\nu^5])$, $Q([1,\nu],[\nu^2,\nu^5])$ and $Q([1,\nu^5])$.

- If $\theta = Q([1,\nu^5])$, it is generic and hence doesn't have a symplectic period.
- If $\theta = Q([1,\nu],[\nu^2,\nu^3],[\nu^4,\nu^5])$, it is the unique irreducible quotient of $Q([\nu^4,\nu^5]) \times Q([\nu^2,\nu^3]) \times Q([1,\nu])$, which doesn't have a symplectic period by Lemma 5.6. Hence it doesn't have one. The other two cases are dealt similarly.
- If $\theta = Q([1,\nu^3],[\nu^4,\nu^5])$, it is the unique irreducible quotient of $Q([\nu^4,\nu^5]) \times Q([1,\nu^3])$. Now this itself is a quotient of $Q([\nu^4,\nu^5]) \times Q([1,\nu]) \times Q([\nu^2,\nu^3])$, which doesn't have a symplectic period by Lemma 5.6. Hence θ doesn't have one.
- If $\theta = Q([1,\nu],[\nu^2,\nu^5])$, it can be obtained by twisting the contragredient of $Q([1,\nu^3],[\nu^4,\nu^5])$, which as shown in the last paragraph, doesn't have a symplectic period.

This concludes case (c).

(d) $\pi = Q([1,\nu]) \times Q([\nu,\nu^2]) \times Q([\nu,\nu^2])$. In this case, all irreducible subquotients of π are $Q([1,\nu],[\nu,\nu^2],[\nu,\nu^2])$ and $Q([1,\nu^2],[\nu],[\nu,\nu^2])$.

- If $\theta = Q([1,\nu^2],[\nu],[\nu,\nu^2]) \cong Q([1,\nu^2]) \times \nu \times Q([\nu,\nu^2])$, it is generic and hence doesn't have a symplectic period.
- If $\theta = Q([1,\nu],[\nu,\nu^2],[\nu,\nu^2])$, it is the sole irreducible quotient of $Q([\nu,\nu^2]) \times Q([\nu,\nu^2]) \times Q([1,\nu])$. Thus it is the unique irreducible quotient of $Q([\nu,\nu^2]) \times Q([1,\nu],[\nu,\nu^2]) \cong Q([\nu,\nu^2]) \times Z([1,\nu],[\nu,\nu^2])$ (by Example 11.4 in [Zelevinsky 1980]). By Theorem 1 of [Badulescu et al. 2012], this representation is

irreducible and so $\theta \cong Q([\nu,\nu^2]) \times Z([1,\nu],[\nu,\nu^2])$. So it is a quotient of $Q([\nu,\nu^2]) \times Z([1,\nu]) \times Z([\nu,\nu^2])$. Now a calculation as in case 1(a), taking $\mu_1 = Q([\nu,\nu^2])$ and $\mu_2 = Z([1,\nu]) \times Z([\nu,\nu^2])$, yields that $Q([\nu,\nu^2]) \times Z([1,\nu]) \times Z([\nu,\nu^2])$, and hence θ, doesn't have a symplectic period.

This concludes case (d).

(e) $\pi = Q([1,\nu]) \times Q([\nu^2,\nu^3]) \times Q([\nu^2,\nu^3])$. In this case, all irreducible subquotients of π are $Q([1,\nu],[\nu^2,\nu^3],[\nu^2,\nu^3])$ and $Q([1,\nu^3],[\nu^2,\nu^3])$.

- If $\theta = Q([1,\nu^3],[\nu^2,\nu^3]) \cong Q([1,\nu^3]) \times Q([\nu^2,\nu^3])$, (by Proposition 8.5 of [Zelevinsky 1980]), it is generic and hence doesn't have a symplectic period (by Theorem 2.5).
- If $\theta = Q([1,\nu],[\nu^2,\nu^3],[\nu^2,\nu^3])$, it is the unique irreducible quotient of $Q([\nu^2,\nu^3]) \times Q([\nu^2,\nu^3]) \times Q([1,\nu])$. This doesn't have a symplectic period by Lemma 5.6 and so θ doesn't have one too.

This concludes case (e).

As before, in cases (f), (g), and (h) all the irreducible subquotients of π are twists of the contragredients of the ones obtained in cases (a), (d), and (e) respectively. Hence the only subquotients with a symplectic period are up to a twist, duals of the ones already obtained previously.

Cases 3–7: The remaining five cases of $\pi_1 \times \pi_2 \times \pi_3$ are dealt similarly, proving Theorem 1.4. We just mention that no new H-distinguished subquotients are obtained from the other cases.

6. Conjectures for the general case

Theorem 1.3 and Theorem 1.4 prompt us to make certain conjectures for the general $2n$ case. In order to do so we need to set up notation.

Define $\mathfrak{G}'$ as the set of all representations of $\mathrm{GL}_{2n}(F)$ of the form $Z(\Delta_1,\dots,\Delta_r)$ that satisfy the following properties:

(1) All the segments are in the same cuspidal line.

(2) Each segment is of even length.

(3) No two segments have the beginning element in common.

(4) Conditions (1) and (3) imply that there is a natural ordering of the segments (with respect to the beginning element). Arrange $\Delta_1,\dots,\Delta_r$ accordingly. We require that the intersection of each segment with its neighbors is odd in length, in particular is nonempty.

The set $\mathfrak{G}'$ is contained in the set of ladder representations as defined in [Badulescu et al. 2012].

Further define $\mathfrak{G} \subset \cup_{i\geq 1} \mathrm{Irr}(\mathrm{GL}_{2i}(F))$ to be the set of all irreducible products of elements in $\mathfrak{G}'$; i.e.,

$$\mathfrak{G} = \{\pi_1 \times \cdots \times \pi_t \mid \pi_1, \dots, \pi_t \in \mathfrak{G}' \text{ and the product is irreducible}\}.$$

Let us now state the conjecture in the general case using the above notation.

Conjecture 6.1. *Let θ be an irreducible representation of $\mathrm{GL}_{2n}(F)$ carrying a symplectic period. Then there exists $\pi_1, \dots, \pi_t \in \mathfrak{G}'$ such that*

$$\theta \cong \pi_1 \times \cdots \times \pi_t.$$

In other words, $\theta \in \mathfrak{G}$.

The following proposition verifies the conjecture for unitary representations.

Proposition 6.2. *Let θ be an irreducible unitary representation having a symplectic period. Then $\theta \in \mathfrak{G}$.*

Proof. Let $\delta = Q([\rho\nu^{\frac{1-d}{2}}, \rho\nu^{\frac{d-1}{2}}])$. By Theorem A.10(iii) of [Tadić 1986], $U(\delta, t) = Z(\Delta_1, \dots, \Delta_d)$, where

$$\begin{aligned}
\Delta_1 &= [(\rho\nu^{\frac{1-d}{2}})\nu^{\frac{1-t}{2}}, (\rho\nu^{\frac{1-d}{2}})\nu^{\frac{t-1}{2}}],\\
\Delta_2 &= [(\rho\nu^{\frac{3-d}{2}})\nu^{\frac{1-t}{2}}, (\rho\nu^{\frac{3-d}{2}})\nu^{\frac{t-1}{2}}], \dots,\\
\Delta_d &= [(\rho\nu^{\frac{d-1}{2}})\nu^{\frac{1-t}{2}}, (\rho\nu^{\frac{d-1}{2}})\nu^{\frac{t-1}{2}}].
\end{aligned}$$

The intersection of each segment with both its neighbors, if they are arranged in the order of precedence, is of length $t-1$. So if t is even, $U(\delta, t) \in \mathfrak{G}'$. The proposition then follows from Theorem 1.2. □

That $U(\delta, 2m) \in \mathfrak{G}'$ leads to an obvious question generalizing Proposition 1.1, which we state as the next conjecture.

Conjecture 6.3 (hereditary property). *Let $\theta \in \mathfrak{G}'$. Then θ has a symplectic period. Moreover, if $\theta_1, \dots, \theta_d \in \mathfrak{G}'$ then $\theta_1 \times \cdots \times \theta_d$ has a symplectic period.*

Conjecture 6.1 and Conjecture 6.3 together imply that $\mathfrak{G}$ is precisely the set of H-distinguished representations of the linear groups. Thus Theorem 1.3 and Theorem 1.4 prove the conjectures for $\mathrm{GL}_4(F)$ and $\mathrm{GL}_6(F)$. Note that the above conjectures together imply that the property of having a symplectic period is dependent only on the combinatorial structure of the segments involved and not on the building blocks, i.e., the cuspidal representations. More precisely:

Conjecture 6.4. *Let $\pi \in \mathrm{Irr}(\mathrm{GL}_{2n}(F))$ be of the form $Z(\Delta_1, \dots, \Delta_r)$ such that all the segments are in the same cuspidal line. Let $\rho \in \mathrm{Irr}(\mathrm{GL}_m(F))$ be an element of the line. Let Δ_i' be the segment obtained from Δ_i by replacing ρ with the trivial representation of $F^\times$ and π' be the representation $Z(\Delta_1', \dots, \Delta_r')$ of $\mathrm{GL}_{2n/m}(F)$.*

(1) *If* $2n/m$ *is even*, π *has a symplectic period if and only if* π' *has a symplectic period.*

(2) *If* $2n/m$ *is odd*, π *doesn't have a symplectic period.*

Acknowledgements

This work is part of the author's thesis. He wishes to thank his advisor, Dipendra Prasad, for suggesting the problem and for his guidance. He thanks Omer Offen, Eitan Sayag and Steven Spallone for their interest and suggestions to improve the paper.

References

[Badulescu et al. 2012] I. Badulescu, E. Lapid, and A. Minguez, "Une condition suffisante pour l'irréducibilité d'une induite parabolique de $\mathrm{GL}(m, D)$", preprint, 2012, http://www.math.huji.ac.il/~erezla/papers/BLM_final.pdf. To appear in *Ann. Inst. Fourier* (*Grenoble*).

[Bernstein and Zelevinsky 1976] I. N. Bernstein and A. V. Zelevinsky, "Representations of the group $GL(n, F)$, where F is a local non–Archimedean field", *Uspehi Mat. Nauk* **31**:3(189) (1976), 5–70. In Russian; transated in *Russ. Math. Surv.* **31**:3 (1976) 1–68. MR 54 #12988 Zbl 0342.43017

[Bernstein and Zelevinsky 1977] I. N. Bernstein and A. V. Zelevinsky, "Induced representations of reductive $\mathfrak{p}$-adic groups, I", *Ann. Sci. École Norm. Sup.* (4) **10**:4 (1977), 441–472. MR 58 #28310 Zbl 0412.22015

[Casselman 1995] W. Casselman, "Notes on p-adic representation theory", unpublished book, 1995, http://www.math.ubc.ca/cass/research/pdf/p-adic-book.pdf.

[Heumos and Rallis 1990] M. J. Heumos and S. Rallis, "Symplectic–Whittaker models for Gl_n", *Pacific J. Math.* **146**:2 (1990), 247–279. MR 91k:22036 Zbl 0752.22012

[Klyachko 1983] A. A. Klyachko, "Models for complex representations of groups $\mathrm{GL}(n, q)$", *Mat. Sb.* (*N.S.*) **120(162)**:3 (1983), 371–386. In Russian; translated in *Mathematicheski Sbornik. Novaya Seriya* **48**:2 (1984), 365–378. MR 84j:20014 Zbl 0526.20033

[Offen 2006] O. Offen, "Residual spectrum of GL_{2n} distinguished by the symplectic group", *Duke Math. J.* **134**:2 (2006), 313–357. MR 2007h:11063 Zbl 1220.11072

[Offen and Sayag 2007a] O. Offen and E. Sayag, "On unitary representations of GL_{2n} distinguished by the symplectic group", *J. Number Theory* **125**:2 (2007), 344–355. MR 2009a:22012 Zbl 1147.11028

[Offen and Sayag 2007b] O. Offen and E. Sayag, "Global mixed periods and local Klyachko models for the general linear group", *Int. Math. Res. Not.* **2007**:1 (2007), Art. ID rnm 136, 25. MR 2009e:22017 Zbl 1158.22021

[Offen and Sayag 2008] O. Offen and E. Sayag, "Uniqueness and disjointness of Klyachko models", *J. Funct. Anal.* **254**:11 (2008), 2846–2865. MR 2009e:22018 Zbl 1143.22013

[Prasad 1990] D. Prasad, "Trilinear forms for representations of GL(2) and local ϵ-factors", *Compositio Math.* **75**:1 (1990), 1–46. MR 91i:22023 Zbl 0731.22013

[Prasad 1993] D. Prasad, "On the decomposition of a representation of GL(3) restricted to GL(2) over a p-adic field", *Duke Math. J.* **69**:1 (1993), 167–177. MR 93m:22019 Zbl 0827.22006

[Tadić 1986] M. Tadić, "Classification of unitary representations in irreducible representations of general linear group (non–Archimedean case)", *Ann. Sci. École Norm. Sup.* (4) **19**:3 (1986), 335–382. MR 88b:22021 Zbl 0614.22005

[Zelevinsky 1980] A. V. Zelevinsky, "Induced representations of reductive $\mathfrak{p}$-adic groups, II: On irreducible representations of GL(n)", *Ann. Sci. École Norm. Sup.* (4) **13**:2 (1980), 165–210. MR 83g:22012 Zbl 0441.22014

Received March 7, 2012. Revised October 21, 2012.

ARNAB MITRA
TATA INSTITUTE OF FUNDAMENTAL RESEARCH
COLABA
MUMBAI 400005
INDIA
00.arnab.mitra@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 268, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.268.465

LINKED TRIPLES OF QUATERNION ALGEBRAS

ALEXANDER S. SIVATSKI

Let F be a field of characteristic different from 2, and let Q_1, Q_2, Q_3 be quaternion algebras over F such that any element in $\mathrm{Br}(F)$ generated by Q_1, Q_2, Q_3 has index at most 2. For the triple $\{Q_1, Q_2, Q_3\}$ we construct a certain invariant, lying in $I^3(F)$, which is a 4-fold Pfister form, provided the algebras Q_1, Q_2, Q_3 have a common slot. Among other results we prove that the algebras Q_1, Q_2, Q_3 have a common slot if and only if the torsion of the group $\mathrm{CH}^2(X_1 \times X_2 \times X_3)$ is zero, where X_i is the projective conic associated with the algebra Q_i.

Let F be a field with char $F \neq 2$, let $Q_1, \ldots, Q_n$ be quaternion algebras over F, and $G \subset {}_2\mathrm{Br}(F)$ the group generated by these quaternions. We call the collection $\{Q_1, \ldots, Q_n\}$ linked if $\mathrm{ind}(\alpha) \leq 2$ for any $\alpha \in G$. We say that the collection $\{Q_1, \ldots, Q_n\}$ has a common slot if there are $a, b_1, \ldots, b_n \in F^*$ such that $Q_i = (a, b_i)$ for each $1 \leq i \leq n$. Obviously, if the collection $\{Q_1, \ldots, Q_n\}$ has a common slot, then it is linked. The opposite statement is true for $n = 2$ [Scharlau 1985], but is not true for $n \geq 3$. Indeed, it was shown in [Peyre 1995] that if $\sqrt{-1} \in F$, $a, b, c \in F^*$, and $a \cup b \cup c \neq 0$ in $H^3(F, \mathbb{Z}/2\mathbb{Z})$, then the triple $\{(a, b), (a, c), (b, c)\}$ is linked, but has no common slot. However, this example does not make clear whether there exists some obstruction that does not permit a given linked triple of quaternion algebras to have a common slot. In this note we construct such an obstruction, which lies in the Witt ring of the field F; more precisely, it lies in $I^3(F)$.

We use the notation usual in the theory of quadratic forms, which can be found, for instance, in the books [Lam 2005; Scharlau 1985; Elman et al. 2008]. The word "form" always means a quadratic form over a field of characteristic different from 2. For a form φ defined on a linear space V over a field F, by $D(\varphi)$ we denote the set of nonzero values $\varphi(v)$, where $v \in V$. The n-fold Pfister form $\langle\langle a_1, \ldots, a_n \rangle\rangle$ is the form $\langle 1, -a_1 \rangle \otimes \cdots \otimes \langle 1, -a_n \rangle$. (Take notice of signs!) Sometimes, slightly abusing notation, we consider regular forms over F as their images in the Witt ring $W(F)$, and, vice versa, elements of $W(F)$ as the associated *anisotropic* forms. In particular, we call the corresponding elements in $W(F)$ n-fold Pfister forms. The signs $+$

MSC2010: primary 16K50; secondary 19D45.
Keywords: quaternion algebra, quadratic form, Chow group.

and $-$ are used for operations in $W(F)$. The dimension of the anisotropic form associated with $\varphi \in W(F)$ is denoted by $\dim\varphi$. By $F(\varphi)$ we denote the function field of the projective quadric associated with the *form* φ over the field F. By $i_1(\varphi)$ we denote the first Witt index of the form φ, which is defined as follows: $i_1(\varphi)=n$, where $\varphi_{F(\varphi)} \simeq \psi \perp n\mathbb{H}$ and ψ is the anisotropic part of $\varphi_{F(\varphi)}$.

Let $\mathfrak{T}=\{Q_1, Q_2, Q_3\}$ be a linked triple. For any quaternion algebra α denote by $c(\alpha)$ the 2-fold Pfister form corresponding to α. Put

$$c(\mathfrak{T})=c(Q_1)+c(Q_2)+c(Q_3)-c(Q_1+Q_2)-c(Q_1+Q_3)\\ -c(Q_2+Q_3)+c(Q_1+Q_2+Q_3)\in W(F).$$

The invariant $c(\mathfrak{T})$ has the following properties:

Proposition 1. (1) $c(\mathfrak{T})\in I^3(F)$, $c(\mathfrak{T})$ *is divisible by each* $c(Q_i)$, *and* $\dim c(\mathfrak{T})$ *equals either* 0, 8, *or* 16. *Moreover*, *if* $\dim c(\mathfrak{T})=16$, *then* $1\in D(c(\mathfrak{T}))$.

(2) $c(\mathfrak{T})$ *is a sum of two elements*, *similar to a* 4*-fold and a* 3*-fold Pfister form*, *respectively. In particular*, $c(\mathfrak{T})\ (\mathrm{mod}\ I^4(F))$ *is a symbol in* $I^3(F)/I^4(F)$.

(3) *If* $\mathfrak{T}$ *has a common slot*, *then* $c(\mathfrak{T})$ *is a* 4*-fold Pfister form.*

Proof. (1) It is obvious that the elements $c(Q_1)+c(Q_2)-c(Q_1+Q_2)$ and $c(Q_3)+c(Q_1+Q_2+Q_3)-c(Q_2+Q_3)-c(Q_1+Q_3)$ both are from $I^3(F)$ and split by the extension $F(SB(Q_1))/F$. Hence they are divisible by $c(Q_1)$ [Scharlau 1985]. Since the dimensions of the elements $c(Q_1)-c(Q_1+Q_2)$, $c(Q_2)-c(Q_2+Q_3)$ and $c(Q_3)-c(Q_1+Q_3)$ are at most 4, we get

$$\dim\big(c(Q_1)+c(Q_2)+c(Q_3)-c(Q_1+Q_2)-c(Q_1+Q_3)-c(Q_2+Q_3)\big)\le 12.$$

Hence $\dim c(\mathfrak{T})\le 16$, and if $\dim c(\mathfrak{T})=16$, then the Pfister form $c(Q_1+Q_2+Q_3)$ is a direct summand of $c(\mathfrak{T})$, and so $1\in D(c(\mathfrak{T}))$ (here we consider $c(\mathfrak{T})$ as the corresponding quadratic form). Furthermore, since $c(\mathfrak{T})\in I^3(F)$ and is divisible by $c(Q_1)$, it follows that $\dim c(\mathfrak{T})$ is divisible by 8. By symmetry $c(\mathfrak{T})$ is divisible by each $c(Q_i)$.

(2) It follows at once from part (1) that

$$\begin{aligned}c(\mathfrak{T})&=c(Q_1)\otimes\langle a_1,a_2,a_3,a_4\rangle\\ &=c(Q_1)\otimes\langle a_1,a_2,a_3,a_1a_2a_3\rangle+c(Q_1)\otimes\langle -a_1a_2a_3,a_4\rangle\end{aligned}$$

for some $a_i\in F^*$. The first summand is similar to a 4-fold Pfister form, and the second one to a 3-fold Pfister form.

(3) Assume that $Q_i=(a,b_i)$ for some $a,b_1,b_2,b_3\in F^*$. Then a straightforward computation shows that

$$c(\mathfrak{T})=\langle\langle a\rangle\rangle\otimes\langle 1,-b_1,-b_2,-b_3,b_1b_2,b_1b_3,b_2b_3,-b_1b_2b_3\rangle=\langle\langle a,b_1,b_2,b_3\rangle\rangle. \quad \square$$

As another example of computation of $c(\mathfrak{T})$, which will be used in the sequel, we have the following:

Proposition 2. *Let F be a field, $a, b_1, b_2 \in F^*$. Let Q be a quaternion algebra over F such that $Q_{F(\sqrt{b_I})} = 0$ for any nonempty $I \subset \{1, 2\}$, where $b_I = \prod_{i \in I} b_i$. Then the triple $\{(a, b_1), (a, b_2), Q\}$ is linked, and $c(\mathfrak{T}) = \langle\langle a\rangle\rangle c(Q)$. Moreover, if $\{(a, b_1), (a, b_2), Q\}$ has a common slot, then $c(\mathfrak{T}) = 0$.*

Proof. For any nonempty $I \subset \{1, 2\}$ there is some element $c_I \in F^*$ such that $Q = (b_I, c_I)$. Obviously,

$$Q + \sum_{i \in I}(a, b_i) = (ac_I, b_I),$$

hence the triple $\mathfrak{T}$ is linked. Furthermore, in view of Proposition 1(3) and the equality $\langle\langle xy, z\rangle\rangle = \langle\langle x, z\rangle\rangle + \langle\langle y, z\rangle\rangle - \langle\langle x, y, z\rangle\rangle$ we get

$$\begin{aligned} c(\mathfrak{T}) &= \langle\langle a, b_1\rangle\rangle + \langle\langle a, b_2\rangle\rangle - \langle\langle a, b_1 b_2\rangle\rangle + Q + \sum_{I \subset \{1,2\}} (-1)^{|I|} \langle\langle ac_I, b_I\rangle\rangle \\ &= \langle\langle a, b_1, b_2\rangle\rangle + Q + \sum_{I \subset \{1,2\}} (-1)^{|I|} \big(\langle\langle a, b_I\rangle\rangle + Q - \langle\langle a, b_I, c_I\rangle\rangle\big) \\ &= \langle\langle a, b_1, b_2\rangle\rangle + \sum_{I \subset \{1,2\}} (-1)^{|I|} \langle\langle a, b_I\rangle\rangle + \sum_{I \subset \{1,2\}} (-1)^{|I|+1} \langle\langle a\rangle\rangle c(Q) \\ &= \langle\langle a, b_1, b_2\rangle\rangle - \langle\langle a, b_1, b_2\rangle\rangle + \langle\langle a\rangle\rangle c(Q) = \langle\langle a\rangle\rangle c(Q). \end{aligned}$$

By Proposition 1(3), the assumption that the triple $\{(a, b_1), (a, b_2), Q\}$ has a common slot implies $c(\mathfrak{T}) = 0$. □

As a consequence we obtain another proof of the example from [Peyre 1995].

Corollary 3. *Let F be a field, $\sqrt{-1} \in F$, $a, b, c \in F^*$. If $a \cup b \cup c \neq 0$ in $H^3(F, \mathbb{Z}/2\mathbb{Z})$, then the triple $\mathfrak{T} = \{(a, b), (a, c), (b, c)\}$ is linked, but does not have a common slot.*

Proof. Since $\langle\langle b, c\rangle\rangle = \langle\langle b, -bc\rangle\rangle = \langle\langle b, bc\rangle\rangle$, we have

$$(a, b) + (a, c) + (b, c) = (a, bc) + (b, bc) = (ab, bc),$$

so the triple is linked. Since

$$(b, c)_{F(\sqrt{b})} = (b, c)_{F(\sqrt{c})} = (b, c)_{F(\sqrt{bc})} = 0,$$

we get by Proposition 2 that $c(\mathfrak{T}) = \langle\langle a, b, c\rangle\rangle \neq 0$. Also by Proposition 2 we get that the triple $\mathfrak{T}$ has no common slot. □

Recall that a field L is called linked if any two quaternion algebras over L have a common slot [Elman et al. 2008]. A natural question arises whether there exists a linked field L and a triple of quaternion algebras $\{Q_1, Q_2, Q_3\}$ over L without a common slot. The answer is in affirmative, and one can construct such a field L even with a few additional properties. To do this we need a couple of lemmas.

Lemma 4. *Let φ be an anisotropic Albert form (i.e., an anisotropic 6-dimensional form with trivial discriminant) over a field F, π the 2-fold Pfister form over $F(\varphi)$ similar to $\varphi_{F(\varphi)}$. Then $\pi \notin h + I^4(F(\varphi))$ for any $h \in I^2(F)$.*

Proof. We will induct on $\dim h$. Assume the converse, i.e., that there is $h \in I^2(F)$ such that $\pi \in h + I^4(F(\varphi))$. Consider a few cases.

(a) $h = 0$. Then $\pi = 0$, hence $\varphi_{F(\varphi)} = 0$. Choose any $a \in F^*$ such that the form $\varphi_{F(\sqrt{a})}$ is isotropic. In particular, $a \notin F^{*2}$. Obviously, $\varphi_{F(\sqrt{a})} = 0$, i.e., $\varphi \simeq \langle\langle a\rangle\rangle\psi$ for some 3-dimensional form ψ. Comparing discriminants, we get $a \in F^{*2}$, a contradiction.

(b) The form $\varphi_{F(h)}$ is isotropic. Then there exists $c \in F^*$ such that either $\varphi = ch$, or $\varphi = c(h - \tilde{h})$ for some form $\tilde{h} \neq 0$ with $\dim h = \dim \tilde{h} = 4$ [Merkurjev 1991]. In both cases by dimension count we have $\pi = h_{F(\varphi)}$. If $\varphi \simeq ch$, then we have $\varphi_{F(\varphi)} = c\pi$; therefore, $\dim(\varphi \perp \langle -c\rangle)_{F(\varphi)} = 3$. If $c \in D(\varphi)$, then the field $F(\varphi)$ splits a proper subform of φ, which is impossible, since $i_1(\varphi) = 1$ [Karpenko and Merkurjev 2003]. (Another, more elementary way to come to a contradiction is to make the form $(\varphi \perp \langle -c\rangle)_{\mathrm{an}}$ a Pfister neighbor, not splitting the form φ [Hoffmann 1995].) If $c \notin D(\varphi)$, then $\dim(\varphi \perp \langle -c\rangle) = 7$. Moreover, $i_1(\varphi \perp \langle -c\rangle) = 1$, since $\varphi_{F(\varphi\perp\langle -c\rangle)}$ is anisotropic [Merkurjev 1991]. Since $c \in D(\varphi_{F(\varphi\perp\langle -c\rangle)})$, changing F for $F(\varphi \perp \langle -c\rangle)$, we get a contradiction again.

If $\varphi = c(h - \tilde{h})$ and $\varphi_{F(\varphi)} = u\pi = uh$ for some $u \in F(\varphi)^*$, then $c\tilde{h} = ch - uh \in I^3(F(\varphi))$, hence $\tilde{h}_{F(\varphi)} = 0$, which is also impossible.

(c) $\varphi_{F(h)}$ is anisotropic, $h \neq 0$. Then, passing to the field $F(h)$, the forms $(h_{F(h)})_{\mathrm{an}}$, $\pi_{F(h)(\varphi)}$ and the Albert form $\varphi_{F(h)}$, we can conclude by induction on $\dim h$ that this case is impossible as well, which finishes the proof. □

Lemma 5. *Let F be a field, Q_1, Q_2, Q_3 pairwise distinct nontrivial quaternion algebras, φ an anisotropic Albert form. Suppose $\{Q_1, Q_2, Q_3\}$ is not a linked triple. Then either the triple $\mathfrak{T} = \{Q_{1F(\varphi)}, Q_{2F(\varphi)}, Q_{3F(\varphi)}\}$ is not linked, or it is linked and $c(\mathfrak{T}_{F(\varphi)}) \notin I^4(F(\varphi))$. In particular, the triple $\mathfrak{T}_{F(\varphi)}$ has no common slot.*

Proof. Notice first that all $\sum_{i\in I} Q_i$ $(I \subset \{1, 2, 3\})$ are pairwise distinct. Suppose the triple $\mathfrak{T}$ is linked. Then there is a unique $\alpha \in G = \langle Q_1, Q_2, Q_3\rangle \in {}_2\operatorname{Br}(F)$ such that $\operatorname{ind}(\alpha) = 4$, and, moreover, φ is an Albert form for α. Assume that $c(\mathfrak{T}) \in I^4(F(\varphi))$. Then we get $c(\alpha)_{F(\varphi)} \in h + I^4(F(\varphi))$ for some form $h \in I^2(F)$, which contradicts Lemma 4. □

Theorem 6. *Suppose that the triple $\{Q_1, Q_2, Q_3\}$ over a field F is not linked. Then there exists a field extension L/F with the following properties*:

(1) *The field L is linked. In particular, the triple $\{Q_1, Q_2, Q_3\}_L$ is linked.*

(2) *The triple $\{Q_1, Q_2, Q_3\}_L$ has no common slot.*

(3) *The field L has no proper odd degree extension.*

(4) *Any 9-dimensional form over L is isotropic.*

(5) *Up to isomorphism there is a unique nontrivial 3-fold Pfister form over L, namely the one similar to $c(\{Q_1, Q_2, Q_3\}_L)$.*

Proof. We will apply a procedure similar to the one used in the construction of fields with prescribed even U-invariant [Merkurjev 1991]. First, applying Lemma 5 a few times, we can construct a field extension K/F such that the triple $\{Q_1, Q_2, Q_3\}_K$ is linked, but has no common slot. Further, splitting if needed some 4-fold Pfister form over K, we pass to the field K_1/K such that $c(\{Q_1, Q_2, Q_3\}_{K_1})$ is similar to some 3-fold Pfister form τ over K_1. Next, splitting all Albert forms, all forms of dimension at least 9, and all 3-fold Pfister forms differing from τ, then passing to a maximal odd degree extension, we construct a tower of fields $K_1 \subset K_2 \subset \cdots$ such that $L = \bigcup_i K_i$ satisfies all the required properties. The point is that for each i the form τ_{K_i} remains nontrivial; this proves the absence of a common slot of the triple $\{Q_1, Q_2, Q_3\}_{K_i}$. □

Theorem 6 has an application for Chow groups of codimension 2 of the product of three projective conics. We give numerous examples of increasing the torsion of these groups when passing to some field extension. We need the following statement, which is an immediate consequence of Theorem 4.1, Proposition 6.1 and Remark 4.1 from [Peyre 1995].

Peyre's Theorem. *Let F be a field, Q_1, Q_2, Q_3 quaternion algebras over F, and X_1, X_2, X_3 the corresponding projective conics. Denote by G the subgroup of ${}_2\operatorname{Br}(F)$ generated by all Q_i. Then*:

(1) *The torsion of the group $\mathrm{CH}^2(X_1 \times X_2 \times X_3)$ is either zero, or $\mathbb{Z}/2\mathbb{Z}$.*

(2) *Denote by d the least common multiple of all the numbers $\operatorname{ind}(\alpha)$, where α runs over all the elements of G. The following two assertions are equivalent*:

 (i) *The algebras Q_i have a common splitting field of degree dm, where m is an odd integer.*

 (ii) *The torsion of the group $\mathrm{CH}^2(X_1 \times X_2 \times X_3)$ equals 0.*

(3) *If the algebras Q_i have a common slot, the sequence*

$$F^*/F^{*2} \otimes G \xrightarrow{\text{cup product}} H^3(F, \mathbb{Z}/2\mathbb{Z}) \xrightarrow{\text{res}} H^3\big(F(X_1 \times X_2 \times X_3), \mathbb{Z}/2\mathbb{Z}\big)$$

is exact.

Corollary 7. *Let F be a field, $\mathfrak{T} = \{Q_1, Q_2, Q_3\}$ a nonlinked triple of quaternion algebras. Suppose that either at least one of the elements $Q_1 + Q_2$, $Q_1 + Q_3$, $Q_2 + Q_3$, $Q_1 + Q_2 + Q_3$ has index* 2, *or* $\operatorname{ind}(Q_1 + Q_2 + Q_3) = 8$. *Denote by X_i the projective conic corresponding to Q_i. Further, let L/F be the field extension constructed in Theorem 6 for the triple $\mathfrak{T}$. Then the torsion of* $\mathrm{CH}^2(X_1 \times X_2 \times X_3)$ *is zero, but the torsion of* $\mathrm{CH}^2(X_{1L} \times X_{2L} \times X_{3L})$ *is* $\mathbb{Z}/2\mathbb{Z}$.

Proof. Let us compute the number d for the algebras Q_1, Q_2, Q_3. Since the triple $\{Q_1, Q_2, Q_3\}$ is not linked, $d \geq 4$. In the first case, since the sum of a few Q_i has index 2, we get $d \neq 8$, i.e., $d = 4$. Assume that either $\operatorname{ind}(Q_1 + Q_2) = 2$ or $\operatorname{ind}(Q_1 + Q_2 + Q_3) = 2$. Choose $a \in F^*$ such that $Q_{3F(\sqrt{a})} = 0$. Then $Q_{1F(\sqrt{a})}$ and $Q_{2F(\sqrt{a})}$ have a common slot, hence for the algebras Q_1, Q_2, Q_3 there exists a common splitting field extension K/F of degree 4. By Peyre's theorem we conclude that the torsion of $\mathrm{CH}^2(X_1 \times X_2 \times X_3)$ is zero. In the second case it is obvious that $d = 8$, and there is a triquadratic common splitting field extension for the Q_i. Hence in this case the torsion of $\mathrm{CH}^2(X_1 \times X_2 \times X_3)$ is zero as well. On the other hand, for the algebras Q_{1L}, Q_{2L}, Q_{3L} we have $d = 2$ in both cases, but there is no common splitting extension E/L of degree $2m$ where m is odd. Indeed, existence of such an extension implies $m = 1$, since L has no proper odd degree extension. This means that the triple $\mathfrak{T}_L$ has a common slot, which is not the case. Applying Peyre's theorem again, we get that the torsion of $\mathrm{CH}^2(X_{1L} \times X_{2L} \times X_{3L})$ is $\mathbb{Z}/2\mathbb{Z}$. □

In the examples above the invariant $c(\mathfrak{T})$ was a Pfister form, either 3-fold or 4-fold. In general this is not the case, as shown by a generic example of a linked triple.

Proposition 8. *Let k be a field, a, b_1, b_2, c, d indeterminates. Further, let φ_1, φ_2, φ_3 be the Albert forms corresponding to the biquaternion algebras*

$$(a, b_1) + (c, d), \quad (a, b_2) + (c, d), \quad (a, b_1 b_2) + (c, d),$$

respectively. Put $F = k(\varphi_1, \varphi_2, \varphi_3)$. Then:

1. *The triple $\mathfrak{T} = \{(a, b_1)_F, (a, b_2)_F, (c, d)_F\}$ is linked.*
2. $\dim c(\mathfrak{T}) = 16$, *and* $c(\mathfrak{T}) \notin I^4(F)$.
3. *The form $c(\mathfrak{T}_{F(c(\mathfrak{T}))})$ is similar, but not equal, to a* 3*-fold Pfister form over $F(c(\mathfrak{T}))$.*

Proof. (1) This is obvious by the definition of linked triple.

(2) We have $(c, d)_{k(\sqrt{ac})} = (a, d)_{k(\sqrt{ac})}$. Put

$$\mathfrak{P} = \left\{(a, b_1)_{k(\sqrt{ac})}, (a, b_2)_{k(\sqrt{ac})}, (c, d)_{k(\sqrt{ac})}\right\}.$$

By Proposition 1 we get $c(\mathfrak{P}) = \langle\langle a, b_1, b_2, d\rangle\rangle_{k(\sqrt{ac})} \neq 0$. Notice that $\mathfrak{T}_{F(\sqrt{ac})} = \mathfrak{P}_{k(\sqrt{ac})(\varphi_1, \varphi_2, \varphi_3)}$, and the forms $\varphi_1, \varphi_2, \varphi_3$ are isotropic over $k(\sqrt{ac})$. Hence the field $k(\sqrt{ac})(\varphi_1, \varphi_2, \varphi_3)$ is purely transcendental over $k(\sqrt{ac})$, which implies that

$$\dim c(\mathfrak{T}_{F(\sqrt{ac})}) = \dim c\big(\mathfrak{P}_{k(\sqrt{ac})(\varphi_1,\varphi_2,\varphi_3)}\big) = 16.$$

By Proposition 1 we get $\dim c(\mathfrak{T}) = 16$. Assume now that $c(\mathfrak{T}) \in I^4(F)$. To come to a contradiction it suffices to construct a field extension L/F such that $\dim c(\mathfrak{T}_L) = 8$. Let

$$\tau_1 \simeq \langle b_1, -c, -d, cd\rangle, \quad \tau_2 \simeq \langle b_2, -c, -d, cd\rangle, \quad \tau_3 \simeq \langle b_1 b_2, -c, -d, cd\rangle.$$

Obviously, the algebra $Q = (c, d)_{k(\tau_1,\tau_2,\tau_3)}$ satisfies the hypothesis of Proposition 2 with respect to the elements $b_1, b_2, b_1 b_2$. Moreover, $Q \neq 0$ [Scharlau 1985] and a is an indeterminate over $k(\tau_1, \tau_2, \tau_3)$. Therefore,

$$c\big(\{(a, b_1), (a, b_2), (c, d)_{k(\tau_1,\tau_2,\tau_3)}\}\big) = \langle\langle a\rangle\rangle Q \neq 0.$$

On the other hand, since the Albert forms $\varphi_{i\,k(\tau_i)}$ are isotropic, the field extension $k(\tau_1, \tau_2, \tau_3, \varphi_1, \varphi_2, \varphi_3)/k(\tau_1, \tau_2, \tau_3)$ is purely transcendental, so

$$\dim c(\mathfrak{T}_L) = 8, \quad \text{where } L = k(\tau_1, \tau_2, \tau_3, \varphi_1, \varphi_2, \varphi_3) = F(\tau_1, \tau_2, \tau_3).$$

(3) Here we consider $c(\mathfrak{T})$ as the corresponding anisotropic form. In view of part (2) we have $\dim c(\mathfrak{T}_{F(c(\mathfrak{T}))}) = 8$, hence the form $\mathfrak{T}_{F(c(\mathfrak{T}))}$ is similar to a 3-fold Pfister form over $F(c(\mathfrak{T}))$. By Proposition 1 we have $1 \in D(c(\mathfrak{T}))$, i.e., $c(\mathfrak{T}) = \langle 1\rangle \perp \psi$ for some 15-dimensional form δ over F. Suppose $1 \in D(c(\mathfrak{T}_{F(c(\mathfrak{T}))}))$. Then $\dim \delta_{F(c(\mathfrak{T}))} = \dim \delta_{F(\delta)} = 7$, i.e., $i_1(\delta) = 4$. Since the first Witt index of an odd dimensional form is odd [Karpenko 2003], we get a contradiction. □

It turns out that existence of a common slot for the triple $\{Q_1, Q_2, Q_3\}$ over the field F is equivalent to isotropicity of a certain 9-dimensional form over $F(t)$.

Proposition 9. *The following conditions are equivalent*:

(1) *The triple* $\{(a_1, b_1), (a_2, b_2), (a_3, b_3)\}$ *has a common slot.*

(2) *The system of quadratic forms*

$$\begin{cases} \varphi_1 = a_1x_1^2 + b_1x_2^2 - a_1b_1x_3^2 - a_2x_4^2 - b_2x_5^2 + a_2b_2x_6^2 = 0 \\ \varphi_2 = -a_2x_4^2 - b_2x_5^2 + a_2b_2x_6^2 + a_3x_7^2 + b_3x_8^2 - a_3b_3x_9^2 = 0 \end{cases}$$

has a nontrivial zero.

(3) *The form*

$$\Phi \simeq \varphi_1 + t\varphi_2 \simeq \langle a_1, b_1, -a_1b_1\rangle \perp (t+1)\langle -a_2, -b_2, a_2b_2\rangle \perp t\langle a_3, b_3, -a_3b_3\rangle$$

is isotropic over $F(t)$.

Proof. By [Brumer 1978] the conditions (2) and (3) are equivalent. If at least one of the algebras (a_i, b_i) is trivial, then it is easy to see that both conditions (1) and (2) hold. Assume that each (a_i, b_i) is nontrivial. Then (2) is equivalent to existence of a nonzero row $(x_1, \dots, x_9)$ such that

$$0 \neq a_1x_1^2 + b_1x_2^2 - a_1b_1x_3^2 = a_2x_4^2 + b_2x_5^2 - a_2b_2x_6^2 = a_3x_7^2 + b_3x_8^2 - a_3b_3x_9^2,$$

which in turn is equivalent to (1). □

Proposition 10. *Let F be a field, $\{Q_1, Q_2, Q_3\}$ a triple over F without a common slot, and L/F a field extension. Then the triple $\{Q_1, Q_2, Q_3\}_L$ has no common slot as well, if L/F is either purely transcendental, or of odd degree, or $L = F(\tau)$, where τ is a form over F,* $\dim \tau \geq 9$.

Proof. We keep the notation of Proposition 9. If L/F is purely transcendental, then the claim is obvious. If L/F is an odd degree extension, then the assertion follows from Proposition 9 and Springer's theorem [Scharlau 1985]. Assume now that τ is a form over F, $\dim \tau \geq 9$, $L = F(\tau)$, and that $\{Q_1, Q_2, Q_3\}_L$ has a common slot. We may assume that τ is anisotropic, for otherwise the extension $F(\tau)/F$ is purely transcendental. By Proposition 9 the form $\Phi_{F(t)(\tau)}$ is isotropic. Since $\dim \Phi = \dim \tau = 9$, it follows that the form $\tau_{F(t)(\Phi)}$ is isotropic as well [Izhboldin 2000]. On the other hand, the form $\Phi_{F(t)(\sqrt{-a_1a_3t})}$ is isotropic, hence the form $\tau_{F(t)(\sqrt{-a_1a_3t})}$ is isotropic, which is impossible, since the extension $F(t)(\sqrt{-a_1a_3t})/F$ is purely transcendental. □

Corollary 11. *Let K/F be a field extension of degree $2m$, where m is odd. Suppose that Q_1, Q_2, Q_3 are quaternion algebras such that $Q_{1K} = Q_{2K} = Q_{3K} = 0$. Then the triple $\{Q_1, Q_2, Q_3\}$ has a common slot.*

Proof. We will induct on m, the case $m = 1$ being trivial. Suppose first that $K = F(\alpha)$, and denote by $p(t)$ the irreducible monic polynomial for α. As earlier, let X_i be the projective conic corresponding to Q_i. By the hypothesis there is a morphism $\operatorname{Spec} K \to X_1 \times X_2 \times X_3$. This map gives rise to the morphism

$$\operatorname{Spec} L[t]/p(t) = \operatorname{Spec} L \otimes_F K \to X_{1L} \times X_{2L} \times X_{3L},$$

where L is a maximal odd-degree field extension of F. Since $\deg p$ is not divisible by 4, and any finite extension of L is 2-primary, there is an irreducible $f \in L[t]$ such that $f | p$ and $\deg f \leq 2$. Therefore, we get a morphism $\operatorname{Spec} L[t]/f(t) \to X_{1L} \times X_{2L} \times X_{3L}$. If $\deg f = 1$, then $Q_{1L} = Q_{2L} = Q_{3L} = 0$, hence, since L/F is an odd degree extension, $Q_1 = Q_2 = Q_3 = 0$. If $\deg f = 2$, then the triple $\{Q_1, Q_2, Q_3\}_L$ has a common slot, hence by Proposition 10, the triple $\{Q_1, Q_2, Q_3\}$ has a common slot as well, and the corollary is proved in the case $K = F(\alpha)$.

In the general case we have a tower of field extensions $F \subset F(\alpha) \subset K$, $F(\alpha) \neq F$, and $[F(\alpha) : F]$ is not divisible by 4. If $[F(\alpha) : F] = 2n$, where n is odd, then

$[K : F(\alpha)]$ is odd, hence $Q_{1F(\alpha)} = Q_{2F(\alpha)} = Q_{3F(\alpha)} = 0$, and by induction $\{Q_1, Q_2, Q_3\}$ has a common slot. If $[F(\alpha) : F]$ is odd, then again by induction with respect to the extension $K/F(\alpha)$ the triple $\{Q_1, Q_2, Q_3\}_{F(\alpha)}$ has a common slot. By Proposition 10, $\{Q_1, Q_2, Q_3\}$ has a common slot as well. □

Corollary 12. *The linked triple* $\{Q_1, Q_2, Q_3\}$ *has a common slot if and only if the torsion of the group* $\mathrm{CH}^2(X_1 \times X_2 \times X_3)$ *is zero.*

Proof. By Peyre's theorem, the group $\mathrm{CH}^2(X_1 \times X_2 \times X_3)$ is zero if and only if there is an extension K/F of degree $2m$, with m odd, such that $Q_{1K} = Q_{2K} = Q_{3K} = 0$. By Corollary 11 this is equivalent to the triple $\{Q_1, Q_2, Q_3\}$ having a common slot. □

Peyre's theorem implies the following curious result on *four* quaternion algebras.

Proposition 13. *Let k be a field, $a_1, a_2, a_3, a_4 \in k^*$, $D \in {}_2\mathrm{Br}(k)$, $F = k(x)$ the rational function field, X_i the projective conics corresponding to the algebras (a_i, x) over F. Then the following conditions are equivalent*:

(1) *There exist $b_1, b_2, b_3, b_4 \in k^*$ such that $D = \sum_{i=1}^4 (a_i, b_i)$.*

(2) *$D \cup (x)_{F(X)} = 0$ in $H^3(F(X), \mathbb{Z}/2\mathbb{Z})$, where $X = X_1 \times X_2 \times X_3 \times X_4$.*

Proof. The implication (1) $\Longrightarrow$ (2) is trivial, since if $D = \sum_{i=1}^4 (a_i, b_i)$, then

$$D \cup (x) = \sum_{i=1}^{4} (a_i, b_i, x),$$

hence $D \cup (x)_{F(X)} = 0$. Suppose now that $D \cup (x)_{F(X)} = 0$. We have $F(X_4) = k(u, v)$, where $u^2 - av^2 = x$. Therefore, $D \cup (u^2 - av^2)_{K(Y_1 \times Y_2 \times Y_3)} = 0$, where $K = k(u, v)$, and Y_i is the conic corresponding to the algebra $(a_i, u^2 - av^2)$. Put $t = u/v$, $a = a_4$. By applying Peyre's theorem to the algebras $(a_i, t^2 - a)$ $(i = 1, 2, 3)$, we get

$$D \cup (t^2 - a) = (a_1, t^2 - a, p_1(t)) + (a_2, t^2 - a, p_2(t)) + (a_3, t^2 - a, p_3(t)) \in H^3(k(t), \mathbb{Z}/2\mathbb{Z})$$

for some $p_i(t) \in k(t)$. Obviously, we can rewrite the last equality as

$$D \cup (t^2 - a) = \sum_{I \subset \{1,2,3\}} (a_I, t^2 - a, f_I),$$

where $a_I = \prod_{i \in I} a_i$, and $f_I \in k[t]$ are some pairwise coprime polynomials. Moreover, we may suppose that each f_I is coprime with $t^2 - a$. Now apply the well-known exact sequence

$$0 \to H^3(k, \mathbb{Z}/2\mathbb{Z}) \xrightarrow{\text{res}} H^3(k(t), \mathbb{Z}/2\mathbb{Z}) \xrightarrow{\coprod \partial_p} \coprod_{p \in \mathbb{P}^1_k} H^2(k_p, \mathbb{Z}/2\mathbb{Z}) \xrightarrow{N} H^2(k, \mathbb{Z}/2\mathbb{Z}) \to 0$$

to the symbols (a_I, t^2-a, f_I). Since the polynomials f_I are pairwise coprime and $\partial_p(D\cup(t^2-a))=0$ for any prime polynomial $p\neq t^2-a$, we have

$$\partial_p(a_I, t^2-a, f_I)=0$$

for such p. Also $\partial_\infty(a_I, t^2-a, f_I)=0$. By the exact sequence above we have $N_{k(\sqrt{a})/k}(\partial_{t^2-a}(a_I, t^2-a, f_I))=0$. Hence $\partial_{t^2-a}(a_I, t^2-a, f_I))=(a_I, b_I)$ for some $b_I\in k^*$. Applying the above exact sequence again, we get

$$(a_I, t^2-a, f_I)=(a_I, b_I, t^2-a).$$

Therefore,

$$\begin{aligned} D_{k(\sqrt{a})} &= \partial_{t^2-a}(D\cup(t^2-a)) \\ &= \partial_{t^2-a}\Bigl(\sum_{I\subset\{1,2,3\}}(a_I, t^2-a, f_I)\Bigr)=\sum_{I\subset\{1,2,3\}}(a_I, b_I)_{k(\sqrt{a})}. \end{aligned}$$

It follows that $D=(a,b)+\sum_{I\subset\{1,2,3\}}(a_I,b_I)$ for some $b\in k^*$, which implies (1). □

Corollary 14. *Assume F is a C_2-field, which means that for each $d\geq 1$ any homogeneous polynomial in d^2+1 variables of degree d over F has a nontrivial zero* [Scharlau 1985]. *Then any triple of quaternion algebras $\{Q_1, Q_2, Q_3\}$ over F has a common slot.*

Proof. Any two 9-dimensional quadratic forms over F have a common nontrivial zero [Scharlau 1985], hence we are done by Proposition 9. □

In the case of global fields we can say more.

Proposition 15. *Let F be a global field. Any finite collection $Q_1,\dots,Q_n$ of quaternion algebras over F has a common slot.*

Proof. It is well known from the class field theory that the natural restriction map $\partial:\mathrm{Br}(F)\to\coprod_v\mathrm{Br}(F_v)$ is injective, where v runs over all valuations of F, and F_v is the completion of F with respect to v. For any finite extension L/F, the diagram

$$\begin{array}{ccc} \mathrm{Br}(F) & \xrightarrow{\partial=\coprod\partial_v} & \coprod_v \mathrm{Br}\,F_v \\ {\scriptstyle\mathrm{res}}\downarrow & & \downarrow{\scriptstyle\mathrm{res}} \\ \mathrm{Br}(L) & \xrightarrow{\partial=\coprod\partial_w} & \coprod_w \mathrm{Br}\,L_w \end{array}$$

is commutative. Let $\{v_1,\dots,v_m\}$ be all valuations over F such that $\partial_{v_i}(Q_j)\neq 0$ for some j. Choose $d\in F^*$ such that $d\notin {F_{v_i}}^{*2}$ and $v_i(d)=0$ for each v_i. Then

$$\partial\circ\mathrm{res}_{F(\sqrt{d})/F}(Q_j)=\mathrm{res}_{F(\sqrt{d})/F}\circ\partial(Q_j)=0,$$

hence $\mathrm{res}_{F(\sqrt{d})/F}(Q_j)=0$, so $Q_j=(d,e_j)$ for some $e_j\in F^*$. □

The above results prompt the following:

Open questions.

(1) Does there exist a linked triple $\mathfrak{T}$ over a field F without common slot such that $c(\mathfrak{T}) \in I^4(F)$? (In view of Proposition 9, an equivalent version of this question is the one where $c(\mathfrak{T}) \in I^4(F)$ is changed for $c(\mathfrak{T}) = 0$.)

(2) Assume that φ is an anisotropic form, $\dim \varphi \geq 5$, $\mathfrak{T}$ is a linked triple without common slot. Is it true that $\mathfrak{T}_{F(\varphi)}$ is without common slot as well? (Notice that if $\dim \varphi = 4$, then the answer is negative in general. For instance, under the notation of Corollary 3 the triple $\mathfrak{T}_{F(\langle a,b,ab,c\rangle)}$ obviously has a common slot, namely c).

(3) Let F be a field such that any form of dimension at least 5 over F is isotropic. Does any triple of quaternion algebras $\{Q_1, Q_2, Q_3\}$ over F have a common slot? (By Corollary 14 and Proposition 15 the answer is positive if F is either a C_2-field or a nonreal global field.)

(4) Suppose $\mathrm{cd}_2(F) = 2$. Is it true that any linked triple has a common slot?

Certainly, these questions are not independent of one another. For instance, if the answer to question (2) is positive, then the answer to question (1) is positive as well, and the answers to questions (3) and (4) are negative.

Acknowledgement

The author express his gratitude to Alexander Merkurjev for reading this paper.

References

[Brumer 1978] A. Brumer, “Remarques sur les couples de formes quadratiques”, *C. R. Acad. Sci. Paris Sér. A-B* **286**:16 (1978), A679–A681. In French. MR 58 #16502 Zbl 0392.10021

[Elman et al. 2008] R. Elman, N. Karpenko, and A. Merkurjev, *The algebraic and geometric theory of quadratic forms*, American Mathematical Society Colloquium Publications **56**, American Mathematical Society, Providence, RI, 2008. MR 2009d:11062 Zbl 1165.11042

[Hoffmann 1995] D. W. Hoffmann, “Isotropy of quadratic forms over the function field of a quadric”, *Math. Z.* **220**:3 (1995), 461–476. MR 96k:11041 Zbl 0840.11017

[Izhboldin 2000] O. T. Izhboldin, “Motivic equivalence of quadratic forms, II”, *Manuscripta Math.* **102**:1 (2000), 41–52. MR 2001g:11051 Zbl 0972.11019

[Karpenko 2003] N. A. Karpenko, “On the first Witt index of quadratic forms”, *Invent. Math.* **153**:2 (2003), 455–462. MR 2004f:11030 Zbl 1032.11016

[Karpenko and Merkurjev 2003] N. Karpenko and A. Merkurjev, “Essential dimension of quadrics”, *Invent. Math.* **153**:2 (2003), 361–372. MR 2004f:11029 Zbl 1032.11015

[Lam 2005] T. Y. Lam, *Introduction to quadratic forms over fields*, Graduate Studies in Mathematics **67**, American Mathematical Society, Providence, RI, 2005. MR 2005h:11075 Zbl 1068.11023

[Merkurjev 1991] A. S. Merkurjev, “Simple algebras and quadratic forms”, *Izv. Akad. Nauk SSSR Ser. Mat.* **55**:1 (1991), 218–224. In Russian; translated in *Math. USSR-Izv.* **38**:1 (1992), 215–221. MR 93b:16025 Zbl 0733.12008

[Peyre 1995] E. Peyre, “Products of Severi–Brauer varieties and Galois cohomology”, pp. 369–401 in *K-theory and algebraic geometry: Connections with quadratic forms and division algebras* (Santa Barbara, CA, 1992), edited by B. Jacob and A. Rosenberg, Proc. Sympos. Pure Math. **58**, Amer. Math. Soc., Providence, RI, 1995. MR 96d:19008 Zbl 0837.14006

[Scharlau 1985] W. Scharlau, *Quadratic and Hermitian forms*, Grundlehren der Mathematischen Wissenschaften **270**, Springer, Berlin, 1985. MR 86k:11022 Zbl 0584.10010

Received July 11, 2012. Revised November 6, 2012.

ALEXANDER S. SIVATSKI
DEPARTAMENTO DE MATEMÁTICA
UNIVERSIDADE FEDERAL DO CEARÁ
AV. HUMBERTO MONTE, CAMPUS DO PICI, BLOCO 914
FORTALEZA, CE 60455-900
BRAZIL

alexander.sivatski@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 268, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.268.477

FINITE NONSOLVABLE GROUPS WITH MANY DISTINCT CHARACTER DEGREES

HUNG P. TONG-VIET

Let G be a finite group and let $\mathrm{Irr}(G)$ denote the set of all complex irreducible characters of G. Let $\mathrm{cd}(G)$ be the set of all character degrees of G. For a degree $d \in \mathrm{cd}(G)$, the multiplicity of d in G, denoted by $m_G(d)$, is the number of irreducible characters of G having degree d. A finite group G is said to be a T_k-group for some integer $k \geq 1$ if there exists a nontrivial degree $d_0 \in \mathrm{cd}(G)$ such that $m_G(d_0) = k$ and that for every $d \in \mathrm{cd}(G) - \{1, d_0\}$, the multiplicity of d in G is trivial, that is, $m_G(d) = 1$. In this paper, we show that if G is a nonsolvable T_k-group for some integer $k \geq 1$, then $k = 2$ and $G \cong \mathrm{PSL}_2(5)$ or $\mathrm{PSL}_2(7)$.

1. Introduction

Let G be a finite group and let $\mathrm{Irr}(G) = \{\chi_1, \chi_2, \cdots, \chi_s\}$ be the set of all complex irreducible characters of G. Let $\mathrm{cd}(G) = \{d_0, d_1, \cdots, d_t\}$, $1 = d_0 < d_1 < \cdots < d_t$, be the set of all character degrees of G. For an integer $d \geq 1$, the multiplicity of d in G, denoted by $m_G(d)$, is the number of irreducible characters of G having degree d; i.e., $m_G(d) = |\{\chi \in \mathrm{Irr}(G) \mid \chi(1) = d\}|$. Let $n_i = \chi_i(1)$ for $1 \leq i \leq s$. We call $\mathrm{mp}(G) = (m_G(d_0), m_G(d_1), \cdots, m_G(d_t))$ the *multiplicity pattern* and $(n_1, n_2, \ldots, n_s)$ the *degree pattern* of G. Let $\mathbb{C}G$ be the complex group algebra of G. We know that $\mathbb{C}G = \bigoplus_{i=1}^{s} \mathrm{M}_{n_i}(\mathbb{C})$ and thus knowing the degree pattern of G is equivalent to knowing the structure of the complex group algebra of G, or, equivalently, the first column of the ordinary character table of G. One of the main questions in character theory of finite groups is Problem 1 in [Brauer 1963], asking for the possible degree patterns of finite groups. It was proved in [Moretó 2007; Craven 2008] that the order of a finite group is bounded in terms of the largest multiplicity of its character degree. This gives a new restriction on the degree patterns of finite groups.

Motivated by this result, we want to explore the relations between the multiplicities of character degrees of finite groups and the structure of the groups.

This research is supported by a Startup Research Fund from the College of Agriculture, Engineering and Science, the University of KwaZulu-Natal.
MSC2010: primary 20C15; secondary 20C33, 20D05.
Keywords: multiplicity, character degrees, nonsolvable groups.

This problem has been the subject of much literature; we mention in particular [Berkovich 1996; Berkovich and Kazarin 1996; Berkovich et al. 1992; Dolfi et al. 2013; Seitz 1968]. In this last reference, G. Seitz classified all finite groups which have exactly one nonlinear irreducible representation. This result was generalized in [Berkovich et al. 1992], where the authors classified all finite groups G in which the multiplicity of every nonlinear irreducible character degree G is trivial. Finite groups in which only two nonlinear irreducible characters have equal degrees have been classified in [Berkovich 1996; Berkovich and Kazarin 1996].

To generalize these results, we consider the following definition. A finite group G is called a *T_k-group* for some integer $k \geq 1$ if there exists a nontrivial degree $d_0 \in \mathrm{cd}(G)$ such that $m_G(d_0) = k$ and that for every nontrivial degree $d \in \mathrm{cd}(G)$ different from d_0, we have $m_G(d) = 1$. Obviously, the finite groups studied in [Berkovich et al. 1992; Seitz 1968] and [Berkovich 1996; Berkovich and Kazarin 1996] are exactly T_1-groups and T_2-groups, respectively. In this paper, we generalize the results in [Berkovich and Kazarin 1996] as follows:

Theorem A. *Let G be a finite nonsolvable group. If G is a T_k-group for some integer $k \geq 1$, then $G \cong \mathrm{PSL}_2(5)$ or $\mathrm{PSL}_2(7)$ and $k = 2$.*

From [Conway et al. 1985] we know that the multiplicity patterns of $\mathrm{PSL}_2(q)$ for $q \in \{5, 7\}$ are $(1, 2, 1, 1)$ and $(1, 2, 1, 1, 1)$, respectively. Suppose that G is a finite group such that $\mathrm{mp}(G) = \mathrm{mp}(\mathrm{PSL}_2(q))$ with $q \in \{5, 7\}$. Since the first entry of $\mathrm{mp}(G)$ is $|G : G'| = 1$, we see that G is perfect and hence a nonsolvable T_2-group. Applying Theorem A, we deduce that $G \cong \mathrm{PSL}_2(5)$ or $\mathrm{PSL}_2(7)$. By comparing the number of distinct character degrees, we deduce that $G \cong \mathrm{PSL}_2(q)$. It follows that $\mathrm{PSL}_2(5)$ and $\mathrm{PSL}_2(7)$ are uniquely determined by the multiplicity patterns.

In [Tong-Viet 2013] we conjectured that every nonabelian simple group is uniquely determined by its multiplicity pattern, and showed that this conjecture holds for every nonabelian simple group with at most 7 distinct character degrees. If true, this conjecture generalizes a result in [Tong-Viet 2012] saying that all nonabelian simple groups are uniquely determined by the structure of their complex group algebras. This latter result is related to Problem 2 in [Brauer 1963], which asks: What does $\mathbb{C}G$ know about G? This is also an important question in character theory and has been studied extensively (see the references in [Tong-Viet 2012]).

Notice that if the degree pattern of a finite group G is given, then both $\mathrm{cd}(G)$ and $\mathrm{mp}(G)$ are known. Thus, apart from being a direct generalization of the results obtained in [Berkovich and Kazarin 1996], Theorem A could be used to study questions raised in [Tong-Viet 2013]. For finite solvable groups, if G is a finite T_k-group of odd order, then $|\mathrm{cd}(G)| \leq 2$ since G has only one real irreducible character (the trivial character) and thus every nontrivial character degree of G has multiplicity at least 2. On the other hand, every finite group with exactly two distinct character degrees is a solvable T_k-group for some integer k (see [Isaacs 1976, Corollary 12.6]) and a compete classification of such finite groups is yet to be found. This together with the fact that there is no explicit upper bound for k makes the classification of solvable T_k-groups quite complicated even for 2-groups.

Throughout this paper, all groups are finite and all characters are complex characters. Let G be a group. If $N \trianglelefteq G$ and $\theta \in \mathrm{Irr}(N)$, the inertia group of θ in G is denoted by $I_G(\theta)$. We write $\mathrm{Irr}(G|\theta)$ for the set of all irreducible constituents of θ^G. The order of an element $x \in G$ is denoted by $|x|$. Denote by $\Phi_k := \Phi_k(q)$ the value of the k-th cyclotomic polynomial evaluated at q. Other notation is standard.

2. Preliminaries

Lemma 2.1. *Let G be a group and let $N \trianglelefteq G$ be such that G/N is cyclic of order $d \geq 2$. Assume that G has a nontrivial degree a with multiplicity m. Suppose that $a > d$ and $m/d \geq 2$. Then N has a nontrivial degree b with multiplicity at least* 2 *and $a/d \leq b \leq a$.*

Proof. Assume first that χ_N is not irreducible for some $\chi \in \mathrm{Irr}(G)$ with $\chi(1) = a$. Let $\theta \in \mathrm{Irr}(N)$ be an irreducible constituent of χ_N. As G/N is cyclic, we deduce from [Isaacs 1976, Corollary 11.22] that θ is not G-invariant. Let $I = I_G(\theta)$ and $t := |G : I|$. Then $t \geq 2$. By Clifford's theorem [Isaacs 1976, Theorem 6.2] and the corollary just cited, we deduce that $\chi_N = \sum_{i=1}^{t} \theta_i$, where $\theta_i \in \mathrm{Irr}(N)$ are distinct conjugates of θ. Hence N has a nontrivial degree a/t with multiplicity at least $t \geq 2$. Since t divides $|G/N| = d$, we deduce that $a/d \leq a/t \leq a$. Assume now that $\chi_N \in \mathrm{Irr}(N)$ for every $\chi \in \mathrm{Irr}(G)$ with $\chi(1) = a$. It follows that N has a nontrivial degree a with multiplicity at least $m/d \geq 2$ as each irreducible character $\chi_N \in \mathrm{Irr}(N)$ has exactly d extensions in G. Therefore, in both cases N has a nontrivial degree b with $1 < a/d \leq b \leq a$ with multiplicity at least 2. The proof is now complete. □

We note that when d in the previous lemma is a prime, then $b \in \{a, a/d\}$. As an application of this lemma, we obtain:

Corollary 2.2. *Let G be a group and let $N \trianglelefteq G$ be such that G/N is cyclic of order $d \geq 2$. Assume that G has two nontrivial degrees a_1, a_2, with multiplicities m_1, m_2. Suppose that $a_2/d > a_1 > d$ and $m_i \geq 2d$ for $i = 1, 2$. Then N is not a T_k-group for any integer $k \geq 1$.*

Proof. By Lemma 2.1, N has two nontrivial character degrees d_1, d_2, each with multiplicity at least 2, such that $a_i/d \leq d_i \leq a_i$, for $i = 1, 2$. Now we have $d_2 \geq a_2/d > a_1 \geq d_1 \geq a_1/d > 1$ by the hypothesis. Hence d_1 and d_2 are character degrees of N and both degrees have nontrivial multiplicity, so N is not a T_k-group for any integer $k \geq 1$. □

The next result is well known. See [Carter 1985, §§ 13.8, 13.9] for the notion of symbols and the classification of unipotent characters of finite groups of Lie type.

Lemma 2.3. *Let S be a nonabelian simple group.*

(1) *If S is a sporadic simple group, the Tits group or an alternating group of degree at least* 7, *then S has two nontrivial irreducible characters, with distinct degrees and both extendible to* $\mathrm{Aut}(S)$.

(2) *If S is a simple group of Lie type in characteristic p and $S \neq {}^2\mathrm{F}_4(2)'$, then the Steinberg character of S, denoted by St_S, of degree $|S|_p$, is extendible to* $\mathrm{Aut}(S)$. *Furthermore, if $S \ncong \mathrm{PSL}_2(3^f)$, then S possesses an irreducible character θ such that $\theta(1) \neq |S|_p$ and θ also extends to* $\mathrm{Aut}(S)$.

Proof. The first statement follows from [Bianchi et al. 2007, Theorems 3 and 4]. For (2), the existence and extendability of the Steinberg character of S is well known. Now assume that $S \ncong \mathrm{PSL}_2(q)$ with $q = p^f$. We can choose θ to be any unipotent character of S which is not one of the exceptions in [Malle 2008, Theorem 2.5] and not the Steinberg character of S, then θ is extendible to $\mathrm{Aut}(S)$ [Carter 1985, §§ 13.8, 13.9]. Finally, assume that $S \cong \mathrm{PSL}_2(q)$ with $q = p^f$ and $p \neq 3$. Then S has an irreducible character θ of degree $q+\delta$, where $q \equiv \delta \pmod 3$ and $\delta \in \{\pm 1\}$ such that θ extends to $\mathrm{Aut}(S)$. Notice that this irreducible character of S corresponds to a semisimple element of order 3 in the dual group $\mathrm{SL}_2(q)$. □

Lemma 2.4 [Zsigmondy 1892]. *Let $q \geq 2$ and $n \geq 3$ be integers such that $(n, q) \neq (6, 2)$. Then $q^n - 1$ has a prime factor ℓ such that $\ell \equiv 1 \pmod n$ and ℓ does not divide $q^m - 1$ for any $m < n$.*

Such an ℓ is called a *primitive prime divisor* and is denoted by $\ell_n(q)$.

The orders of two maximal tori and the corresponding primitive prime divisors of the finite classical groups are given in Table 1, taken from [Malle 1999, Table 3.5]. Table 2 lists the degrees of some unipotent characters of the simple exceptional groups of Lie type. This can be found in [Carter 1985, §13.9].

3. Simple T_k-groups

The main purpose of this section is to classify all simple T_k-groups. As we will see shortly, there are only two simple T_k-groups and they are exactly the simple T_2-groups. Let $\mathscr{L}$ be the set consisting of the following simple groups:

$$\mathrm{PSL}_2(q),\ \mathrm{PSL}_3(q),\ \mathrm{PSU}_3(q),\ \mathrm{PSp}_4(q),$$
$$\mathrm{PSL}_6(2),\ \mathrm{PSL}_7(2),\ \mathrm{PSU}_4(2),\ \mathrm{PSp}_6(2),\ \mathrm{PSp}_8(2),\ \mathrm{P}\Omega_8^{\pm}(2)$$

and

$$\mathrm{PSL}_4(2),\ \mathrm{PSU}_4(3),\ \mathrm{PSU}_5(2),\ \mathrm{PSp}_6(3),$$
$$\Omega_7(3),\ \mathrm{PSp}_8(3),\ \Omega_9(3),\ \mathrm{P}\Omega_8^{\pm}(3),\ \mathrm{P}\Omega_{10}^{+}(2),\ \mathrm{P}\Omega_{10}^{-}(3).$$

$G = G(q)$	$\lvert T_1\rvert$	$\lvert T_2\rvert$	ℓ_1	ℓ_2
A_n	$(q^{n+1}-1)/(q-1)$	q^n-1	$\ell_{n+1}(q)$	$\ell_n(q)$
${}^2\mathrm{A}_n,\ (n\equiv 0(4))$	$(q^{n+1}+1)/(q+1)$	q^n-1	$\ell_{2n+2}(q)$	$\ell_n(q)$
${}^2\mathrm{A}_n,\ (n\equiv 1(4))$	$(q^{n+1}-1)/(q+1)$	q^n+1	$\ell_{(n+1)/2}(q)$	$\ell_{2n}(q)$
${}^2\mathrm{A}_n,\ (n\equiv 2(4))$	$(q^{n+1}+1)/(q+1)$	q^n-1	$\ell_{2n+2}(q)$	$\ell_{n/2}(q)$
${}^2\mathrm{A}_n,\ (n\equiv 3(4))$	$(q^{n+1}-1)/(q+1)$	q^n+1	$\ell_{n+1}(q)$	$\ell_{2n}(q)$
$\mathrm{B}_n,\ \mathrm{C}_n\ (n\ge 3\text{ odd})$	q^n+1	q^n-1	$\ell_{2n}(q)$	$\ell_n(q)$
$\mathrm{B}_n,\ \mathrm{C}_n\ (n\ge 2\text{ even})$	q^n+1	$(q^{n-1}+1)(q+1)$	$\ell_{2n}(q)$	$\ell_{2n-2}(q)$
$\mathrm{D}_n,\ (n\ge 5\text{ odd})$	$(q^{n-1}+1)(q+1)$	q^n-1	$\ell_{2n-2}(q)$	$\ell_n(q)$
$\mathrm{D}_n,\ (n\ge 4\text{ even})$	$(q^{n-1}+1)(q+1)$	$(q^{n-1}-1)(q-1)$	$\ell_{2n-2}(q)$	$\ell_{n-1}(q)$
${}^2\mathrm{D}_n$	q^n+1	$(q^{n-1}+1)(q-1)$	$\ell_{2n}(q)$	$\ell_{2n-2}(q)$

Table 1. Two tori for classical groups.

The following result will be needed when dealing with simple classical groups of Lie type. We refer to [Dolfi et al. 2013, §4.3] and [Larsen et al. 2013, Theorem 4.7] for some related results.

Lemma 3.1. *Let $\mathcal{G}$ be a simply connected simple algebraic group of classical type and let F be a suitable Frobenius map such that $S\cong\mathcal{G}^F/\mathrm{Z}(\mathcal{G}^F)$ is a simple classical group of Lie type defined over a finite field of size q with $S\notin\mathcal{L}$. Let the pair $(\mathcal{G}^*, F^*)$ be dual to $(\mathcal{G}, F)$ and let $G=(\mathcal{G}^*)^{F^*}$. For $i=1,2$, let T_i be the maximal tori of G with order given in Table 1. Then for each i, there exist two regular semisimple elements $s_i, t_i\in T_i$ such that $s_i, t_i\in T_i\cap G'$ and that s_i and t_i are not G-conjugate.*

Proof. Since $\mathcal{G}$ is of simply connected type, the dual group $\mathcal{G}^*$ is of adjoint type and thus by using the identifications with classical groups in [Carter 1985, page 40], G/S is either a cyclic or an elementary abelian group of order 4. In all cases, G/S is abelian and so $G'=S$. For each $i=1,2$, let $T_i'=T_i\cap G'$. Since $G'\trianglelefteq T_iG'\le G$, we obtain that $|T_i'|=|T_i\cap G'|\ge|T_i|/d$ with $d:=|G:G'|$. Since $S\notin\mathcal{L}$, the primitive prime divisors ℓ_1 and ℓ_2 in Table 1 both exist.

Claim 1. *For $i=1,2$, every element $s_i\in T_i$ of order ℓ_i is a regular semisimple element and $s_i\in T_i\cap G'=T_i'$.*

Observe that the two maximal tori of G with order given in Table 1 have the properties that they are uniquely determined up to conjugation by their orders. Furthermore, for each $i=1,2$, the conjugacy class of maximal tori containing T_i is the only class of maximal tori whose order is divisible by ℓ_i. Also, the Sylow ℓ_i-subgroups of G are cyclic. Let $s_i\in T_i$ be a semisimple element of order ℓ_i. As in the proof of [Malle 2010, Proposition 2.4], if $\mathrm{C}_G(s_i)$ is not a torus, then its semisimple rank is at least 1, and thus it contains two maximal tori of different orders. Both of these tori must have orders divisible by ℓ_i, which is impossible. Hence we obtain that $C_G(s_i) = T_i$. Since $\gcd(\ell_i, |G : G'|) = 1$, we deduce that $s_i \in G'$ and so $s_i \in T_i' = T_i \cap G'$.

Claim 2. *For every $x \in T_i$, if $|x|$ is divisible by ℓ_i, then x is a regular semisimple element.*

Assume that $x \in T_i$ such that $|x| = m\ell_i$ where $m \geq 1$ is an integer. Let $s = x^m \in T_i$. Then $|s| = \ell_i$ and so by Claim 1, we know that $C_G(s) = T_i$, hence $T_i \leq C_G(x) \leq C_G(x^m) = C_G(s) = T_i$. So, $C_G(x) = T_i$ and x is a regular semisimple element.

Claim 3. *If $T_i' = T_i \cap G'$ has no element whose order is a proper multiple of ℓ_i, then T_i' contains two distinct G-conjugacy classes of regular semisimple elements of order ℓ_i.*

For $i = 1, 2$, let T_i'' be a cyclic subgroup of T_i' whose order is divisible by ℓ_i. By our assumption, we must have that $|T_i''| = \ell_i$ and so $T_i'' = \langle s_i \rangle$, where s_i is an element of order ℓ_i. As s_i is regular semisimple by Claim 1, we deduce that $N_G(\langle s_i \rangle) \leq N_G(T_i)$ and so as the Sylow ℓ_i-subgroup of T_i is cyclic and $|s_i| = \ell_i$, we obtain that $N_G(\langle s_i \rangle) = N_G(T_i)$. Since $C_G(s_i) = T_i = C_G(T_i)$ and $|N_G(T_i)/T_i| \leq m(S)$, we deduce that $|N_G(\langle s_i \rangle)/C_G(s_i)| \leq m(S)$, where $m(S)$ is the dimension of the natural module for $S = G'$ over $\mathbb{F}_q$. (Notice that the fact $|N_G(T_i)/T_i| \leq m(S)$ can be deduced from [Babai et al. 2009, Lemma 4.7].) It follows that s_i is G-conjugate to at most $m(S)$ of its powers and thus $T_i'' = \langle s_i \rangle$ contains at least $\varphi(|s_i|)/m(S)$ G-conjugacy classes of regular semisimple elements of order ℓ_i, where φ is the Euler φ-function. Since $|s_i| = \ell_i$, we deduce that $\varphi(|s_i|)/m(S) = (|s_i| - 1)/m(S) \geq (|T_i''| - 1)/m(S)$. We now verify that for each possibility of S, we have that $(|T_i''| - 1)/m(S) \geq 2$, which implies that T_i' contains at least two distinct G-conjugacy classes of regular semisimple elements of order ℓ_i.

(a) Assume first that $S \cong \mathrm{PSL}_n(q)$. Then $m(S) = n$ and $d = \gcd(n, q-1)$. Since T_i, $i = 1, 2$, are cyclic, we deduce that both T_i' are also cyclic of order at least $|T_i|/d$. Hence we can choose $T_i'' = T_i'$ for $i = 1, 2$. Then $|T_1''| \geq (q^n - 1)/(d(q-1))$ and $|T_2''| \geq (q^{n-1} - 1)/d$. As $S \notin \mathscr{L}$, it is routine to check that $q^{n-1} - 1 \geq (2n+1)(q-1)$ and so since $q - 1 \geq d = \gcd(n, q-1)$, we obtain that $q^{n-1} - 1 \geq (2n+1)d$ or equivalently $(|T_2''| - 1)/n \geq 2$. Similarly, we can check that $(|T_1''| - 1)/n \geq 2$.

(b) Assume that $S \cong \mathrm{PSU}_n(q)$ and $n \geq 5$ is odd. We have that $m(S) = n$ and $d = \gcd(n, q+1)$. As in the previous case, we see that both T_i are cyclic and so are T_i', hence we can choose $T_i'' = T_i'$. Then $|T_1''| \geq (q^n + 1)/(d(q+1))$ and $|T_2''| \geq (q^{n-1} - 1)/d$. Since $q^{n-1} - 1 > (q^n + 1)/(q+1)$ and $d = \gcd(n, q+1) \leq q+1$, it suffices to show that $(q^n + 1) \geq (2n+1)(q+1)^2$ with $n \geq 5$ odd. Since $S \notin \mathscr{L}$, we can check that the previous inequality holds so that $(|T_i''| - 1)/m(S) \geq 2$ for $i = 1, 2$, as required.

(c) Assume that $S \cong \mathrm{PSU}_n(q)$ and $n \geq 4$ is even. Arguing as in the case n is odd, we have that $|T_1''| \geq (q^n-1)/(d(q+1))$ and $|T_2''| \geq (q^{n-1}+1)/d$. Since $q^{n-1}+1 > (q^n-1)/(q+1)$ and $d \leq q+1$, it suffices to show that $q^n-1 \geq (2n+1)(q+1)^2$ where $n \geq 4$ is even. As $S \notin \mathcal{L}$, we can check that the previous inequality holds so that $(q^n-1)/(n(q+1)^2) \geq 2$ and thus $(|T_i''|-1)/m(S) \geq 2$ for $i = 1, 2$, as required.

(d) Assume that $S \cong \mathrm{PSp}_{2n}(q)$ or $\Omega_{2n+1}(q)$ and $n \geq 3$ is odd. We have that $m(S) \leq 2n+1$ and $d = \gcd(2, q-1)$. In this case both T_i are cyclic, so we can choose $T_i'' = T_i'$ and hence $|T_1''| \geq (q^n+1)/d$ and $|T_2''| \geq (q^n-1)/d$. Since $q^n+1 > q^n-1$, it suffices to show that $q^n-1 \geq (4n+3)d$, where $n \geq 3$ is odd and $d = \gcd(2, q-1)$. Since $S \notin \mathcal{L}$, we can check that the latter inequality holds, so for $i = 1, 2$, we obtain that $(|T_i''|-1)/m(S) \geq 2$.

(e) Assume that $S \cong \mathrm{PSp}_{2n}(q)$ or $\Omega_{2n+1}(q)$ and $n \geq 4$ is even. We can choose $|T_1''| = |T_1'| \geq (q^n+1)/d$ and $|T_2''| \geq (q^{n-1}+1)/d$. Since $q^n+1 > q^{n-1}+1$, it suffices to show that $q^{n-1}+1 \geq (4n+3)d$, where $n \geq 4$ is even and $d = \gcd(2, q-1)$. Since $S \notin \mathcal{L}$, we can check that the latter inequality holds, so for $i = 1, 2$, we have that $(|T_i''|-1)/m(S) \geq 2$.

(f) Assume that $S \cong P\Omega_{2n}^+(q)$ where $n \geq 5$ is odd. Then $m(S) = 2n$ and $d = \gcd(4, q^n-1)$. We have $|T_1''| \geq (q^{n-1}+1)/d$ and $|T_2''| \geq (q^n-1)/d$. Since $q^n-1 > q^{n-1}+1$, it suffices to show that $q^{n-1}+1 \geq (4n+1)d$, where $n \geq 5$ is odd. Since $S \notin \mathcal{L}$, we can check that the latter inequality holds, so for $i = 1, 2$, we obtain that $(|T_i''|-1)/m(S) \geq 2$.

(g) Assume that $S \cong P\Omega_{2n}^+(q)$ where $n \geq 4$ is even. Then $|T_1''| \geq (q^{n-1}+1)/d$ and $|T_2''| \geq (q^{n-1}-1)/d$. Since $q^{n-1}+1 > q^{n-1}-1$, it suffices to show that $q^{n-1}-1 \geq (4n+1)d$, where $n \geq 4$ is even. Since $S \notin \mathcal{L}$, we can check that the latter inequality holds, so for $i = 1, 2$, $(|T_i''|-1)/m(S) \geq 2$.

(h) Assume that $S \cong P\Omega_{2n}^-(q)$ where $n \geq 4$. Then $|T_1''| \geq (q^n+1)/d$ and $|T_2''| \geq (q^{n-1}+1)/d$. Since $q^n+1 > q^{n-1}+1$, it suffices to show that $q^{n-1}+1 \geq (4n+1)d$, where $n \geq 4$. Since $S \notin \mathcal{L}$, we can check that the latter inequality holds, so for $i = 1, 2$, $(|T_i''|-1)/m(S) \geq 2$ as wanted. This completes the proof of Claim 3.

Finally, by Claim 1, to finish the proof of the lemma, we only need to find a regular semisimple element $t_i \in T_i'$ such that t_i is not G-conjugate to s_i for $i = 1, 2$. Now, for each i, if T_i' contains an element whose order is a proper multiple of ℓ_i, then this element is a regular semisimple element by Claim 2 and clearly it is not G-conjugate to s_i as the orders of these two semisimple elements are distinct. Otherwise, if no such elements exists, then by Claim 3 we can find a regular semisimple element $t_i \in T_i'$ with the same order as that of s_i and they are not G-conjugate. The proof is now complete. □

We now prove the main result of this section.

Theorem 3.2. *Let S be a nonabelian simple group. If S is a T_k-group for some integer $k \geq 1$, then $k = 2$ and $S \cong \mathrm{PSL}_2(5)$ or $\mathrm{PSL}_2(7)$.*

Proof. Using the classification of finite simple groups, we consider several cases:

(1) S is a sporadic simple group or the Tits group. It is routine to check using [Conway et al. 1985] that S has at least two nontrivial distinct degrees, each with multiplicity at least 2. Hence S is not a T_k-group for any integer $k \geq 1$.

(2) $S \cong \mathrm{A}_n$ with $n \geq 5$. If $n = 5$, then $\mathrm{cd}(\mathrm{A}_5) = \{1, 3, 4, 5\}$ and every degree of A_5 has multiplicity 1, except for the degree 3 with multiplicity 2. Hence A_5 is a T_2-group. Now assume that $n \geq 6$. For $6 \leq n \leq 13$, we can check that A_n is not a T_k-group by using [Conway et al. 1985]. Thus we can assume that $n \geq 14$. Let λ be a self-conjugate partition of n and denote by χ^λ the irreducible character of S_n labeled by λ. Then χ^λ when restricted to A_n will split into the sum of two irreducible characters having the same degree $\chi^\lambda(1)/2$. Thus $\chi^\lambda(1)/2 \in \mathrm{cd}(A_n)$ has multiplicity at least two. Therefore, in order to show that A_n is not a T_k-group for any $k \geq 1$, it suffices to find two distinct self-conjugate partitions λ_i, $i = 1, 2$, of n such that $\chi^{\lambda_i}(1)/2$, $i = 1, 2$, are distinct and nontrivial.

Assume first that $n \geq 15$ is odd. We can write $n = 2k + 9 = 2(k + 4) + 1$. Then $\lambda_1 = (k+5, 1^{k+4})$ and $\lambda_2 = (k+3, 3^2, 1^k)$ are two distinct self-conjugate partitions of n. Assume next that $n \geq 14$ is even. Write $n = 2k+8$. Then $\lambda_1 = (k+4, 2, 1^{k+2})$ and $\lambda_2 = (k+3, 3^2, 1^k)$ are two distinct self-conjugate partitions of n. Using Hook formula, we can easily check that $\chi^{\lambda_i}(1)/2$ are distinct and nontrivial for $i = 1, 2$. Thus A_n is not a T_k-group for $n \geq 14$.

(3) S is a simple exceptional group of Lie type in characteristic p.

Assume first that $S \cong {}^2\mathrm{B}_2(q^2)$, where $q^2 = 2^{2m+1}$ and $m \geq 1$. By [Suzuki 1962], S has irreducible characters of degree $\sqrt{2}q(q^2 - 1)/2$ and $q^4 + 1$, with multiplicity 2 and $(q^2 - 2)/2$, respectively. Hence S is not a T_k-group for any $k \geq 1$.

Assume next that $S \cong {}^3\mathrm{D}_4(q)$. If $q = 2$, then ${}^3\mathrm{D}_4(2)$ is not a T_k-group for any integer $k \geq 1$ by using [Conway et al. 1985]. Hence we can assume that $q \geq 3$. By [Deriziotis and Michler 1987, Table 4.4], S has degrees given by

$$(q^3 + \delta)(q^2 - \delta q + 1)(q^4 - q^2 + 1)$$

with multiplicity $\frac{1}{2}q(q + \delta)$, where $\delta = \pm 1$. Since $q \geq 3$, we deduce that these two degrees are distinct and nontrivial and $q(q + 1)/2 \geq q(q - 1)/2 \geq 3$, so S is not a T_k-group.

Assume that $S \cong \mathrm{E}_6(q)$ and let $G = \mathrm{E}_6(q)_{\mathrm{ad}}$. Let $d = |G : S| = \gcd(3, q - 1)$. By [Lübeck 2007], G has an irreducible character χ of degree

$$\chi(1) = \tfrac{1}{2}q^3\Phi_1^2\Phi_3^2\Phi_4\Phi_5\Phi_6^2\Phi_8\Phi_9\Phi_{12}, \quad \text{with } m_G(\chi(1)) \geq q(q - 1).$$

$S = S(q)$	Symbol	Degree
${}^2\mathrm{G}_2(q^2)$	cuspidal	$\frac{1}{\sqrt{3}}q\Phi_1\Phi_2\Phi_4$
	cuspidal	$\frac{1}{2\sqrt{3}}q\Phi_1\Phi_2\Phi_{12}'$
${}^2\mathrm{F}_4(q^2)$	$({}^2\mathrm{B}_2[a], 1)$, $({}^2\mathrm{B}_2[b], 1)$	$\frac{1}{\sqrt{2}}q\Phi_1\Phi_2\Phi_4^2\Phi_6$
	$({}^2\mathrm{B}_2[a], \epsilon)$, $({}^2\mathrm{B}_2[b], \epsilon)$	$\frac{1}{\sqrt{2}}q^{13}\Phi_1\Phi_2\Phi_4^2\Phi_6$
$\mathrm{G}_2(q)$	$\phi_{1,3}'$, $\phi_{1,3}''$	$\frac{1}{3}q\Phi_3\Phi_6$
	$\mathrm{G}_2[\theta]$, $\mathrm{G}_2[\theta^2]$	$\frac{1}{3}q\Phi_1^2\Phi_2^2$
$\mathrm{F}_4(q)$	$\phi_{8,3}'$, $\phi_{8,3}''$	$q^3\Phi_4^2\Phi_8\Phi_{12}$
	$\phi_{8,9}'$, $\phi_{8,9}''$	$q^9\Phi_4^2\Phi_8\Phi_{12}$
${}^2\mathrm{E}_6(q)$	${}^2\mathrm{E}_6[\theta]$, ${}^2\mathrm{E}_6[\theta^2]$	$\frac{1}{3}q^7\Phi_1^4\Phi_2^6\Phi_4^2\Phi_8\Phi_{10}$
$\mathrm{E}_6(q)$	$\mathrm{E}_6[\theta]$, $\mathrm{E}_6[\theta^2]$	$\frac{1}{3}q^7\Phi_1^6\Phi_2^4\Phi_4^2\Phi_5\Phi_8$
$\mathrm{E}_7(q)$	$\mathrm{E}_7[\xi]$, $\mathrm{E}_7[-\xi]$	$\frac{1}{2}q^{11}\Phi_1^7\Phi_3^3\Phi_4^2\Phi_5\Phi_7\Phi_8\Phi_9\Phi_{12}$
	$\mathrm{E}_6[\theta]$, $\mathrm{E}_6[\theta^2]$	$\frac{1}{3}q^7\Phi_1^6\Phi_2^6\Phi_4^2\Phi_5\Phi_7\Phi_8\Phi_{10}\Phi_{14}$
$\mathrm{E}_8(q)$	$(\mathrm{E}_7[\xi], 1)$, $(\mathrm{E}_7[-\xi], 1)$	$\frac{1}{2}q^{11}\Phi_1^7\Phi_3^4\Phi_4^4\Phi_5^2\Phi_7\Phi_8^2\Phi_9\Phi_{12}^2\Phi_{15}\Phi_{20}\Phi_{24}$
	$(\mathrm{E}_7[\xi], \epsilon)$, $(\mathrm{E}_7[-\xi], \epsilon)$	$\frac{1}{2}q^{26}\Phi_1^7\Phi_3^4\Phi_4^4\Phi_5^2\Phi_7\Phi_8^2\Phi_9\Phi_{12}^2\Phi_{15}\Phi_{20}\Phi_{24}$

Table 2. Some unipotent characters of simple exceptional groups of Lie type.

Obviously, $\chi(1) > d$ and $m_G(\chi(1)) \geq 2d$. By Lemma 2.1, S has a nontrivial degree $b \in \{\chi(1), \chi(1)/d\}$ with nontrivial multiplicity. By Table 2, S also has a nontrivial degree $\psi(1)$ with multiplicity at least 2. Observe that $\psi(1) \notin \{\chi(1), \chi(1)/d\}$. Therefore, S has two distinct nontrivial degrees, each with multiplicity at least 2, so S is a not a T_k-group.

The same argument applies to the simple group $S \cong {}^2\mathrm{E}_6(q)$ with $q > 2$ since $G = {}^2\mathrm{E}_6(q)_{\mathrm{ad}}$ has a degree

$$\chi(1) = \tfrac{1}{2}q^3\Phi_2^2\Phi_3\Phi_4\Phi_6^3\Phi_8\Phi_{10}\Phi_{12}\Phi_{18}, \quad \text{with } m_G(\chi(1)) \geq (q+1)(q-2),$$

$|G : S| = \gcd(3, q+1) =: d$; and S has a nontrivial degree $\psi(1) \notin \{\chi(1), \chi(1)/d\}$ with multiplicity at least 2 by Table 2. For the case $q = 2$, we can check that ${}^2\mathrm{E}_6(2)$ is not a T_k-group by using [Conway et al. 1985].

Finally, for the remaining simple exceptional groups of Lie type, by Table 2 each simple group S has two distinct nontrivial degrees, each with multiplicity at least 2, so S is not a T_k-group.

(4) Assume that S is a simple classical group in characteristic p.

(4_1) Assume first that $S \notin \mathscr{L}$. We consider the following setup. Let $\mathscr{G}$ be a simply connected simple algebraic group of classical type and let F be a suitable Frobenius map such that $L/Z(L) \cong S$, where $L = \mathcal{G}^F$. Let the pair $(\mathcal{G}^*, F^*)$ be dual to $(\mathcal{G}, F)$ and let $G = (\mathcal{G}^*)^{F^*}$. Let $T \le G$ be a maximal torus of G. By Deligne–Lusztig theory, for each G-conjugacy class of regular semisimple element $s \in T$, there exists a semisimple character $\chi_s \in \mathrm{Irr}(L)$ with degree $|G:T|_{p'}$ and if $s \in G'$, then $\mathrm{Z}(L) \subseteq \ker\chi_s$, so χ_s is an irreducible character of $L/\mathrm{Z}(L) \cong S$. Moreover, if $t \in T \cap G'$ is also a regular semisimple element which is not G-conjugate to s, then the semisimple character $\chi_t \in \mathrm{Irr}(L)$ is an irreducible character of S with the same degree as that of χ_s and thus the nontrivial degree $|G:T|_{p'} \in \mathrm{cd}(S)$ has multiplicity at least 2.

Since $S \notin \mathcal{L}$, by Lemma 3.1 G contains two maximal tori T_i, $i = 1, 2$, such that each $T_i' = T_i \cap G'$ possesses two regular semisimple elements s_i and t_i which are not G-conjugate. By the discussion above, we deduce that each nontrivial degree $|G:T_i|_{p'} \in \mathrm{cd}(S)$ has multiplicity at least 2. Since $|G:T_i|_{p'}$, $i = 1, 2$, are distinct and nontrivial, we deduce that S is not a T_k-group for any integer $k \ge 1$.

(4_2) Assume next that $S \in \mathcal{L}$.

(a) Assume first that $S \cong \mathrm{PSL}_2(q)$ with $q \ge 4$. As $\mathrm{PSL}_2(4) \cong \mathrm{PSL}_2(5) \cong \mathrm{A}_5$, we can assume that $q \ge 7$. If $q = 7$, then $\mathrm{PSL}_2(7)$ is a T_2-group by using [Conway et al. 1985]. Hence we assume that $q \ge 8$. If q is even, then $S \cong \mathrm{SL}_2(q)$ has degrees $q-1$ and $q+1$ with multiplicity $q/2$ and $q/2-1$, respectively. Since $q \ge 8$, we can see that $q/2 > q/2-1 \ge 3$, so S is not a T_k-group. Now assume that $q \ge 9$ is odd. Since $\mathrm{PSL}_2(9) \cong \mathrm{A}_6$, we can assume that $q \ge 11$. We know that S has two irreducible characters of degree $(q+\epsilon)/2$ where $q \equiv \epsilon \pmod 4$ and $\epsilon \in \{1, -1\}$. Furthermore, S has a nontrivial degree $q-1$ with multiplicity $(q-\delta)/4$, where $q \equiv \delta \pmod 4$ and $\delta \in \{1, 3\}$. As $(q-1)/4 > (q-3)/4 \ge (11-3)/4 = 2$ and $q-1 > (q+1)/2 \ge (q-1)/2$, S has two distinct nontrivial degrees, each with multiplicity at least 2 and thus S is not a T_k-group.

(b) Assume that $S \cong \mathrm{PSL}_3(q)$. Since $\mathrm{PSL}_3(2) \cong \mathrm{PSL}_2(7)$, we can assume that $q \ge 3$. For $3 \le q \le 11$, we can check that $\mathrm{PSL}_3(q)$ is not a T_k-group by using [Conway et al. 1985]. So, we assume that $q \ge 13$. In this case, by [Simpson and Frame 1973] S has degrees $d_1 = q^2+q+1$ and $d_2 = q(q^2+q+1)$, both with multiplicity $(q-1)/d - 1$. Since $q \ge 13$ and $d = \gcd(3, q-1) \le 3$, it follows that $(q-1)/d - 1 \ge (q-1)/3 - 1 \ge 2$ and hence S is not a T_k-group in these cases.

(c) Assume that $S \cong \mathrm{PSU}_3(q)$. Since $\mathrm{PSU}_3(2)$ is not simple, we can assume that $q \ge 3$. For $3 \le q \le 9$, we can check that $\mathrm{PSU}_3(q)$ is not a T_k-group by using [Conway et al. 1985]. So, we assume that $q \ge 11$. In this case, by [Simpson and Frame 1973] S has degrees $d_1 = q^2-q+1$ and $d_2 = q(q^2-q+1)$, both with multiplicity $(q+1)/d-1$. Since $q \geq 11$ and $d = \gcd(3, q+1) \leq 3$, it follows that $(q+1)/d-1 \geq (q+1)/3-1 \geq 2$. Thus S is not a T_k-group.

(d) Assume that $S \cong \mathrm{PSp}_4(q)$ with $q \geq 3$. If $q \geq 4$ is even, then S possesses two distinct nontrivial degrees $q(q^2+1)/2$ and $(q-1)(q^2+1)$ with multiplicity 2 and q respectively by using [Lübeck 2007]. Now assume that $q \geq 3$ is odd. Using [Lübeck 2007] again, G has two distinct nontrivial character degrees $a_1 := 2q(q^2+1)/2$ and $a_2 := (q+1)(q^2+1)$ with multiplicity 4 and $3(q-3)/2$, respectively. Since $d = |G:S| = 2$, we deduce that $a_2/d > a_1 > d$. If $q \geq 5$, then $3(q-3)/2 \geq 2d = 4$ and $4 \geq 2d$, so it follows from Corollary 2.2 that S is not a T_k-group. For the remaining cases, we can check directly using [Conway et al. 1985] that S is not a T_k-group.

(e) Finally, for the remaining simple groups in $\mathscr{L}$, it is routine to check using [GAP 2012] that S is not a T_k-group for any $k \geq 1$. □

4. Nonsolvable T_k-groups

We first prove a special case of the main theorem. In fact, we show that nonperfect T_k-groups must be solvable. We note that if G is a T_k-group for some integer $k \geq 1$ and $N \trianglelefteq G$, then since $\mathrm{Irr}(G/N) \subseteq \mathrm{Irr}(G)$, we can easily see that G/N is also a T_m-group for some integer $m \leq k$.

Theorem 4.1. *If G is a nonperfect T_k-group for some $k \geq 1$, then G is solvable.*

Proof. Let G be a counterexample to the theorem with minimal order. Then $G' \neq G$ and G is a T_k-group for some $k \geq 1$ but G is nonsolvable. Let M be the last term of the derived series of G and let $N \trianglelefteq G$ such that M/N is a chief factor of G. Since G is nonsolvable, we see that M is nontrivial and hence it is perfect, so M/N is nonabelian and $M/N \cong W^t$ for some nonabelian simple group W and some integer $t \geq 1$. Then M/N is a minimal normal subgroup of G/N and that

$$|G/N : (G/N)'| = |G/N : G'/N| = |G : G'| > 1$$

as G is nonperfect and $N \leq M \leq G'$. It follows that G/N is a nonperfect nonsolvable group and since $\mathrm{Irr}(G/N) \subseteq \mathrm{Irr}(G)$, we deduce that G/N is a nonperfect nonsolvable T_m-group for some integer $m \geq 1$. If N is nontrivial, then $|G/N| < |G|$, which contradicts the minimality of $|G|$. Therefore, we conclude that N must be trivial and $M \cong W^t$.

Claim 1. $M \cong \mathrm{PSL}_2(3^f)$ *for some* $f \geq 2$.

We first show that $W \cong \mathrm{PSL}_2(3^f)$ with $f \geq 2$. Suppose by contradiction that $W \ncong \mathrm{PSL}_2(3^f)$ with $f \geq 2$. Then there exist two irreducible characters $\theta_i \in \mathrm{Irr}(W)$ such that $\theta_1(1) \neq \theta_2(1)$ and both θ_i extend to $\mathrm{Aut}(W)$ by Lemma 2.3. Let $\varphi_i = \theta_i^t \in \mathrm{Irr}(M)$ for $i = 1, 2$. By [Bianchi et al. 2007, Lemma 5], we deduce that both φ_i extend to

$\chi_i \in \operatorname{Irr}(G)$. Furthermore, by Gallagher's theorem [Isaacs 1976, Corollary 6.17] we know that each φ_i has exactly $|G/M : (G/M)'| = |G : G'|$ extensions. For each i, all extensions of φ_i have the same degree which is $\varphi_i(1) = \theta_i^t(1) > 1$. So, G has two distinct nontrivial degrees $\theta_i^t(1)$, $i = 1, 2$, both with nontrivial multiplicity, which is a contradiction. Hence $W \cong \mathrm{PSL}_2(3^f)$ with $f \geq 2$ as we wanted.

We now claim that $t = 1$ and thus $M \cong \mathrm{PSL}_2(3^f)$ with $f \geq 2$. By way of contradiction, assume that $t \geq 2$. Let θ be the Steinberg character of W. Then $\varphi = \theta^t \in \operatorname{Irr}(M)$ extends to $\varphi_0 \in \operatorname{Irr}(G)$ by [Bianchi et al. 2007, Lemma 5]. Thus by Gallagher's theorem [Isaacs 1976, Corollary 6.17] again, we have that $\varphi_0(1) = \varphi(1) = \theta(1)^t$ is a nontrivial degree with nontrivial multiplicity. It follows that if $d \in \operatorname{cd}(G)$ with $1 < d \neq \theta(1)^t = 3^{tf}$, then the multiplicity of d is trivial. It is well know that $\mathrm{PSL}_2(3^f)$ has two irreducible characters of degree $(3^f + \epsilon)/2$, where $3^f \equiv \epsilon \pmod 4$ and $\epsilon \in \{\pm 1\}$, hence $\mathrm{PSL}_2(3^f)$ has a nontrivial degree $(3^f + \epsilon)/2 < 3^f$ with multiplicity 2. Denote these two irreducible characters by α_i, $i = 1, 2$. For $i = 1, 2$, let

$$\varphi_i = 1 \times 1 \times \cdots \times \alpha_i \in \operatorname{Irr}(M) \quad \text{and} \quad \psi = 1 \times 1 \times \cdots \times \theta \in \operatorname{Irr}(M).$$

Let I, I_1 and I_2 be the inertia groups of ψ, φ_1 and φ_2, respectively. Obviously, we have that $M \leq I_i \leq I \leq G$ for $i = 1, 2$. By the representation theory of wreath products, we know that ψ extends to $\psi_0 \in \operatorname{Irr}(I)$ and $|G : I| = t$. Since G/M is solvable, we deduce that I/M is solvable. If I/M is nontrivial, then I/M has $j > 1$ linear characters and thus ψ has j distinct extensions to I, which are $\psi_0\lambda$ with $\lambda \in \operatorname{Irr}(I/M)$ and $\lambda(1) = 1$, so by Clifford's theorem [Isaacs 1976, Theorem 6.11] we have that $(\psi_0\lambda)^G \in \operatorname{Irr}(G)$ are distinct irreducible characters of G having the same degree. Furthermore, for $\lambda \in \operatorname{Irr}(I/M)$ with $\lambda(1) = 1$, we have

$$(\psi_0\lambda)^G(1) = \psi_0^G(1) = |G : I|\psi_0(1) = 3^f \cdot t < 3^{tf}.$$

The last inequality holds since $t \geq 2$. But then this is a contradiction since the multiplicity of the nontrivial degree $3^f \cdot t$ is at least $|I/M : (I/M)'|$ which is nontrivial by our assumption. Therefore, we conclude that I/M is trivial and so $I = M$. It follows that for $i = 1, 2$, we have $I_i = I = M$ since $M \leq I_i \leq I = M$. Thus for each i, we have $\varphi_i^G \in \operatorname{Irr}(G)$ and

$$\varphi_i^G(1) = |G : M|\alpha_i(1) = |G : M|(3^f + \epsilon)/2$$

which is nontrivial and different from 3^{ft}. Clearly, $\varphi_1^G \neq \varphi_2^G$, so we deduce that G has a nontrivial degree $|G : M|(3^f + \epsilon)/2 \neq 3^{ft}$ with multiplicity at least 2, which is impossible. This contradiction proves our claim.

Claim 2. *G is an almost simple group with socle M.*

By the previous claim, we know that $M \cong \mathrm{PSL}_2(3^f)$ with $f \geq 2$. Let $C = C_G(M)$.

Then $C \trianglelefteq G$ and G/C is an almost simple group with socle MC/C. Assume first that G/C is perfect. Then $G = MC$ and since M is nonabelian simple, we must have that $M \cap C = 1$ and so $G = M \times C$, where $G/M \cong C$ is solvable. If C is nontrivial, then $|C : C'| > 1$ and so for each nontrivial irreducible character $\mu \in \mathrm{Irr}(M)$ of M, we see that μ has $|C : C'|$ extensions to $G = M \times C$ and thus G cannot be a T_k-group for any $k \geq 1$. Hence C must be trivial and so G is simple, which is impossible as $G \neq G'$. Assume next that G/C is nonperfect. Then G/C is a nonperfect nonsolvable T_m-group for some $m \geq 1$. By the minimality of $|G|$, we must have that $C = 1$ and thus G is an almost simple group with socle $M \cong \mathrm{PSL}_2(3^f)$.

The final contradiction. Let $q = 3^f$, with $f \geq 2$. Let α be an irreducible character of M with $\alpha(1) = (q+\epsilon)/2$ where $q \equiv \epsilon \pmod 4$ and $\epsilon \in \{\pm 1\}$, let δ be the diagonal automorphism of M and let φ be the field automorphism of M of order f. Then $\mathrm{Out}(M) = \langle \delta \rangle \times \langle \varphi \rangle$. Since G is nonperfect, we deduce that $|G : M|$ is nontrivial. Observe that the Steinberg character St_M of M is extendible to $\mathrm{Aut}(M)$ and so it extends to G by [Bianchi et al. 2007, Lemma 5], hence by Gallagher's theorem [Isaacs 1976, Corollary 6.17] the degree $\mathrm{St}_M(1) = |M|_3$ has multiplicity at least $|G : M| > 1$. Thus the multiplicity of every nontrivial degree of G different from $|M|_3 = q$ must be trivial. By [White 2013, Lemma 4.6], α is φ-invariant. Now if $G \leq M\langle \varphi \rangle$, then α is G-invariant and since G/M is cyclic, α extends to G and so G has a nontrivial degree $(q+\epsilon)/2$ with multiplicity at least $|G : M| \geq 2$, which is impossible. Hence $G \not\leq M\langle \varphi \rangle$. By [White 2013, Theorem 6.5], we have $I_G(\alpha) = G \cap M\langle \varphi \rangle$ and $|G : I_G(\alpha)| = 2$. If $|I_G(\alpha) : M| > 1$, then α has $|I_G(\alpha) : M|$ extensions to $I_G(\alpha)$ as $I_G(\alpha)/M$ is cyclic and thus by inducing these characters to G, we see that G has at least $|I_G(\alpha) : M| \geq 2$ irreducible characters of degree $q+\epsilon$, which is a contradiction. Thus, we conclude that $I_G(\alpha) = M$ and $|G : M| = 2$. By [White 2013, Corollary 6.2], we have that either $G \cong \mathrm{PGL}_2(q)$ or $G \cong M\langle \delta\varphi^{f/2} \rangle$, where f is even. Clearly, the first case cannot happen. For the latter case, since f is even, we obtain that $q \equiv 1 \pmod 4$, so M has exactly $(q-1)/4$ irreducible characters of degree $q-1$. As $|G : M| = 2$, by [White 2013, Theorem 6.6] all irreducible characters of G lying over an irreducible character of M of degree $q-1$ have degree $2(q-1)$. Therefore, G has at least $(q-1)/8$ irreducible characters of degree $2(q-1)$. If $q = 9$, then we can check directly using [Conway et al. 1985] that all almost simple groups with socle $\mathrm{PSL}_2(9) \cong \mathrm{A}_6$ are not T_k-groups for any integer $k \geq 1$. Thus we can assume that $q \geq 81$ as f is even, so $(q-1)/8 \geq 10$, hence G is not a T_k-group. This final contradiction proves our theorem. □

We are now ready to prove the main theorem.

Proof of Theorem A. Let G be a counterexample to the theorem with minimal order. Then G is a nonsolvable T_k-group for some integer $k \geq 1$ but G is isomorphic to neither $\mathrm{PSL}_2(5)$ nor $\mathrm{PSL}_2(7)$. If $G' \neq G$, then G is solvable by Theorem 4.1, which is a contradiction. Thus we can assume that G is perfect. Let M be a maximal normal subgroup of G. Then G/M is a nonabelian simple group. Since $\mathrm{Irr}(G/M) \subseteq \mathrm{Irr}(G)$, we deduce that G/M is a T_m-group for some $m \geq 1$. Now by Theorem 3.2, we have that $G/M \cong \mathrm{PSL}_2(q)$ with $q \in \{5, 7\}$ and G/M is a T_2-group. Since G is a counterexample to the theorem, we deduce that M is nontrivial.

By [Conway et al. 1985], we know that $\mathrm{cd}(\mathrm{PSL}_2(5)) = \{1, 3, 4, 5\}$ with multiplicity 1, 2, 1, 1 and $\mathrm{cd}(\mathrm{PSL}_2(7)) = \{1, 3, 6, 7, 8\}$ with multiplicity 1, 2, 1, 1, 1. Since G is a T_k-group, we deduce that the degree $3 \in \mathrm{cd}(G/M)$ is the unique nontrivial character degree in $\mathrm{cd}(G)$ with nontrivial multiplicity and also $k \geq 2$. If $k = 2$, then G must be isomorphic to either $\mathrm{PSL}_2(5)$ or $\mathrm{PSL}_2(7)$ by [Berkovich and Kazarin 1996, Main Theorem], which is a contradiction. Therefore, we must have that $k \geq 3$. Hence there exists $\chi \in \mathrm{Irr}(G|M)$ with $\chi(1) = 3$ and thus $\chi \in \mathrm{Irr}(G|\theta)$ for some nontrivial irreducible character θ of M. By Clifford's theorem [Isaacs 1976, Theorem 6.2], we have that $\chi_M = e(\theta_1 + \theta_2 + \cdots + \theta_t)$, where all θ_i are conjugate to θ in G, $e \geq 1$ is the degree of an irreducible projective representation of $I_G(\theta)/M$ and $t = |G : I_G(\theta)|$. Since $\chi(1) = 3$, we have that $3 = et\theta(1)$ and hence $t \leq 3$. As the index of a proper subgroup of G/M with $G/M \cong \mathrm{PSL}_2(5)$ or $\mathrm{PSL}_2(7)$ is at least 5, we must have that $t = 1$, so $\chi_M = e\theta$ and θ is G-invariant. If θ is extendible to $\theta_0 \in \mathrm{Irr}(G)$, then $\theta_0(1) = \theta(1) \geq 2$ since G is perfect; also by Gallagher's theorem [Isaacs 1976, Corollary 6.17], we obtain that $\mathrm{Irr}(G|\theta) = \{\theta_0\,\lambda \mid \lambda \in \mathrm{Irr}(G/M)\}$. It follows that $\chi = \mu\theta_0$ for some $\mu \in \mathrm{Irr}(G/M)$. As $3 = \chi(1) = \mu(1)\theta(1)$ and $\theta(1) \geq 2$, we must have that $\mu(1) = 1$ and $\theta(1) = 3$. Since $3 \in \mathrm{cd}(G/M)$ has multiplicity 2, there exist two distinct irreducible characters λ_i, $i = 1, 2$, of G/M with $\lambda_i(1) = 3$ and so $\theta_0\lambda_i$, $i = 1, 2$, are two distinct irreducible characters of G, both have degree 9. Thus $9 \in \mathrm{cd}(G)$ has multiplicity at least 2 which is a contradiction as $3 \in \mathrm{cd}(G)$ already has nontrivial multiplicity. Thus θ is G-invariant but it is not extendible to G. As the Schur multiplier of $\mathrm{PSL}_2(q)$ with $q \in \{5, 7\}$ is cyclic of order 2, by the theory of character triple isomorphism [Isaacs 1976, Chapter 11], the triple (G, M, θ) must be isomorphic to $(\mathrm{SL}_2(q), A, \lambda)$, where $q \in \{5, 7\}$, $A = \mathrm{Z}(\mathrm{SL}_2(q))$ and μ is a nontrivial irreducible character of A. Since $\mathrm{cd}(\mathrm{SL}_2(q)|\lambda) = \{(q-\epsilon)/2, q-1, q+1\}$, with $q \equiv \epsilon \pmod 4$ and $\epsilon \in \{\pm 1\}$, we deduce that $\mathrm{cd}(G|\theta) = \{\theta(1)(q-\epsilon)/2, (q-1)\theta(1), (q+1)\theta(1)\}$. However, we can check that all degrees in $\mathrm{cd}(G|\theta)$ are even and thus $3 \notin \mathrm{cd}(G|\theta)$. This contradiction proves the theorem. □

Acknowledgments

The author is indebted to G. Malle for his help with the proof of Lemma 3.1. The author is also grateful to the anonymous referee for useful comments and suggestions.

References

[Babai et al. 2009] L. Babai, P. P. Pálfy, and J. Saxl, "On the number of p-regular elements in finite simple groups", *LMS J. Comput. Math.* **12** (2009), 82–119. MR 2010f:20017 Zbl 1225.20013

[Berkovich 1996] Y. Berkovich, "Finite solvable groups in which only two nonlinear irreducible characters have equal degrees", *J. Algebra* **184**:2 (1996), 584–603. MR 97j:20005b Zbl 0861.20008

[Berkovich and Kazarin 1996] Y. Berkovich and L. Kazarin, "Finite nonsolvable groups in which only two nonlinear irreducible characters have equal degrees", *J. Algebra* **184**:2 (1996), 538–560. MR 97j:20005a Zbl 0861.20009

[Berkovich et al. 1992] Y. Berkovich, D. Chillag, and M. Herzog, "Finite groups in which the degrees of the nonlinear irreducible characters are distinct", *Proc. Amer. Math. Soc.* **115**:4 (1992), 955–959. MR 92j:20006 Zbl 0822.20004

[Bianchi et al. 2007] M. Bianchi, D. Chillag, M. L. Lewis, and E. Pacifici, "Character degree graphs that are complete graphs", *Proc. Amer. Math. Soc.* **135**:3 (2007), 671–676. MR 2007i:20014 Zbl 1112.20006

[Brauer 1963] R. Brauer, "Representations of finite groups", pp. 133–175 in *Lectures on modern mathematics, I*, edited by T. L. Saaty, Wiley, New York, 1963. MR 31 #2314 Zbl 0124.26504

[Carter 1985] R. W. Carter, *Finite groups of Lie type: Conjugacy classes and complex characters*, Pure and Applied Mathematics **5**, Wiley, New York, 1985. MR 87d:20060 Zbl 0567.20023

[Conway et al. 1985] J. H. Conway, R. T. Curtis, S. P. Norton, R. A. Parker, and R. A. Wilson, *Atlas of finite groups*, Oxford University Press, Eynsham, 1985. MR 88g:20025 Zbl 0568.20001

[Craven 2008] D. A. Craven, "Symmetric group character degrees and hook numbers", *Proc. Lond. Math. Soc.* (3) **96**:1 (2008), 26–50. MR 2009f:20011 Zbl 1165.20008

[Deriziotis and Michler 1987] D. I. Deriziotis and G. O. Michler, "Character table and blocks of finite simple triality groups ${}^3D_4(q)$", *Trans. Amer. Math. Soc.* **303**:1 (1987), 39–70. MR 88j:20011 Zbl 0628.20014

[Dolfi et al. 2013] S. Dolfi, G. Navarro, and P. H. Tiep, "Finite groups whose same degree characters are Galois conjugate", *Israel J. Math.* **198**:1 (2013), 283–331. MR 3096641

[GAP 2012] The GAP Group, *GAP – groups, algorithms, and programming, version 4.5.6*, 2012, http://www.gap-system.org.

[Isaacs 1976] I. M. Isaacs, *Character theory of finite groups*, Pure and Applied Mathematics **69**, Academic Press, New York, 1976. Reprinted Chelsea RI, 2006. MR 57 #417 Zbl 0337.20005

[Larsen et al. 2013] M. Larsen, G. Malle, and P. H. Tiep, "The largest irreducible representations of simple groups", *Proc. London Math. Soc.* **106**:1 (2013), 65–96. Zbl 06139422

[Lübeck 2007] F. Lübeck, "Character degrees and their multiplicities for some groups of Lie type rank < 9", Online data, 2007, http://www.math.rwth-aachen.de/~Frank.Luebeck/chev/DegMult/index.html.

[Malle 1999] G. Malle, "Almost irreducible tensor squares", *Comm. Algebra* **27**:3 (1999), 1033–1051. MR 99m:20007 Zbl 0931.20009

[Malle 2008] G. Malle, "Extensions of unipotent characters and the inductive McKay condition", *J. Algebra* **320**:7 (2008), 2963–2980. MR 2009e:20028 Zbl 1163.20003

[Malle 2010] G. Malle, "Zeros of Brauer characters and the defect zero graph", *J. Group Theory* **13**:2 (2010), 171–187. MR 2011c:20016 Zbl 1208.20009

[Moretó 2007] A. Moretó, "Complex group algebras of finite groups: Brauer's problem 1", *Adv. Math.* **208**:1 (2007), 236–248. MR 2008c:20011 Zbl 1109.20008

[Seitz 1968] G. Seitz, "Finite groups having only one irreducible representation of degree greater than one", *Proc. Amer. Math. Soc.* **19** (1968), 459–461. MR 36 #5212 Zbl 0244.20010

[Simpson and Frame 1973] W. A. Simpson and J. S. Frame, "The character tables for $SL(3, q)$, $SU(3, q^2)$, $PSL(3, q)$, $PSU(3, q^2)$", *Canad. J. Math.* **25** (1973), 486–494. MR 49 #398 Zbl 0264.20010

[Suzuki 1962] M. Suzuki, "On a class of doubly transitive groups", *Ann. of Math.* (2) **75** (1962), 105–145. MR 25 #112 Zbl 0106.24702

[Tong-Viet 2012] H. P. Tong-Viet, "Simple classical groups of Lie type are determined by their character degrees", *J. Algebra* **357** (2012), 61–68. MR 2905242 Zbl 06104526

[Tong-Viet 2013] H. P. Tong-Viet, "Characterization of some simple groups by the multiplicity pattern", *Monatsh. Math.* **172**:2 (2013), 189–206. MR 3117188

[White 2013] D. L. White, "Character degrees of extensions of $PSL_2(q)$ and $SL_2(q)$,", *J. Group Theory* **16**:1 (2013), 1–33.

[Zsigmondy 1892] K. Zsigmondy, "Zur Theorie der Potenzreste", *Monatsh. Math. Phys.* **3**:1 (1892), 265–284. MR 1546236 JFM 24.0176.02

Received October 25, 2012. Revised December 5, 2012.

HUNG P. TONG-VIET
SCHOOL OF MATHEMATICS, STATISTICS AND COMPUTER SCIENCE
UNIVERSITY OF KWAZULU-NATAL
PRIVATE BAG X01
SCOTTSVILLE, 3209
SOUTH AFRICA
Current address:
FAKULTÄT FÜR MATHEMATIK
UNIVERSITÄT BIELEFELD POSTFACH 10 01 31
D-33501 BIELEFELD
GERMANY

ptongviet@math.uni-bielefeld.de

PACIFIC JOURNAL OF MATHEMATICS
Vol. 268, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.268.493

ERRATA TO "DYNAMICS OF ASYMPTOTICALLY HYPERBOLIC MANIFOLDS"

JULIE ROWLETT

Volume **242**:2 (2009), 377–397

Theorem 1.4 of the article in question — its main result — is a prime orbit theorem for the geodesic flow of asymptotically hyperbolic manifolds with negative sectional curvatures. We correct a typo and supply a missing technical assumption to make the result and its proof correct. We also supply a remainder term required in the dynamical wave trace formula given in Theorem 1.1, relating the length spectrum to the regularized trace of the wave group. This correction does not affect the main terms in the trace formula nor its application by Borthwick and Perry (2011).

1. Corrections

Joint work with P. Suarez-Serrato and S. Tapie [Rowlett et al. 2011] has led to the discovery of a missing term in the dynamical wave trace formula given in Theorem 1.1 of [Rowlett 2009], due to a subtle gap in the proof. A correct version follows:

Theorem 1.1. *Suppose (X, g) is an asymptotically hyperbolic $(n+1)$-dimensional manifold with negative sectional curvatures. Let $0\text{-tr}\cos(t\sqrt{\Delta - n^2/4})$ denote the regularized trace of the wave group, and let $t_0 > 0$. Let $\mathscr{L}_p$ denote the set of primitive closed geodesics of (X, g), and for $\gamma \in \mathscr{L}_p$, let $l(\gamma)$ denote the length of γ. Then*

$$0\text{-tr}\cos\big(t\sqrt{\Delta - n^2/4}\big) = \sum_{\gamma\in\mathscr{L}_p}\sum_{k\in\mathbb{N}} \frac{l(\gamma)\delta(|t| - kl(\gamma))}{2\sqrt{|\det(I - \mathscr{P}_\gamma^k)|}} + R(t), \tag{1-1}$$

as a distributional equality in $\mathscr{D}'([t_0, \infty))$[1], where $\mathscr{P}_\gamma^k$ is the k-times Poincaré map around γ in the cotangent bundle. The remainder $R(t)$ is continuous and can be written as the sum of two continuous terms

$$R(t) = A(t) + B(t),$$

which satisfy the following.

MSC2010: 37D40, 53C22, 58J50.

Keywords: asymptotically hyperbolic, wave trace, prime orbit theorem, renormalized trace.

[1] $\mathscr{D}'(X)$ is the dual of $\mathscr{C}_0^\infty(X)$.

(1) *There exist constants* $\varepsilon, C > 0$ *such that*

$$\left| \int_0^\infty A(t)\cos(\lambda t)\rho(t)\,dt \right| \le C \tag{1-2}$$

for all $\lambda > 1$ *and* $\rho \in \mathscr{C}_0^\infty([t_0, \varepsilon \ln \lambda])$. *The constants* C *and* ε *depend only on* t_0 *and* $\|\rho\|_\infty$; *they are independent of* λ.

(2) $B(t)$ *is independent of the set of closed geodesics, and it is possible that* $B(t)$ *has exponential growth for large time.*

Theorem 1.1 of [Rowlett 2009] was restated in Theorem 5.1 of [Rowlett 2010], which is therefore subject to the same correction; it was also applied in [Borthwick and Perry 2011], but without use of the remainder estimate for large time, so the results there are unaffected.

Corollary 1.2 of [Rowlett 2009] (restated in [Rowlett 2010] as Corollary 5.2) should be corrected as follows.

Corollary 1.2. *Let* (X, g) *be a manifold with negative sectional curvatures that is hyperbolic near infinity. Then, for any* $t_0 > 0$, *as an element of* $\mathscr{D}'([t_0, \infty))$, *we have the distributional equality*

$$\sum_{s \in \mathscr{R}^{sc}} e^{(s-n/2)|t|} = \sum_{\gamma \in \mathscr{L}_p} \sum_{k \in \mathbb{N}} \frac{l(\gamma)\delta(|t| - kl(\gamma))}{\sqrt{|\det(I - \mathscr{P}_\gamma^k)|}} + A(t) + B(t).$$

Here the resonances $\mathscr{R}^{sc}$ *of the scattering operator* (see [Rowlett 2009]) *are summed* with multiplicity, *and the remainders* $A(t)$ *and* $B(t)$ *have the same properties as in Theorem 1.1.*

In Theorem 1.4 of [Rowlett 2009], a factor of h was omitted in the numerator; this was corrected in Theorem 4.4 of [Rowlett 2010]. An additional technical assumption, that the length-spectrum is nonarithmetic (see [Rowlett et al. 2011, §4]), was also missing. (Experts believe that this assumption always holds, but a proof is not known.) Thus the correct statement is this:

Theorem 1.4. *Suppose* (X, g) *is an asymptotically hyperbolic* $(n+1)$*-dimensional manifold with negative sectional curvatures and nonarithmetic length spectrum. Let* h *be the topological entropy of the geodesic flow, and assume* $h > 0$. *The dynamical zeta function*

$$Z(s) = \exp\left(\sum_{\gamma \in \mathscr{L}_p} \sum_{k \in \mathbb{N}} \frac{e^{-ksl_p(\gamma)}}{k} \right)$$

has a nowhere vanishing analytic extension to an open neighborhood of $\Re(s) \ge h$ *except for a simple pole at* $s = h$. *Let* $\mathscr{L}$ *denote the set of all closed geodesics, and* *for $\gamma \in \mathscr{L}$, let $l(\gamma)$ denote the length of γ. The length spectrum counting function*

$$N(T) := \#\{\gamma \in \mathscr{L} : l(\gamma) \leq T\} \tag{1-3}$$

satisfies $\lim_{T\to\infty} \frac{hTN(T)}{e^{hT}} = 1$.

The corrected statements of Theorem 1.1, Corollary 1.2 and Theorem 1.4 follow from the proofs in [Rowlett 2009], with the exception of statement (2) above, about the remainder term $B(t)$, which shall be proved in Section 2.

As we shall demonstrate below, the remainder term $B(t)$ may grow exponentially for large time. This was unexpected because in the compact case there is no such term, and in the noncompact model case of conformally compact hyperbolic manifolds, the remainder term has exponential decay as $t \to \infty$; see [Guillarmou and Naud 2006, Theorem 1.1].

Corollary 1.5 of [Rowlett 2009] (and the analogous Corollary 5.3 in [Rowlett 2010]) should both be replaced by the version below. Before stating it, we recall the standard big-O notation. For a function $f : \mathbb{R} \to \mathbb{C}$ and a function $F : \mathbb{R} \to \mathbb{R}$ we use the notation

$$f(t) = O(F(t)) \quad \text{as } t \to \infty$$

if there exist constants $T, C > 0$ such that $|f(t)| \leq CF(t)$ for all $t > T$.

Proposition 1.5. *Let (M^{n+1}, g) be a Riemannian manifold with negative sectional curvatures that is hyperbolic near infinity, and whose length spectrum is nonarithmetic. Assume the topological entropy of the geodesic flow is positive. Let W be the Sinai–Bowen–Ruelle potential, and let $\wp$ denote topological pressure.*

(1) *Assume the remainder $B(t)$ in Corollary 1.2 also satisfies a statement identical to (1) of Theorem 1.1, with the substitution of $B(t)$ for $A(t)$. If*

$$\wp(-W/2) > 0,$$

then the discrete spectrum of the Laplacian is nonempty and the infimum Λ_1 of the spectrum of the Laplacian satisfies the estimate

$$\Lambda_1 = \min \sigma_{\mathrm{pp}}(\Delta) \leq \frac{n^2}{4} - \big(\wp(-W/2)\big)^2.$$

(2) *If both remainder terms are uniformly bounded as $t \to \infty$, then*

$$\wp(-W/2) > 0 \iff \sigma_{\mathrm{pp}} \neq \varnothing,$$

and these equivalent conditions imply

$$\Lambda_1 = \frac{n^2}{4} - \big(\wp(-W/2)\big)^2.$$

(3) *The remainder satisfies*

$$|R(t)| = O\big(\sup\{e^{(s_1-n/2)t},\ e^{\wp(-W/2)t},\ t\}\big) \quad as\ t \to \infty.$$

(If $\sigma_{\text{pp}} = \varnothing$, the term $e^{(s_1-n/2)t}$ is omitted from the supremum.)

(4) *The assumption about $B(t)$ in part (1) does not always hold.*

The proof of this proposition is based on Theorem 1.1, Corollary 1.2, and the following result, which will appear in [Rowlett et al. 2011].

Theorem 1.6. *Let (M, g) be a conformally compact manifold whose sectional curvatures satisfy $-b^2 \le K_g \le -a^2 < 0$, with nonabelian fundamental group*[2] *and nonarithmetic length spectrum. The weighted zeta function*

$$\tilde{Z}(s) = \exp\bigg(\sum_{\gamma\in\mathcal{L}_p}\sum_{k\in\mathbb{N}} \frac{e^{-ksl_p(\gamma)}}{k\sqrt{|\det(I-\mathcal{P}_\gamma^k)|}}\bigg)$$

is an analytic nonzero function on the half-plane $\Re(s) > \wp(-W/2)$. It admits a meromorphic extension to the half plane

$$\Re(s) > \wp(-W/2) - \inf\Big\{\frac{\lambda a}{b}, \frac{\lambda}{2}\Big\},$$

where λ is the expansion factor of the geodesic flow on the nonwandering set. Moreover, with the exception of a simple pole at $\wp(-W/2)$, this extension is analytic and nonvanishing in an open neighborhood of $\{\Re(s) \ge \wp(-W/2)\}$. If $\wp(-W/2) > 0$, then we have the counting estimate

$$\sum_{\gamma\in\mathcal{L}_T} |\det(I-\mathcal{P}_\gamma)|^{-1/2} \sim \frac{\exp\big(\wp(-W/2)T\big)}{\wp(-W/2)T} \quad as\ T \to \infty,$$

where $\mathcal{L}_T$ denotes the set of closed geodesics of length at most T.

2. Proofs

The relationship between the support of the test function and its oscillation in (1-2) is known in the context of semiclassical analysis as *Ehrenfest time.* Therefore, we will say that $(\lambda, T) \in (1,\infty)^2$ is an *ε-Ehrenfest pair* if it satisfies $T \le \varepsilon \ln\lambda$, where $\varepsilon > 0$ is a constant. Due to the oscillation of test functions needed to control the remainder, we shall use the "Dirichlet box principle" technique of Jakobson, Polterovich, and Toth:

[2]Note that this is equivalent to the positivity of the topological entropy of the geodesic flow (see [Rowlett et al. 2011, Proposition 3.12]).

Proposition 2.1 (Dirichlet box principle [Jakobson et al. 2008]). *There are infinitely many ε-Ehrenfest pairs $(\lambda, T(\lambda))$ such that*

(2-1) *for each $\gamma \in \mathscr{L}$ with $l(\gamma) \le T$,*
there exists $k(\gamma) \in \mathbb{Z}$ such that $|\lambda l(\gamma) - 2\pi k(\gamma)| \le \frac{1}{2}$.

Remark. For such an ε-Ehrenfest pair $(\lambda, T(\lambda))$, for each $\gamma \in \mathscr{L}$ with $l(\gamma) \le T$,

$$\cos(\lambda l(\gamma)) \ge \tfrac{1}{2}.$$

Proof of Proposition 1.5. From the assumption of negative sectional curvatures and hyperbolicity near infinity, it follows that the manifold is asymptotically hyperbolic and has pinched negative sectional curvatures; see [Rowlett et al. 2011, §2]. We may therefore apply Corollary 1.2 and Theorem 1.6.

We first prove part (1). Let the assumptions in that statement hold and set

$$\eta := \wp(-W/2) > 0.$$

The spectral side of the trace formula can be reformulated in terms of the resonances of the resolvent as follows:

$$\sum_{s\in\mathcal{R}} m(s) e^{(s-n/2)|t|} + \sum_{k\in\mathbb{N}} d_k e^{-k|t|}.$$

There are at most finitely many terms in the first sum with $\Re(s) > n/2$, and these are in bijection with the pure point (discrete) spectrum (see [Mazzeo and Melrose 1987]) via

$$s(n-s) = \Lambda \in \sigma_{\mathrm{pp}} \subset \Bigl(0, \frac{n^2}{4}\Bigr).$$

Write

$$\mathcal{R} = \mathcal{R}_{\mathrm{pp}} \cup \mathcal{R}_0,$$

where

$$\mathcal{R}_{\mathrm{pp}} = \{s \in \mathcal{R} \mid \Re(s) > n/2\}, \quad \mathcal{R}_0 = \{s \in \mathcal{R} \mid \Re(s) \le n/2\}.$$

We shall require the counting estimate demonstrated in Theorem 1.1 of [Borthwick 2008], which implies

(2-2) $$\mathcal{N}(R) := \sum_{\substack{s\in\mathcal{R}\\ |s|\le R}} m(s) = O(R^{n+1}), \quad \sum_{k=1}^{R} d_k = O(R^{n+1}).$$

Let $(\lambda, T(\lambda))$ be an ε-Ehrenfest pair that satisfies (2-1). Let

$$l_0 := \inf\{l(\gamma) \mid \gamma \in \mathscr{L}_p\}.$$

For $T > 0$, let $\rho_T \in \mathscr{C}^\infty((0, \infty))$ be a smooth, nonnegative function such that

$$\rho_T(t) = \begin{cases} 0 & \text{if } t \le l_0/2, \\ 1 & \text{if } l_0 \le t \le T, \\ 0 & \text{if } T+1 \le t. \end{cases} \tag{2-3}$$

Let

$$f(t) := \cos(\lambda t)\rho_{T(\lambda)}(t), \quad \rho(t) := \rho_{T(\lambda)}(t). \tag{2-4}$$

By Corollary 1.2 and the assumption of statement (1), we have $I = II$, where

$$I := \sum_{s \in \mathcal{R}} m(s) \int_0^\infty e^{(s-n/2)t} f(t)\, dt + \sum_{k=1}^\infty d_k \int_0^\infty e^{-kt} f(t)\, dt$$

and

$$II := \sum_{\gamma \in \mathcal{L}_p} \sum_{k=1}^\infty \frac{l(\gamma) f(kl(\gamma))}{\sqrt{|\det(I - \mathcal{P}_\gamma^k)|}} + O(1).$$

By the Dirichlet box principle, for all $\gamma \in \mathcal{L}_p$ and for any $k \in \mathbb{N}$ with $kl(\gamma) = \ell(k\gamma) \le T(\lambda)$, there exists $j(\gamma) \in \mathbb{Z}$ such that

$$|\lambda l(k\gamma) - 2\pi j(\gamma)| \le \tfrac{1}{2}.$$

It follows that

$$f(kl(\gamma)) = \cos(\lambda k l(\gamma))\rho(kl(\gamma)) \ge \tfrac{1}{2}.$$

Therefore,

$$II \ge \sum_{\substack{\gamma \in \mathcal{L}_p,\, k \in \mathbb{N} \\ kl(\gamma) \le T}} \frac{l(\gamma)}{2\sqrt{|\det(I - \mathcal{P}_\gamma^k)|}} + O(1).$$

By Theorem 1.6, there exists a constant $C > 0$, independent of T, such that

$$II \ge C \frac{e^{\eta T}}{T} + O(1).$$

On the other hand, we can estimate I from above. Define

$$F(s) := \int_0^\infty e^{(s-n/2)t} f(t)\, dt.$$

Since the pure point spectrum is finite or empty, and $0 < \Lambda_1 \le \Lambda_2 \le \cdots < n^2/4$, with $\Lambda_j = s_j(n - s_j)$, it follows that

$$s_1 \ge s_2 \ge \cdots > \frac{n}{2}.$$

Therefore, for each $s \in \mathcal{R}_{\text{pp}}$,

$$|F(s)| \le \int_0^\infty e^{(s_1 - n/2)t} \rho(t)\, dt = O(e^{(s_1 - n/2)T}).$$

Next we estimate for $s \in \mathcal{R}_0$. Since

$$F(s) = \tfrac{1}{2}\int_0^\infty e^{st} e^{-(n/2)t}(e^{i\lambda t} + e^{-i\lambda t})\rho(t)\,dt,$$

for $s = 0$ and $s = n/2 \pm i\lambda$, we have

$$F(s) = O(T).$$

For $s \neq 0$, $s \neq n/2 \pm i\lambda$, integrating by parts k times (compare [Guillarmou and Naud 2006, §3]), we have

$$\begin{aligned} F(s) = \quad & \frac{(-1)^k}{2(s-n/2+i\lambda)^k}\int_0^\infty e^{(s-n/2+i\lambda)t}\rho^{(k)}(t)\,dt \\ & + \frac{(-1)^k}{2(s-n/2-i\lambda)^k}\int_0^\infty e^{(s-n/2-i\lambda)t}\rho^{(k)}(t)\,dt. \end{aligned}$$

We therefore have the estimate

$$|F(s)| = O\left(\frac{T}{|s|^k + |\lambda|^k}\right) = O\left(\frac{T}{|s|^k}\right).$$

For $s \neq 0$, $s \neq n/2 \pm i\lambda$, using the counting estimate (2-2) and the above with $k = n+2$, we estimate (compare [Guillarmou and Naud 2006, §3])

$$\sum_{s\in\mathcal{R}_0} |F(s)| = O(T) + \int_1^\infty \frac{d\mathcal{N}(R)}{R^{n+2}} = O(T).$$

By the counting estimate (2-2), it follows that there is a constant $c > 0$ (independent of k) such that

$$d_k \leq ck^{n+1} \quad \text{for all } k \in \mathbb{N}.$$

Therefore the sum

$$\sum_{k=1}^\infty d_k e^{-kt}$$

converges uniformly on $[l_0/2, \infty)$ and is uniformly bounded. Consequently

$$\left|\int_0^\infty f(t)\sum_{k=1}^\infty d_k e^{-kt}\,dt\right| = O(T).$$

Putting this together with the estimate for $\mathcal{R}_{\mathrm{pp}}$, we have

$$I = O(e^{(s_1-n/2)T}) + O(T).$$

If the pure point spectrum of the Laplacian is empty, then there is no contribution to I from $\mathcal{R}_{\mathrm{pp}} = \varnothing$, and we have

$$|I| = O(T) = |II| \geq C\frac{e^{\eta T}}{T}.$$

This implies that we cannot have $\eta > 0$. We have thus shown that $\sigma_{\mathrm{pp}} = \varnothing \Longrightarrow \eta \leq 0$.

By the contrapositive, if $\eta > 0$ then $\sigma_{\mathrm{pp}}(\Delta) \neq \varnothing$. In this case, denote the infimum of the spectrum σ_{pp} by Λ_1, with

$$s_1(n - s_1) = \Lambda_1.$$

By our above estimates,

$$O(e^{(s_1 - n/2)T}) + O(T) = |I| = |II| \geq C\frac{e^{\eta T}}{T}. \tag{2-5}$$

By the Dirichlet box principle, there exist infinitely many Ehrenfest pairs, so we may let $T \to \infty$, from which it follows that

$$s_1 - \frac{n}{2} \geq \eta, \quad \text{and hence} \quad s_1 \geq \eta + \frac{n}{2}.$$

This in turn implies the upper bound

$$\Lambda_1 \leq \left(\eta + \frac{n}{2}\right)\left(\frac{n}{2} - \eta\right) = \frac{n^2}{4} - \eta^2.$$

We now turn to part (2), and thus assume that $A(t)$ and $B(t)$ are both uniformly bounded as $t \to \infty$. In this case we do not need the Dirichlet box principle and may simply use the test function ρ. If $\eta > 0$, then by part (1) the pure point spectrum is nonempty and

$$s_1 \geq \eta + \frac{n}{2}.$$

We have the estimate

$$\int_0^\infty e^{(s_1 - n/2)t}\rho(t)\,dt \geq Ce^{(s_1 - n/2)T}, \quad C > 0.$$

Combining this with our estimates above, we have

$$I \geq Ce^{(s_1 - n/2)T} - C'T,$$

for some constant C'. Since $\eta > 0$, Theorem 1.6 implies that

$$II \leq C''\frac{e^{\eta T}}{T}$$

for some constant $C'' > 0$. Putting this together, we have

$$Ce^{(s_1 - n/2)T} - C'T \leq I = II \leq C''\frac{e^{\eta T}}{T}.$$

It follows that

$$s_1 - \frac{n}{2} \le \eta,$$

which, combined with the inequality from (1) (note that the condition in (2) implies the one in (1)), gives

$$\eta > 0 \implies \sigma_{\mathrm{pp}} \neq \varnothing \text{ and } s_1 - \frac{n}{2} = \eta.$$

The last equality implies $\Lambda_1 = n^2/4 - \eta^2$.

Conversely, if $\sigma_{\mathrm{pp}} \neq \varnothing$, we may similarly estimate I from below, which shows that the dynamical side must also grow exponentially as $T \to \infty$. Since the remainder terms are bounded, by Theorem 1.6 we must have $\eta > 0$. This concludes the proof that $\sigma_{\mathrm{pp}} \neq \varnothing \iff \eta > 0$, and as we have seen we have $\Lambda_1 = n^2/4 - \eta^2$ in this case.

We shall prove (3) by contradiction. Assume that for any $C > 0$, for each $N \in \mathbb{N}$ there is $T_N > N$ with

$$|R(T_N)| > C \sup\{e^{(s_1 - n/2)T_N}, e^{\wp(-W/2)T_N}, T_N\}.$$

By the continuity of R, for each $N \in \mathbb{N}$ there exists a nonnegative test function $\rho_N \in \mathscr{C}_0^\infty(0, T_N + 1)$ such that

$$\int_0^\infty \rho_N(t)\, dt = 1,$$

and

$$\left| \int_0^\infty R(t)\rho_N(t)\, dt \right| > C \sup\{e^{(s_1 - n/2)T_N}, e^{\wp(-W/2)T_N}, T_N\}.$$

By Corollary 1.2,

$$\sum_{s \in \mathscr{R}} m(s) \int_0^\infty e^{(s-n/2)t} \rho_N(t)\, dt + \sum_{k=1}^{\infty} d_k \int_0^\infty e^{-kt} \rho_N(t)\, dt - \sum_{\gamma \in \mathscr{L}_p} \sum_{k=1}^{\infty} \frac{l(\gamma)\rho_N(kl(\gamma))}{\sqrt{|\det(I - \mathscr{P}_\gamma^k)|}} = \int_0^\infty R(t)\rho_N(t)\, dt.$$

By Theorem 1.6 and our above estimates, the norm of the left side is bounded above by

$$C' \sup\{e^{(s_1 - n/2)T_N}, e^{\wp(-W/2)T_N}, T_N\} \tag{2-6}$$

for a fixed constant $C' > 0$ that is independent of N and ρ_N. This in turn implies the same upper bound for the right side. This is a contradiction. Therefore, there exist constants $C > 0$ and $T > 0$ such that

$$|R(t)| \le C \sup\{e^{(s_1 - n/2)t}, e^{\wp(-W/2)t}, t\} \quad \text{for all } t > T,$$

which is the conclusion of (3).

To prove (4), we will describe a counterexample suggested by Gilles Carron and Samuel Tapie. Let $S = \mathbb{H}^2/\Gamma$ be a convex cocompact hyperbolic surface whose topological entropy satisfies $h = h(g_H) > \frac{1}{2}$, where g_H is the hyperbolic metric on S, and assume that the length spectrum is nonarithmetic. Then by [Sullivan 1979] its Laplacian admits an isolated first eigenvalue $\Lambda_H = h(1-h) \in \left(0, \frac{1}{4}\right)$. We shall assume for the sake of contradiction that for any manifold that has negative sectional curvatures and is hyperbolic near infinity, the remainder term $B(t)$ in Corollary 1.2 also satisfies the estimate (1-2).

Let Ω be a compact, convex subset of S that properly contains the nonwandering set,[3] and let

$$\Omega \subset B_R \subset B_{R'} \subset B_{R''},$$

with B_R, $B_{R'}$, $B_{R''}$ balls of radii $0 < R < R' < R''$.

Let $\alpha : S \to [1-\varepsilon, 1]$ be a smooth function satisfying

$$\alpha(x) = \begin{cases} 1 & \text{for } x \in \Omega, \\ 1-\varepsilon & \text{for } x \in B_{R'} \setminus B_R, \\ 1 & \text{for } x \in S \setminus B_{R''}. \end{cases} \tag{2-7}$$

Let g be the Riemannian metric defined for all $x \in S$ by $g(x) = \alpha(x) g_H(x)$. For $\varepsilon > 0$ sufficiently small, the sectional curvatures of g remain negative. Thus (S, g) is again hyperbolic near infinity, and a function is in $\mathscr{L}^2(S, g)$ if and only if it is in $\mathscr{L}^2(S, g_H)$. Since $\alpha \equiv 1$ on the nonwandering set, $g = g_H$ on Ω. It follows that the length spectrum of (S, g) is identical to the length spectrum of (S, g_H), and hence is also nonarithmetic. Moreover, $h_g = h_{g_H}$, and therefore the topological entropy of (S, g) is also positive; equivalently, the fundamental group of (S, g) is nonabelian. Since $W(g) = W(g_H)$ along all closed geodesics,

$$\eta := \wp\left(\frac{-W(g)}{2}\right) = \wp\left(\frac{-W(g_H)}{2}\right) = h - \tfrac{1}{2} > 0.$$

Since $B(t)$ satisfies the estimate (1-2) and the hypotheses of Proposition 1.5, it follows from (1) that the discrete spectrum of (S, g) is nonempty and

$$\Lambda_1 \le \tfrac{1}{4} - \eta^2 = \Lambda_H.$$

It follows that there exists $\phi : S \to [0, \infty)$ (not identically zero) such that

$$\frac{\int_S \|\nabla\phi\|_g^2 \, dv_g}{\int_S \|\phi\|_g^2 \, dv_g} = \Lambda_1 \le \Lambda_H.$$

[3] Note that since S is convex cocompact and hyperbolic, it is asymptotically hyperbolic as well as convex cocompact with pinched negative curvatures and therefore the nonwandering set is a compact subset; see [Joshi and Sá Barreto 2001; Rowlett et al. 2011].

For any positive smooth function ϕ,

$$\int_S \|\nabla\phi\|_g^2\,dv_g = \int_S g^{-1}(d\phi, d\phi)\alpha^2\,dv_{g_H} = \int_S \alpha^{-2} g_H^{-1}(d\phi, d\phi)\alpha^2\,dv_{g_H}$$
$$= \int_S g_H^{-1}(d\phi, d\phi)\,dv_{g_H} = \int_S \|\nabla\phi\|_{g_H}^2\,dv_{g_H}.$$

By the maximum principle, ϕ cannot vanish identically on $B_{R'} \setminus B_R$, so by definition of α we have

$$\int_S \|\phi\|_{g_H}^2 \alpha^2\,dv_{g_H} < \int_S \|\phi\|_{g_H}^2\,dv_{g_H}.$$

Since $0 < \Lambda_1 \le \Lambda_H$, we have

$$\int_S \|\nabla\phi\|_{g_H}^2\,dv_{g_H} = \int_S \|\nabla\phi\|_g^2\,dv_g = \Lambda_1 \int_S \|\phi\|_g^2\,dv_g$$
$$= \Lambda_1 \int_S \|\phi\|_{g_H}^2 \alpha^2\,dv_{g_H} < \Lambda_1 \int_S \|\phi\|_{g_H}^2\,dv_{g_H} \le \Lambda_H \int_S \|\phi\|_{g_H}^2\,dv_{g_H},$$

which leads to the estimate

$$\frac{\int_S \|\nabla\phi\|_{g_H}^2\,dv_{g_H}}{\int_S \|\phi\|_{g_H}^2\,dv_{g_H}} < \Lambda_H.$$

This contradicts the definition of Λ_H as the infimum of the spectrum of the Laplacian on S with respect to the hyperbolic metric. □

For manifolds of higher dimension, it is also possible to build counterexamples to the long-time estimate (1-2) for $B(t)$ using conformal deformations.

Proof of Theorem 1.1(2). The remainder term $A(t)$ is *defined* to depend only on the set of closed geodesics. We shall use the preceding example to prove statement (2) of Theorem 1.1. Since the set of closed geodesics is contained in the nonwandering set, which is a compact, convex subset of the manifold (see [Rowlett et al. 2011]), the estimate (1-2) follows from the Ehrenfest estimate for compact manifolds with pinched negative curvature as in [Jakobson et al. 2008] and in the original proof of this estimate in [Rowlett 2009]. The unexpected news is that the difference

$$B(t) := R(t) - A(t),$$

may have exponential growth for large time. To prove this, we shall use the example used to prove part (4) of Proposition 1.5. Let (S, g_H), (S, g), Λ_H, and Λ_1 be defined as above. Since the assumption that

$$\Lambda_1 \le \Lambda_H$$

leads to a contradiction, we have $\Lambda_1 > \Lambda_H$, and hence

$$s_1 < s_H = h = \eta + \tfrac{1}{2} = \wp\big(-W(g)/2\big) + \tfrac{1}{2},$$

where

$$s_1 = \tfrac{1}{2} + \sqrt{\tfrac{1}{4} - \Lambda_1}.$$

For the constant curvature metric (S, g_H), the Guillarmou–Naud trace formula [2006, Theorem 1.1, Theorem 1.2] implies

$$\sum_{s \in \mathcal{R}} m(s) e^{(s-n/2)|t|} + \sum_{k \in \mathbb{N}} d_k e^{-k|t|} = \sum_{\gamma \in \mathcal{L}_p} \sum_{k=1}^{\infty} \frac{l(\gamma)\delta(|t| - kl(\gamma))}{\sqrt{|\det(I - \mathcal{P}_\gamma^k)|}} + R_H(t),$$

where

$$R_H(t) = O(e^{-t/2}) \quad \text{as } t \to \infty.$$

Let us write

$$R_H(t) = A_H(t) + B_H(t),$$

where A_H is defined to be the contribution from closed geodesics, and

$$B_H := R_H - A_H.$$

The perturbation (S, g) is hyperbolic near infinity and has negative sectional curvatures. Let $R(t)$ denote the remainder in the trace formula Theorem 1.1 for (S, g), and similarly write

$$R(t) = A(t) + B(t).$$

Since the perturbation did not change the set of closed geodesics,

$$A = A_H.$$

Estimating as in the proof of Proposition 1.5, by (2-5),

$$O(e^{(s_1 - 1/2)T}) + O(T) = |I| = |II| \geq C \frac{e^{\eta T}}{T} - \left| \int_0^\infty R(t) f(t)\, dt \right|.$$

Rearranging, we have

$$\left| \int_0^\infty R(t) f(t)\, dt \right| \geq C \frac{e^{\eta T}}{T} - O(e^{(s_1 - 1/2)T}). \tag{2-8}$$

Since

$$s_1 - \tfrac{1}{2} < s_H - \tfrac{1}{2} = h - \tfrac{1}{2} = \eta,$$

both sides of (2-8) have exponential growth as $t \to \infty$. For the test function f as defined in (2-3), (2-4),

$$\left| \int_0^\infty R_H(t) f(t)\, dt \right| = \left| \int_0^\infty A_H(t) f(t)\, dt + \int_0^\infty B_H(t) f(t)\, dt \right| = O(e^{-t/2}).$$

This shows that the perturbation of the metric $g_H \mapsto g$, which did not affect $A_H = A$, had a rather drastic effect on the second part of the remainder term, B_H.

In particular, for the original metric, since A_H satisfies the estimate (1-2), and the total remainder R_H decays exponentially, the remainder term B_H must also satisfy (1-2). However, this clearly cannot be the case for B. Hence, the remainder in the trace formula Theorem 1.1 may include a contribution that is independent of the set of closed geodesics and that grows exponentially for large time. □

3. Concluding remarks

The estimate in (1-2) [Rowlett 2009] arises solely from the nonwandering set, which is a compact, convex subset of the manifold (see [Rowlett et al. 2011]), and therefore reduces to the Ehrenfest estimate for compact manifolds with pinched negative curvature as in [Jakobson et al. 2008]. This estimate corresponds to the remainder term $A(t)$. The proof of Proposition 1.5 and Theorem 1.1(2) shows that one may perturb the manifold away from the nonwandering set, which does not change the leading term in the dynamical side of the trace formula nor the remainder $A(t)$, but which does change the bottom of the spectrum. It follows that this perturbation affects the long-time asymptotics of the dynamical side of the trace formula and therefore must change the long-time estimate of the remainder term. Since the perturbation does not affect the nonwandering set, which contains all closed geodesics, the remainder term must have an additional contribution arising from nonclosed geodesics; this is the term $B(t)$ in Theorem 1.1 and Corollary 1.2 above. This unexpected contribution is not seen in either the compact case or the model case of convex cocompact hyperbolic manifolds. It would be interesting to identify this contribution more precisely.

References

[Borthwick 2008] D. Borthwick, “Upper and lower bounds on resonances for manifolds hyperbolic near infinity”, *Comm. Partial Differential Equations* **33**:7-9 (2008), 1507–1539. MR 2009i:58039 Zbl 1168.58012

[Borthwick and Perry 2011] D. Borthwick and P. A. Perry, “Inverse scattering results for manifolds hyperbolic near infinity”, *J. Geom. Anal.* **21**:2 (2011), 305–333. MR 2012i:58023 Zbl 1229.58024

[Guillarmou and Naud 2006] C. Guillarmou and F. Naud, “Wave 0-trace and length spectrum on convex co-compact hyperbolic manifolds”, *Comm. Anal. Geom.* **14**:5 (2006), 945–967. MR 2008f:58032 Zbl 1127.58028

[Jakobson et al. 2008] D. Jakobson, I. Polterovich, and J. A. Toth, “A lower bound for the remainder in Weyl’s law on negatively curved surfaces”, *Int. Math. Res. Not.* **2008**:2 (2008), Art. ID rnm142. MR 2009f:58038 Zbl 1161.58010

[Joshi and Sá Barreto 2001] M. S. Joshi and A. Sá Barreto, “The wave group on asymptotically hyperbolic manifolds”, *J. Funct. Anal.* **184**:2 (2001), 291–312. MR 2002m:58046 Zbl 0997.58010

[Mazzeo and Melrose 1987] R. R. Mazzeo and R. B. Melrose, “Meromorphic extension of the resolvent on complete spaces with asymptotically constant negative curvature”, *J. Funct. Anal.* **75**:2 (1987), 260–310. MR 89c:58133 Zbl 0636.58034

[Rowlett 2009] J. Rowlett, "Dynamics of asymptotically hyperbolic manifolds", *Pacific J. Math.* **242**:2 (2009), 377–397. MR 2011e:58045 Zbl 1198.37036

[Rowlett 2010] J. Rowlett, "On the spectral theory and dynamics of asymptotically hyperbolic manifolds", *Ann. Inst. Fourier* (*Grenoble*) **60**:7 (2010), 2461–2492. MR 2012j:58044 Zbl 1252.37025

[Rowlett et al. 2011] J. Rowlett, P. Suárez-Serrato, and S. Tapie, "Dynamics and zeta functions on conformally compact manifolds", preprint, 2011. To appear in *Trans. Amer. Math. Soc*. arXiv 1106.1857

[Sullivan 1979] D. Sullivan, "The density at infinity of a discrete group of hyperbolic motions", *Inst. Hautes Études Sci. Publ. Math.* 50 (1979), 171–202. MR 81b:58031 Zbl 0439.30034

Received January 4, 2013. Revised October 29, 2013.

JULIE ROWLETT
INSTITUT FÜR ANALYSIS
LEIBNIZ UNIVERSITÄT HANNOVER
WELFENGARTEN 1
30167 HANNOVER
GERMANY

rowlett@math.uni-hannover.de

PACIFIC JOURNAL OF MATHEMATICS
Vol. 268, No. 2, 2014
dx.doi.org/10.2140/pjm.2014.268.507

ERRATUM TO "SINGULARITIES OF THE PROJECTIVE DUAL VARIETY"

ROLAND ABUAF

Volume **253**:1 (2011), 1–17

We give a counterexample to a proposition claimed to be proven in an earlier paper of ours and used in the proof of its main theorem. We also show how to salvage the main result of that paper under additional hypotheses.

Let $X \subset \mathbb{P}^N$ be a nondegenerate smooth projective variety such that X^* is a hypersurface. Let $L \subset \mathbb{P}^N$ be a linear subspace such that for general $x \in X$ we have $\langle L, T_{X,x}\rangle \neq \mathbb{P}^N$. We say that $L^\perp$ is an *unexpected equisingular space* in X^* (see Definition 3.2.1 of [Abuaf 2011], hereafter cited as [A]) if the general hyperplane containing $\langle L, T_{X,x}\rangle$ has the same multiplicity in X^* as a general hyperplane containing L. In [A], the following side-result, whose aim was to discuss a necessary hypothesis in our main theorem, was stated in Section 3 ("Open question and corollaries"):

Theorem 3.2.2 of [A]. *Let $X \subset \mathbb{P}^N$ be an irreducible, smooth, nondegenerate projective variety such that X^* is a hypersurface. Let $L \subset X$ be a linear space with* $\dim(L) = \dim(X) - 1$. *Assume that $L^\perp$ is an unexpected equisingular linear space in X^* such that* $\operatorname{mult}_{L^\perp} X^* = 2$. *Then X is the cubic scroll surface in $\mathbb{P}^4$.*

Its proof was based on this proposition:

Proposition 3.2.3 of [A]. *Let $X \subset \mathbb{P}^N$ be a smooth, irreducible, nondegenerate projective variety such that X^* is a hypersurface. Let $[h] \in X^*$ be such that* $\operatorname{mult}_{[h]} X^* = 2$. *The scheme-theoretic tangency locus of H with X is one of the following*:

- *An irreducible hyperquadric and in this case* $|\mathscr{C}_{[h]}(X^*)|^* = \operatorname{Tan}(H, X)$.
- *The union of two (not necessarily distinct) linear spaces.*
- *A linear space with at least one embedded component.*

This proposition is false as shown by the following example.

MSC2010: primary 14B05; secondary 14N15.

Keywords: projective geometry, singularities, dual variety.

Example 1. Let V be a vector space of dimension 6 and let $W = \mathbb{G}(3, V) \subset \mathbb{P}(\bigwedge^3 V)$ be the Grassmannian of $\mathbb{C}^3 \subset V$ in its Plücker embedding. The dual of X is a quartic hypersurface in $\mathbb{P}(\bigwedge^3 V^*)$. We can decompose $\bigwedge^3 V^*$ as

$$\mathbb{C} \oplus U \oplus U^* \oplus \mathbb{C},$$

where U is identified with the space of 3×3 matrices (see [Landsberg and Manivel 2001, Section 5] for more details). We denote by C the determinant on U, which can be seen as a map $S^3U \to \mathbb{C}$ or as a map $S^2U \to U^*$. We also denote by C^* the determinant on U^*.

It is shown in (ibid.) that an equation (up to an automorphism of $\mathbb{P}(\bigwedge^3 V^*)$) of W^* is

$$Q(x, X, Y, y) = \bigl(3xy - \tfrac{1}{2}\langle X, Y\rangle\bigr)^2 + \tfrac{1}{3}\bigl(yC(X^{\otimes 3}) + xC^*(Y^{\otimes 3})\bigr) - \tfrac{1}{6}\bigl\langle C^*(Y^{\otimes 2}), C(X^{\otimes 2})\bigr\rangle,$$

where $\langle \cdot, \cdot \rangle$ is the standard pairing between U and U^*. The partial derivatives of Q give the equations of the variety of "stationary secants" to $W^\perp := \mathbb{G}(3, V^*) \subset \mathbb{P}(\bigwedge^3 V^*)$, which we denote by $\sigma_+(W^\perp)$. The Jacobian criterion shows that the variety $\sigma_+(W^\perp)$ is singular precisely along $W^\perp$. However, a simple Taylor expansion of Q around the point $[1, 0, \ldots, 0] \in W^\perp$ shows that, contrary to what is claimed in Proposition 5.10 of (ibid.), $W^\perp$ is not defined by all the second derivatives of Q. The orbit structure of the action of SL_6 on $\mathbb{P}(\bigwedge^3 V^*)$ is

$$W^\perp \subset \sigma_+(W^\perp) \subset W^* \subset \mathbb{P}\bigl(\textstyle\bigwedge^3 V^*\bigr).$$

Since $W^\perp$ is the deepest strata in W^* and all the second derivatives of the equation of W^* do not vanish on $W^\perp$, we conclude that there are no point of multiplicity bigger than 2 in W^*. However one can prove (see (ibid.) for instance) that a point in $W^\perp$ is tangent to W along a cone over $\mathbb{P}^2 \times \mathbb{P}^2$. This gives a counterexample to the above proposition. Note that an easy computation shows that if $p = (p_0, P_0, P_1, p_1) \in \mathbb{P}(\bigwedge^3 V^*)$ is a generic point then the cubic hypersurface (which we denote by $\mathcal{P}(Q, p)$) defined by the equation $p_0 \frac{\partial Q}{\partial x} + P_0 \frac{\partial Q}{\partial X} + P_1 \frac{\partial Q}{\partial Y} + p_1 \frac{\partial Q}{\partial y}$ is smooth. Moreover the polar $\mathcal{P}(W^*, p) := W^* \cap \mathcal{P}(Q, p)$ has multiplicity 3 along $W^\perp$.

In [A] I claim that I "prove" Proposition 3.2.3 in the appendix. This proof relies on the following statement:

Lemma A.3 of [A]. *Let $Z \subset \mathbb{P}^N$ be an irreducible and reduced hypersurface, whose defining equation is denoted by f_Z. Let $z \in Z$ and let $k \in \{-1, \ldots, N-2\}$. Then one of the following holds for general $D \in \mathbb{G}(k, N)$:*

- $z \notin P(Z, D)$.
- $$\operatorname{mult}_z P(Z, D) = \operatorname{mult}_z Z . \operatorname{mult}_z P(f_Z, D)$$

if $\dim(Z^{(z)}_{sing}) < \dim P(Z, D)$, *where* $Z^{(z)}_{sing}$ *is an irreducible component of* Z_{sing} *of maximal dimension passing through* z.

- $$\operatorname{mult}_z P(Z, D) < \operatorname{mult}_z Z . \operatorname{mult}_z P(f_Z, D)$$

if $\dim(Z^{(z)}_{sing}) \geq \dim P(Z, D)$, *where* $Z^{(z)}_{sing}$ *is an irreducible component of* Z_{sing} *of maximal dimension passing through* z.

This lemma is also false as shown by Example 1. Indeed the hypersurface $\mathcal{P}(Q, p)$ is smooth for generic p, the hypersurface W^* has multiplicity 2 along $W^\perp$, but the polar $\mathcal{P}(W^*, p) := W^* \cap \mathcal{P}(Q, p)$ has multiplicity 3 along $W^\perp$. The mistake in the proof of the lemma can be easily found. On line 5, page 14 of [A], I write “Let $(Z_i)_{i \in I}$ be a stratification of Z such that Z_i is smooth and Z is normally flat along Z_i for all $i \in I$. Such a stratification exists, due to the open nature of normal flatness [...]. Consider the Gauss map $G : Z \to (\mathbb{P}^N)^*$. It restricts to a map $G_i : Z_i \to (\mathbb{P}^N)^*$ ... .” This last sentence is nonsense since the Gauss map is not defined on the singular locus of Z.

I used Lemma A.3 of [A] in the form of the following corollary:

Corollary A.4 of [A]. *Let* $Z \subset \mathbb{P}^N$ *be an hypersurface and let* $z \in Z$ *such that* $\operatorname{mult}_z Z = 2$ *and let* $k \in \{-1, \dots N-2\}$. *Then, for generic* $D \in \mathbb{G}(k, N)$, *we have* $\operatorname{mult}_z \mathcal{P}(Z, D) \leq 2$.

This corollary is again false as shown in Example 1, but it seems natural to use its conclusion as an hypothesis. Indeed the rest of the proof of Proposition 3.2.3 of [A] is correct, and thus we get the following result:

Proposition 2 (replacement for Proposition 3.2.3 of [A]). *Let* $X \subset \mathbb{P}^N$ *be a smooth, irreducible, nondegenerate projective variety such that* X^* *is a hypersurface. Let* $[h] \in X^*$ *be such that* $\operatorname{mult}_{[h]} X^* = 2$ *and that for all* $k \in \{-1, \dots, N-2\}$ *and generic* $D \in \mathbb{G}(k, N)$, *we have* $\operatorname{mult}_{[h]} \mathcal{P}(X^*, D) \leq 2$. *The scheme theoretic tangency locus of* H *with* X *is one of the following*:

- *An irreducible hyperquadric and in this case* $|\mathcal{C}_{[h]}(X^*)|^* = \operatorname{Tan}(H, X)$.
- *The union of two (not necessarily distinct) linear spaces.*
- *A linear space with at least one embedded component.*

Finally, we can formulate a version of Theorem 3.2.2 of [A], whose proof relies on the above proposition:

Theorem 3 (replacement for Theorem 3.2.2 of [A]). *Let $X \subset \mathbb{P}^N$ be an irreducible, smooth, nondegenerate projective variety such that X^* is a hypersurface. Let $L \subset X$ be a linear space with* $\dim(L) = \dim(X) - 1$. *Assume that $L^\perp$ is an unexpected equisingular linear space in X^* such that* $\operatorname{mult}_{L^\perp}(X^*) = 2$. *Assume moreover that for all $[h] \in L^\perp$, for all $k \in \{-1, \dots, N-2\}$ and generic $D \in \mathbb{G}(k, N)$, we have* $\operatorname{mult}_{[h]} \mathcal{P}(X^*, D) \leq 2$. *Then X is the cubic scroll surface in $\mathbb{P}^4$.*

References

[Abuaf 2011] R. Abuaf, "Singularities of the projective dual variety", *Pacific J. Math.* **253**:1 (2011), 1–17. MR 2869431 Zbl 1252.14035

[Landsberg and Manivel 2001] J. M. Landsberg and L. Manivel, "The projective geometry of Freudenthal's magic square", *J. Algebra* **239**:2 (2001), 477–512. MR 2002g:14070 Zbl 1064.14053

Received September 11, 2012.

Roland Abuaf
Institut Fourier
100 rue des maths, BP 74
38402 Saint-Martin d'Hères cedex
France
abuaf@ujf-grenoble.fr

CONTENTS

Volume 268, no. 1 and no. 2

Guidelines for Authors

Authors may submit manuscripts at msp.berkeley.edu/pjm/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use LaTeX, but papers in other varieties of TeX, and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used but tags will be converted to the house format (see a current issue for examples).

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to pacific@math.berkeley.edu.

Each figure should be captioned and numbered, so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text ("the curve looks like this:"). It is acceptable to submit a manuscript will all figures at the end, if their placement is specified in the text by means of comments such as "Place Figure 1 here". The same considerations apply to tables, which should be used sparingly.

Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Page proofs will be made available to authors (or to the designated corresponding author) at a website in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

PACIFIC JOURNAL OF MATHEMATICS

Volume 268 No. 2 April 2014