Pacific Journal of Mathematics

Volume 289 No. 2 August 2017

PACIFIC JOURNAL OF MATHEMATICS

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

msp.org/pjm

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

The subscription price for 2017 is US $450/year for the electronic version, and $625/year for print and electronic.
Subscriptions, requests for back issues and changes of subscriber 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. 289, No. 2, 2017
dx.doi.org/10.2140/pjm.2017.289.257

REGULAR REPRESENTATIONS OF COMPLETELY BOUNDED MAPS

B. V. RAJARAMA BHAT, NIRUPAMA MALLICK AND K. SUMESH

We study properties and the structure of some special classes of homomorphisms on C^*-algebras. These maps are $*$-preserving up to conjugation by a symmetry. Making use of these homomorphisms, we prove a new structure theorem for completely bounded maps from a unital C^*-algebra into the algebra of all bounded linear maps on a Hilbert space. Finally we provide alternative proofs for some of the known results about completely bounded maps and improve on them.

1. Introduction

Completely positive maps and completely bounded maps on C^* algebras are well-studied objects (see [Arveson 1969; Paulsen 1986; Pisier 2001]). We look at two well-known structure theorems for completely bounded maps. The first one, which we call the *fundamental representation theorem* of completely bounded maps (Paulsen 2002; Wittstock 1981; 1984; Haagerup 1980) says that all completely bounded maps from a unital C^*-algebra $\mathcal{A}$ into the algebra $\mathcal{B}(\mathcal{H})$ of bounded operators on a Hilbert space $\mathcal{H}$ can be obtained from a unital representation of $\mathcal{A}$ on another Hilbert space, by composing it with two bounded operators. Unlike Stinespring's [1955] representation theorem for completely positive maps, there is no minimality condition on the representing Hilbert space, and hence the fundamental representation is not unique. The second structure theorem, namely *commutant representation theorem* (proved by Paulsen and Suen [1985]) says that all such completely bounded maps can be obtained from a unital representation, by first multiplying by an element in the commutant and then conjugating by a bounded operator. Later, using the theory of Hilbert C^*-modules, Heo [1999] proved analogues of these structure theorems for the case when the range algebra is an injective C^*-algebra.

We prove a new structure theorem for completely bounded maps (Theorems 3.2, 3.6). The idea is that to represent completely positive maps in Stinespring's theorem,

MSC2010: primary 46L07, 46L08; secondary 46L05.
Keywords: C^*-algebra, completely bounded map, Hilbert C^*-module, Stinespring's theorem, $*$-homomorphism.

one requires $*$-homomorphisms, and for a similar representation of completely bounded maps we need to consider homomorphisms which are not necessarily $*$-preserving. This is necessary because conjugation of a $*$-homomorphism by a bounded operator is always a completely positive map. However, it is possible to impose symmetry or some other additional restrictions on the homomorphisms in order to realize all completely bounded maps. That is what we do here.

To begin with, symmetries are self-adjoint unitaries. A homomorphism is symmetric if it preserves adjoints modulo conjugation by a symmetry. We define *regular homomorphisms* and *ternary homomorphisms*. We demonstrate that these homomorphisms are symmetric. Regular homomorphisms are essentially direct sums or direct integrals of ternary homomorphisms.

In Section 3A we prove that (Theorem 3.2) every completely bounded map from $\mathcal{A}$ into $\mathcal{B}(\mathcal{H})$ can be obtained by composing a regular homomorphism with a single bounded operator. We will call this the *regular representation theorem* for completely bounded maps. Moreover, we show that there exists a universal regular representation which can be used to generate all completely bounded maps from $\mathcal{A}$ into $\mathcal{B}(\mathcal{H})$. Since all representations (i.e., $*$-homomorphisms) are regular homomorphisms, we may consider the regular representation theorem as an immediate generalization of Stinespring's representation theorem for completely positive maps. We look at Hilbert C^*-module versions of these results. We also provide (Section 3B) new proofs of some results due to Paulsen and Suen [1985] and Heo [1999]. In Section 3C we study natural relationships between different representation theorems of completely bounded maps.

1A. ***Basic definitions and results:*** Throughout $\mathcal{H}$, $\mathcal{K}$ are complex Hilbert spaces. Our inner products are linear in the second variable and conjugate linear in the first variable. Typically $\mathcal{A}, \mathcal{B}, \mathcal{C}$ denote unital C^*-algebras. For an element a of a unital C^*-algebra $\mathcal{A}$, $\sigma(a)$ denote the spectrum of a. For a subset $\mathcal{X}$ of $\mathcal{A}$, the "commutant" is defined as $\mathcal{X}' = \{a \in \mathcal{A} : xa = ax \quad \text{for all } x \in \mathcal{X}\}$.

Suppose $\mathcal{A}, \mathcal{B}$ are C*-algebras. A multiplicative linear map $\pi : \mathcal{A} \to \mathcal{B}$ is called a *homomorphism*. By a $*$*-homomorphism* we mean a homomorphism which is also $*$-preserving, i.e., $\pi(a^*) = \pi(a)^*$ for all $a \in \mathcal{A}$. If a $*$-homomorphism is mapping into the algebra of all bounded operators on a Hilbert space, or into the algebra of all bounded adjointable operators on a Hilbert C^*-module we call it a representation. If $\mathcal{A}, \mathcal{B}$ are unital C^*-algebras and $\pi(1_{\mathcal{A}}) = 1_{\mathcal{B}}$, then π is said to be *unital*.

Recall that a linear map $\varphi : \mathcal{A} \to \mathcal{B}$ between two C^*-algebras is said to be

(i) a *completely positive map* (CP-map) if for all $n \geq 1$, $\varphi_n : M_n(\mathcal{A}) \to M_n(\mathcal{B})$ defined by $\varphi_n([a_{ij}]) = [\varphi(a_{ij})]$ is positive, i.e., $\varphi_n([a_{ij}]) \geq 0$ for all $0 \leq [a_{ij}] \in M_n(\mathcal{A})$,

(ii) a *completely bounded map* (CB-map) if $\|\varphi\|_{cb} := \sup\|\varphi_n\| < \infty$,

(iii) a *completely contractive map* (CC-map) if $\|\varphi\|_{cb} \leq 1$.

We let $CB(\mathcal{A}, \mathcal{B})$ denote the space of all CB-maps from $\mathcal{A}$ into $\mathcal{B}$. Given a map $\psi : \mathcal{A} \to \mathcal{B}$ define $\psi^* : \mathcal{A} \to \mathcal{B}$ by $\psi^*(a) := \psi(a^*)^*$ for all $a \in \mathcal{A}$. Note that if $\psi \in CB(\mathcal{A}, \mathcal{B})$, then $\psi^* \in CB(\mathcal{A}, \mathcal{B})$ with $\|\psi^*\|_{cb} = \|\psi\|_{cb}$. See [Paulsen 2002] for details on basic results.

Stinespring [1955] proved that if $\mathcal{A}$ is a unital C^*-algebra and $\varphi : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ is a CP-map, then there exists a triple $(\mathcal{K}, \pi, V)$, called Stinespring's dilation for φ, consisting of a unital representation $\pi : \mathcal{A} \to \mathcal{B}(\mathcal{K})$ of $\mathcal{A}$ on a Hilbert space $\mathcal{K}$ and a bounded linear map $V : \mathcal{H} \to \mathcal{K}$ with $\|\varphi\|_{cb} = \|V\|^2$ such that $\varphi(\,\cdot\,) = V^*\pi(\,\cdot\,)V$. Moreover, $\mathcal{K}$ can be chosen to be "minimal" in the sense that $\mathcal{K} = \overline{\operatorname{span}}\, \pi(\mathcal{A})V\mathcal{H}$, and in such case the triple is unique up to unitary equivalence. Conversely, $\varphi(\,\cdot\,) := V^*\pi(\,\cdot\,)V$ is a CP-map if π is a representation. If π is a J-homomorphism (but not a $*$-homomorphism), then $\varphi(\,\cdot\,) := V^*\pi(\,\cdot\,)V$ need not be CP. In fact it need not even be positive. But if π is in a special class of J-homomorphisms, called regular homomorphisms, then φ is a CB-map. Theorem 3.2 says that all CB-maps from $\mathcal{A}$ into $\mathcal{B}(\mathcal{H})$ are of this form.

Here we recall basics of Hilbert C^*-module theory we will use. Given a C^*-algebra $\mathcal{B}$, by an *inner product $\mathcal{B}$-module* we mean a complex vector space E with a right $\mathcal{B}$-module structure and a $\mathcal{B}$-valued inner product $\langle\,\cdot\,,\,\cdot\,\rangle : E \times E \to \mathcal{B}$ satisfying

(i) $\langle x, x\rangle \geq 0$,

(ii) $\langle x, x\rangle = 0$ if and only if $x = 0$,

(iii) $\langle x, \lambda y + z\rangle = \lambda\langle x, y\rangle + \langle x, z\rangle$,

(iv) $\langle x, y\rangle = \langle y, x\rangle^*$,

(v) $\langle x, yb\rangle = \langle x, y\rangle b$,

for all $x, y, z \in E,\ b \in \mathcal{B},\ \lambda \in \mathbb{C}$. If E is complete with respect to the norm $\|x\| := \|\langle x, x\rangle\|^{1/2}$, then E is called a *Hilbert $\mathcal{B}$-module*. An inner product $\mathcal{B}$-module for which the condition (ii) does not hold is called a *semi-inner product $\mathcal{B}$-module*. For a semi-inner product $\mathcal{B}$-module we have

$$\langle x, y\rangle^*\langle x, y\rangle \leq \|\langle x, x\rangle\|\langle y, y\rangle.$$

A linear map $T : E \to F$ between two Hilbert $\mathcal{B}$-modules is said to be *adjointable* if there exists a linear map $T^* : F \to E$ such that $\langle Tx, y\rangle = \langle x, T^*y\rangle$ for all $x \in E,\ y \in F$. Adjointable maps are bounded and $\mathcal{B}$-linear (i.e., $T(\lambda x_1 + x_2 b) = \lambda T(x_1) + T(x_2)b$ for all $\lambda \in \mathbb{C},\ x_i \in E,\ b \in \mathcal{B}$). But the converse may not be true. We denote the space of all bounded and adjointable maps from E into F by $\mathcal{B}^{\mathrm{a}}(E, F)$, which is a Banach space under the operator norm. In particular $\mathcal{B}^{\mathrm{a}}(E) := \mathcal{B}^{\mathrm{a}}(E, E)$ forms a C^*-algebra with natural algebraic operations.

In the following E, F are Hilbert C^*-modules. An inner product preserving map $V : E \to F$ is called an *isometry*, and a surjective isometry is called a *unitary*. An isometry with complemented range is adjointable. In general, closed submodules of Hilbert C^*-modules are complemented only if there is an adjointable projection onto that submodule.

If x, y are the elements of a Hilbert $\mathcal{B}$-module E we let $|x\rangle\langle y|$ denote the adjointable operator $z \mapsto x\langle y, z\rangle$. Note that $(|x\rangle\langle y|)^* = |y\rangle\langle x|$. We let $\mathcal{K}(E)$ denote the completion of $\mathcal{F}(E) := \overline{\mathsf{span}}\{|x\rangle\langle y| : x, y \in E\}$ (Always in the module context "span" would mean $\mathcal{B}$-linear span). A Hilbert $\mathcal{B}$-module E is said to be a *Hilbert $\mathcal{A}$-$\mathcal{B}$-module* if there exists (a left module action of $\mathcal{A}$ on E, i.e.,) a $*$-homomorphism $\vartheta : \mathcal{A} \to \mathcal{B}^{\mathrm{a}}(E)$ such that $\overline{\mathsf{span}}\{\vartheta(a)x : a \in \mathcal{A}, x \in E\} = E$ (or equivalently ϑ is unital if $\mathcal{A}$ is unital). Clearly any Hilbert $\mathcal{B}$-module E is a Hilbert $\mathcal{B}^{\mathrm{a}}(E)$-$\mathcal{B}$-module with left module action given by the identity map. In fact, E is a $\mathcal{K}(E)$-$\mathcal{B}$-module with inclusion map as the left module action. Given $x \in E$ we let $x^* \in \mathcal{B}^{\mathrm{a}}(E, \mathcal{B})$ denote the adjointable map $y \mapsto \langle x, y\rangle$. Note that $E^* := \overline{\mathsf{span}}\{x^* : x \in E\}$ forms a Hilbert $\mathcal{B}^{\mathrm{a}}(E)$-module with inner-product $\langle x_1^*, x_2^*\rangle := |x_1\rangle\langle x_2|$ and right module action $x^*a := (a^*x)^*$ for all $a \in \mathcal{B}^{\mathrm{a}}(E)$. Moreover, $\vartheta : \mathcal{B} \to \mathcal{B}^{\mathrm{a}}(E^*)$ given by $\vartheta(b)x^* := (xb^*)^*$ is a $*$-homomorphism such that $E^* = \overline{\mathsf{span}}\{\vartheta(\mathcal{B})E^*\}$ so that E^* forms a Hilbert $\mathcal{B}$-$\mathcal{B}^{\mathrm{a}}(E)$-module. If F is a Hilbert $\mathcal{B}$-$\mathcal{C}$-module, then by $\mathcal{B}^{\mathrm{a,bil}}(F)$ we mean the space of all adjointable, bilinear (i.e., preserves both left and right module actions) maps on F.

Suppose E is a Hilbert $\mathcal{B}$-module and F is a Hilbert $\mathcal{B}$-$\mathcal{C}$-module with left action given by the $*$-homomorphism $\vartheta : \mathcal{B} \to \mathcal{B}^{\mathrm{a}}(F)$. We let $E \odot_\vartheta F$ (or $E \odot_{\mathcal{B}} F$ or simply $E \odot F$) denote the completion of the algebraic tensor product $E \otimes F$ with respect to the $\mathcal{C}$-valued semi-inner product

$$\langle x_1 \otimes y_1, x_2 \otimes y_2\rangle := \langle y_1, \vartheta(\langle x_1, x_2\rangle)y_2\rangle.$$

We let $x \odot y \in E \odot_\vartheta F$ denote the equivalence class containing $x \otimes y \in E \otimes F$. Note that $E \odot_\vartheta F$ forms a Hilbert $\mathcal{C}$-module with right module action $(x \odot y)c := x \odot yc$. In addition if E has a left action of $\mathcal{A}$ via a $*$-homomorphism $\vartheta' : \mathcal{A} \to \mathcal{B}^{\mathrm{a}}(E)$, then $E \odot_\vartheta F$ also has a left module action $\tilde{\vartheta} : \mathcal{A} \to \mathcal{B}^{\mathrm{a}}(E \odot_\vartheta F)$ given by $\tilde{\vartheta}(a)(x \odot y) := \vartheta'(a)x \odot y$. Thus $E \odot_\vartheta F$ forms a Hilbert $\mathcal{A}$-$\mathcal{C}$-module. Note that $\|xb \odot y - x \odot \vartheta(b)y\| = 0$ so that $xb \odot y = x \odot \vartheta(b)y$ for all $x \in E, y \in F, b \in \mathcal{B}$. We may identify $\mathcal{B} \odot_\vartheta F = F$ via the unitary isomorphism $b \odot y \mapsto \vartheta(b)y$. If $a \in \mathcal{B}^{\mathrm{a}}(E)$ and $\mathfrak{a} \in \mathcal{B}^{\mathrm{a,bil}}(F)$, then $a \odot I_F \in \mathcal{B}^{\mathrm{a}}(E \odot F)$ and $I_E \odot \mathfrak{a} \in \mathcal{B}^{\mathrm{a}}(E \odot F)$ are the maps defined by $x \odot y \mapsto ax \odot y$ and $x \odot y \mapsto x \odot \mathfrak{a}y$, respectively.

Suppose ρ is a representation of a C^*-algebra $\mathcal{B}$ on a Hilbert space $\mathcal{G}$. Given a Hilbert $\mathcal{B}$-module E, by considering $\mathcal{G}$ as a Hilbert $\mathcal{B}$-$\mathbb{C}$-module with left module action given by ρ, we let $\eta : \mathcal{B}^{\mathrm{a}}(E) \to \mathcal{B}(E \odot_\rho \mathcal{G})$ denote the unital $*$-homomorphism $a \mapsto a \odot I_{\mathcal{G}}$, that is, $\eta(a)(x \odot g) := ax \odot g$ for all $a \in \mathcal{B}^{\mathrm{a}}(E),\ x \in E,\ g \in \mathcal{G}$.

We refer to [Lance 1994; Paschke 1973; Skeide 2000] for the basic theory of Hilbert C^*-modules.

2. Symmetric homomorphisms

In this section we study homomorphisms between C^*-algebras which are not necessarily $*$-homomorphisms. These homomorphisms need not be contractive.

2A. *Symmetries.*

Definition 2.1. An element J in a unital C^*-algebra $\mathcal{B}$ satisfying $J = J^* = J^{-1}$ is called a *symmetry*. A homomorphism $\tau : \mathcal{A} \to \mathcal{B}$ is called a *J-homomorphism* if $J\tau(a)^*J = \tau(a^*)$ for all $a \in \mathcal{A}$. A homomorphism $\tau : \mathcal{A} \to \mathcal{B}$ is said to be a *symmetric homomorphism* if τ is a J-homomorphism for some symmetry $J \in \mathcal{B}$.

Clearly all $*$-homomorphisms are symmetric homomorphisms. But the converse is not true. For example, $\tau : \mathbb{C} \to \mathcal{B}(\mathbb{C}^2)$ given by $a \mapsto \left[\begin{smallmatrix} a/2 & a/4 \\ a & a/2 \end{smallmatrix}\right]$ is a J-homomorphism where $J = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$. But τ is not $*$-preserving. It is easily seen that τ is neither positive nor contractive.

Example 2.2. Define $\tau : M_2(\mathbb{C}) \to M_2(\mathbb{C})$ by $\tau(a) = sas^{-1}$, where $s = \left[\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right]$. Clearly τ is a homomorphism. But it is not a symmetric homomorphism. For, suppose $J \in M_2(\mathbb{C})$ is a symmetry; then $J\tau(a^*)J = \tau(a)^*$ implies that $(s^*Js)a^* = a^*(s^*Js)$ for all $a \in M_2(\mathbb{C})$. Hence there exists $\lambda \in \mathbb{C}$ such that $s^*Js = \lambda I$, so that $J = \lambda(ss^*)^{-1}$ which is not a symmetry. Exactly the same situation arises when s is an invertible element which is not a scalar multiple of a unitary.

The next proposition answers the question of uniqueness of symmetry J in a symmetric homomorphism.

Proposition 2.3. *Suppose $\tau : \mathcal{A} \to \mathcal{B}$ is a symmetric homomorphism. If there exist symmetries $J, J' \in \mathcal{B}$ such that τ is both a J- and J'-homomorphism, then there exists a unitary $U \in \tau(\mathcal{A})' \subseteq \mathcal{B}$ such that $J = UJU$ and $J' = JU$.*

Proof. We have $J\tau(a)J = \tau(a^*)^* = J'\tau(a)J'$ for all $a \in \mathcal{A}$. Multiplying on the left and right side of this equation by J and J' respectively, we get $\tau(a)JJ' = JJ'\tau(a)$ for all $a \in \mathcal{A}$. Hence there exists a $U \in \tau(\mathcal{A})'$ such that $JJ' = U$. Clearly $U^*U = I = UU^*$ and $J' = JU$. Further, $(J')^* = J'$, yields $J = UJU$. □

Now we show that given any homomorphism, we can associate a symmetry J in a very natural way. The usefulness of this symmetry will be seen later.

Proposition 2.4. *Suppose $\tau : \mathcal{A} \to \mathcal{B}$ is a homomorphism. Then there exists a symmetry $J \in \mathcal{B}$ such that: $J\tau(a)J = \tau(a)^*$ for all $a \in \mathcal{A}$ satisfying $\tau(a)^*\tau(1) = \tau(1)^*\tau(a)$ and $\tau(a)\tau(1)^* = \tau(1)\tau(a)^*$.*

Proof. Suppose $T := \tau(1) = R + iS$ is the cartesian decomposition of $\tau(1)$. Since $\tau(1)^2 = \tau(1)$ we have $R^2 - S^2 = R$ and $RS + SR = S$. Also since $R = R^*$ with $R^2 - R = S^2 \geq 0$ we have $\sigma(R) \subseteq \mathbb{R} \setminus (0, 1)$. Define $\hat{f} : \mathbb{R} \to \mathbb{R}$ by

$$\hat{f}(t) = \begin{cases} -1 & \text{if } t \leq 0, \\ 2t - 1 & \text{if } 0 < t < 1, \\ 1 & \text{if } t \geq 1, \end{cases}$$

which is clearly a continuous function. Set $f = \hat{f}|_{\sigma(R)} \in C(\sigma(R)) \cong C^*(\{1, R\})$ and $J = f(R) \in \mathcal{B}$. Clearly $J^2 = I$ and $J = J^*$.

Step 1: First we prove that $JTJ = T^*$. It is enough to show that $JR = RJ$ and $JS = -SJ$. Clearly $JR = RJ$. Now

$$RS + SR = S \Longrightarrow RS = S(1 - R) \Longrightarrow R^n S = S(1 - R)^n \quad \text{for all} \quad n \geq 0.$$

Approximating f by polynomials, from $C(\sigma(R))$ we get $f(R)S = Sf(1 - R)$. But since

$$\hat{f}(1 - x) = \begin{cases} 1 & \text{if } x < 1, \\ -1 & \text{if } x \geq 1, \end{cases}$$

we have $\hat{f}(1 - x) = -\hat{f}(x)$ for all $x \in \sigma(R)$, and hence $f(1 - R) = -f(R)$. Thus

$$JS = f(R)S = Sf(1 - R) = -Sf(R) = -SJ.$$

Step 2: Fix $a \in \mathcal{A}$. Let $\tau(a) = X + iY$ be the Cartesian decomposition of $\tau(a)$. Then

$$X = \tfrac{1}{2}(\tau(a) + \tau(a)^*) = \tfrac{1}{2}(\tau(a)T + (T\tau(a))^*) = XR - YS,$$
$$Y = \tfrac{1}{2i}(\tau(a) - \tau(a)^*) = \tfrac{1}{2i}(T\tau(a) - (\tau(a)T)^*) = RY + SX.$$

Since X and Y are self-adjoint we have

$$XR - YS = X = RX - SY, \tag{2-1}$$

$$XS + YR = Y = RY + SX. \tag{2-2}$$

Now if $\tau(a)^*\tau(1) = \tau(1)^*\tau(a)$ and $\tau(a)\tau(1)^* = \tau(1)\tau(a)^*$, then we get

$$(XR + YS) + i(XS - YR) = (RX + SY) + i(RY - SX),$$
$$(XR + YS) - i(XS - YR) = (RX + SY) - i(RY - SX).$$

Adding and subtracting above two equations we get

$$XR + YS = RX + SY, \tag{2-3}$$

$$XS - YR = RY - SX. \tag{2-4}$$

Adding equations (2-1) and (2-3) we get $XR = RX$, hence $Xf(R) = f(R)X$, i.e., $XJ = JX$. Adding equations (2-2) and (2-4) we get $XS = RY$. Now since $JS = -SJ$, from equation (2-2), we have

$$YJ = (XS+YR)J = (XS+SX)J = -J(XS+SX) = -J(XS+YR) = -JY.$$

Now a direct computation shows that $J\tau(a)J = \tau(a)^*$. □

Now we introduce two subfamilies of symmetric homomorphisms, and study their structure and properties. Later, in terms of these maps we prove a new structure theorem for completely bounded maps. Before proceeding, we give a definition.

Definition 2.5. A linear map $\tau : \mathcal{A} \to \mathcal{B}^{\mathrm{a}}(E)$ is said to be

(i) *nondegenerate* if $\overline{\mathsf{span}}\{\tau(a)x : a \in \mathcal{A},\ x \in E\} = E$;

(ii) $*$-*nondegenerate* if $\overline{\mathsf{span}}\{\tau(a_1)x_1, \tau(a_2)^*x_2 : x_i \in E,\ a_i \in \mathcal{A},\ i = 1, 2\} = E$.

Remark 2.6. (i) If τ is a homomorphism, then $\tau(a) = \tau(1)\tau(a)$ and $\tau(a)^* = \tau(1)^*\tau(a)^*$, therefore τ is $*$-nondegenerate if and only if

$$\overline{\mathsf{span}}\{\tau(1)x, \tau(1)^*x' : x, x' \in E\} = E.$$

A $*$-homomorphism $\tau : \mathcal{A} \to \mathcal{B}^{\mathrm{a}}(E)$ is $*$-nondegenerate if and only if τ is nondegenerate or equivalently τ is unital.

(ii) Suppose a Hilbert space $\mathcal{H}$ plays the role of E. Then a linear map $\tau : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ is $*$-nondegenerate if and only if $\{h \in \mathcal{H} : \tau(a)h = 0 = \tau(a)^*h$ for all $a \in \mathcal{A}\} = \{0\}$. If τ is a homomorphism, then the above conditions are equivalent to the condition $\{h \in \mathcal{H} : \tau(1)h = 0 = \tau(1)^*h\} = \{0\}$.

2B. *Regular homomorphisms.*

Definition 2.7. A map $\tau : \mathcal{A} \to \mathcal{B}$ is said to be *regular* if $\tau(u)^*\tau(u) = \tau(1)^*\tau(1)$ and $\tau(u)\tau(u)^* = \tau(1)\tau(1)^*$ for all unitary $u \in \mathcal{A}$.

Example 2.8. (i) The map $\tau : M_2(\mathbb{C}) \to M_2(\mathbb{C})$ defined in Example 2.2 is a homomorphism but it is not regular. Because $\tau(u)^*\tau(u) \neq \tau(1)^*\tau(1)$ for the unitary $u = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

(ii) All $*$-homomorphisms are regular. But the converse is not true. For example $\tau : \mathcal{A} \to M_2(\mathcal{A})$ given by $\tau(a) = \begin{bmatrix} a & 0 \\ a & 0 \end{bmatrix}$ is a regular homomorphism but it is not $*$-preserving.

Proposition 2.9. *Suppose $\tau : \mathcal{A} \to \mathcal{B}$ is a unital homomorphism. Then τ is regular if and only if it is $*$-preserving.*

Proof. Suppose τ is a unital regular homomorphism. Then for all unitary $u \in \mathcal{A}$,

$$\tau(u)^* = \tau(u)^*\tau(uu^*) = \tau(u)^*\tau(u)\tau(u^*) = \tau(1)^*\tau(1)\tau(u^*) = \tau(u^*).$$

Since any $a \in \mathcal{A}$ is a linear combination of at most four unitaries it follows that $\tau(a)^* = \tau(a^*)$. The converse is obvious. □

The following theorem says that all regular homomorphisms preserve conjugation $*$ up to a symmetry. This is one of the reasons to study regular homomorphisms.

Theorem 2.10. *Every regular homomorphism $\tau : \mathcal{A} \to \mathcal{B}$ is symmetric.*

Proof. Suppose $J \in \mathcal{B}$ is the symmetry given by Proposition 2.4. Since τ is regular, given any unitary $u \in \mathcal{A}$, we have

$$\tau(u)^*\tau(1) = \tau(u)^*\tau(u)\tau(u^*) = \tau(1)^*\tau(1)\tau(u^*) = \tau(1)^*\tau(u^*).$$

Since u^* is also a unitary, we also get $\tau(u^*)^*\tau(1) = \tau(1)^*\tau(u)$. In a similar fashion, by regularity,

$$\tau(1)\tau(u)^* = \tau(u^*)\tau(u)\tau(u)^* = \tau(u^*)\tau(1)\tau(1)^* = \tau(u^*)\tau(1)^*,$$

and replacing u by u^*, $\tau(1)\tau(u^*)^* = \tau(u)\tau(1)^*$. So if we let $u_1 := u + u^*$ and $u_2 := u - u^*$, then $\tau(u_1)^*\tau(1) = \tau(1)^*\tau(u_1)$ and $\tau(u_1)\tau(1)^* = \tau(1)\tau(u_1)^*$, so that Proposition 2.4 is applicable and we get $J\tau(u_1)J = \tau(u_1)^*$. On the other hand, as $u_2^* = -u_2$ and Proposition 2.4 can be applied to iu_2. Then we get $J\tau(u_2)J = -\tau(u_2)^*$. Thus $J\tau(u^*)J = \tau(u)^*$ for every unitary u. Since every element in a C^*-algebra can be written as a linear combination of at most four unitaries it follows that $J\tau(a^*)J = \tau(a)^*$ for all $a \in \mathcal{A}$. □

Example 2.11. Suppose $\tau : \mathcal{A} \to \mathcal{B}$ is a J-homomorphism for some symmetry $J \in \mathcal{B}$. Then it can be seen that $\tau(u)^*\tau(u) = \tau(1)^*\tau(1)$ for all unitary $u \in \mathcal{A}$ if and only if $\tau(u)\tau(u)^* = \tau(1)\tau(1)^*$ for all unitary $u \in \mathcal{A}$. But for general homomorphisms this is not true. For example, let $\mathcal{A} = M_2(\mathbb{C})$ and let $v = \left[\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right] \in \mathcal{A}$. Define a homomorphism $\tau : \mathcal{A} \to M_2(\mathcal{A})$ by $\tau(a) = \left[\begin{smallmatrix} a & av \\ 0 & 0 \end{smallmatrix}\right]$. Then τ satisfies $\tau(u)^*\tau(u) = \tau(1)^*\tau(1)$ for all unitary u. But $\tau(u)\tau(u)^* \neq \tau(1)\tau(1)^*$ for the unitary $u = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$.

Example 2.12. Suppose $v \in \mathcal{A} = \mathcal{B}(\mathcal{H})$ is a nonscalar unitary. Define a homomorphism $\tau : \mathcal{A} \to M_2(\mathcal{A})$ by $\tau(a) = \left[\begin{smallmatrix} a & \sqrt{3}(va-av) \\ 0 & a \end{smallmatrix}\right]$. Then τ is symmetric with symmetry $J = \frac{1}{2}\left[\begin{smallmatrix} -1 & \sqrt{3}v \\ \sqrt{3}v^* & 1 \end{smallmatrix}\right]$. But τ is not regular since $\tau(u)^*\tau(u) \neq \tau(1)^*\tau(1)$ for any unitary $u \in \mathcal{A}$ not commuting with v.

Proposition 2.13. *Let $\tau : \mathcal{A} \to \mathcal{B}$ be a homomorphism. Then τ is regular if and only if for all $a, b \in \mathcal{A}$,*

(i) $\tau(a)^*\tau(b) = \tau(1)^*\tau(a^*b) = \tau(b^*a)^*\tau(1)$,

(ii) $\tau(a)\tau(b)^* = \tau(ab^*)\tau(1)^* = \tau(1)\tau(ba^*)^*$.

Proof. Assume that τ is regular. Suppose u is a unitary. Then for any $b, c \in \mathcal{A}$,

$$\tau(u)^*\tau(b) = \tau(u)^*\tau(u)\tau(u^*b) = \tau(1)^*\tau(1)\tau(u^*b) = \tau(1)^*\tau(u^*b), \tag{2-5}$$

$$\tau(c)\tau(u)^* = \tau(cu^*)\tau(u)\tau(u)^* = \tau(cu^*)\tau(1)\tau(1)^* = \tau(cu^*)\tau(1)^*. \tag{2-6}$$

Since any element in $\mathcal{A}$ is a linear combination of at most four unitaries from equations (2-5) and (2-6) we have $\tau(a)^*\tau(b) = \tau(1)^*\tau(a^*b)$ and $\tau(c)\tau(d)^* = \tau(cd^*)\tau(1)^*$, respectively for all $a, d \in \mathcal{A}$. Taking adjoints of these equalities proves (i) and (ii). The converse is obvious. $\square$

Note that given a (unital) $*$-homomorphism $\vartheta : \mathcal{A} \to \mathcal{B}^a(E)$ and an idempotent operator $T \in \vartheta(\mathcal{A})' \subseteq \mathcal{B}^a(E)$ the map $a \mapsto \vartheta(a)T$ always defines a bounded ($*$-nondegenerate) regular homomorphism from $\mathcal{A}$ into $\mathcal{B}^a(E)$. Now we will prove that all $*$-nondegenerate regular homomorphisms can be represented this way.

Theorem 2.14. *Suppose $\tau : \mathcal{A} \to \mathcal{B}^a(E)$ is a $*$-nondegenerate, regular homomorphism. Then there exists a unique unital $*$-homomorphism $\vartheta : \mathcal{A} \to \mathcal{B}^a(E)$ such that $\tau(a) = \vartheta(a)\tau(1) = \tau(1)\vartheta(a)$ for all $a \in \mathcal{A}$. Consequently τ is completely bounded with $\|\tau\|_{cb} = \|\tau(1)\|$.*

Proof. If τ is unital, then it is $*$-preserving and in such case we take $\vartheta = \tau$. Otherwise let $E_0 = \mathsf{span}\{\tau(\mathcal{A})E, \tau(\mathcal{A})^*E\} = \mathsf{span}\{\tau(1)E, \tau(1)^*E\}$. Now for each unitary $u \in \mathcal{A}$, define $\vartheta(u) : E_0 \to E_0$ by $\vartheta(u)\big(\sum_i \tau(1)x_i + \tau(1)^*y_i\big) = \sum_i (\tau(u)x_i + \tau(u^*)^*y_i)$ for all $x_i, y_i \in E$. Since

$$
\begin{aligned}
&\Big\|\vartheta(u)\Big(\sum_i \tau(1)x_i + \tau(1)^*y_i\Big)\Big\|^2 \\
&\quad= \Big\|\Big\langle \sum_i \tau(u)x_i + \tau(u^*)^*y_i, \sum_j \tau(u)x_j + \tau(u^*)^*y_j \Big\rangle\Big\| \\
&\quad= \Big\|\sum_{i,j} (\langle x_i, \tau(u)^*\tau(u)x_j\rangle + \langle x_i, \tau(u)^*\tau(u^*)^*y_j\rangle \\
&\qquad\qquad + \langle y_i, \tau(u^*)\tau(u)x_j\rangle + \langle y_i, \tau(u^*)\tau(u^*)^*y_j\rangle)\Big\| \\
&\quad= \Big\|\sum_{i,j} (\langle x_i, \tau(1)^*\tau(1)x_j\rangle + \langle x_i, \tau(1)^*y_j\rangle + \langle y_i, \tau(1)x_j\rangle + \langle y_i, \tau(1)\tau(1)^*y_j\rangle)\Big\| \\
&\quad= \Big\|\Big\langle \sum_i \tau(1)x_i + \tau(1)^*y_i, \sum_j \tau(1)x_j + \tau(1)^*y_j \Big\rangle\Big\| \\
&\quad= \Big\|\sum_i \tau(1)x_i + \tau(1)^*y_i\Big\|^2,
\end{aligned}
$$

we see that $\vartheta(u)$ is well defined and norm preserving on E_0. It is also $\mathcal{B}$-linear. Hence $\vartheta(u)$ is an isometry. Note that $\mathrm{ran}(\vartheta(u)) = E_0$ so that $\vartheta(u) : E_0 \to E_0$ is a unitary. Now given a linear combination of unitaries, say $a = \sum \lambda_i u_i \in \mathcal{A}$, we define $\vartheta(a) := \sum \lambda_i \vartheta(u_i)$. Note that if $\sum \lambda_i u_i = 0$, then

$$\begin{aligned}
&\vartheta\Big(\sum_i \lambda_i u_i\Big)\Big(\sum_j \tau(1)x_j + \tau(1)^* y_j\Big) \\
&\qquad\qquad = \sum_i \lambda_i \vartheta(u_i)\Big(\sum_j \tau(1)x_j + \tau(1)^* y_j\Big) \\
&\qquad\qquad = \sum_{i,j} (\lambda_i \tau(u_i)x_j + \lambda_i \tau(u_i^*)^* y_j) \\
&\qquad\qquad = \sum_j \Big(\tau\Big(\sum_i \lambda_i u_i\Big)\tau(1)x_j + \tau\Big(\sum_i \bar{\lambda}_i u_i^*\Big)^* \tau(1)^* y_j\Big) \\
&\qquad\qquad = 0,
\end{aligned}$$

so that $\vartheta\big(\sum \lambda_i u_i\big) = 0$. Thus the definition of $\vartheta(a) : E_0 \to E_0$ is independent of the choice of linear combination $\sum \lambda_i u_i$. Note that if $a = \sum \lambda_i u_i \in \mathcal{A}$, then

$$\begin{aligned}
(2\text{-}7)\quad &\vartheta(a)(\tau(a_1)x_1 + \tau(a_2)^* x_2) \\
&\qquad = \sum \lambda_i \vartheta(u_i)\tau(1)\tau(a_1)x_1 + \sum \lambda_i \vartheta(u_i)\tau(1)^*\tau(a_2)^* x_2 \\
&\qquad = \sum \lambda_i \tau(u_i a_1)x_1 + \sum \lambda_i \tau(a_2 u_i^*)^* x_2 \\
&\qquad = \tau\Big(\sum \lambda_i u_i a_1\Big)x_1 + \tau\Big(a_2 \sum \bar{\lambda}_i u_i^*\Big)^* x_2 \\
&\qquad = \tau(a a_1)x_1 + \tau(a_2 a^*)^* x_2
\end{aligned}$$

for all $x_i \in E$. Now it follows that $\vartheta(1) = I_{E_0}$, $\vartheta(a+b) = \vartheta(a) + \vartheta(b)$ and $\vartheta(a)\vartheta(b) = \vartheta(ab)$ for all $a, b \in \mathcal{A}$. Since any element in $\mathcal{A}$ is a linear combination of at most four unitaries and $\|\vartheta(u)\| \le 1$ for all unitary $u \in \mathcal{A}$, we have $\|\vartheta(a)\| < \infty$ for all $a \in \mathcal{A}$. Hence each $\vartheta(a) : E_0 \to E_0$ can be extended to a bounded operator, again denoted by $\vartheta(a)$, on $E = \overline{E_0}$. Also, from Proposition 2.13, for all $a \in \mathcal{A}$, $x_i \in E$ we have

$$\begin{aligned}
&\langle \vartheta(a)(\tau(1)x_1 + \tau(1)^* x_2), \tau(1)x_3 + \tau(1)^* x_4\rangle \\
&\quad = \langle x_1, \tau(a)^*\tau(1)x_3\rangle + \langle x_1, \tau(a)^*\tau(1)^* x_4\rangle + \langle x_2, \tau(a^*)\tau(1)x_3\rangle + \langle x_2, \tau(a^*)\tau(1)^* x_4\rangle \\
&\quad = \langle x_1, \tau(1)^*\tau(a^*)x_3\rangle + \langle x_1, \tau(a)^* x_4\rangle + \langle x_2, \tau(a^*)x_3\rangle + \langle x_2, \tau(1)\tau(a)^* x_4\rangle \\
&\quad = \langle \tau(1)x_1 + \tau(1)^* x_2, \tau(a^*)x_3 + \tau(a)^* x_4\rangle \\
&\quad = \langle \tau(1)x_1 + \tau(1)^* x_2, \vartheta(a^*)(\tau(1)x_3 + \tau(1)^* x_4)\rangle,
\end{aligned}$$

so that $\langle \vartheta(a)x, x'\rangle = \langle x, \vartheta(a^*)x'\rangle$ for all $x, x' \in E$, that is, $\vartheta(a)$ is adjointable with $\vartheta(a)^* = \vartheta(a^*)$. Thus $a \mapsto \vartheta(a)$ defines a unital $*$-homomorphism $\vartheta : \mathcal{A} \to \mathcal{B}^{a}(E)$. Moreover, from (2-7) we have $\vartheta(a)\tau(1)x = \tau(a)x$ and $\vartheta(a)\tau(1)^* x = \tau(a^*)^* x$ for all $x \in E$. Hence we get $\vartheta(a)\tau(1) = \tau(a) = \tau(1)\vartheta(a)$ for all $a \in \mathcal{A}$, therefore $\|\tau\|_{cb} \le \|\vartheta\|_{cb}\|\tau(1)\| \le \|\tau(1)\|$.

Uniqueness: If $\vartheta' : \mathcal{A} \to \mathcal{B}^a(E)$ is any other such $*$-homomorphism, then

$$\vartheta'(a)\tau(1)x = \tau(a)x = \vartheta(a)\tau(1)x,$$
$$\vartheta'(a)\tau(1)^*x' = (\tau(1)\vartheta'(a^*))^*x' = \tau(a^*)^*x' = (\tau(1)\vartheta(a^*))^*x' = \vartheta(a)\tau(1)^*x'.$$

Hence $\vartheta(a)x = \vartheta'(a)x$ for all $a \in \mathcal{A},\ x \in E = \overline{\mathsf{span}}\{\tau(1)E, \tau(1)^*E\}$. □

Remark 2.15. Suppose $\tau : \mathcal{A} \to \mathcal{B}^a(E)$ is a regular homomorphism, not necessarily $*$-nondegenerate. Then also $\vartheta : \mathcal{A} \to \mathcal{B}^a(\overline{E_0})$ given as in the proof is a well-defined unital $*$-homomorphism. Note that $\overline{E_0}$ is a $\tau(a)$-reducing closed $\mathcal{B}$-submodule of E. Thus the proof says that: *If $\tau : \mathcal{A} \to \mathcal{B}^a(E)$ is a regular homomorphism, then there exists a closed $\mathcal{B}$-submodule $E_0 \subseteq E$, which reduces all $\tau(a)$; and a unique unital $*$-homomorphism $\vartheta : \mathcal{A} \to \mathcal{B}^a(E_0)$ such that $\tau(a)|_{E_0} = \tau(1)\vartheta(a) = \vartheta(a)\tau(1)|_{E_0}$ for all $a \in \mathcal{A}$. Moreover, if E_0 is complemented in E, then $\tilde{\vartheta} = \left[\begin{smallmatrix}\vartheta & 0\\ 0 & 0\end{smallmatrix}\right] : \mathcal{A} \to \mathcal{B}^a(E)$ is a $*$-homomorphism such that $\tau(a) = \tau(1)\tilde{\vartheta}(a) = \tilde{\vartheta}(a)\tau(1)$.*

Corollary 2.16. *Suppose $\tau : \mathcal{A} \to \mathcal{B}^a(E)$ is a $*$-nondegenerate regular homomorphism with $\tau(1) = \tau(1)^*$. Then τ is a $*$-homomorphism.*

Proof. Suppose $\vartheta : \mathcal{A} \to \mathcal{B}^a(E)$ is the unique unital $*$-homomorphism such that $\tau(a) = \vartheta(a)\tau(1) = \tau(1)\vartheta(a)$. Then

$$\tau(a)^* = (\vartheta(a)\tau(1))^* = \tau(1)^*\vartheta(a^*) = \tau(1)\vartheta(a^*) = \tau(a^*). \qquad \square$$

Note that a unital C^*-algebra $\mathcal{B}$ is a Hilbert $\mathcal{B}$-module with inner product $\langle b, b'\rangle := b^*b'$. Moreover, $\mathcal{B} \cong \mathcal{B}^a(\mathcal{B})$ as C^*-algebras under the unital isometric $*$-isomorphism $b \mapsto T_b$ where $T_b \in \mathcal{B}^a(\mathcal{B})$ is given by $T_b(b') = bb'$ for all $b \in \mathcal{B}$. (Note that adjointable maps preserves module action so that $T(b) = T(1b) = T(1)b$ for all $b \in \mathcal{B},\ T \in \mathcal{B}^a(\mathcal{B})$.) So given a linear map $\tau : \mathcal{A} \to \mathcal{B}$ we say that τ is $*$-nondegenerate if $\overline{\mathsf{span}}\{\tau(\mathcal{A})\mathcal{B}, \tau(\mathcal{A})^*\mathcal{B}\} = \mathcal{B}$.

Corollary 2.17. *Suppose $\tau : \mathcal{A} \to \mathcal{B}$ is a $*$-nondegenerate regular homomorphism. Then there exists a unique unital $*$-homomorphism $\vartheta : \mathcal{A} \to \mathcal{B}$ such that $\tau(a) = \vartheta(a)\tau(1) = \tau(1)\vartheta(a)$ for all $a \in \mathcal{A}$. Consequently τ is completely bounded with $\|\tau\|_{cb} = \|\tau(1)\|$. Moreover, if $\tau(1) = \tau(1)^*$, then τ is a $*$-homomorphism.*

Example 2.18. Let $\mathcal{A}$ be the C^*-algebra of continuous functions on the interval $[0, 1]$. Let $\mathcal{B} = M_2(\mathcal{A})$ and let $E = \mathcal{B}$, with usual inner product, so that $B^a(E) \cong \mathcal{B}$. Let $g : [0, 1] \to \mathbb{C}$ be the function defined by $g(x) = x$ for all $x \in [0, 1]$. Define $\tau : \mathcal{A} \to B^a(E)$ by

$$\tau(f) = \begin{bmatrix} f & gf \\ 0 & 0 \end{bmatrix}.$$

Then it is easily seen that τ is a regular homomorphism, E_0 defined as above is

$$\begin{aligned} E_0 &= \mathsf{span}\{\tau(\mathcal{A})E, \tau(\mathcal{A})^*E\} \\ &= \left\{ \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix} : f_{ij} \in \mathcal{B},\ 1 \le i,\ j \le 2, \quad \text{where } f_{2j}(0) = 0 \right\}, \end{aligned}$$

and is not complemented in E.

In the following E, F are Hilbert C^*-modules over possibly different C^*-algebras $\mathcal{B}$, $\mathcal{C}$ respectively. We wish to obtain a structure theorem for strictly continuous regular homomorphisms from $\mathcal{B}^{\mathrm{a}}(E)$ to $\mathcal{B}^{\mathrm{a}}(F)$. Recall (see [Lance 1994]) that a net $\{a_\alpha\}$ in $\mathcal{B}^{\mathrm{a}}(E)$ is said to converge *strictly* (or $*$-*strongly*) to $a \in \mathcal{B}^{\mathrm{a}}(E)$ if, for all $x, x' \in E$, the nets $\{a_\alpha x\}$ and $\{a_\alpha^* x'\}$ converge to ax and a^*x', respectively. Note that $\{a_\alpha\}$ converges to a strictly if and only if $\{a_\alpha^*\}$ converges to a^* strictly. A bounded linear map $\tau : \mathcal{B}^{\mathrm{a}}(E) \to \mathcal{B}^{\mathrm{a}}(F)$ is said to be *strict* if $\tau(a_\alpha)$ converges strictly to $\tau(a)$ in $\mathcal{B}^{\mathrm{a}}(F)$ whenever a net $\{a_\alpha\}$ in the unit ball of $\mathcal{B}^{\mathrm{a}}(E)$ converges strictly to $a \in \mathcal{B}^{\mathrm{a}}(E)$.

Remark 2.19. If the map τ given in Theorem 2.14 is also a strict map (so bounded by definition), then ϑ is a strict map. For, suppose $\{a_\alpha\}$ is a net in the unit ball of $\mathcal{A} = \mathcal{B}^{\mathrm{a}}(\mathcal{A})$ which converges strictly to $a \in \mathcal{A}$. Then for all $x_1, x_2 \in E$ we have

$$\vartheta(a_\alpha)\tau(1)x_1 = \tau(a_\alpha)x_1 \xrightarrow{\alpha} \tau(a)x_1 = \vartheta(a)\tau(1)x_1,$$

and

$$\begin{aligned} \vartheta(a_\alpha)\tau(1)^*x_2 = (\tau(1)\vartheta(a_\alpha^*))^*x_2 = \tau(a_\alpha^*)^*x_2 \xrightarrow{\alpha} \tau(a^*)^*x_2 &= (\tau(1)\vartheta(a^*))^*x_2 \\ &= \vartheta(a)\tau(1)^*x_2, \end{aligned}$$

so that $\{\vartheta(a_\alpha)x\}$ and $\{\vartheta(a_\alpha)^*x'\}$ converge to $\vartheta(a)x$ and $\vartheta(a)^*x'$ respectively, for all $x, x' \in E$. Thus ϑ is a strict unital $*$-homomorphism. In particular, if $\tau : \mathcal{B}^{\mathrm{a}}(E) \to \mathcal{B}^{\mathrm{a}}(F)$ is a $*$-nondegenerate, strict, regular homomorphism then the strict, unital $*$-homomorphism $\vartheta : \mathcal{B}^{\mathrm{a}}(E) \to \mathcal{B}^{\mathrm{a}}(F)$ has a factorization $\vartheta(a) = U(a \odot I)U^*$ where $U : E \odot F_\vartheta \to F$ is a unitary on a suitable Hilbert $\mathcal{B}$-$\mathcal{C}$-module F_ϑ (see [Muhly et al. 2006, Theorem 1.4]). In fact, if we consider F as a Hilbert $\mathcal{B}^{\mathrm{a}}(E)$-$\mathcal{C}$-module with left action given by ϑ, then $F_\vartheta = E^* \odot_\vartheta F$ and $U \in \mathcal{B}^{\mathrm{a,bil}}(E \odot F_\vartheta, F)$. A more generalized version says that: if $\vartheta : \mathcal{B}^{\mathrm{a}}(E) \to \mathcal{B}^{\mathrm{a}}(F)$ is a strict CP-map, then $\vartheta(a) = W(a \odot I)W^*$ for some bounded adjointable operator $W : E \odot F_\vartheta \to F$ on a suitable Hilbert $\mathcal{B}$-$\mathcal{C}$-module F_ϑ (see [Skeide and Sumesh 2014, Theorem 3.2]). In this case, $F_\vartheta = E^* \odot \mathcal{E} \odot F$ where $\mathcal{E}$ is the Hilbert $\mathcal{B}^{\mathrm{a}}(E)$-$\mathcal{B}^{\mathrm{a}}(F)$-module obtained from the *GNS construction* ([Paschke 1973, Theorem 5.2]) of the CP-map ϑ.

Recall that a Hilbert $\mathcal{B}$-module E is said to be *full* if $\overline{\mathsf{span}}\{\langle x, y\rangle : x, y \in E\} = \mathcal{B}$. The following lemma is a known result. But for the sake of completeness of the note we include a proof here.

Lemma 2.20. *Suppose F is a Hilbert $\mathcal{B}$-$\mathcal{C}$-module. Then for any full Hilbert $\mathcal{B}$-module E the relative commutant of $\mathcal{B}^{\mathrm{a}}(E) \odot I_F$ in $\mathcal{B}^{\mathrm{a}}(E \odot F)$ is $I_E \odot \mathcal{B}^{\mathrm{a,bil}}(F)$.*

Proof. If $T \in \mathcal{B}^{\mathrm{a,bil}}(F)$, then $I_E \odot T \in \mathcal{B}^{\mathrm{a}}(E \odot F)$ commutes with all elements of the form $a \odot I_F$ for all $a \in \mathcal{B}^{\mathrm{a}}(E)$ and hence we have $I_E \odot \mathcal{B}^{\mathrm{a,bil}}(F) \subseteq (\mathcal{B}^{\mathrm{a}}(E) \odot I_F)'$. For the reverse inclusion assume that $\mathfrak{a} \in (\mathcal{B}^{\mathrm{a}}(E) \odot I_F)' \subseteq \mathcal{B}^{\mathrm{a}}(E \odot F)$. Since E is full and $F = \overline{\mathrm{span}}\, \mathcal{B}F = \overline{\mathrm{span}}\{\langle x_1, x_2\rangle y : x_i \in E,\ y \in F\}$ we have $F = E^* \odot_{\mathcal{B}^{\mathrm{a}}(E)} E \odot_{\mathcal{B}} F$ under the identification $\langle x_1, x_2\rangle y \mapsto x_1^* \odot x_2 \odot y$. Set $T = (I_{E^*} \odot \mathfrak{a}) \in \mathcal{B}^{\mathrm{a,bil}}(F)$. Then, since $E \odot_{\mathcal{B}} E^* \cong \mathcal{K}(E)$ via $x_1 \odot x_2^* \mapsto |x_1\rangle\langle x_2|$, and $\mathcal{K}(E) \odot_{\mathcal{K}(E)} E \odot_{\mathcal{B}} F \cong E \odot_{\mathcal{B}} F$ via $|x_1\rangle\langle x_2| \odot x \odot y \mapsto x_1\langle x_2, x\rangle \odot y$, we get

$$
\begin{aligned}
(I_E \odot T)(x_1 \odot \langle x_2, x_3\rangle y) &= (I_E \odot I_{E^*} \odot \mathfrak{a})(x_1 \odot x_2^* \odot x_3 \odot y) \\
&= x_1 \odot x_2^* \odot \mathfrak{a}(x_3 \odot y) \\
&= |x_1\rangle\langle x_2| \odot \mathfrak{a}(x_3 \odot y) \\
&= |x_1\rangle\langle x_2|\mathfrak{a}(x_3 \odot y) \\
&= (|x_1\rangle\langle x_2| \odot id_F)\mathfrak{a}(x_3 \odot y) \\
&= \mathfrak{a}(|x_1\rangle\langle x_2| \odot id_F)(x_3 \odot y) \\
&= \mathfrak{a}(|x_1\rangle\langle x_2|x_3 \odot y) \\
&= \mathfrak{a}(x_1 \odot \langle x_2, x_3\rangle y).
\end{aligned}
$$

for all $x_i \in E,\ y \in F$. Thus $T \in \mathcal{B}^{\mathrm{a,bil}}(F)$ is such that $\mathfrak{a} = id_E \odot T$. Hence $(\mathcal{B}^{\mathrm{a}}(E) \odot I_F)' \subseteq I_E \odot \mathcal{B}^{\mathrm{a,bil}}(F)$. □

Theorem 2.21. *Suppose E is a full Hilbert $\mathcal{B}$-module, F is a Hilbert $\mathcal{B}$-$\mathcal{C}$-module and $\tau : \mathcal{B}^{\mathrm{a}}(E) \to \mathcal{B}^{\mathrm{a}}(F)$ is a $*$-nondegenerate, strict, regular homomorphism. Then there exists a Hilbert $\mathcal{B}$-$\mathcal{C}$-module F_τ, an idempotent operator $T \in \mathcal{B}^{\mathrm{a,bil}}(F_\tau)$ and a unitary $U : E \odot F_\tau \to F$ such that*

$$\tau(a) = U(a \odot T)U^*$$

for all $a \in \mathcal{B}^{\mathrm{a}}(E)$. Moreover, the triple (F_τ, T, U) is unique up to a unitary isomorphism.

Proof. Suppose $\vartheta : \mathcal{B}^{\mathrm{a}}(E) \to \mathcal{B}^{\mathrm{a}}(F)$ is the unique unital $*$-homomorphism such that $\tau(a) = \vartheta(a)\tau(1) = \tau(1)\vartheta(a)$. Since τ is strict we have ϑ is strict, and hence there exists a Hilbert $\mathcal{B}$-$\mathcal{C}$-module F_ϑ and a unitary $U \in \mathcal{B}^{\mathrm{a,bil}}(E \odot F_\vartheta, F)$ such that $\vartheta(a) = U(a \odot I)U^*$. Take $F_\tau = F_\vartheta$. Then $\vartheta(a)\tau(1) = \tau(1)\vartheta(a)$ implies that $(a \odot I)U^*\tau(1)U = U^*\tau(1)U(a \odot I)$ for all $a \in \mathcal{B}^{\mathrm{a}}(E)$ so that $U^*\tau(1)U \in (\mathcal{B}^{\mathrm{a}}(E) \odot I_{F_\tau})'$. Hence there exists a $T \in \mathcal{B}^{\mathrm{a,bil}}(F_\tau)$ such that $\tau(1) = U(I_E \odot T)U^*$.

Clearly, $\tau(a) = \vartheta(a)\tau(1) = U(a \odot T)U^*$. Now

$$\begin{aligned}\tau(1)^2 = \tau(1) &\Longrightarrow I_E \odot T^2 = I_E \odot T \\ &\Longrightarrow \langle (I_E \odot T^2)(x_1 \odot y_1), x_2 \odot y_2\rangle = \langle (I_E \odot T)(x_1 \odot y_1), x_2 \odot y_2\rangle \\ &\Longrightarrow \langle T^2 y_1, \langle x_1, x_2\rangle y_2\rangle = \langle T y_1, \langle x_1, x_2\rangle y_2\rangle\end{aligned}$$

for all $x_1, x_2 \in E$ and $y_1, y_2 \in F_\tau$. But since E is full and F_τ has a nondegenerate left action of $\mathcal{B}$, from equation above, we have $\langle T^2 y, y'\rangle = \langle T y, y'\rangle$ for all $y, y' \in F_\tau$, so that $T^2 = T$.

Uniqueness: Suppose (F'_τ, T', U') is another such triple. Then we have $E \odot F_\tau \cong F \cong E \odot F'_\tau$ via the unitary isomorphism U'^*U. Since E is full we identify $E^* \odot_{\mathcal{B}^a(E)} E = \mathcal{B}$ via the unitary isomorphism $x^* \odot x' \mapsto \langle x, x'\rangle$. Then

$$F_\tau = \mathcal{B} \odot F_\tau = E^* \odot E \odot F_\tau \cong E^* \odot E \odot F'_\tau = \mathcal{B} \odot F'_\tau = F'_\tau,$$

where the isomorphism is given by the unitary $\widehat{U} = (I_{E^*} \odot U'^*)(I_{E^*} \odot U) : F_\tau \to F'_\tau$. Observe that

$$U = I_E \odot I_{E^*} \odot U = I_E \odot (I_{E^*} \odot U')\widehat{U} = (I_E \odot I_{E^*} \odot U')(I_{E^*} \odot \widehat{U}) = U'(I_{E^*} \odot \widehat{U}).$$

Also since $U(I_E \odot T)U^* = \tau(I_E) = U'(I_E \odot T')U'^*$ we have

$$T = I_{E^*} \odot I_E \odot T = I_{E^*} \odot \{U^*U'(I_E \odot T')U'^*U\} = \widehat{U}^*(I_{E^*} \odot I_E \odot T')\widehat{U} = \widehat{U}^*T'\widehat{U}.$$

Thus $\widehat{U}$ gives the required unitary equivalence. □

Corollary 2.22. *Suppose τ and T are as in the theorem above. Then τ is a $*$-homomorphism if and only if $T = T^*$.*

Proof. Clearly if $T = T^*$, then τ is a $*$-homomorphism. Conversely assume that τ is a $*$-homomorphism. Then $U(I \odot T)U^* = \tau(I) = \tau(I)^* = U(I \odot T^*)U^*$. Since U is a unitary we get $I \odot T = I \odot T^*$. Since E is full, this implies $T = T^*$. □

Proposition 2.23. *Suppose $\tau : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ is a regular homomorphism. Then there exists a $*$-homomorphism $\vartheta : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ such that $\tau(a) = \vartheta(a)\tau(1) = \tau(1)\vartheta(a)$ for all $a \in \mathcal{A}$. Consequently τ is completely bounded with $\|\tau\|_{cb} = \|\tau(1)\|$. If $\tau(1) = \tau(1)^*$, then τ is a $*$-homomorphism. If τ is $*$-nondegenerate, then ϑ is unique and it is unit-preserving.*

Proof. Follows from Remark 2.15. □

Suppose $\mathcal{A}$ is a von Neumann algebra and $\tau : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ is a normal, regular homomorphism. Then it can be verified that ϑ given by the Proposition 2.23 is a normal $*$-homomorphism. In particular if $\mathcal{A} = \mathcal{B}(\mathcal{H}')$, where $\mathcal{H}'$ is another Hilbert space, then it is well known that ϑ has a factorization $\vartheta(a) = V(a \odot I)V^*$ for some isometry $V : \mathcal{H}' \odot \mathcal{H}_\vartheta \to \mathcal{H}$ on a suitable Hilbert space $\mathcal{H}_\vartheta$. Again

$\vartheta(a)\tau(1) = \tau(1)\vartheta(a)$ for all $a \in \mathcal{B}(\mathcal{H}')$ implies that $\tau(1) = V(I \odot T)V^*$ for some $T \in \mathcal{B}(\mathcal{K}_\vartheta)$. Thus we have:

Theorem 2.24. *Suppose $\tau : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ is a normal, regular homomorphism. Then there exists a Hilbert space $\mathcal{K}_\tau$, an idempotent operator $T \in \mathcal{B}(\mathcal{K}_\tau)$ and an isometry $V : \mathcal{H} \odot \mathcal{K}_\tau \to \mathcal{K}$ such that*

$$\tau(a) = V(a \odot T)V^*.$$

Moreover, there exists a symmetry $J_0 \in \mathcal{B}(\mathcal{K}_\tau)$ such that τ is a J-homomorphism with $J = V(I \odot J_0)V^*$. If τ is $*$-nondegenerate, then V is a unitary and (K_τ, T, V) is unique up to unitary equivalence. Further, τ is a $*$-homomorphism if and only if $T = T^*$.

2C. *Ternary homomorphisms.*

Definition 2.25. Let $t \in \mathbb{R}$. A map $\tau : \mathcal{A} \to \mathcal{B}$ is said to be *t-ternary* if

$$\tau(a)\tau(b)^*\tau(c) = t\tau(ab^*c) \quad \text{for all } a, b, c \in \mathcal{A}.$$

A 1-ternary map is simply called a *ternary* map. Note that all $*$-homomorphisms are ternary maps. In fact if $\mathcal{A}, \mathcal{B}$ are unital C^*-algebras, then a unital linear map $\tau : \mathcal{A} \to \mathcal{B}$ is ternary if and only if τ is a $*$-homomorphism. Here is a typical example of a t-ternary homomorphism:

Example 2.26. Clearly, $\tau : \mathcal{A} \to M_2(\mathcal{A})$ given by $\tau(a) = \left[\begin{smallmatrix} a & 0 \\ (\sqrt{t-1})a & 0 \end{smallmatrix}\right]$ is a t-ternary homomorphism for all $t \in (1, \infty)$.

We are only interested in t-ternary maps which are homomorphisms. In this context, we have the following basic observation.

Proposition 2.27. *Let $\mathcal{A}, \mathcal{B}$ be unital C^*-algebras and let $\tau : \mathcal{A} \to \mathcal{B}$ be a nonzero t-ternary homomorphism. Then $1 \le t = \|\tau(1)\|^2$. If $t = 1$, then τ is a $*$-homomorphism.*

Proof. For convenience, without loss of generality, we assume $\mathcal{B} \subseteq \mathcal{B}(\mathcal{H})$ for some Hilbert space $\mathcal{H}$. Take $T = \tau(1)$. Let $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_0^\perp$ be the orthogonal decomposition of $\mathcal{H}$, where $\mathcal{H}_0 = \overline{T(\mathcal{H})}$. Since $T^2 = T$, $Th = h$ for $h \in \mathcal{H}_0$, as a consequence the operator T decomposes as

$$T = \begin{bmatrix} I & N \\ 0 & 0 \end{bmatrix}$$

for some N, with respect to the decomposition $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_0^\perp$. Now computing, $TT^*T = tT$, we see $I + NN^* = tI$. In particular $t \ge 1$. Also since $P = (1/t)T^*T$ is a nonzero projection we have $\|T\|^2 = t$.

If $t = 1$, we get $N = 0$ and hence $\tau(1)^* = \tau(1)$. Taking $a = c = 1$ in the definition of 1-ternary, we get $\tau(b)^* = \tau(1)^*\tau(b)^*\tau(1)^* = \tau(1)\tau(b)^*\tau(1) = 1.\tau(1.b^*.1) = \tau(b^*)$. Therefore, τ is a $*$-homomorphism. $\square$

Proposition 2.28. *All t-ternary homomorphisms $\tau : \mathcal{A} \to \mathcal{B}$ are regular.*

Proof. An easy computation using t-ternary and homomorphism properties yields $(\tau(1)^*\tau(1) - \tau(u)^*\tau(u))^2 = 0$ and $(\tau(1)\tau(1)^* - \tau(u)\tau(u)^*)^2 = 0$ for any unitary $u \in \mathcal{A}$. □

The converse of this proposition is not true. For example, the direct sum of t_1 and t_2-ternaries as in Example 2.26, is easily seen to be regular but not a t-ternary for any t. The direct sum of two t-ternary homomorphisms on a common domain algebra is again a t-ternary homomorphism.

From the proposition above, since all regular homomorphisms are symmetric, all t-ternary homomorphisms $\tau : \mathcal{A} \to \mathcal{B}$ are symmetric homomorphisms. Since 1-ternary homomorphisms are already $*$-homomorphisms, we will assume that $t > 1$. Now we will show that for a $*$-nondegenerate t-ternary homomorphism $\tau : \mathcal{A} \to \mathcal{B}^{\mathrm{a}}(E)$, $t \in (1, \infty)$, a possible choice of symmetry can be written down explicitly as $(1/\sqrt{t})(\tau(1) + \tau(1)^* - I)$.

Proposition 2.29. *Suppose $t \in (1, \infty)$ and $\tau : \mathcal{A} \to \mathcal{B}$ is a t-ternary homomorphism. Take $T = \tau(1)$ and $J_t = (1/\sqrt{t})(T + T^* - I)$. Then*

(i) *$J_t\tau(a)^*J_t = \tau(a^*)$ and $J_t\tau(a^*)J_t = \tau(a)^*$ for all $a \in \mathcal{A}$;*

(ii) *$\sigma(J_t) \subseteq \{1, -1, -1/\sqrt{t}\}$.*

Proof. We have $T^2 = T$ and $TT^*T = tT$. To prove (i),

$$\begin{aligned} J_t\tau(a)^*J_t =& \frac{1}{\sqrt{t}}(T + T^* - I)\tau(a)^*\frac{1}{\sqrt{t}}(T + T^* - I) \\ =& \frac{1}{t}(T\tau(a)^* + T^*\tau(a)^* - \tau(a)^*)(T + T^* - I) \\ =& \frac{1}{t}T\tau(a)^*(T + T^* - I) \\ =& \frac{1}{t}T\tau(a)^*T \\ =& \tau(a^*). \end{aligned}$$

Similarly we can prove that $J_t\tau(a^*)J_t = \tau(a)^*$. To see (ii), observe,

$$\begin{aligned} (J_t + I)(J_t - I)(\sqrt{t}J_t + I) &= \sqrt{t}J_t^3 + J_t^2 - \sqrt{t}J_t - I \\ &= J_t(T + T^* - I)J_t + J_t^2 - \sqrt{t}J_t - I \\ &= (T^* + T - J_t^2) + J_t^2 - \sqrt{t}J_t - I \\ &= 0. \end{aligned}$$

Since $J_t = J_t^*$ the proof is complete. □

In this Proposition, as $J_t = J_t^*$ by spectral theorem $J_t = P_1 - P_2 + (-1/\sqrt{t})P_3$, where P_1, P_2, P_3 are orthogonal projections with $P_1 + P_2 + P_3 = I$. Note that due to finiteness of the spectrum of J_t, P_1, P_2, P_3 are in the C^*-algebra $\mathcal{B}$. Furthermore, J_t is a symmetry if and only if $P_3 = 0$.

Proposition 2.30. *Suppose $\tau : \mathcal{A} \to \mathcal{B}^a(E)$ is a t-ternary homomorphism with $t \in (1, \infty)$. Then $E_0 = \overline{\mathsf{span}}\{\tau(a_1)x_1, \tau(a_2)^*x_2 : x_i \in E,\ a_i \in \mathcal{A}, i = 1, 2\}$ is complemented in E. Moreover, the following are equivalent*:

(i) *τ is $*$-nondegenerate*;

(ii) $\ker(\tau(1) + \tau(1)^*) = \{0\}$;

(iii) *J_t is a symmetry.*

Proof. Set $T = \tau(1)$, $E_0 = \mathsf{span}\{T(E), T^*(E)\}$. We wish to construct an orthogonal projection onto E_0. Take $Q = (1/(t-1))[TT^* + T^*T - T - T^*]$. From $T^2 = T$ and $TT^*T = tT$, simple algebra shows $QT = T$, $QT^* = T^*$ and $Q = Q^* = Q^2$. So Q is a projection whose range contains $T(E)$ and $T^*(E)$. From the definition of Q, $Q(E) \subseteq E_0$. This proves $Q(E) = E_0$. Clearly then $(I - Q)$ is the projection onto $E_0^\perp$ and $E = E_0 \oplus E_0^\perp$.

To show the equivalence of (i) to (ii), we show $\ker(T + T^*) = \ker Q$. If $(T + T^*)x = 0$, then

$$\begin{aligned} Qx = \frac{1}{t-1}[TT^* + T^*T - T - T^*]x &= \frac{1}{t-1}[TT^*x + T^*Tx] \\ &= \frac{1}{t-1}[T(-Tx) + T^*(-T^*)x] \\ &= \frac{-1}{t-1}[Tx + T^*x] \\ &= 0. \end{aligned}$$

Conversely if $Qx = 0$, then $x \in E_0^\perp$; hence $(T + T^*)x = 0$. The equivalence of (ii) and (iii) is obvious as $\ker(T+T^*) = \ker(\sqrt{t}J_t + 1) = \left\{x : J_t x = \frac{-1}{\sqrt{t}}x\right\} = \operatorname{ran}(P_3)$. $\square$

Theorem 2.31. *Suppose $t \in (1, \infty)$ and $\tau : \mathcal{A} \to \mathcal{B}^a(E)$ is a $*$-nondegenerate linear map. Take $J_t = (1/\sqrt{t})[\tau(1) + \tau(1)^* - 1]$. Then the following are equivalent*:

(i) *τ is a t-ternary homomorphism.*

(ii) *τ is a J_t-homomorphism.*

Proof. (i) $\Rightarrow$ (ii) : We have already seen this.

(ii) $\Rightarrow$ (i) : For all $a, b, c \in \mathcal{A}$ we have

$$\begin{aligned} t\tau(ab^*c) &= t\tau(a)\tau(b^*)\tau(c) \\ &= t\tau(a)J_t\tau(b)^*J_t\tau(c) \\ &= t\tau(a)\frac{\tau(1)\tau(b)^*\tau(1)}{t}\tau(c) \\ &= \tau(a)\tau(b)^*\tau(c). \end{aligned}$$

□

Theorem 2.32. *Suppose $\tau : \mathcal{A} \to \mathcal{B}^{a}(E)$ is a t-ternary homomorphism, where $t \in (1, \infty)$. Then there exists a closed, complemented, $\mathcal{B}$-submodule $E_1 \subseteq E$, a unital $*$-homomorphism $\pi : \mathcal{A} \to \mathcal{B}^{a}(E_1)$ and isometries $V_1, V_2 \in \mathcal{B}^{a}(E_1, E)$ with $V_2^* V_1 = (1/\sqrt{t}) I_{E_1}$ such that*

$$\tau(\,\cdot\,) = \sqrt{t}\, V_1 \pi(\,\cdot\,) V_2^*. \tag{2-8}$$

Consequently τ is completely bounded with $\|\tau\|_{cb} = \|\tau(1)\| = \sqrt{t}$. Moreover, (2-8) *always defines a t-ternary homomorphism.*

Proof. Let E_1 be the range of the orthogonal projection $P = (1/t)T^*T$ where $T = \tau(1)$. Define linear maps $V_i : E_1 \to E$ by $V_1 = (1/\sqrt{t})T|_{E_1}$ and $V_2 = I|_{E_1}$. Note that the V_i's are adjointable isometries with $V_1^* = (1/\sqrt{t})PT^*$ and $V_2^* = P$. Now for each $a \in \mathcal{A}$ define $\pi(a) : E_1 \to E_1$ by $\pi(a) = P\tau(a)|_{E_1}$. Clearly $\pi(1) = P\tau(1)|_{E_1} = P|_{E_1} = I_{E_1}$. Also for all $a, b \in \mathcal{A}$,

$$\begin{aligned}\pi(a)\pi(b) &= P\tau(a)\frac{\tau(1)^*\tau(1)}{t}\tau(b)|_{E_1}\\ &= P\frac{\tau(a)\tau(1)^*\tau(b)}{t}|_{E_1}\\ &= P\tau(ab)|_{E_1}\\ &= \pi(ab).\end{aligned}$$

Now since

$$P\tau(a)^*P = \frac{1}{t^2}\tau(1)^*\tau(1)\tau(a)^*\tau(1) = \frac{1}{t}\tau(1)^*\tau(a^*) = P\tau(a^*)$$

for all $x, x' \in E_1$ we have

$$\langle \pi(a)x, x'\rangle = \langle P\tau(a)x, x'\rangle = \langle x, \tau(a)^*Px'\rangle = \langle x, P\tau(a)^*Px'\rangle = \langle x, \pi(a^*)x'\rangle,$$

so that $\pi(a)^* = \pi(a^*)$. Thus $a \mapsto \pi(a)$ defines a unital $*$-homomorphism $\pi : \mathcal{A} \to \mathcal{B}^{a}(E_1)$. Also for $a \in \mathcal{A}$ we have

$$\sqrt{t}V_1\pi(a)V_2^* = TP\tau(a)P = \frac{1}{t^2}\tau(1)\tau(1)^*\tau(1)\tau(a)\tau(1)^*\tau(1) = \tau(a).$$

Since the V_i are isometries, τ is completely bounded with $\|\tau\|_{cb} \le \sqrt{t} = \|\tau(1)\|$. □

Now we show that every regular homomorphism is essentially a direct sum or direct integral of t-ternary homomorphisms through spectral integration. Let $\tau : \mathcal{A} \to \mathcal{B} \subseteq \mathcal{B}(\mathcal{H})$ be a $*$-nondegenerate regular homomorphism. In view of Proposition 2.23, it suffices to know the structure of $\tau(1)$. As before, take $T = \tau(1) = R + iS$ and let $J = f(R)$ be the symmetry constructed in Proposition 2.4. In the following, we decompose the Hilbert space $\mathcal{H}$ as $\mathcal{H} = \mathcal{H}_+ \oplus \mathcal{H}_-$, where

$\mathcal{H}_+ = \{h \in \mathcal{H} : Jh = h\}$ and $\mathcal{H}_- = \{h \in \mathcal{H} : Jh = -h\}$. With respect to this decomposition, decompose the operator $T = \tau(1)$ as

$$T = \begin{bmatrix} X & Y \\ Z & W \end{bmatrix}.$$

Now $JTJ = T^*$ yields $X = X^*$, $W = W^*$, $Z = -Y^*$. Furthermore, from $T^2 = T$, we get $X^2 - YY^* = X$, $XY + YW = Y$, $-Y^*Y + W^2 = W$. So $Y^*Y = W(W - I)$. Let $Y = V[W(W-I)]^{1/2}$, be the polar decomposition of Y. Suppose $0 \in \sigma(W)$. Let $0 \neq h_- \in \mathcal{H}_-$ such that $Wh_- = 0$. Set $h = \left[\begin{smallmatrix} 0 \\ h_- \end{smallmatrix}\right]$. Then $Yh_- = 0$, hence $\tau(1)h = 0$. Since τ is $*$-nondegenerate this implies that $h = 0$, which is a contradiction. Again, suppose $1 \in \sigma(W)$, choose $0 \neq h_- \in \mathcal{H}_-$ such that $Wh_- = h_-$. Then $Yh_- = 0$. Since τ is t-ternary, $\tau(1)\tau(1)^*\tau(1)h = t\tau(1)h$ for $t \in (1, \infty)$. But $\tau(1)\tau(1)^*\tau(1)h = \tau(1)h$. Thus $\tau(1)h = t\tau(1)h \Longrightarrow \tau(1)h = 0$. Since τ is $*$-nondegenerate this implies that $h = 0$, which is a contradiction. Thus $0, 1 \notin \sigma(W)$. To prove V is unitary it is enough to show that $\overline{\operatorname{ran}(Y)} = \mathcal{H}_+$, i.e, we have to show that $\ker(Y^*) = 0$. Let $0 \neq h_+ \in \mathcal{H}_+$. We have $X^2 = X - YY^*$ and $Y^*h_+ = 0$, which implies $X^2h_+ = Xh_+$. Now $-ZXh_+ = Y^*X^*h_+ = (XY)^*h_+ = (Y^* - W^*Y^*)h_+ = 0$. Set $h = \left[\begin{smallmatrix} h_+ \\ 0 \end{smallmatrix}\right]$. Then $\tau(1)\tau(1)^*\tau(1)h = \tau(1)h$. Since τ is t-ternary, where $t \in (1, \infty)$, we will get $\tau(1)h = t\tau(1)h$. Thus $Xh_+ = 0$ and hence $\tau(1)h = 0 = \tau(1)^*h$. Since τ is $*$-nondegenerate, $h = 0$. Thus to avoid degenerate cases assume that V is a unitary and $0, 1 \notin \sigma(W)$. Now $X^2 - X = YY^* = V[W(W - I)]V^*$. Also from $XY = Y(I - W)$, we get $XV[W(W - I)]^{1/2} = V[W(W - I)]^{1/2}(I - W)$, which yields, $X = V(I - W)V^*$. Now T decomposes as

$$T = \begin{bmatrix} V & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} (I - W) & [W(W - I)]^{1/2} \\ -[W(W - I)]^{1/2} & W \end{bmatrix} \begin{bmatrix} V^* & 0 \\ 0 & I \end{bmatrix}.$$

Observe that, for any real number $w \notin [0, 1]$

$$T_w = \begin{bmatrix} (1 - w) & [w(w - 1)]^{1/2} \\ -[w(w - 1)]^{1/2} & w \end{bmatrix}$$

satisfies $T_w = T_w^2$ and $J_2 T_w J_2 = T_w^*$, where

$$J_2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

and also $T_w T_w^* T_w = (1 - 2w)^2 T_w$. In other words $z \mapsto zT_w$ is a $(1 - 2w)^2$-ternary of complex numbers in $M_2(\mathbb{C})$.

3. Representations of completely bounded maps

In this section we give a new structure theorem for CB-maps from $\mathcal{A}$ into $\mathcal{B}(\mathcal{H})$. We study some known structure theorems (see [Paulsen and Suen 1985; Suen 1991]) and make a comparison. We repeatedly use the following well-known theorem (see [Paulsen 2002, Theorem 8.3]):

Theorem 3.1. *Suppose $\mathcal{A}$ is a unital C^*-algebra and $\psi : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ is a CB-map. Then there exist CP-maps $\varphi_i : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ with $\|\varphi_i\|_{cb} = \|\psi\|_{cb}$ such that $\Phi : M_2(\mathcal{A}) \to \mathcal{B}(\mathcal{H} \oplus \mathcal{H})$ defined by*

$$\Phi\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} \varphi_1(a) & \psi(b) \\ \psi^*(c) & \varphi_2(d) \end{bmatrix}$$

is a CP-map. Moreover, if $\|\psi\|_{cb} \le 1$, it is possible to take $\varphi_i(1) = I_{\mathcal{H}}$.

3A. ***Regular representations.*** Observe that if $\mathcal{K}$ is another Hilbert space and $\tau : \mathcal{A} \to \mathcal{B}(\mathcal{K})$ is a regular homomorphism, then $\psi(\,\cdot\,) := W^*\tau(\,\cdot\,)W$ defines a CB-map from $\mathcal{A}$ into $\mathcal{B}(\mathcal{H})$ for all $W \in \mathcal{B}(\mathcal{H}, \mathcal{K})$. We prove all CB-maps arise this way.

Theorem 3.2. *Suppose $\psi : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ is a CB-map. Then there exists a Hilbert space $\mathcal{K}$, a $*$-nondegenerate regular homomorphism $\tau : \mathcal{A} \to \mathcal{B}(\mathcal{K})$ and a bounded linear map $W : \mathcal{H} \to \mathcal{K}$ such that $\psi(\,\cdot\,) = W^*\tau(\,\cdot\,)W$. Moreover, given any $t \in (1, \infty)$ we can choose τ and W such that τ is t-ternary and W satisfies*

$$\left(\frac{(t-1)\sqrt{t-1}}{2\sqrt{t-1}+2t-1}\right)\|W\|^2 \le \|\psi\|_{cb} \le \sqrt{t}\,\|W\|^2.$$

Proof. Since ψ is a CB-map, by Theorem 3.1, there exists CP-maps $\varphi_i : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ such that $\Phi = \left[\begin{smallmatrix} \varphi_1 & \psi \\ \psi^* & \varphi_2 \end{smallmatrix}\right] : M_2(\mathcal{A}) \to \mathcal{B}(\mathcal{H} \oplus \mathcal{H})$ is a CP-map. Suppose $(\mathcal{K}, \Pi, V)$ is the (minimal) Stinespring dilation for Φ. Given $t \in (1, \infty)$ set $t' = \sqrt{t-1}$. Then for $a \in \mathcal{A}$ we have

$$\begin{aligned}
\psi(a) &= \begin{bmatrix} I_{\mathcal{H}} & I_{\mathcal{H}} \end{bmatrix} \begin{bmatrix} 0 & \psi(a) \\ 0 & 0 \end{bmatrix} \begin{bmatrix} I_{\mathcal{H}} \\ I_{\mathcal{H}} \end{bmatrix} \\
&= \begin{bmatrix} I_{\mathcal{H}} & I_{\mathcal{H}} \end{bmatrix} \Phi\left(\begin{bmatrix} 0 & a \\ 0 & 0 \end{bmatrix}\right) \begin{bmatrix} I_{\mathcal{H}} \\ I_{\mathcal{H}} \end{bmatrix} \\
&= \begin{bmatrix} I_{\mathcal{H}} & I_{\mathcal{H}} \end{bmatrix} V^*\Pi\left(\begin{bmatrix} 0 & a \\ 0 & 0 \end{bmatrix}\right) V \begin{bmatrix} I_{\mathcal{H}} \\ I_{\mathcal{H}} \end{bmatrix} \qquad (3\text{-}1) \\
&= \begin{bmatrix} I_{\mathcal{H}} & I_{\mathcal{H}} \end{bmatrix} V^*\Pi\left(\begin{bmatrix} 0 & \frac{1}{\sqrt{t'}} \\ \frac{1}{\sqrt{t'}} & \frac{-1}{t'\sqrt{t'}} \end{bmatrix}\right) \Pi\left(\begin{bmatrix} a & 0 \\ t'a & 0 \end{bmatrix}\right) \Pi\left(\begin{bmatrix} 0 & \frac{1}{\sqrt{t'}} \\ \frac{1}{\sqrt{t'}} & \frac{-1}{t'\sqrt{t'}} \end{bmatrix}\right) V \begin{bmatrix} I_{\mathcal{H}} \\ I_{\mathcal{H}} \end{bmatrix}. \qquad (3\text{-}2)
\end{aligned}$$

Define $\tau : \mathcal{A} \to \mathcal{B}(\mathcal{K})$ by $\tau(a) = \Pi\left(\left[\begin{smallmatrix} a & 0 \\ t'a & 0 \end{smallmatrix}\right]\right)$, which is a t-ternary homomorphism. Note that if $k \in \mathcal{K}$ is such that $\tau(1)k = 0 = \tau(1)^*k$, then

$$\begin{aligned} k = \Pi(1)k &= \Pi\left(\begin{bmatrix} 0 & \frac{1}{t'} \\ 0 & \frac{-1}{t'^2} \end{bmatrix}\begin{bmatrix} 1 & 0 \\ t' & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \frac{1}{t'} & 0 \end{bmatrix}\begin{bmatrix} 1 & t' \\ 0 & 0 \end{bmatrix}\right)k \\ &= \Pi\left(\begin{bmatrix} 0 & \frac{1}{t'} \\ 0 & \frac{-1}{t'^2} \end{bmatrix}\right)\tau(1)k + \Pi\left(\begin{bmatrix} 0 & 0 \\ \frac{1}{t'} & 0 \end{bmatrix}\right)\tau(1)^*k \\ &= 0. \end{aligned}$$

Thus τ is a $*$-nondegenerate t-ternary homomorphism. Set

$$W = \Pi\left(\begin{bmatrix} 0 & 1/\sqrt{t'} \\ 1/\sqrt{t'} & -1/t'\sqrt{t'} \end{bmatrix}\right) V \begin{bmatrix} I_{\mathcal{H}} \\ I_{\mathcal{H}} \end{bmatrix} \in \mathcal{B}(\mathcal{H}, \mathcal{K}).$$

Then from (3-2) we get $\psi(\,\cdot\,) = W^*\tau(\,\cdot\,)W$; hence $\|\psi\|_{cb} \le \|W\|^2\|\tau\|_{cb} = \sqrt{t}\|W\|^2$. Also note that

$$\begin{aligned} \|W\|^2 &= \|W^*W\| \\ &= \left\| \begin{bmatrix} I_{\mathcal{H}} & I_{\mathcal{H}} \end{bmatrix} V^*\Pi\left(\begin{bmatrix} 0 & 1/\sqrt{t'} \\ 1/\sqrt{t'} & -1/t'\sqrt{t'} \end{bmatrix}\begin{bmatrix} 0 & 1/\sqrt{t'} \\ 1/\sqrt{t'} & -1/t'\sqrt{t'} \end{bmatrix}\right) V \begin{bmatrix} I_{\mathcal{H}} \\ I_{\mathcal{H}} \end{bmatrix}\right\| \\ &= \left\| \begin{bmatrix} I_{\mathcal{H}} & I_{\mathcal{H}} \end{bmatrix} \Phi\left(\begin{bmatrix} 1/t' & -1/t'^2 \\ -1/t'^2 & (t'^2+1)/t'^3 \end{bmatrix}\right) \begin{bmatrix} I_{\mathcal{H}} \\ I_{\mathcal{H}} \end{bmatrix}\right\| \\ &= \left\| \begin{bmatrix} I_{\mathcal{H}} & I_{\mathcal{H}} \end{bmatrix} \begin{bmatrix} \varphi_1(1/t') & \psi(-1/t'^2) \\ \psi^*(-1/t'^2) & \varphi_2((t'^2+1)/t'^3) \end{bmatrix} \begin{bmatrix} I_{\mathcal{H}} \\ I_{\mathcal{H}} \end{bmatrix}\right\| \\ &= \left\| \varphi_1\left(\frac{1}{t'}\right) - \psi\left(\frac{1}{t'^2}\right) - \psi^*\left(\frac{1}{t'^2}\right) + \varphi_2\left(\frac{t'+1}{t'^2}\right)\right\| \\ &\le \frac{1}{t'}\|\varphi_1\|_{cb} + \frac{1}{t'^2}\|\psi\|_{cb} + \frac{1}{t'^2}\|\psi\|_{cb} + \left(\frac{t'^2+1}{t'^3}\right)\|\varphi_2\|_{cb} \\ &= \left(\frac{2t'+2t'^2+1}{t'^3}\right)\|\psi\|_{cb} \\ &= \frac{2\sqrt{t-1}+2t-1}{(t-1)\sqrt{t-1}}\|\psi\|_{cb}. \end{aligned}$$

□

Theorem 3.3. *Suppose $\mathcal{A}$ is a unital C^*-algebra and $\mathcal{H}$ is a Hilbert space. Then there exists a Hilbert space $\mathcal{K}$ and a $*$-nondegenerate regular homomorphism $\tau : \mathcal{A} \to \mathcal{B}(\mathcal{K})$ such that given any $\psi \in CB(\mathcal{A}, \mathcal{B}(\mathcal{H}))$ there exists an operator $W_\psi \in \mathcal{B}(\mathcal{H}, \mathcal{K})$ such that $\psi(\,\cdot\,) = W_\psi^*\tau(\,\cdot\,)W_\psi$. Moreover, given any $t \in (1, \infty)$ we can choose τ and W_ψ such that τ is t-ternary and W_ψ satisfies*

$$\left(\frac{(t-1)\sqrt{t-1}}{2\sqrt{t-1}+2t-1}\right)\|W_\psi\|^2 \le \|\psi\|_{cb} \le \sqrt{t}\|W_\psi\|^2.$$

Proof. Suppose $t \in (1, \infty)$. For each $\psi \in CB(\mathcal{A}, \mathcal{B}(\mathcal{H}))$ fix a $*$-nondegenerate regular representation $(\mathcal{K}_\psi, \tau_\psi, W_\psi)$ as in Theorem 3.2. Take $\mathcal{K} = \oplus_\psi \mathcal{K}_\psi$ and $\tau = \oplus_\psi \tau_\psi$. Note that $\tau : \mathcal{A} \to \mathcal{B}(\mathcal{K})$ is a well defined $*$-nondegenerate t-ternary homomorphism since each τ_ψ is a $*$-nondegenerate t-ternary homomorphism with $\|\tau_\psi\| = \sqrt{t}$. Now given any CB-map ψ we have the corresponding $W_\psi \in \mathcal{B}(\mathcal{H}, \mathcal{K}_\psi)$. Considering $\mathcal{K}_\psi \subseteq \mathcal{K}$ via the natural inclusion map we have $W_\psi \in \mathcal{B}(\mathcal{H}, \mathcal{K})$ with

$$\left(\frac{(t-1)\sqrt{t-1}}{2\sqrt{t-1}+2t-1}\right)\|W_\psi\|^2 \leq \|\psi\|_{cb} \leq \sqrt{t}\|W_\psi\|^2$$

and $\psi(\cdot) = W_\psi^* \tau_\psi(\cdot) W_\psi = W_\psi^* \tau(\cdot) W_\psi$. □

Theorem 3.4. *Suppose $\psi : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ is a CC-map. Then there exists a Hilbert space $\mathcal{K}$, a (not necessarily $*$-nondegenerate) regular homomorphism $\tau : \mathcal{A} \to \mathcal{B}(\mathcal{K})$ and an isometry $V \in \mathcal{B}(\mathcal{H}, \mathcal{K})$ such that $\psi(\cdot) = V^* \tau(\cdot) V$. Moreover, we can choose τ to be t-ternary for t large enough* ($t \geq 18$).

Proof. As in the proof of Theorem 3.2 consider the CP-extension of ψ given by $\Phi = \left[\begin{smallmatrix} \varphi_1 & \psi \\ \psi^* & \varphi_2 \end{smallmatrix}\right] : M_2(\mathcal{A}) \to \mathcal{B}(\mathcal{H} \oplus \mathcal{H})$ and the corresponding (minimal) Stinespring dilation $(\mathcal{K}', \Pi', V')$ for Φ. Given $t \in (1, \infty)$ set $t' = \sqrt{t-1}$ and define

$$W' = \Pi'\left(\begin{bmatrix} 0 & 1/\sqrt{t'} \\ 1/\sqrt{t'} & -1/(t'\sqrt{t'}) \end{bmatrix}\right) V' \begin{bmatrix} I_{\mathcal{H}} \\ I_{\mathcal{H}} \end{bmatrix} \in \mathcal{B}(\mathcal{H}, \mathcal{K}')$$

and define $\tau' : \mathcal{A} \to \mathcal{B}(\mathcal{K}')$ by $\tau'(a) = \Pi'\left(\left[\begin{smallmatrix} a & 0 \\ t'a & 0 \end{smallmatrix}\right]\right)$, which is a $*$-nondegenerate t-ternary homomorphism. Clearly $\psi(a) = W'^* \tau'(a) W'$. Note that since ψ is CC-map V' can be chosen to be an isometry. Hence

$$\|W'\|^2 \leq \left\|\begin{bmatrix} 0 & 1/\sqrt{t'} \\ 1/\sqrt{t'} & -1/(t'\sqrt{t'}) \end{bmatrix}\right\|^2 \left\|\begin{bmatrix} I_{\mathcal{H}} \\ I_{\mathcal{H}} \end{bmatrix}\right\|^2 \leq 2\left(\frac{1}{t'} + \frac{1}{t'} + \frac{1}{t'^3}\right),$$

because $\|[a_{ij}]\|^2 \leq \sum_{ij} |a_{ij}|^2$. So we can assume that $W' \in \mathcal{B}(\mathcal{H}, \mathcal{K}')$ is a contraction by taking t large enough ($t \geq 18$). Let $\mathcal{K}'' = \mathcal{H} \oplus \mathcal{K}'$ and $W'' := \left[\begin{smallmatrix} 0 & 0 \\ W' & 0 \end{smallmatrix}\right] \in \mathcal{B}(\mathcal{K}'')$. Define $\tau'' : \mathcal{A} \to \mathcal{B}(\mathcal{K}'')$ by $a \mapsto \left[\begin{smallmatrix} 0 & 0 \\ 0 & \tau'(a) \end{smallmatrix}\right]$, which is a t-ternary homomorphism. Suppose U is Halmos's unitary dilation of W'', that is,

$$U = \begin{bmatrix} W'' & (1 - W''W''^*)^{1/2} \\ (1 - W''^*W'')^{1/2} & -W''^* \end{bmatrix} \in \mathcal{B}(\mathcal{K}'' \oplus \mathcal{K}'').$$

Set $\mathcal{K} = \mathcal{K}'' \oplus \mathcal{K}''$ and define $\tau : \mathcal{A} \to \mathcal{B}(\mathcal{K})$ by $\tau(a) = \left[\begin{smallmatrix} \tau''(a) & 0 \\ 0 & 0 \end{smallmatrix}\right]$ which is a t-ternary homomorphism. Let V_0 be the inclusion map of $\mathcal{H}$ in $\mathcal{K} = \mathcal{K}'' \oplus \mathcal{K}'' = \mathcal{H} \oplus \mathcal{K}' \oplus \mathcal{K}''$.

Then $V_0 = \left[\begin{smallmatrix} I_{\mathcal{K}''} \\ 0_{\mathcal{K}''} \end{smallmatrix}\right]\left[\begin{smallmatrix} I_{\mathcal{H}} \\ 0_{\mathcal{K}'} \end{smallmatrix}\right]$. Set $V = UV_0 \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, which is an isometry. Then

$$\begin{aligned}
V^*\tau(a)V &= V_0^*U^*\tau(a)UV_0 \\
&= V_0^*\begin{bmatrix} W''^*\tau''(a)W'' & * \\ * & * \end{bmatrix} V_0 \\
&= \begin{bmatrix} I_{\mathcal{H}} & 0_{\mathcal{K}'} \end{bmatrix} W''^*\tau''(a)W'' \begin{bmatrix} I_{\mathcal{H}} \\ 0_{\mathcal{K}'} \end{bmatrix} \\
&= \begin{bmatrix} I_{\mathcal{H}} & 0_{\mathcal{K}'} \end{bmatrix} \begin{bmatrix} 0 & W'^* \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & \tau'(a) \end{bmatrix} \begin{bmatrix} 0 & 0 \\ W' & 0 \end{bmatrix} \begin{bmatrix} I_{\mathcal{H}} \\ 0_{\mathcal{K}'} \end{bmatrix} \\
&= W'^*\tau'(a)W' \\
&= \psi(a).
\end{aligned}$$

Note that τ is a t-ternary for $t \geq 18$. □

Theorem 3.5. *Suppose $\mathcal{A}$ is a unital C^*-algebra and $\mathcal{H}$ is a Hilbert space. There exists a Hilbert space $\mathcal{K}$ and a regular homomorphism $\tau : \mathcal{A} \to \mathcal{B}(\mathcal{K})$ such that given any completely contractive map $\psi : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ there exists an isometry $V_\psi \in \mathcal{B}(\mathcal{H}, \mathcal{K})$ such that*

$$\psi(\,\cdot\,) = V_\psi^*\tau(\,\cdot\,)V_\psi.$$

Moreover, we can choose τ to be t-ternary for t large enough ($t \geq 18$).

Proof. This follows by considering the direct sum of all representations given by Theorem 3.4. □

Now we prove analogues of above theorems for the case when the range algebra is an injective C^*-algebra.

Theorem 3.6. *Suppose $\mathcal{B}$ is an injective C^*-algebra and $\psi : \mathcal{A} \to \mathcal{B}$ is a CB-map. Then there exists a Hilbert $\mathcal{B}$-module E, a $*$-nondegenerate regular homomorphism $\tau : \mathcal{A} \to \mathcal{B}^{\mathrm{a}}(E)$ and a vector $z \in E$ such that*

$$\psi(\,\cdot\,) = \langle z, \tau(\,\cdot\,)z\rangle.$$

Moreover, given any $t \in (1, \infty)$ we can choose τ and z such that τ is t-ternary and z satisfies $((t-1)\sqrt{t-1}/(2\sqrt{t-1}+2t-1))\|z\|^2 \leq \|\psi\|_{cb} \leq \sqrt{t}\|z\|^2$.

Proof. Suppose $t \in (1, \infty)$. Let $\rho : \mathcal{B} \to \mathcal{B}(\mathcal{G})$ be a faithful unital $*$-homomorphism $\rho : \mathcal{B} \to \mathcal{B}(\mathcal{G})$ of $\mathcal{B}$ on some Hilbert space $\mathcal{G}$ satisfying $\overline{\mathrm{span}}\,\rho(\mathcal{B})\mathcal{G} = \mathcal{G}$. As $\mathcal{B}$ is injective there exists a conditional expectation $P_\mathcal{B} : \mathcal{B}(\mathcal{G}) \to \rho(\mathcal{B})$, i.e., $P_\mathcal{B}$ is a CP-map satisfying $P_\mathcal{B}(b_1Tb_2) = b_1P_\mathcal{B}(T)b_2$ for all $b_i \in \rho(\mathcal{B})$, $T \in \mathcal{B}(\mathcal{G})$. Consider the CB-map $\tilde{\psi} = \rho \circ \psi : \mathcal{A} \to \mathcal{B}(\mathcal{G})$ with a $*$-nondegenerate regular representation

$(\mathcal{K}, \tilde{\tau}, W)$, as in Theorem 3.2, where $\tilde{\tau} : \mathcal{A} \to \mathcal{B}(\mathcal{K})$ is a $*$-nondegenerate t-ternary homomorphism. Set $E_0 = \mathcal{B}(\mathcal{G}, \mathcal{K})$. Given $x, y \in E_0$ and $b \in \mathcal{B}$ define

$$xb := x \circ \rho(b) \in E_0 \text{ and } \langle x, y\rangle := \rho^{-1} P_{\mathcal{B}}(x^* \circ y) \in \mathcal{B}.$$

Here x^* is the adjoint of $x \in \mathcal{B}(\mathcal{G}, \mathcal{K})$. It is easy to verify that with the above operations E_0 forms a semi-inner product $\mathcal{B}$-module. Let E be the completion of the inner product $\mathcal{B}$-module E_0/N where

$$N := \{x \in E_0 : \langle x, x\rangle = 0\} = \{x \in E_0 : \langle x, y\rangle = 0 \text{ for all } y \in E_0\}.$$

Note that for $x+N, x'+N \in E$ the inner product is given by $\langle x+N, x'+N\rangle := \langle x, x'\rangle$. We denote the equivalence classes $x + N$ by x itself. Now for each $a \in \mathcal{A}$ define $\tau(a) : E \to E$ by $\tau(a)x := \tilde{\tau}(a) \circ x$ for all $x \in E_0$. Note that

$$\begin{aligned}
\|\tau(a)x\|^2 &= \|\langle \tilde{\tau}(a) \circ x, \tilde{\tau}(a) \circ x\rangle\| \\
&= \|\rho^{-1} P_{\mathcal{B}}(x^* \circ \tilde{\tau}(a)^* \circ \tilde{\tau}(a) \circ x)\| \\
&\le \|\tilde{\tau}(a)\|^2 \|\rho^{-1} P_{\mathcal{B}}(x^* \circ x)\| \\
&= \|\tilde{\tau}(a)\|^2 \|\langle x, x\rangle\| \\
&= \|\tilde{\tau}(a)\|^2 \|x\|^2,
\end{aligned}$$

so that $\tau(a)$ is a well defined bounded linear map. Also for all $x, y \in E$ we have

$$\langle \tau(a)x, y\rangle = \rho^{-1} P_{\mathcal{B}}((\tilde{\tau}(a) \circ x)^* \circ y) = \rho^{-1} P_{\mathcal{B}}(x^* \circ \tilde{\tau}(a)^* \circ y) = \langle x, \tilde{\tau}(a)^* \circ y\rangle,$$

so that $\tau(a) \in \mathcal{B}^{\mathrm{a}}(E)$ with $\tau(a)^* y = \tilde{\tau}(a)^* \circ y$. Also for all $a, b, c \in \mathcal{A}$ and $x \in E$ we have

$$\tau(a)\tau(b)x = \tau(a)(\tilde{\tau}(b) \circ x) = \tilde{\tau}(a) \circ \tilde{\tau}(b) \circ x = \tilde{\tau}(ab) \circ x = \tau(ab)x$$

and

$$\tau(a)\tau(b)^*\tau(c)x = \tilde{\tau}(a) \circ \tilde{\tau}(b)^* \circ \tilde{\tau}(c) \circ x = t\tilde{\tau}(ab^*c) \circ x = t\tau(ab^*c)x,$$

so that $\tau(a)\tau(b) = \tau(ab)$ and $\tau(a)\tau(b)^*\tau(c) = t\tau(ab^*c)$. Thus $a \mapsto \tau(a)$ defines a t-ternary homomorphism $\tau : \mathcal{A} \to \mathcal{B}^{\mathrm{a}}(E)$. Now if we set $z = W \in E$, then for $a \in \mathcal{A}$ we have

$$\begin{aligned}
\langle z, \tau(a)z\rangle &= \rho^{-1} P_{\mathcal{B}}(W^* \circ \tilde{\tau}(a) \circ W) \\
&= \rho^{-1} P_{\mathcal{B}}(\tilde{\psi}(a)) \\
&= \rho^{-1} P_{\mathcal{B}}(\rho \circ \psi(a)) \\
&= \rho^{-1} \circ \rho \circ \psi(a) \\
&= \psi(a)
\end{aligned}$$

and hence $\|\psi\|_{cb} \le \|\tau\|_{cb}\|z\|^2 = \sqrt{t}\|z\|^2$. Also

$$\left(\frac{(t-1)\sqrt{t-1}}{2\sqrt{t-1}+2t-1}\right)\|z\|^2 = \left(\frac{(t-1)\sqrt{t-1}}{2\sqrt{t-1}+2t-1}\right)\|W\|^2 \le \|\tilde{\psi}\|_{cb} \le \|\psi\|_{cb}.$$

Also since $\tilde{\tau}$ is a $*$-nondegenerate t-ternary homomorphism, from Proposition 2.30 it follows that τ is also $*$-nondegenerate. □

In this case also we can have a universal representation. Fixing one regular representation for each $\psi \in CB(\mathcal{A}, \mathcal{B})$ and considering the direct sum of all such representations as in the proof of Theorem 3.3 we can have the following.

Theorem 3.7. *Suppose $\mathcal{B}$ is an injective C^*-algebra. There exists a Hilbert $\mathcal{B}$-module E, and a $*$-nondegenerate regular homomorphism $\tau : \mathcal{A} \to \mathcal{B}^{a}(E)$ such that given any $\psi \in CB(\mathcal{A}, \mathcal{B})$ there exists a vector $z_\psi \in E$ such that*

$$\psi(\,\cdot\,) = \langle z_\psi, \tau(\,\cdot\,)z_\psi\rangle.$$

Moreover, given any $t \in (1, \infty)$ we can choose τ and z_ψ such that τ is t-ternary and z_ψ satisfies $((t-1)\sqrt{t-1}/(2\sqrt{t-1}+2t-1))\|z_\psi\|^2 \le \|\psi\|_{cb} \le \sqrt{t}\|z_\psi\|^2$.

3B. *Commutant representations.* In this section we provide new and possibly simpler proofs of some known results for completely bounded maps. To begin with we give a different proof of the following result due to Paulsen and Suen [1985, Theorem 2.2]. Our proof involves mainly matrix manipulation.

Theorem 3.8. *Suppose $\psi : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ is a CB-map. Then there exists a Hilbert space $\mathcal{K}$, a unital representation $\pi : \mathcal{A} \to \mathcal{B}(\mathcal{K})$, an isometry $V : \mathcal{H} \to \mathcal{K}$ and a unique operator $T \in \pi(\mathcal{A})' \subseteq \mathcal{B}(\mathcal{K})$ such that*

$$\psi(\,\cdot\,) = V^*T\pi(\,\cdot\,)V \quad \text{and} \quad \overline{\mathsf{span}}\,\pi(\mathcal{A})V\mathcal{H} = \mathcal{K}.$$

Furthermore, $\|\psi\|_{cb} \le \|T\| \le 2\|\psi\|_{cb}$. If $\psi = \psi^*$, then $T = T^*$ and $\|\psi\|_{cb} = \|T\|$.

Proof. For nonzero ψ, replacing ψ by $\psi/\|\psi\|_{cb}$ if necessary, we may assume that $\|\psi\|_{cb} = 1$. Construct Φ as in Theorem 3.1 and let $(\widetilde{\mathcal{K}}, \Pi, \widetilde{V})$ be the minimal Stinespring dilation for Φ with $\widetilde{V}$ an isometry. Then from equation (3-1) we have

$$\psi(a) = \begin{bmatrix} \frac{I_{\mathcal{H}}}{\sqrt{2}} & \frac{I_{\mathcal{H}}}{\sqrt{2}} \end{bmatrix} \widetilde{V}^*\Pi\left(\begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}\right)\Pi\left(\begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}\right)\widetilde{V}\begin{bmatrix} I_{\mathcal{H}}/\sqrt{2} \\ I_{\mathcal{H}}/\sqrt{2} \end{bmatrix} = V^*\widetilde{T}\tilde{\pi}(a)V, \tag{3-3}$$

where $V = \widetilde{V}\begin{bmatrix} I_{\mathcal{H}}/\sqrt{2} \\ I_{\mathcal{H}}/\sqrt{2} \end{bmatrix} \in \mathcal{B}(\mathcal{H}, \widetilde{\mathcal{K}})$ is an isometry, $\widetilde{T} = \Pi\left(\left[\begin{smallmatrix} 0 & 2 \\ 0 & 0 \end{smallmatrix}\right]\right) \in \mathcal{B}(\mathcal{K})$ and $\tilde{\pi} : \mathcal{A} \to \mathcal{B}(\widetilde{\mathcal{K}})$ is the unital representation given by $\tilde{\pi}(a) := \Pi\left(\left[\begin{smallmatrix} a & 0 \\ 0 & a \end{smallmatrix}\right]\right)$. Clearly

$$\widetilde{T}\tilde{\pi}(a) = \Pi\left(\begin{bmatrix} 0 & 2a \\ 0 & 0 \end{bmatrix}\right) = \tilde{\pi}(a)\widetilde{T}$$

for all $a \in \mathcal{A}$ so that $\widetilde{T} \in \tilde{\pi}(\mathcal{A})' \subseteq \mathcal{B}(\widetilde{\mathcal{K}})$. Set $\mathcal{K} = \overline{\operatorname{span}}\, \tilde{\pi}(\mathcal{A})V\mathcal{H} \subseteq \widetilde{\mathcal{K}}$ and $T = P_{\mathcal{K}}\widetilde{T}|_{\mathcal{K}} \in \mathcal{B}(\mathcal{K})$ where $P_{\mathcal{K}}$ is the orthogonal projection of $\widetilde{\mathcal{K}}$ onto $\mathcal{K}$. Note that $\tilde{\pi}(a)$ reduces $\mathcal{K}$ for all $a \in \mathcal{A}$. Then $\pi : \mathcal{A} \to \mathcal{B}(\mathcal{K})$ given by $\pi(a) = \tilde{\pi}(a)|_{\mathcal{K}}$ defines a unital representation such that

$$\tilde{\pi}(a) = \begin{bmatrix} \pi(a) & 0 \\ 0 & * \end{bmatrix} \in \mathcal{B}(\widetilde{\mathcal{K}}) = \mathcal{B}(\mathcal{K} \oplus \mathcal{K}^{\perp}).$$

So $\widetilde{T} = \left[\begin{smallmatrix} T & * \\ * & * \end{smallmatrix}\right] \in \tilde{\pi}(\mathcal{A})' \subseteq \mathcal{B}(\mathcal{K} \oplus \mathcal{K}^{\perp})$ implies that $T\pi(a) = \pi(a)T$ for all $a \in \mathcal{A}$. That is, $T \in \pi(\mathcal{A})' \subseteq \mathcal{B}(\mathcal{K})$. Since $\tilde{\pi}$ is unital we have $V\mathcal{H} \subseteq \mathcal{K}$, i.e., $V \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, hence $\mathcal{K} = \overline{\operatorname{span}}\, \tilde{\pi}(\mathcal{A})V\mathcal{H} = \overline{\operatorname{span}}\, \pi(\mathcal{A})V\mathcal{H}$. Also,

$$\psi(a) = V^*\widetilde{T}\tilde{\pi}(a)V = \begin{bmatrix} V^* & 0 \end{bmatrix} \begin{bmatrix} T & * \\ * & * \end{bmatrix} \begin{bmatrix} \pi(a) & 0 \\ 0 & * \end{bmatrix} \begin{bmatrix} V \\ 0 \end{bmatrix} = V^*T\pi(a)V$$

for all $a \in \mathcal{A}$. Clearly $\|\psi\|_{cb} \leq \|T\| \leq \|\widetilde{T}\| \leq 2$ gives the required bounds.

To see uniqueness of T, suppose there exists another operator $S \in \pi(\mathcal{A})'$ such that $\psi(\cdot) = V^*S\pi(\cdot)V$. Then

$$\begin{aligned} \langle \pi(a_1)Vh_1, (T-S)\pi(a_2)Vh_2\rangle &= \langle h_1, V^*\pi(a_1^*)T\pi(a_2)Vh_2\rangle - \langle h_1, V^*\pi(a_1^*)S\pi(a_2)Vh_2\rangle \\ &= \langle h_1, V^*T\pi(a_1^*a_2)Vh_2\rangle - \langle h_1, V^*S\pi(a_1^*a_2)Vh_2\rangle \\ &= \langle h_1, \psi(a_1^*a_2)h_2\rangle - \langle h_1, \psi(a_1^*a_2)h_2\rangle \\ &= 0 \end{aligned}$$

for all $a_i \in \mathcal{A},\ h_i \in \mathcal{H}$ so that $T - S = 0$.

Finally if $\psi = \psi^*$, observe

$$V^*T\pi(a)V = \psi(a) = \psi^*(a) = \psi(a^*)^* = (V^*T\pi(a^*)V)^* = V^*T^*\pi(a)V,$$

and by the uniqueness property we have $T = T^*$. Note that $\frac{1}{2}(\widetilde{T} + \widetilde{T}^*) = \left[\begin{smallmatrix} T & * \\ * & * \end{smallmatrix}\right]$, so that $\|T\| \leq \left\|\frac{1}{2}(\widetilde{T} + \widetilde{T}^*)\right\| = \frac{1}{2}\left\|\Pi\left(\left[\begin{smallmatrix} 0 & 2 \\ 0 & 0 \end{smallmatrix}\right]\right) + \Pi\left(\left[\begin{smallmatrix} 0 & 0 \\ 2 & 0 \end{smallmatrix}\right]\right)\right\| = 1 = \|\psi\|_{cb}$. $\square$

For CB-maps from unital C^*-algebras into injective C^*-algebras Heo [1999] gave an analogue of Theorem 3.8. For a Hilbert $\mathcal{B}$ module E, consider $E^{\sharp} := \mathcal{B}_{\mathcal{B}}(E, \mathcal{B})$, the set of all bounded $\mathcal{B}$-module maps from E into $\mathcal{B}$. It forms a right $\mathcal{B}$-module with the following operations:

$$(\phi_1 + \phi_2)(x) := \phi_1(x) + \phi_2(x), \quad (\lambda\phi)(x) := \bar{\lambda}\phi(x), \quad (\phi b)(x) := b^*\phi(x)$$

for all $x \in E, b \in \mathcal{B}, \lambda \in \mathbb{C}$ and $\phi, \phi_i \in E^{\sharp}$. Also the operator norm makes $E^{\sharp}$ a Banach $\mathcal{B}$-module. With this notation, the theorem of Heo states the following.

Theorem 3.9. *Suppose $\mathcal{B}$ is an injective C^*-algebra and $\psi : \mathcal{A} \to \mathcal{B}$ is a CB-map. Then there exists a Hilbert $\mathcal{B}$-module E, a vector $z \in E$, a representation*

$\pi : \mathcal{A} \to \mathcal{B}^{\mathrm{a}}(E)$ and a unique operator $T \in \mathcal{B}_{\mathcal{B}}(E, E^{\sharp}) \cap \pi(\mathcal{A})'$ such that

$$\psi(\,\cdot\,) = \langle z, T\pi(\,\cdot\,)z\rangle \quad \text{and} \quad \overline{\operatorname{span}}\,\pi(\mathcal{A})z\mathcal{B} = E.$$

Heo proved this result using a structure theorem for so-called "*completely multi-positive*" linear maps. Our proof is straightforward and we also get some norm estimates.

Theorem 3.10. *Suppose $\mathcal{B}$ is an injective C^*-algebra and $\psi : \mathcal{A} \to \mathcal{B}$ is a CB-map. Then there exists a quadruple (E, z, π, T) consisting of a Hilbert $\mathcal{B}$-module E, a unit vector $z \in E$, a unital representation $\pi : \mathcal{A} \to \mathcal{B}^{\mathrm{a}}(E)$, and an operator $T \in \pi(\mathcal{A})' \subseteq \mathcal{B}^{\mathrm{a}}(E)$ with $\|\psi\|_{cb} \le \|T\| \le 2\|\psi\|_{cb}$ such that*

$$\psi(\,\cdot\,) = \langle z, T\pi(\,\cdot\,)z\rangle.$$

If $\psi = \psi^*$, then $T = T^*$ and $\|\psi\|_{cb} = \|T\|$. Furthermore, if $\overline{\operatorname{span}}\,\pi(\mathcal{A})z\mathcal{B} = E$, then T is unique.

Proof. Consider a unital faithful representation $\rho : \mathcal{B} \to \mathcal{B}(\mathcal{G})$ of $\mathcal{B}$ on some Hilbert space $\mathcal{G}$ satisfying $\overline{\operatorname{span}}\,\rho(\mathcal{B})\mathcal{G} = \mathcal{G}$. By the assumption of injectivity, there is a conditional expectation map $P_{\mathcal{B}} : \mathcal{B}(\mathcal{G}) \to \rho(\mathcal{B})$. Suppose $(\mathcal{K}, \tilde{\pi}, \widetilde{T}, V)$ is a commutant representation of the CB-map $\tilde{\psi} := \rho \circ \psi : \mathcal{A} \to \mathcal{B}(\mathcal{G})$ given by Theorem 3.8. From $\mathcal{B}(\mathcal{G}, \mathcal{K})$, with $z = V$, construct the triple (E, π, z) as in the proof of Theorem 3.6. Note that since $\tilde{\pi}$ is a unital representation so is π. Also z is a unit vector since

$$\langle z, z\rangle = \rho^{-1}P_{\mathcal{B}}(V^* \circ V) = \rho^{-1}P_{\mathcal{B}}(I) = \rho^{-1}P_{\mathcal{B}}\rho(1) = \rho^{-1}\rho(1) = 1.$$

Define $T : E \to E$ by $T(x) = \widetilde{T} \circ x$ for all $x \in E$. It can be verified that T is well defined and $T \in \mathcal{B}^{\mathrm{a}}(E)$ with $T^*(x) = \widetilde{T}^* \circ x$ for all $a \in E$. Note that

$$T\pi(a)x = \widetilde{T} \circ \tilde{\pi}(a) \circ x = \tilde{\pi}(a) \circ \widetilde{T} \circ x = \pi(a)Tx$$

for all $a \in \mathcal{A},\ x \in E$ so that $T \in \pi(\mathcal{A})' \subseteq \mathcal{B}^{\mathrm{a}}(E)$. Also,

$$\langle z, T\pi(a)z\rangle = \rho^{-1}P_{\mathcal{B}}(V^* \circ \widetilde{T} \circ \tilde{\pi}(a) \circ V) = \rho^{-1}P_{\mathcal{B}}(\tilde{\psi}(a)) = \psi(a)$$

for all $a \in \mathcal{A}$. Now it follows that $\|\psi\|_{cb} \le \|T\| \le \|\widetilde{T}\| \le 2\|\tilde{\psi}\|_{cb} \le 2\|\psi\|_{cb}$ since z is a unit vector and π is a unital representation. Now if $\psi = \psi^*$, then for all $a \in \mathcal{A}$,

$$\tilde{\psi}^*(a) = \tilde{\psi}(a^*)^* = \rho(\psi(a^*))^* = \rho(\psi(a^*)^*) = \rho(\psi^*(a)) = \rho(\psi(a)) = \tilde{\psi}(a),$$

so that $\tilde{\psi} = \tilde{\psi}^*$ and hence $\widetilde{T} = \widetilde{T}^*$. Therefore $T = T^*$. Also $\|T\|_{cb} \le \|\widetilde{T}\|_{cb} = \|\tilde{\psi}\|_{cb} = \|\psi\|_{cb}$.

Uniqueness: Suppose $\overline{\operatorname{span}}\,\pi(\mathcal{A})z\mathcal{B} = E$. Now if $S \in \pi(A)'$ any other operator such that $\pi(\,\cdot\,) = \langle z, S\pi(\,\cdot\,)z\rangle$, then $\langle \pi(a)zb, (T - S)\pi(a')zb'\rangle = 0$ for all $a, a' \in \mathcal{A}$ and $b, b' \in \mathcal{B}$, so that $T - S = 0$. $\square$

3C. ***Representations of CB-maps: One from another.*** In this section we see how different representations of CB-maps $\psi : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ are related each other. Since the results are straightforward we do not provide proofs.

Proposition 3.11 (Commutant representation from regular representation I). *Let $\psi : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ be a CB-map with a regular representation $(\mathcal{K}, \tau, W)$, that is, $\tau : \mathcal{A} \to \mathcal{B}(\mathcal{K})$ is a $*$-nondegenerate regular homomorphism and $W \in \mathcal{B}(\mathcal{H}, \mathcal{K})$ such that $\psi(\,\cdot\,) = W^*\tau(\,\cdot\,)W$. Suppose $\vartheta : \mathcal{A} \to \mathcal{B}(\mathcal{K})$ is the unique unital $*$-homomorphism such that $\tau(\,\cdot\,) = \vartheta(\,\cdot\,)T = T\vartheta(\,\cdot\,)$, where $T = \tau(1)$. Then $(\mathcal{K}, \vartheta, T, W)$ is a commutant representation for ψ.*

Note that W of this proposition may not be an isometry. This can be taken care of as follows:

Proposition 3.12 (Commutant representation from regular representation II). *Let $\psi : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ be a nonzero CB-map. Let $(\mathcal{K}, \tau, V)$ be a regular representation for $\hat{\psi} = \psi/\|\psi\|_{cb}$ with V as an isometry. Choose a (not necessarily unital) $*$-homomorphism $\vartheta : \mathcal{A} \to \mathcal{B}(\mathcal{K})$ such that $\tau(\,\cdot\,) = \vartheta(\,\cdot\,)\tau(1) = \tau(1)\vartheta(\,\cdot\,)$. Then $T = \|\psi\|_{cb}\tau(1) \in \tau(\mathcal{A})' \subseteq \mathcal{B}(\mathcal{K})$ is such that $\psi(\,\cdot\,) = \|\psi\|_{cb}V^*\tau(\,\cdot\,)V = V^*T\vartheta(\,\cdot\,)V$, so that $(\mathcal{K}, \vartheta, T, V)$ is a commutant representation for ψ.*

The drawback of the previous representation is that the $*$-homomorphism ϑ may not be unital.

Proposition 3.13 (Regular representation from commutant representation I). *Suppose (K, π, T, V) is a commutant representation of a CB-map $\psi : \mathcal{A} \to \mathcal{B}(\mathcal{H})$. Set $\widehat{\mathcal{K}} = \mathcal{K} \oplus \mathcal{K}$. Define $\tau : \mathcal{A} \to \mathcal{B}(\widehat{\mathcal{K}})$ by*

$$\tau(a) = \begin{bmatrix} \pi(a) & (2T - I)\pi(a) \\ 0 & 0 \end{bmatrix} \quad \text{and set} \quad W = \begin{bmatrix} V/\sqrt{2} \\ V/\sqrt{2} \end{bmatrix} \in \mathcal{B}(\mathcal{H}, \widehat{\mathcal{K}}).$$

Then τ is a regular homomorphism and W is an isometry such that $\psi(\,\cdot\,) = W^*\tau(\,\cdot\,)W$.

We may prefer to get a t-ternary representation instead of just a regular representation. This can be achieved as follows:

Proposition 3.14 (Regular representation from commutant representation II). *Suppose (K, π, T, V) is a commutant representation of a CB-map $\psi : \mathcal{A} \to \mathcal{B}(\mathcal{H})$. Set $\widehat{\mathcal{K}} = \mathcal{K} \oplus \mathcal{K}$. Given any $t \in (1, \infty)$ define $\tau : \mathcal{A} \to \mathcal{B}(\widehat{\mathcal{K}})$ by*

$$\tau(a) = \begin{bmatrix} \pi(a) & 0 \\ -\sqrt{t-1}\,\pi(a) & 0 \end{bmatrix} \quad \text{and set} \quad W = \begin{bmatrix} 0 & I \\ \frac{I-T^*}{\sqrt{t-1}} & 0 \end{bmatrix} \begin{bmatrix} V \\ V \end{bmatrix} \in \mathcal{B}(\widehat{\mathcal{K}}).$$

Then τ is a $*$-nondegenerate t-ternary homomorphism. Also $\psi(\,\cdot\,) = W^*\tau(\,\cdot\,)W$, so that $(\widehat{\mathcal{K}}, \tau, W)$ is a regular representation of ψ.

Finally we show that any regular representation also gives another familiar representation called the fundamental representation for completely bounded maps (Theorem 8.4 of [Paulsen 2002]):

Proposition 3.15 (Fundamental representation from regular representation). *Suppose $\psi : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ is a CB-map with regular representation $(\mathcal{K}, \tau, W)$. Let $\vartheta : \mathcal{A} \to \mathcal{B}(\mathcal{K})$ be the $*$-homomorphism such that $\tau(\,\cdot\,) = \vartheta(\,\cdot\,)\tau(1) = \tau(1)\vartheta(\,\cdot\,)$. Then $V_1 := W$ and $V_2 := \tau(1)W$ are elements of $\mathcal{B}(\mathcal{H}, \mathcal{K})$ such that*

$$\psi(\,\cdot\,) = W^*\tau(\,\cdot\,)W = V_1^*\vartheta(\,\cdot\,)V_2.$$

Acknowledgement

Bhat and Mallick are supported by UGC under India–Israel Joint Research Project 2014: *E_0-semigroups: classification and invariants.* Mallick is thankful to the National Board for Higher Mathematics (NBHM), India for a senior research fellowship. This work is a part of the PhD thesis she submitted to the Indian Statistical Institute. This research work was done when Sumesh was at IMSc Chennai, and he thanks the Department of Atomic Energy (DAE), Government of India for financial support and IMSc Chennai for providing necessary facilities.

References

[Arveson 1969] W. B. Arveson, "Subalgebras of C^*-algebras", *Acta Math.* **123** (1969), 141–224. MR Zbl

[Haagerup 1980] U. Haagerup, "Decomposition of completely bounded maps on operator algebras", unpublished manuscript, 1980.

[Heo 1999] J. Heo, "Completely multi-positive linear maps and representations on Hilbert C^*-modules", *J. Operator Theory* **41**:1 (1999), 3–22. MR Zbl

[Lance 1994] E. C. Lance, "Unitary operators on Hilbert C^*-modules", *Bull. London Math. Soc.* **26**:4 (1994), 363–366. MR Zbl

[Muhly et al. 2006] P. S. Muhly, M. Skeide, and B. Solel, "Representations of $\mathcal{B}^a(E)$", *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* **9**:1 (2006), 47–66. MR Zbl

[Paschke 1973] W. L. Paschke, "Inner product modules over B^*-algebras", *Trans. Amer. Math. Soc.* **182** (1973), 443–468. MR Zbl

[Paulsen 1986] V. I. Paulsen, *Completely bounded maps and dilations*, Pitman Research Notes in Mathematics Series **146**, Longman Scientific, Harlow, UK, 1986. MR Zbl

[Paulsen 2002] V. Paulsen, *Completely bounded maps and operator algebras*, Cambridge Studies in Advanced Mathematics **78**, Cambridge Univ. Press, 2002. MR Zbl

[Paulsen and Suen 1985] V. I. Paulsen and C. Y. Suen, "Commutant representations of completely bounded maps", *J. Operator Theory* **13**:1 (1985), 87–101. MR Zbl

[Pisier 2001] G. Pisier, *Similarity problems and completely bounded maps*, 2nd ed., Lecture Notes in Mathematics **1618**, Springer, Berlin, 2001. MR Zbl

[Skeide 2000] M. Skeide, "Generalised matrix C^*-algebras and representations of Hilbert modules", *Math. Proc. R. Ir. Acad.* **100A**:1 (2000), 11–38. MR Zbl

[Skeide and Sumesh 2014] M. Skeide and K. Sumesh, “CP-H-extendable maps between Hilbert modules and CPH-semigroups”, *J. Math. Anal. Appl.* **414**:2 (2014), 886–913. MR Zbl

[Stinespring 1955] W. F. Stinespring, “Positive functions on C^*-algebras”, *Proc. Amer. Math. Soc.* **6** (1955), 211–216. MR Zbl

[Suen 1991] C. Y. Suen, “On representations of completely bounded maps”, *J. Math. Anal. Appl.* **154**:1 (1991), 212–225. MR Zbl

[Wittstock 1981] G. Wittstock, “Ein operatorwertiger Hahn–Banach Satz”, *J. Funct. Anal.* **40**:2 (1981), 127–150. MR Zbl

[Wittstock 1984] G. Wittstock, “On matrix order and convexity”, pp. 175–188 in *Functional analysis: surveys and recent results, III* (Paderborn, 1983), edited by B. Fuchssteiner, North-Holland Math. Stud. **90**, North-Holland, Amsterdam, 1984. MR Zbl

Received October 6, 2015.

B. V. RAJARAMA BHAT
INDIAN STATISTICAL INSTITUTE BANGALORE
R.V. COLLEGE POST
8TH MILE MYSORE ROAD
BANGALORE 560059
INDIA
bhat@isibang.ac.in

NIRUPAMA MALLICK
INDIAN STATISTICAL INSTITUTE BANGALORE
R.V. COLLEGE POST
8TH MILE MYSORE ROAD
BANGALORE 560059
INDIA
niru.mallick@gmail.com

K. SUMESH
DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY MADRAS
SARDAR PATEL ROAD, ADYAR
IIT POST
CHENNAI 600036
INDIA
sumeshkpl@iitm.ac.in

PACIFIC JOURNAL OF MATHEMATICS
Vol. 289, No. 2, 2017
dx.doi.org/10.2140/pjm.2017.289.287

BALL CONVEX BODIES IN MINKOWSKI SPACES

THOMAS JAHN, HORST MARTINI AND CHRISTIAN RICHTER

To our teachers, colleagues and friends, Prof. Dr. Johannes Böhm, on the occasion of his 90th birthday, and Prof. Dr. Eike Hertel, on the occasion of his 75th birthday.

The notion of ball convexity, considered in finite-dimensional real Banach spaces, is a natural and useful extension of usual convexity; one replaces intersections of half-spaces by suitable intersections of balls. A subset S of a normed space is called ball convex if it coincides with its ball hull, which is obtained as the intersection of all balls (of fixed radius) containing S. Ball convex sets are closely related to notions like ball polytopes, complete sets, bodies of constant width, and spindle convexity. We will study geometric properties of ball convex bodies in normed spaces, for example deriving separation theorems, characterizations of strictly convex norms, and an application to complete sets. Our main results refer to minimal representations of ball convex bodies in terms of their ball exposed faces, to representations of ball hulls of sets via unions of ball hulls of finite subsets, and to ball convexity of increasing unions of ball convex bodies.

1. Introduction

It is well known that generalized convexity notions are helpful for solving various (metrical) problems from non-Euclidean geometries in an elegant way. For example, Menger's notion of d-segments, yielding that of d-convex sets (see Chapter II of [Boltyanski et al. 1997]), is a useful tool for solving location problems in finite-dimensional real Banach spaces (see [Martini et al. 2002]). Another example, also referring to normed spaces, is the notion of ball convexity: usual convexity is extended by considering suitably defined intersections of balls instead of intersections of half-spaces. The ball hull of a given point set S is the intersection of all balls (of fixed radius) which contain S, and S is called ball convex if it coincides with its ball hull. Ball convex sets are strongly related to notions from several recent research topics, such as ball polytopes, applications of spindle convexity, bodies of

MSC2010: 46B20, 52A01, 52A20, 52A21, 52A35.
Keywords: ball convex body, ball hull, ball polytope, b-exposed point, b-face, Carathéodory's theorem, circumball, complete set, exposed b-face, Minkowski space, normed space, separation theorem, spindle convexity, strictly convex norm, supporting sphere.

constant width, and diametrically maximal (or complete) sets. In this article we study geometric properties and (minimal) representations of ball convex bodies in normed spaces. In terms of ball convexity and related notions, we derive separation properties of ball convex bodies, various characterizations of strictly convex norms, and an application for diametrically maximal sets, which answers a question from [Martini et al. 2014]. Introducing suitable notions describing the boundary structure of ball convex bodies, our main results refer to minimal representations of ball convex bodies, particularly in terms of their ball exposed faces. More precisely, we extend the formula $K = \mathrm{cl}(\mathrm{conv}(\exp(K)))$ from classical convexity (where K is a convex body in $\mathbb{R}^n$) to the concept of ball convexity in normed spaces. On the other hand, we derive theorems on the representation of ball convex bodies "from inside". That is, we show that unions of increasing sequences of ball convex bodies are, essentially, ball convex, and we present ball hulls of sets by unions of ball hulls of finite subsets. In that context we solve a problem from [Lángi et al. 2013]. We finish with some open questions inspired by the notions of ball hull and ball convexity; they refer to spindle convex sets and generalized Minkowski spaces (whose unit balls need not be centered at the origin).

We will give now a brief survey on what has been done regarding ball convexity and related notions. Intersections of finitely many congruent Euclidean balls were studied in [Bieberbach 1955; 1970; Martini and Swanepoel 2004], and in three dimensions in [Heppes 1956; Heppes and Révész 1956; Straszewicz 1957; Grünbaum 1956]. The notions of ball hull and ball convexity have been considered by various authors, defining them via intersections of balls of some fixed radius $R > 0$ and calling this concept also R-convexity; see, e.g., [Bezdek et al. 2006; Bezdek and Naszódi 2006; Kupitz et al. 2010; Lángi et al. 2013]. In view of this concept, bodies of constant width, Hausdorff limits, Minkowski sums, and approximation properties of R-convex sets (see [Montejano 1991; Polovinkin 1996a; 1996b; Polovinkin and Balashov 2007], respectively) are investigated. Analogues of the Krein–Milman theorem and of Carathéodory's theorem (see [Polovinkin 1996b; 1997]) are also considered, but only for the Euclidean norm. Not much has been done for normed spaces; however, for related results we refer to [Balashov 2002] for Hilbert spaces and to [Balashov and Polovinkin 2000; Alimov 2012; Balashov and Ivanov 2006; Martini and Spirova 2009] for normed spaces. A recent contribution is [Lángi et al. 2013], referring, e.g., to the Banach–Mazur distance and Hadwiger illumination numbers of sets being ball convex in the sense described here.

Closely related is the concept of ball polytopes. It was investigated in [Bezdek et al. 2007; Kupitz et al. 2010; Papez 2010] (but see also [Polovinkin 1996b; Bezdek 2010, Chapter 6; Bezdek 2013, Chapter 5]). The boundary structure of ball polytopes is interesting (digonal facets can occur, and hence their edge-graphs are different from usual polyhedral edge-graphs), their properties are also useful for constructing bodies of constant width, and analogues of classical theorems like those of Carathéodory and Steinitz on linear convex hulls are proved in these papers.

The study of the related notion of spindle convexity (also called hyperconvexity or K-convexity) was initiated by Mayer [1935]; see also [Meissner 1911] and, for Minkowski spaces, [Valentine 1964, p. 99]. The definition is given in Section 8 below. For a discussion of this notion we refer to the survey [Danzer et al. 1963, p. 160] and, for further results and references in the spirit of abstract convexity and combinatorial geometry, to [Bezdek et al. 2007; Lángi et al. 2013; Papez 2010; Bezdek 2012; Fodor and Vígh 2012; Bezdek 2013, Chapters 5 and 6]. In [Bezdek and Naszódi 2015] this notion was extended to analogues of star-shaped sets.

To avoid confusion, we briefly mention another concept which is also called ball convexity. Namely, in [Lassak 1977] a set is called ball convex if, with any finite number of points, it contains the intersection of all balls (of arbitrary radii) containing the points. The ball convex hull of a set S is again defined as the intersection of all ball convex sets containing S. In [Lassak 1977; 1979] this notion was investigated for normed spaces, and in [Lassak 1982] the relations of these notions to metric or d-convexity were investigated. The ball hull mapping studied for Banach spaces in [Moreno and Schneider 2007a; 2007b] is also related.

2. Definitions and notations

Let $\mathcal{K}^n = \{S \subseteq \mathbb{R}^n : S$ is compact, convex, and nonempty$\}$ be the set of all *convex bodies* in $\mathbb{R}^n$ (thus, in our terminology, a convex body need not have interior points). Let $B \in \mathcal{K}^n$ be centered at the origin o of $\mathbb{R}^n$ and have nonempty interior. We denote by $(\mathbb{R}^n, \|\cdot\|)$ the n-dimensional *normed* or *Minkowski space* with unit ball B, i.e., the n-dimensional real Banach space whose *norm* is given by $\|x\| = \min\{\lambda \geq 0 : x \in \lambda B\}$. Any homothetic copy $B(x, r)$, $x \in \mathbb{R}^n$ and $r \geq 0$, of B is a *closed ball* of $(\mathbb{R}^n, \|\cdot\|)$ with center x and radius r; therefore we write $B(o, 1)$ from now on for the *unit ball of* $(\mathbb{R}^n, \|\cdot\|)$. The boundary of the ball $B(x, r)$ is the *sphere* $S(x, r)$, and therefore $S(o, 1)$ denotes the *unit sphere* of our Minkowski space. Note that we will use the symbol S for an arbitrarily given point set in $\mathbb{R}^n$. For a compact S, we write $\mathrm{dist}(x, S) = \min\{\|x - y\| : y \in S\}$ for the *distance of* x *and* S, and we denote by $\mathrm{rad}(S)$ the *circumradius* of S, i.e., the radius of any *circumball* (or minimal enclosing ball) of S, whose existence is assured by the boundedness of S. The *diameter* of S is given by $\mathrm{diam}(S) = \max\{\|x - y\| : x, y \in S\}$. The triangle inequality yields the left-hand side of

$$\tfrac{1}{2}\,\mathrm{diam}(S) \leq \mathrm{rad}(S) \leq n/(n+1)\ \mathrm{diam}(S), \tag{1}$$

and we refer to [Bohnenblust 1938, Theorem 6] for the right-hand side.

As usual, we use the abbreviations $\mathrm{int}(S)$, $\mathrm{cl}(S)$, $\mathrm{bd}(S)$, $\mathrm{conv}(S)$, and $\mathrm{aff}(S)$ for the *interior, closure, boundary, convex hull*, and *affine hull* of S, respectively. We write $[x_1, x_2]$ for the *closed segment* with endpoints $x_1, x_2 \in \mathbb{R}^n$, and, analogously, (a, b), $(a, b]$, $[a, b]$ are used for the *open*, *half-open* or *closed interval* with $a, b \in \mathbb{R}$, respectively. We use $|\cdot|$ for the *cardinality* of a set.

A convex body is called *strictly convex* if its boundary does not contain proper segments; analogously, $\|\cdot\|$ is called a *strictly convex norm* if the respective unit ball is strictly convex.

Since we want to derive results for generalized convexity notions, the following definitions yield direct analogues of notions from classical convexity; see [Schneider 1993]. The first of them is an analogue of the (closed) convex hull. Namely, the *ball hull* of a set S is defined by

$$\mathrm{bh}_1(S) = \bigcap_{S \subseteq B(x,1)} B(x, 1).$$

A formally clearer expression would be $\mathrm{bh}_1(S) = \bigcap_{x \in \mathbb{R}^n \colon S \subseteq B(x,1)} B(x, 1)$, but we assume that the above shorter notation, as well as similar ones in the sequel, will not cause confusion. (We underline once more that here and below we use balls of radius 1.) A *ball convex* (*b-convex*) *set* S is characterized by $S = \mathrm{bh}_1(S)$ or, equivalently, by the property that S is an intersection of closed balls of radius 1 (then S is necessarily closed and convex). A *b-convex body* K is a bounded nonempty b-convex set (the analogue of a convex body in classical convexity); $\varnothing$ and $\mathbb{R}^n$ are the only b-convex sets that are not b-convex bodies. (Note that $\mathbb{R}^n$ is b-convex, since we want to understand the intersection of an empty family of sets as $\mathbb{R}^n$.) A *supporting sphere* $S(x, 1)$ of K is characterized by $K \subseteq B(x, 1)$ and $K \cap S(x, 1) \neq \varnothing$; the corresponding *exposed b-face* (or *b-support set*) is $K \cap S(x, 1)$ (note that nonempty *facets* from [Kupitz et al. 2010, Definition 5.3] are a special case).

If an exposed b-face is a singleton $\{x_0\}$, then x_0 is called a *b-exposed point of* K, and $\text{b-exp}(K)$ denotes the set of all b-exposed points. We note that several such concepts, referring to the analogous notions for ball polytopes, their boundary structure, separation properties with respect to spheres etc., can be found in [Bezdek 2012; Kupitz et al. 2010], but are defined there only for the subcase of the Euclidean norm. Finally, a set S is called *b-bounded* if $\mathrm{rad}(S) < 1$. This means that S is inside a ball of radius 1 and separated from its bounding sphere, which plays the role of a hyperplane in classical convexity.

We close this section by summarizing several basic facts about ball hulls and circumradii, and we give a lemma on intersections of compact sets with the boundaries of their circumballs.

Lemma 1. *Let* $(\mathbb{R}^n, \|\cdot\|)$ *be a Minkowski space. The following are satisfied for all* $S, T \subseteq \mathbb{R}^n$ *and* $x \in \mathbb{R}^n$:

(a) $S \subseteq \mathrm{cl}(S) \subseteq \mathrm{cl}(\mathrm{conv}(S)) \subseteq \mathrm{bh}_1(S) = \mathrm{bh}_1(\mathrm{cl}(S)) = \mathrm{bh}_1(\mathrm{conv}(S)) = \mathrm{bh}_1(\mathrm{bh}_1(S))$.

(b) *If* $S \subseteq T$, *then* $\mathrm{bh}_1(S) \subseteq \mathrm{bh}_1(T)$.

(c) $B(x, r)$ *is a b-convex body for every* $r \in [0, 1]$.

(d) *If* $\mathrm{rad}(S) \leq 1$, *then* $\mathrm{rad}(\mathrm{bh}_1(S)) = \mathrm{rad}(S)$. *In particular*, $\mathrm{bh}_1(S)$ *is b-bounded if S is b-bounded.*

(e) *If S is closed and* $S \subseteq \mathrm{int}(B(x, r))$ *for some* $r > 0$, *then* $S \subseteq B(x, r')$ *for some* $r' \in (0, r)$ *and* $\mathrm{rad}(S) < r$. *In particular, a closed subset of* $\mathbb{R}^n$ *is b-bounded if and only if it is covered by an open ball of radius* 1.

Proof. Parts (a) and (b) are obvious; see [Martini et al. 2013, Lemma 1] for a collection of related statements.

For (c), the triangle inequality gives the following representation of $B(x, r)$ as an intersection of balls of radius 1: $B(x, r) = \bigcap_{\|y-x\| \leq 1-r} B(y, 1)$.

To see (d), first note that $\mathrm{rad}(S) \leq \mathrm{rad}(\mathrm{bh}_1(S))$ by (a). If $B(x, \mathrm{rad}(S))$ is a circumball of S, then $\mathrm{bh}_1(S) \subseteq \mathrm{bh}_1(B(x, \mathrm{rad}(S))) = B(x, \mathrm{rad}(S))$ by (b) and (c). Hence $\mathrm{rad}(\mathrm{bh}_1(S)) \leq \mathrm{rad}(B(x, \mathrm{rad}(S))) = \mathrm{rad}(S)$.

For (e), suppose that S contains at least two points. Consider the continuous function $f : S \to \mathbb{R}$, $f(y) = \mathrm{dist}(y, S(x, r)) = \mathrm{dist}(y, \mathbb{R}^n \setminus B(x, r))$. Since S is compact, f attains its minimum: $f(y) \geq f(y_0) \in (0, r)$ for all $y \in S$. This shows that $\mathrm{dist}(y, \mathbb{R}^n \setminus B(x, r)) \geq f(y_0)$ for all $y \in S$; i.e., $S \subseteq B(x, r')$, where $r' = r - f(y_0) \in (0, r)$. □

Lemma 2. *Let* $B(x_0, \mathrm{rad}(S))$ *be a circumball of a nonempty compact subset S of a Minkowski space* $(\mathbb{R}^n, \|\cdot\|)$. *Then* $\mathrm{rad}(S \cap S(x_0, \mathrm{rad}(S))) = \mathrm{rad}(S)$. *In particular, there exist* $x, x' \in S \cap S(x_0, \mathrm{rad}(S))$ *such that* $\|x - x'\| \geq (n+1)/n \ \mathrm{rad}(S)$.

Proof. Without loss of generality, we set $B(x_0, \mathrm{rad}(S)) = B(o, 1)$. Assume that, contrary to our claim, $\mathrm{rad}(S \cap S(o, 1)) < 1$. Then there exists $x_1 \in \mathbb{R}^n$ such that

$$S \cap S(o, 1) \subseteq \mathrm{int}(B(x_1, 1)). \tag{2}$$

Since $\mathrm{rad}(S) = 1$, Lemma 1(e) gives points

$$y_i \in S \setminus \mathrm{int}\Big(B\Big(\frac{1}{i}x_1, 1\Big)\Big), \quad i = 1, 2, \ldots. \tag{3}$$

Note that

$$(y_i)_{i=1}^{\infty} \subseteq S \setminus \mathrm{int}(B(x_1, 1)),$$

as $y_i \in S \setminus \mathrm{int}\big(B\big(\frac{1}{i}x_1, 1\big)\big) \subseteq B(o, 1) \setminus \mathrm{int}\big(B\big(\frac{1}{i}x_1, 1\big)\big)$ gives $\|y_i\| \leq 1$, $\|y_i - \frac{1}{i}x_1\| \geq 1$, and in turn

$$\begin{aligned}\|y_i - x_1\| &= \Big\| i\Big(y_i - \frac{1}{i}x_1\Big) - (i-1)y_i \Big\| \\ &\geq i \Big\| y_i - \frac{1}{i}x_1 \Big\| - (i-1)\|y_i\| \\ &\geq i - (i-1) = 1.\end{aligned}$$

Because $S \setminus \operatorname{int}(B(x_1, 1))$ is compact, $(y_i)_{i=1}^{\infty}$ has an accumulation point $y_0 \in S \setminus \operatorname{int}(B(x_1, 1))$. We know that $\|y_0\| \le 1$ from $S \subseteq B(o, 1)$, whereas (3) gives $\left\| y_i - \frac{1}{i} x_1 \right\| \ge 1$ and, by $i \to \infty$, $\|y_0\| \ge 1$. This way we see that

$$y_0 \in S \cap S(o, 1) \setminus \operatorname{int}(B(x_1, 1)),$$

which contradicts (2) and completes the proof of $\operatorname{rad}(S \cap S(x_0, \operatorname{rad}(S))) = \operatorname{rad}(S)$.

Now the additionally claimed existence of $x, x' \in S \cap S(x_0, \operatorname{rad}(S))$ such that $\|x - x'\| \ge (n+1)/n \operatorname{rad}(S)$ is a consequence of the right-hand estimate in (1) and the compactness of S. □

3. Separation properties

The following results on the separation of b-convex bodies and points by spheres are analogues of theorems on the separation by hyperplanes in classical convexity.

Proposition 3. *Let K be a b-convex body in a Minkowski space* $(\mathbb{R}^n, \|\cdot\|)$.

(a) *For every $x_0 \in \operatorname{bd}(K)$, there exists a supporting sphere $S(y_0, 1)$ of K such that $x_0 \in S(y_0, 1)$.*

(b) *For every $x_0 \in \mathbb{R}^n \setminus K$, there exists a supporting sphere $S(y_0, 1)$ of K such that $x_0 \notin B(y_0, 1)$.*

(c) *If K is b-bounded then, for every $x_0 \in \mathbb{R}^n \setminus K$, there exists a sphere of unit radius $S(y_0, 1)$ such that $K \subseteq \operatorname{int}(B(y_0, 1))$ and $x_0 \notin B(y_0, 1)$. In particular, $K \subseteq B(y_0, r)$ for some $r \in (0, 1)$.*

Proof. For proving (a), note that the assumption

$$x_0 \in \operatorname{bd}(K) = \operatorname{bd}(\operatorname{bh}_1(K)) = \operatorname{bd}\Big(\bigcap_{K \subseteq B(y,1)} B(y, 1)\Big)$$

yields the existence of a sequence $(y_i)_{i=1}^{\infty} \subseteq \mathbb{R}^n$ such that $K \subseteq B(y_i, 1)$ for all i and

$$0 = \lim_{i \to \infty} \operatorname{dist}(x_0, \mathbb{R}^n \setminus B(y_i, 1)) = \lim_{i \to \infty} (1 - \|x_0 - y_i\|).$$

By compactness, $(y_i)_{i=1}^{\infty}$ has a convergent subsequence, and we can assume that $\lim_{i \to \infty} y_i = y_0$ without loss of generality. Then the above observations imply $K \subseteq B(y_0, 1)$ and $\|x_0 - y_0\| = 1$, i.e., $x_0 \in S(y_0, 1)$. This is our claim.

For (b), we have $x_0 \notin K = \bigcap_{K \subseteq B(y,1)} B(y, 1)$. Hence there is $y_1 \in \mathbb{R}^n$ such that $K \subseteq B(y_1, 1)$ and $x_0 \notin B(y_1, 1)$. We consider the translated balls $B_\lambda := B(y_1 + \lambda(y_1 - x_0), 1)$, $\lambda \ge 0$. We know that $K \subseteq B_0$. Let $\lambda_0 \ge 0$ be maximal such that

$$K \subseteq B_\lambda \quad \text{for } 0 \le \lambda \le \lambda_0.$$

By the maximality of λ_0, $\mathrm{bd}(B_{\lambda_0}) = S(y_1+\lambda_0(y_1-x_0), 1) =: S(y_0, 1)$ is a supporting sphere of K. Moreover, $x_0 \notin B(y_0, 1)$, because $x_0 \notin B(y_1, 1)$ gives

$$\|x_0 - y_0\| = \|x_0 - (y_1 + \lambda_0(y_1 - x_0))\| = (1 + \lambda_0)\|x_0 - y_1\| > 1 + \lambda_0 \geq 1.$$

This proves (b).

For the proof of (c), the b-boundedness of K gives $y_1 \in \mathbb{R}^n$ such that $K \subseteq \mathrm{int}(B(y_1, 1))$. By (b), there is $y_2 \in \mathbb{R}^n$ with $K \subseteq B(y_2, 1)$ and $x_0 \notin B(y_2, 1)$. We can pick $\varepsilon \in (0, 1)$ small enough such that

$$x_0 \notin B(y_0, 1), \qquad \text{where } y_0 := y_2 + \varepsilon(y_1 - y_2).$$

Then we obtain

$$K \subseteq \mathrm{int}(B(y_0, 1)), \tag{4}$$

because, for arbitrary $x \in K$, the inclusions $K \subseteq \mathrm{int}(B(y_1, 1))$ and $K \subseteq B(y_2, 1)$ imply $\|x - y_1\| < 1$, $\|x - y_2\| \leq 1$, and in turn

$$\begin{aligned}\|x - y_0\| &= \|x - (y_2 + \varepsilon(y_1 - y_2))\| \\ &= \|\varepsilon(x - y_1) + (1 - \varepsilon)(x - y_2)\| \\ &\leq \varepsilon\|x - y_1\| + (1 - \varepsilon)\|x - y_2\| \\ &< \varepsilon + (1 - \varepsilon) = 1.\end{aligned}$$

Finally, (4) yields $K \subseteq B(y_0, r)$ for suitable $r \in (0, 1)$ by Lemma 1(e). □

Corollary 4. *Every b-convex body in a Minkowski space* $(\mathbb{R}^n, \|\cdot\|)$ *satisfies*

$$\mathrm{bd}(K) = \bigcup\bigl\{F : F \textit{ is an exposed b-face of } K\bigr\}.$$

Proof. Proposition 3(a) gives "$\subseteq$". The converse inclusion is implied by the definition of exposed b-faces. □

Proposition 3 gives rise to alternative representations of ball hulls.

Corollary 5. *Every b-bounded subset S of a Minkowski space* $(\mathbb{R}^n, \|\cdot\|)$ *satisfies*

$$\mathrm{bh}_1(S) = \bigcap_{S \subseteq \mathrm{int}(B(x,1))} B(x, 1) = \bigcap_{S \subseteq B(x,r),\, r<1} B(x, 1) = \bigcap_{S \subseteq B(x,r),\, r<1} B(x, r).$$

Proof. We assume that $S \neq \varnothing$ and put $A := \mathrm{bh}_1(S) = \bigcap_{S \subseteq B(x,1)} B(x, 1)$, $B := \bigcap_{S \subseteq \mathrm{int}(B(x,1))} B(x, 1)$, $C := \bigcap_{S \subseteq B(x,r),\, r<1} B(x, 1)$ and $D := \bigcap_{S \subseteq B(x,r),\, r<1} B(x, r)$. The inclusions $A \subseteq B \subseteq C$ and $D \subseteq C$ are trivial. It suffices to prove that $C \subseteq A$ and $A \subseteq D$.

For proving $C \subseteq A$, we consider an arbitrary $x_0 \in \mathbb{R}^n \setminus A$ and have to show that $x_0 \notin C$. Application of Proposition 3(c) to A, which is b-bounded by Lemma 1(d), and to x_0 gives $y_0 \in \mathbb{R}^n$ and $r \in (0, 1)$ such that $S \subseteq A \subseteq B(y_0, r)$ and $x_0 \notin B(y_0, 1)$. This yields $x_0 \notin C$.

For $A \subseteq D$, note that

$$S \subseteq B(x, r) \quad \Leftrightarrow \quad A \subseteq B(x, r).$$

Indeed, if $S \subseteq B(x, r)$, then $A = \mathrm{bh}_1(S) \subseteq \mathrm{bh}_1(B(x, r)) = B(x, r)$ by (b) and (c) of Lemma 1. Conversely, if $A \subseteq B(x, r)$, then $S \subseteq \mathrm{bh}_1(S) = A \subseteq B(x, r)$ by Lemma 1(a).

The above equivalence yields

$$A \subseteq \bigcap_{A \subseteq B(x,r),\, r<1} B(x, r) = \bigcap_{S \subseteq B(x,r),\, r<1} B(x, r) = D. \qquad \square$$

Corollary 6. *Every b-bounded closed subset S of a Minkowski space* $(\mathbb{R}^n, \|\cdot\|)$ *satisfies*

$$\mathrm{bh}_1(S) = \bigcap_{S \subseteq \mathrm{int}(B(x,1))} \mathrm{int}(B(x, 1)).$$

Proof. By Corollary 5,

$$\mathrm{bh}_1(S) = \bigcap_{S \subseteq \mathrm{int}(B(x,1))} B(x, 1) \supseteq \bigcap_{S \subseteq \mathrm{int}(B(x,1))} \mathrm{int}(B(x, 1)).$$

For the converse inclusion, note that

$$S \subseteq \mathrm{int}(B(x, 1)) \quad \Rightarrow \quad \mathrm{bh}_1(S) \subseteq \mathrm{int}(B(x, 1)).$$

Indeed, if $S \subseteq \mathrm{int}(B(x, 1))$, then $S \subseteq B(x, r)$ for some $r \in (0, 1)$ by Lemma 1(e), and, by Lemma 1(b) and (c), $\mathrm{bh}_1(S) \subseteq \mathrm{bh}_1(B(x, r)) = B(x, r) \subseteq \mathrm{int}(B(x, 1))$.

The above implication yields

$$\bigcap_{S \subseteq \mathrm{int}(B(x,1))} \mathrm{int}(B(x, 1)) \supseteq \bigcap_{\mathrm{bh}_1(S) \subseteq \mathrm{int}(B(x,1))} \mathrm{int}(B(x, 1)) \supseteq \mathrm{bh}_1(S). \qquad \square$$

The assumption of b-boundedness is essential in Corollaries 5 and 6. For example, if S is a closed ball of radius 1, then $\mathrm{bh}_1(S) = S$, whereas the four other intersections represent $\mathbb{R}^n$, since they are intersections over empty index sets.

To see that the assumption of closedness in Corollary 6 cannot be dropped, consider the example $S = \mathrm{int}(B(x_0, r_0))$ with $x_0 \in \mathbb{R}^n$ and $r_0 \in (0, 1)$. Then

$$\mathrm{bh}_1(\mathrm{int}(B(x_0, r_0))) = \mathrm{bh}_1(B(x_0, r_0)) = B(x_0, r_0)$$

by Lemma 1(a) and (c). In contrast to that,

$$\bigcap_{\mathrm{int}(B(x_0,r_0)) \subseteq \mathrm{int}(B(x,1))} \mathrm{int}(B(x, 1)) = \bigcap_{\|x - x_0\| \leq 1 - r_0} \mathrm{int}(B(x, 1)) = \mathrm{int}(B(x_0, r_0)),$$

as can be checked by the triangle inequality.

In classical convexity, two disjoint convex sets can be separated by a hyperplane. The analogous claim for ball convexity would say that, given two disjoint b-convex bodies $K_1, K_2 \subseteq \mathbb{R}^n$, there exists a *separating sphere* $S(x_0, 1)$ for K_1 and K_2; i.e., $K_1 \subseteq B(x_0, 1)$ and $K_2 \cap B(x_0, 1) = \varnothing$. In fact, one knows even more if the underlying Minkowski space $(\mathbb{R}^n, \|\cdot\|)$ is a Euclidean space (see [Bezdek et al. 2007, Lemma 3.1 and Corollary 3.4]), if its unit ball is a cube (see [Lángi et al. 2013, Corollary 3.15]), or if it is two-dimensional (see [Lángi et al. 2013, Theorem 4]). Then, for every b-convex body K and every supporting hyperplane H of K, there exists a sphere $S(x_0, 1)$ such that $K \subseteq B(x_0, 1)$ and $\operatorname{int}(B(x_0, 1)) \cap H = \varnothing$. However, the last statement fails in general (see [Lángi et al. 2013, Example 3.9] for an example in a generalized Minkowski space whose unit ball is not centrally symmetric). Here we show that even the (formally weaker) separation of two b-convex bodies by a unit sphere may fail in a (symmetric) Minkowski space.

Example 7. Let l_1^3 be the three-dimensional Minkowski space with unit ball $B(o, 1) = \operatorname{conv}(\{(\pm 1, 0, 0), (0, \pm 1, 0), (0, 0, \pm 1)\})$, let $0 < \varepsilon < \frac{1}{2}$, and consider the segments $K_1 = \left[\left(\frac{1}{4}, \frac{1}{4}, 0\right), \left(-\frac{1}{4}, -\frac{1}{4}, 0\right)\right]$ and $K_2 = \left[\left(\frac{1}{4}, -\frac{1}{4}, \varepsilon\right), \left(-\frac{1}{4}, \frac{1}{4}, \varepsilon\right)\right]$. Then K_1 and K_2 are disjoint b-bounded b-convex bodies in l_1^3, and there is no unit sphere $S(x_0, 1)$ such that $K_1 \subseteq B(x_0, 1)$ and $K_2 \cap \operatorname{int}(B(x_0, 1)) = \varnothing$.

Proof. K_1 is b-convex, because $K_1 = B\left(\left(-\frac{3}{4}, \frac{1}{4}, 0\right), 1\right) \cap B\left(\left(\frac{3}{4}, -\frac{1}{4}, 0\right), 1\right)$, and b-bounded, since $\operatorname{rad}(K_1) = \frac{1}{2}$. Similarly, K_2 is b-bounded and b-convex.

If $K_1 \subseteq B(x_0, 1)$, then $B(x_0, 1)$ contains at least one of the points of the segment $\left[\left(\frac{1}{4}, -\frac{1}{4}, \frac{1}{2}\right), \left(-\frac{1}{4}, \frac{1}{4}, \frac{1}{2}\right)\right]$, and we get $K_2 \cap \operatorname{int}(B(x_0, 1)) \neq \varnothing$, since $0 < \varepsilon < \frac{1}{2}$. $\square$

4. Characterizations of strict convexity

Some of our results will require strict convexity of the norm $\|\cdot\|$. On the other hand, strict convexity can be reflected by numerous properties related to concepts introduced in Section 2. The current section is devoted to characterizations of strict convexity. We start with characterizations by properties of balls, circumballs and circumradii; for (iv) and (v) in the following lemma we also refer to [Amir and Ziegler 1980; Martini et al. 2001].

Lemma 8. *Let* $(\mathbb{R}^n, \|\cdot\|)$ *be a Minkowski space. The following are equivalent*:

(i) *The norm* $\|\cdot\|$ *is strictly convex.*

(ii) *Each supporting hyperplane of a closed ball meets that ball in exactly one point.*

(iii) *The circumradius of the intersection of any two distinct balls of the same radius* $r > 0$ *is smaller than* r.

(iv) *Every bounded nonempty subset of* $\mathbb{R}^n$ *has a unique circumball.*

(v) *For any two distinct points* $x_1, x_2 \in \mathbb{R}^n$, $\{x_1, x_2\}$ *has a unique circumball.*

Proof. (i)⇒(ii): If (ii) fails, then some ball meets one of its supporting hyperplanes in at least two distinct points x_1, x_2. Then the segment $[x_1, x_2]$ is contained in the boundary of that ball, contradicting (i).

(ii)⇒(i): If (i) fails, then the boundary of $B(o, 1)$ contains a line segment L of positive length. The disjoint convex sets $\mathrm{int}(B(o, 1))$ and L can be separated by a hyperplane H. Then H is a supporting hyperplane of $B(o, 1)$ and contains L, contradicting (ii).

(i)⇒(iii): If $x_1, x_2 \in \mathbb{R}^n$ are distinct points and if $r > 0$, then

$$B(x_1, r) \cap B(x_2, r) \subseteq \mathrm{int}\Big(B\Big(\frac{x_1+x_2}{2}, r\Big)\Big), \tag{5}$$

which implies our claim $\mathrm{rad}(B(x_1, r) \cap B(x_2, r)) < r$ by Lemma 1(e). To verify (5), assume the contrary; i.e., $\|x - (x_1 + x_2)/2\| \geq r$ for some $x \in B(x_1, r) \cap B(x_2, r)$. Then

$$r \leq \Big\| x - \frac{x_1+x_2}{2} \Big\| \leq \tfrac{1}{2}\big(\|x - x_1\| + \|x - x_2\|\big) \leq \tfrac{1}{2}(r + r) = r,$$

hence all terms in the above estimate agree and we obtain $\|x - x_1\| = \|x - x_2\| = \big\|x - \frac{1}{2}(x_1 + x_2)\big\| = r$. This shows that $[x_1, x_2]$ is a segment in $S(x, r)$, contradicting (i) and proving (5).

(iii)⇒(iv): If (iv) fails, then there is a bounded set S with circumradius $\mathrm{rad}(S) > 0$ that has two circumballs $B(x_1, \mathrm{rad}(S))$ and $B(x_2, \mathrm{rad}(S))$, $x_1 \neq x_2$. This implies $B(x_1, \mathrm{rad}(S)) \cap B(x_2, \mathrm{rad}(S)) \supseteq S$ and $\mathrm{rad}(B(x_1, \mathrm{rad}(S)) \cap B(x_2, \mathrm{rad}(S))) \geq \mathrm{rad}(S)$, contradicting (iii).

For (iv)⇔(v)⇔(i), see [Amir and Ziegler 1980, Lemma 1.2]. □

Now we come to characterizations of strict convexity of norms in terms of concepts related to b-convexity that are defined in Section 2.

Proposition 9. *Let* $(\mathbb{R}^n, \|\cdot\|)$ *be a Minkowski space. The following are equivalent*:

(i) *The norm* $\|\cdot\|$ *is strictly convex.*

(vi) *Every b-convex body that is not b-bounded is a closed ball of radius* 1.

(vii) *Every b-convex body that is not b-bounded has only one supporting sphere.*

(viii) *For every boundary point x of a b-convex body K that is not b-bounded, there exists only one supporting sphere of K that contains x.*

(ix) *For every $x \in \mathbb{R}^n$, every $r \in (0, 1)$ and every $x_0 \in \mathrm{bd}(B(x, r))$, $B(x, r)$ has only one supporting sphere that contains x_0.*

(x) *There exist $x \in \mathbb{R}^n$ and $r \in (0, 1)$ such that, for every $x_0 \in \mathrm{bd}(B(x, r))$, $B(x, r)$ has only one supporting sphere that contains x_0.*

(xi) *For every $x \in \mathbb{R}^n$ and every $r \in (0, 1)$, each supporting sphere of $B(x, r)$ meets $B(x, r)$ in only one point.*

(xii) *There exist* $x \in \mathbb{R}^n$ *and* $r \in (0, 1)$ *such that each supporting sphere of* $B(x, r)$ *meets* $B(x, r)$ *in only one point.*

(xiii) *For every* $x \in \mathbb{R}^n$ *and every* $r \in (0, 1)$*,* $\text{b-exp}(B(x, r)) = S(x, r)$.

(xiv) *There exist* $x \in \mathbb{R}^n$ *and* $r \in (0, 1)$ *such that* $\text{b-exp}(B(x, r)) = S(x, r)$.

(xv) *Every b-convex body is strictly convex.*

(xvi) *For any two distinct points* $x_1, x_2 \in \mathbb{R}^n$*,* $\text{bh}_1(\{x_1, x_2\})$ *is strictly convex.*

(xvii) *Every b-convex body that contains at least two points has nonempty interior.*

(xviii) *For any two distinct points* $x_1, x_2 \in \mathbb{R}^n$*,* $\text{int}(\text{bh}_1(\{x_1, x_2\}))$ *is nonempty.*

Proof. The implications (vi)⇒(vii)⇒(viii), (ix)⇒(x), (xi)⇒(xii), (xiii)⇒(xiv), (xv)⇒(xvi), and (xvii)⇒(xviii) are obvious.

(i)⇒(vi): Every b-convex body K is a nonempty intersection of a nonempty family of closed balls of radius 1. If the family consisted of more than one ball, then its intersection K would be b-bounded by Lemma 8(i)⇒(iii). Hence the only b-convex bodies that are not b-bounded are closed balls of radius 1.

(viii)⇒(i): Suppose that (i) fails. Then condition (iii) from Lemma 8 fails as well, and there are two points $x_1 \neq x_2$ such that $\text{rad}(B(x_1, 1) \cap B(x_2, 1)) = 1$. The body $K = B(x_1, 1) \cap B(x_2, 1)$ shows that (viii) fails as well, because every $x \in \text{bd}(B(x_1, 1)) \cap \text{bd}(B(x_2, 1))$ belongs to $\text{bd}(K)$ and has the two supporting spheres $S(x_1, 1)$ and $S(x_2, 1)$.

(i)⇒(ix) and (i)⇒(xi): We use the fact that *if two balls* $B(y, s)$ *and* $B(y', s')$ *of positive radii in a strictly convex Minkowski space are on the same side of a common supporting hyperplane* H *with respective touching points* y_0 *and* y_0'*, then the dilatation* φ *that is uniquely determined by* $\varphi(y_0) = y_0'$ *and the dilatation factor* s'/s *maps* $B(y, s)$ *onto* $B(y', s')$. To see this, consider the homotheties δ and δ' that map $B(y, s)$ and $B(y', s')$ onto $B(o, 1)$, respectively. Then $\delta(H) = \delta'(H)$, because $B(o, 1)$ has only one supporting hyperplane with the same outer normal vector as H, and $\delta(y_0) = \delta'(y_0')$, since $\delta(H) = \delta'(H)$ has only one touching point with $B(o, 1)$ (see Lemma 8). Now $\varphi = (\delta')^{-1} \circ \delta$, and the fact is verified.

Coming back to the proof of (i)⇒(ix) and (i)⇒(xi), we consider an arbitrary supporting sphere $S(y, 1)$ of $B(x, r)$ and suppose that x_0 belongs to the b-support set $B(x, r) \cap S(y, 1)$. The supporting hyperplane of $B(y, 1)$ at x_0 supports $B(x, r)$ as well. Now the above fact says that $B(y, 1)$ is the image of $B(x, r)$ under the dilatation φ with fixed point x_0 and factor $1/r$. This shows in particular that the supporting sphere $S(y, 1)$ is uniquely determined by the touching point x_0, which proves (ix) (because there exists at least one supporting sphere at x_0 according to Proposition 3(a)). To show (xi), we must prove that every point $x_1 \in B(x, r) \cap S(y, 1)$ coincides with x_0. By the same argument as above, $B(y, 1)$ is the image of $B(x, r)$

under the dilatation ψ with fixed point x_1 and factor $1/r$. We obtain

$$y = \varphi(x) = x_0 + \frac{1}{r}(x - x_0) \quad \text{and} \quad y = \psi(x) = x_1 + \frac{1}{r}(x - x_1),$$

which gives $x_0 = x_1$ and completes the proof of (xi).

(x)⇒(i): Suppose that (i) fails. Then, for every $x \in \mathbb{R}^n$ and every $r \in (0, 1)$, $S(x, r)$ contains a line segment $[x_0, x_1] \subseteq S(x, r)$, $x_0 \neq x_1$. If φ_0 and φ_1 are dilatations with factor $1/r$ and fixed points x_0 and x_1, respectively, then $\varphi_0(S(x, r))$ and $\varphi_1(S(x, r))$ are distinct supporting spheres of $B(x, r)$. Clearly, we have

$$x_0 = \varphi_0(x_0) \in \varphi_0([x_0, x_1]) \subseteq \varphi_0(S(x, r)).$$

Moreover,

$$x_0 = \varphi_1(r x_0 + (1 - r)x_1) \in \varphi_1([x_0, x_1]) \subseteq \varphi_1(S(x, r)).$$

Therefore $\varphi_0(S(x, r))$ and $\varphi_1(S(x, r))$ are both supporting spheres of $B(x, r)$ at $x_0 \in \mathrm{bd}(B(x, r))$, and (x) is disproved.

(xi)⇒(xiii) and (xii)⇒(xiv) follow from Proposition 3(a).

(xiv)⇒(i): If (i) fails, then every ball $B(x, r)$, $x \in \mathbb{R}^n$, $r \in (0, 1)$, contains a segment $[x_1, x_2]$, $x_1 \neq x_2$, in its boundary $S(x, r)$. We shall show that $\frac{1}{2}(x_1 + x_2)$ is not contained in b-exp$(B(x, r))$, thus disproving (xiv). Indeed, if $S(y, 1)$ is a supporting sphere of $B(x, r)$ with touching point $\frac{1}{2}(x_1 + x_2) \in S(y, 1)$, then $[x_1, x_2] \subseteq B(x, r) \subseteq B(y, 1)$, and the point $\frac{1}{2}(x_1 + x_2)$ (from the relative interior) of $[x_1, x_2]$ is in $S(y, 1) = \mathrm{bd}(B(y, 1))$. Hence $[x_1, x_2] \subseteq S(y, 1)$. This shows that the exposed b-face $B(x, r) \cap S(y, 1)$ that contains $\frac{1}{2}(x_1 + x_2)$ must necessarily cover the whole segment $[x_1, x_2]$. Thus $\frac{1}{2}(x_1 + x_2) \notin$ b-exp$(B(x, r))$.

(i)⇒(xv): Assume that (xv) fails; i.e., there are a b-convex body K and two points $x_1 \neq x_2$ such that $[x_1, x_2] \subseteq \mathrm{bd}(K)$. By Proposition 3(a), there is a supporting sphere $S(y, 1)$ of K such that $\frac{1}{2}(x_1 + x_2) \in S(y, 1)$. As above, we have $[x_1, x_2] \subseteq K \subseteq B(y, 1)$ and $\frac{1}{2}(x_1 + x_2) \in S(y, 1) = \mathrm{bd}(B(y, 1))$, which yields $[x_1, x_2] \subseteq S(y, 1)$ and contradicts (i).

(xv)⇒(xvii) and (xvi)⇒(xviii): If a convex body K contains two distinct points x_1 and x_2 and has empty interior, then K is not strictly convex, because $[x_1, x_2] \subseteq K = \mathrm{bd}(K)$. This yields the two implications.

(xviii)⇒(i): If (i) fails, then there are $x_1 \neq x_2$ such that $[x_1, x_2] \subseteq S(o, 1)$. Hence

$$\begin{aligned}\left[\frac{x_1 - x_2}{2}, \frac{x_2 - x_1}{2}\right] &= \left([x_1, x_2] - \frac{x_1 + x_2}{2}\right) \cap \left(-[x_2, x_1] + \frac{x_1 + x_2}{2}\right) \\ &\subseteq \left(S(o, 1) - \frac{x_1 + x_2}{2}\right) \cap \left(S(o, 1) + \frac{x_1 + x_2}{2}\right) \\ &= S\left(-\frac{x_1 + x_2}{2}, 1\right) \cap S\left(\frac{x_1 + x_2}{2}, 1\right) \\ &= B\left(-\frac{x_1 + x_2}{2}, 1\right) \cap B\left(\frac{x_1 + x_2}{2}, 1\right).\end{aligned}$$

The last equation is a consequence of $\left\|\frac{1}{2}(x_1+x_2)-(-\frac{1}{2}(x_1+x_2))\right\|=2$. We obtain

$$\begin{aligned}\operatorname{int}\Bigl(\mathrm{bh}_1\Bigl(\Bigl\{\tfrac{x_1-x_2}{2},\,\tfrac{x_2-x_1}{2}\Bigr\}\Bigr)\Bigr) &\subseteq \operatorname{int}\Bigl(\mathrm{bh}_1\Bigl(B\Bigl(-\tfrac{x_1+x_2}{2},1\Bigr)\cap B\Bigl(\tfrac{x_1+x_2}{2},1\Bigr)\Bigr)\Bigr)\\ &= \operatorname{int}\Bigl(B\Bigl(-\tfrac{x_1+x_2}{2},1\Bigr)\cap B\Bigl(\tfrac{x_1+x_2}{2},1\Bigr)\Bigr)\\ &= \operatorname{int}\Bigl(S\Bigl(-\tfrac{x_1+x_2}{2},1\Bigr)\cap S\Bigl(\tfrac{x_1+x_2}{2},1\Bigr)\Bigr)\\ &= \varnothing,\end{aligned}$$

which contradicts (xviii). □

5. Representation of ball hulls "from inside"

In this section we deal with b-convexity of unions of increasing sequences of b-convex bodies and with the representation of ball hulls of sets by unions of ball hulls of finite subsets. We start with an auxiliary statement.

Lemma 10. *Let $K\in\mathcal{K}^n$, let $H\subseteq\mathbb{R}^n$ be a hyperplane, and let $(y_i)_{i=1}^{\infty}\subseteq\mathbb{R}^n$ be such that $y_i\xrightarrow{i\to\infty}y_0\in\mathbb{R}^n$ and $H\cap(K+y_i)\neq\varnothing$ for all $i=1,2,\ldots.$ Then $H\cap(K+y_i)\xrightarrow{i\to\infty}H\cap(K+y_0)$ in the Hausdorff distance.*

Proof. First note that $H\cap(K+y_0)\neq\varnothing$, and in turn $H\cap(K+y_0)\in\mathcal{K}^n$. Indeed, for every $i\geq 1$, we can pick $z_i\in H\cap(K+y_i)$, i.e., $z_i=x_i+y_i$ with $x_i\in K$. By the compactness of K we see that, without loss of generality, $x_i\xrightarrow{i\to\infty}x_0\in K$. We get $z_0:=x_0+y_0=\lim_{i\to\infty}z_i\in H\cap(K+y_0)$, because H is closed.

By [Schneider 1993, Theorem 1.8.7], our claim $H\cap(K+y_i)\xrightarrow{i\to\infty}H\cap(K+y_0)$ is now equivalent to the following conditions taken together:

(a) for every $t_0\in H\cap(K+y_0)$, there exist $t_i\in H\cap(K+y_i)$, $i=1,2,\ldots$, such that $t_i\xrightarrow{i\to\infty}t_0$;

(b) if $(t_{i_j})_{j=1}^{\infty}$ is a sequence with $i_1<i_2<\cdots$, $t_{i_j}\xrightarrow{j\to\infty}t_0\in\mathbb{R}^n$, and $t_{i_j}\in H\cap(K+y_{i_j})$, then $t_0\in H\cap(K+y_0)$.

Proof of (a). Suppose that $H=\{x\in\mathbb{R}^n:\langle u,x\rangle=c\}$, with $\langle\cdot,\cdot\rangle$ denoting the usual scalar product. Then fix $t_0\in H\cap(K+y_0)$, i.e., $t_0=x_0+y_0$ with $x_0\in K$ and

$$\langle u,t_0\rangle=c,\ \text{ i.e., } \langle u,x_0\rangle=c-\langle u,y_0\rangle. \tag{6}$$

Pick $x^*,x^{**}\in K$ such that

$$\langle u,x^*\rangle=\min\{\langle u,x\rangle:x\in K\},\qquad \langle u,x^{**}\rangle=\max\{\langle u,x\rangle:x\in K\}.$$

For every $i=1,2,\ldots,$ $H\cap(K+y_i)\neq\varnothing$ gives $\tilde{x}_i\in K$ such that

$$\langle u,\tilde{x}_i+y_i\rangle=c. \tag{7}$$

We choose $t_i = x_i + y_i \in H \cap (K + y_i)$ as follows: We know from $\tilde{x}_i \in K$ that $\langle u, \tilde{x}_i \rangle \in [\langle u, x^* \rangle, \langle u, x^{**} \rangle]$.

Case 1: $\langle u, \tilde{x}_i \rangle = \langle u, x_0 \rangle$. In this case we put

$$x_i := x_0.$$

Then $t_i = x_0 + y_i \in K + y_i$ and $t_i \in H$, because $\langle u, t_i \rangle = \langle u, x_0 + y_i \rangle = \langle u, \tilde{x}_i + y_i \rangle \overset{(7)}{=} c$.

Case 2: $\langle u, \tilde{x}_i \rangle \in [\langle u, x^* \rangle, \langle u, x_0 \rangle)$. Then

$$x_i := \frac{\langle u, \tilde{x}_i \rangle - \langle u, x^* \rangle}{\langle u, x_0 \rangle - \langle u, x^* \rangle} x_0 + \frac{\langle u, x_0 \rangle - \langle u, \tilde{x}_i \rangle}{\langle u, x_0 \rangle - \langle u, x^* \rangle} x^* \tag{8}$$

satisfies $x_i \in [x_0, x^*] \subseteq K$, hence $t_i = x_i + y_i \in K + y_i$, and

$$\langle u, x_i \rangle = \frac{\langle u, \tilde{x}_i \rangle - \langle u, x^* \rangle}{\langle u, x_0 \rangle - \langle u, x^* \rangle} \langle u, x_0 \rangle + \frac{\langle u, x_0 \rangle - \langle u, \tilde{x}_i \rangle}{\langle u, x_0 \rangle - \langle u, x^* \rangle} \langle u, x^* \rangle = \langle u, \tilde{x}_i \rangle.$$

This gives $t_i \in H$, because $\langle u, t_i \rangle = \langle u, x_i + y_i \rangle = \langle u, \tilde{x}_i + y_i \rangle \overset{(7)}{=} c$.

Case 3: $\langle u, \tilde{x}_i \rangle \in (\langle u, x_0 \rangle, \langle u, x^{**} \rangle]$. Then

$$x_i := \frac{\langle u, x^{**} \rangle - \langle u, \tilde{x}_i \rangle}{\langle u, x^{**} \rangle - \langle u, x_0 \rangle} x_0 + \frac{\langle u, \tilde{x}_i \rangle - \langle u, x_0 \rangle}{\langle u, x^{**} \rangle - \langle u, x_0 \rangle} x^{**} \tag{9}$$

satisfies $x_i \in [x_0, x^{**}] \subseteq K$, hence $t_i = x_i + y_i \in K + y_i$, and

$$\langle u, x_i \rangle = \frac{\langle u, x^{**} \rangle - \langle u, \tilde{x}_i \rangle}{\langle u, x^{**} \rangle - \langle u, x_0 \rangle} \langle u, x_0 \rangle + \frac{\langle u, \tilde{x}_i \rangle - \langle u, x_0 \rangle}{\langle u, x^{**} \rangle - \langle u, x_0 \rangle} \langle u, x^{**} \rangle = \langle u, \tilde{x}_i \rangle.$$

This yields $t_i \in H$, because $\langle u, t_i \rangle = \langle u, x_i + y_i \rangle = \langle u, \tilde{x}_i + y_i \rangle \overset{(7)}{=} c$.

Finally, for proving $t_i \xrightarrow{i \to \infty} t_0$, we use the following arguments. We get $x_i \xrightarrow{i \to \infty} x_0$ by partitioning the sequence $(x_i)_{i=1}^\infty$ into three subsequences corresponding to Cases 1–3, where each of the subsequences (if it is infinite) converges to x_0. In Case 1 this is trivial, and in the other two cases it follows from

$$\langle u, \tilde{x}_i \rangle \overset{(7)}{=} c - \langle u, y_i \rangle \xrightarrow{i \to \infty} c - \langle u, y_0 \rangle \overset{(6)}{=} \langle u, x_0 \rangle$$

and from the definitions (8) and (9). This yields $t_i = x_i + y_i \xrightarrow{i \to \infty} x_0 + y_0 = t_0$.

Proof of (b). The inclusion $t_{i_j} \in H \cap (K + y_{i_j})$ gives $x_{i_j} := t_{i_j} - y_{i_j} \in K$. Hence $x_{i_j} = t_{i_j} - y_{i_j} \xrightarrow{j \to \infty} t_0 - y_0 \in K$, because K is closed. Thus $t_0 \in K + y_0$. Moreover, $t_{i_j} \in H$ yields $t_{i_j} \xrightarrow{i \to \infty} t_0 \in H$, because H is closed. Hence $t_0 \in H \cap (K + y_0)$. □

Remark 11. Note that H cannot be replaced by an affine subspace L of arbitrary dimension in Lemma 10. See the following example, where

$$K := \left\{ (\xi_1, \xi_2, \xi_3) \in \mathbb{R}^3 : |\xi_3| \leq 1 - \sqrt{\xi_1^2 + \xi_2^2} \right\}.$$

(Thus $K = \operatorname{conv}(\{(\cos\varphi, \sin\varphi, 0) : 0 \le \varphi < 2\pi\} \cup \{(0, 0, \pm 1)\})$ is a compact double cone.) Consider the affine subspace $L := \operatorname{aff}(\{(1, 0, 0), (0, 0, 1)\})$ of $\mathbb{R}^3$, and let $y_i := (1, 0, 0) - \bigl(\cos\frac{1}{i}, \sin\frac{1}{i}, 0\bigr) \xrightarrow{i\to\infty} y_0 = (0, 0, 0)$. Then $L \cap (K + y_i) = \{(1, 0, 0)\}$ for any i, and $L \cap (K + y_0) = L \cap K = [(1, 0, 0), (0, 0, 1)]$. Hence $L \cap (K + y_i)$ does not converge to $L \cap (K + y_0)$ in the described way.

Theorem 12. *Let $C_1 \subseteq C_2 \subseteq \cdots$ be an increasing sequence of b-convex bodies in a Minkowski space $(\mathbb{R}^n, \|\cdot\|)$ such that*

$$\dim\Bigl(\operatorname{aff}\Bigl(\operatorname{bh}_1\Bigl(\bigcup_{i=1}^{\infty} C_i\Bigr)\Bigr)\Bigr) \in \{0, 1, n-1, n\}. \tag{10}$$

Then

$$\operatorname{cl}\Bigl(\bigcup_{i=1}^{\infty} C_i\Bigr) = \operatorname{bh}_1\Bigl(\bigcup_{i=1}^{\infty} C_i\Bigr). \tag{11}$$

In particular, $\operatorname{cl}\bigl(\bigcup_{i=1}^{\infty} C_i\bigr)$ *is a b-convex body.*

Proof. Since "$\subseteq$" in (11) is obvious, we prove now "$\supseteq$".

Case 1: $\dim\bigl(\operatorname{aff}\bigl(\operatorname{bh}_1\bigl(\bigcup_{i=1}^{\infty} C_i\bigr)\bigr)\bigr) = 0$. Here, $\bigcup_{i=1}^{\infty} C_i = \{x_0\}$ is a singleton. Thus $C_i = \{x_0\}$ for all $i = 1, 2, \ldots$, and (11) is trivial.

Case 2: $\dim\bigl(\operatorname{aff}\bigl(\operatorname{bh}_1\bigl(\bigcup_{i=1}^{\infty} C_i\bigr)\bigr)\bigr) = 1$. Here, $\operatorname{bh}_1\bigl(\bigcup_{i=1}^{\infty} C_i\bigr)$ is a line segment. Since every closed line segment of the same direction and having smaller length is also b-convex, each C_i is a closed segment of that kind (perhaps of length 0), $\bigcup_{i=1}^{\infty} C_i$ is a segment of that direction (not necessarily closed), and

$$\operatorname{bh}_1\Bigl(\bigcup_{i=1}^{\infty} C_i\Bigr) = \operatorname{cl}\Bigl(\bigcup_{i=1}^{\infty} C_i\Bigr).$$

Case 3: $\dim\bigl(\operatorname{aff}\bigl(\operatorname{bh}_1\bigl(\bigcup_{i=1}^{\infty} C_i\bigr)\bigr)\bigr) \ge n - 1 > 0$. Assume that we have "$\not\supseteq$" in (11). Then there exists $x_0 \in \operatorname{bh}_1\bigl(\bigcup_{i=1}^{\infty} C_i\bigr) \setminus \operatorname{cl}\bigl(\bigcup_{i=1}^{\infty} C_i\bigr)$, and we find $\varepsilon_0 > 0$ such that $B(x_0, \varepsilon_0) \cap \bigl(\bigcup_{i=1}^{\infty} C_i\bigr) = \varnothing$. Thus, there exists $x_1 \in \operatorname{relint}\bigl(\operatorname{bh}_1\bigl(\bigcup_{i=1}^{\infty} C_i\bigr)\bigr) \setminus \bigl(\bigcup_{i=1}^{\infty} C_i\bigr)$, that is,

$$x_1 \in \operatorname{relint}\Bigl(\operatorname{bh}_1\Bigl(\bigcup_{i=1}^{\infty} C_i\Bigr)\Bigr) \tag{12}$$

and

$$x_1 \notin C_i \quad \text{for } i = 1, 2, \ldots. \tag{13}$$

Property (13) and Proposition 3(b) give $y_i \in \mathbb{R}^n$ such that

$$x_1 \notin B(y_i, 1) \quad \text{and} \quad C_i \subseteq B(y_i, 1) \quad \text{for } i = 1, 2, \ldots.$$

There exists a convergent subsequence $y_{i_j} \xrightarrow{j\to\infty} y_0 \in \mathbb{R}^n$, and, by continuity of the norm and the inclusions $C_1 \subseteq C_2 \subseteq \cdots$, we have

$$x_1 \notin \operatorname{int}(B(y_0, 1)) \quad \text{and} \quad \bigcup_{i=1}^{\infty} C_i \subseteq B(y_0, 1). \tag{14}$$

Subcase 3.1: $\dim\bigl(\operatorname{aff}\bigl(\operatorname{bh}_1\bigl(\bigcup_{i=1}^{\infty} C_i\bigr)\bigr)\bigr) = n$. With (14) we have $\operatorname{bh}_1\bigl(\bigcup_{i=1}^{\infty} C_i\bigr) \subseteq B(y_0, 1)$ and

$$x_1 \notin \operatorname{int}\Bigl(\operatorname{bh}_1\Bigl(\bigcup_{i=1}^{\infty} C_i\Bigr)\Bigr) = \operatorname{relint}\Bigl(\operatorname{bh}_1\Bigl(\bigcup_{i=1}^{\infty} C_i\Bigr)\Bigr),$$

a contradiction to (12).

Subcase 3.2: $\dim\bigl(\operatorname{aff}\bigl(\operatorname{bh}_1\bigl(\bigcup_{i=1}^{\infty} C_i\bigr)\bigr)\bigr) = n-1$. Let $H := \operatorname{aff}\bigl(\operatorname{bh}_1\bigl(\bigcup_{i=1}^{\infty} C_i\bigr)\bigr)$. From $y_{i_j} \xrightarrow{j\to\infty} y_0$ we get $B(y_{i_j}, 1) \xrightarrow{j\to\infty} B(y_0, 1)$ in the Hausdorff metric, and with Lemma 10 we get $B(y_{i_j}, 1)\cap H \xrightarrow{j\to\infty} B(y_0, 1)\cap H$ as well. Then, by $x_1 \notin B(y_{i_j}, 1)$, we obtain $x_1 \notin \operatorname{int}_H(B(y_0, 1)\cap H)$ (where $\operatorname{int}_H(\,\cdot\,)$ is the interior in the natural topology of H), whereas $\operatorname{bh}_1\bigl(\bigcup_{i=1}^{\infty} C_i\bigr) \subseteq B(y_0, 1)\cap H$ by (14) and the choice of H. With this and the choice of H we obtain $x_1 \notin \operatorname{int}_H\bigl(\operatorname{bh}_1\bigl(\bigcup_{i=1}^{\infty} C_i\bigr)\bigr) = \operatorname{relint}\bigl(\operatorname{bh}_1\bigl(\bigcup_{i=1}^{\infty} C_i\bigr)\bigr)$, a contradiction to (12). The proof of "$\supseteq$" in (11) is complete.

To show that $\operatorname{cl}\bigl(\bigcup_{i=1}^{\infty} C_i\bigr) = \operatorname{bh}_1\bigl(\bigcup_{i=1}^{\infty} C_i\bigr)$ is a b-convex body, it is enough to verify that $\operatorname{bh}_1\bigl(\bigcup_{i=1}^{\infty} C_i\bigr) \neq \mathbb{R}^n$, i.e., that $\bigcup_{i=1}^{\infty} C_i$ is contained in some ball of radius 1. This is obvious by the second part of (14) (which can be shown analogously in Cases 1 and 2). □

Remark 13. Note that the technical assumption (10) is satisfied in each of the following situations:

(i) $n \leq 3$,

(ii) $(\mathbb{R}^n, \|\cdot\|)$ is strictly convex,

(iii) $\dim\bigl(\operatorname{aff}\bigl(\bigcup_{i=1}^{\infty} C_i\bigr)\bigr) \geq n-1$ or, equivalently, $\dim(\operatorname{aff}(C_{i_0})) \geq n-1$ for some i_0.

Proof. Situation (i) is trivial. In situation (ii), the equivalence (i) $\Leftrightarrow$ (xvii) from Proposition 9 shows that $\dim\bigl(\operatorname{aff}\bigl(\operatorname{bh}_1\bigl(\bigcup_{i=1}^{\infty} C_i\bigr)\bigr)\bigr) = n$ as soon as $\bigcup_{i=1}^{\infty} C_i$ is not a singleton. Condition (iii) implies (10), because $\bigcup_{i=1}^{\infty} C_i \subseteq \operatorname{bh}_1\bigl(\bigcup_{i=1}^{\infty} C_i\bigr)$. □

In Example 16 we shall see that assumption (10) cannot be dropped in Theorem 12.

Theorem 14. *Let $S \neq \varnothing$ be a subset of a Minkowski space $(\mathbb{R}^n, \|\cdot\|)$ such that*

$$\dim(\operatorname{aff}(\operatorname{bh}_1(S))) \in \{0, 1, n-1, n\}. \tag{15}$$

Then

$$\operatorname{bh}_1(S) = \operatorname{cl}\Bigl(\bigcup_{F\subseteq S \text{ finite}} \operatorname{bh}_1(F)\Bigr). \tag{16}$$

Proof. The inclusion "$\supseteq$" is evident. For "$\subseteq$", first note that S, considered as a metric subspace of the separable metric space $(\mathbb{R}^n, \|\cdot\|)$, is separable itself; i.e., there are $x_1, x_2, \ldots \in S$ such that $\operatorname{cl}(S) = \operatorname{cl}(\{x_1, x_2, \ldots\})$. (To be more constructive, let $r_1, r_2, \ldots \in \mathbb{R}^n$ be the vectors with only rational coordinates and pick $x_i \in S$ such that $\|x_i - r_i\| < \inf\{\|x - r_i\| : x \in S\} + \frac{1}{i}$.) Putting $C_i = \operatorname{bh}_1(\{x_1, \ldots, x_i\})$ for $i = 1, 2, \ldots$, we obtain $C_1 \subseteq C_2 \subseteq \cdots$. If C_{i_0} is not a b-convex body for some i_0, then $\mathrm{bh}_1(\{x_1, \dots, x_{i_0}\}) = C_{i_0} = \mathbb{R}^n$, and (16) is obvious (both sides are $\mathbb{R}^n$). Hence we can assume that all C_i are b-convex bodies. Moreover,

$$\begin{aligned}(17)\qquad \mathrm{bh}_1\Big(\bigcup_{i=1}^{\infty} C_i\Big) &= \mathrm{bh}_1\Big(\bigcup_{i=1}^{\infty} \mathrm{bh}_1(\{x_1,\dots,x_i\})\Big) = \mathrm{bh}_1\Big(\bigcup_{i=1}^{\infty}\{x_1,\dots,x_i\}\Big)\\ &= \mathrm{bh}_1\big(\{x_1,x_2,\dots\}\big) = \mathrm{bh}_1\big(\mathrm{cl}(\{x_1,x_2,\dots\})\big)\\ &= \mathrm{bh}_1\big(\mathrm{cl}(S)\big) = \mathrm{bh}_1(S),\end{aligned}$$

which gives

$$\dim\Big(\mathrm{aff}\Big(\mathrm{bh}_1\Big(\bigcup_{i=1}^{\infty} C_i\Big)\Big)\Big) = \dim(\mathrm{aff}(\mathrm{bh}_1(S))) \overset{(15)}{\in} \{0, 1, n-1, n\}.$$

Now we can apply Theorem 12 and obtain

$$\begin{aligned}\mathrm{bh}_1(S) &\overset{(17)}{=} \mathrm{bh}_1\Big(\bigcup_{i=1}^{\infty} C_i\Big) = \mathrm{cl}\Big(\bigcup_{i=1}^{\infty} C_i\Big) = \mathrm{cl}\Big(\bigcup_{i=1}^{\infty} \mathrm{bh}_1(\{x_1, \dots, x_i\})\Big)\\ &\subseteq \mathrm{cl}\Big(\bigcup_{F\subseteq S \text{ finite}} \mathrm{bh}_1(F)\Big). \qquad \square\end{aligned}$$

Remark 15. As in Remark 13, we see that (15) holds in each of the following situations:

(i) $n \le 3$,

(ii) $(\mathbb{R}^n, \|\cdot\|)$ is strictly convex,

(iii) $\dim(\mathrm{aff}(S)) \ge n-1$.

The claim of Theorem 14 is shown in [Lángi et al. 2013, Theorem 1] under the stronger assumption that $\dim(\mathrm{aff}(\mathrm{bh}_1(S))) = n$. The authors ask in [Lángi et al. 2013, Problem 2.6] if the assumption can be dropped. Our generalization to the additional cases $\dim(\mathrm{aff}(\mathrm{bh}_1(S))) \in \{0, 1, n-1\}$ shows that the assumption can be weakened; however, Example 16 illustrates that the restrictions (10) in Theorem 12 and (15) in Theorem 14 are essential, which shows that the answer to Problem 2.6 from [Lángi et al. 2013] is negative.

Example 16. We denote the Euclidean norm by $\|\cdot\|_2$ and consider convex bodies

$$\begin{aligned}K &= \{(\kappa_1, \kappa_2, \kappa_3, 0) : \|(\kappa_1, \kappa_2)\|_2 \le 1,\ \|(\kappa_2, \kappa_3)\|_2 \le 1\},\\ L &= \{(\lambda_1, 0, \lambda_3, \lambda_4) : \|(\lambda_1, \lambda_4)\|_2 \le 1,\ |\lambda_3| \le 1\}\end{aligned}$$

in $\mathbb{R}^4$. We define the unit ball $B(o, 1) = B$ of a Minkowski space $(\mathbb{R}^4, \|\cdot\|)$ by

$$\begin{aligned}(18)\quad B &= \mathrm{conv}(K \cup L)\\ &= \Big\{(\kappa_1 + \lambda_1, \xi_2, \kappa_3 + \lambda_3, \xi_4) :\\ &\qquad \max\{\|(\kappa_1, \xi_2)\|_2, \|(\xi_2, \kappa_3)\|_2\} + \max\{\|(\lambda_1, \xi_4)\|_2, |\lambda_3|\} \le 1\Big\}.\end{aligned}$$

For that space we shall see that

(19) $\mathrm{bh}_1(\{(\alpha,0,0,0),(-\alpha,0,0,0)\}) = [(\alpha,0,0,0),(-\alpha,0,0,0)]$ for $\alpha \in (0,1)$

and

(20) $$\mathrm{bh}_1(\{(1,0,0,0),(-1,0,0,0)\}) = \{(\xi_1,0,0,\xi_4) : \|(\xi_1,\xi_4)\|_2 \le 1\}.$$

Consequently, the segments $C_i = \big[\big(1-\frac{1}{i},0,0,0\big),\big(-1+\frac{1}{i},0,0,0\big)\big]$, $i=1,2,\dots,$ form an increasing sequence of b-convex bodies, and we obtain

$$\mathrm{cl}\Big(\bigcup\nolimits_{i=1}^{\infty} C_i\Big) = [(1,0,0,0),(-1,0,0,0)],$$
$$\mathrm{bh}_1\Big(\bigcup\nolimits_{i=1}^{\infty} C_i\Big) = \{(\xi_1,0,0,\xi_4) : \|(\xi_1,\xi_4)\|_2 \le 1\}.$$

Hence (11) fails and $\mathrm{cl}(\bigcup_{i=1}^{\infty} C_i)$ is not b-convex.

Similarly, the relatively open segment $S = ((1,0,0,0),(-1,0,0,0))$ satisfies

$$\mathrm{bh}_1(S) = \{(\xi_1,0,0,\xi_4) : \|(\xi_1,\xi_4)\|_2 \le 1\},$$
$$\mathrm{cl}\Big(\bigcup\nolimits_{F\subseteq S \text{ finite}} \mathrm{bh}_1(F)\Big) = [(1,0,0,0),(-1,0,0,0)],$$

and (16) fails.

Proof of (19) *and* (20). Step 1: verification of (19). Let $\alpha \in (0,1)$ be fixed. We use the linear functional

$$f(\xi_1,\xi_2,\xi_3,\xi_4) = \sqrt{1-\alpha^2}\,\xi_2 + \alpha\xi_3 = \big\langle\big(\sqrt{1-\alpha^2},\alpha\big),(\xi_2,\xi_3)\big\rangle$$

of Euclidean norm $\|f\|_2 = \big\|\big(\sqrt{1-\alpha^2},\alpha\big)\big\|_2 = 1$, and we define related level and sublevel sets $f_{=\mu},\ f_{\le\mu},\ f_{\ge\mu}$ by

$$f_{=/\le/\ge\mu} = \{x \in \mathbb{R}^4 : f(x) = / \le / \ge \mu\}.$$

We obtain, partially based on the Cauchy–Schwarz inequality,

$$K \subseteq f_{\ge -1} \cap f_{\le 1}, \quad L \subseteq f_{\ge -\alpha} \cap f_{\le \alpha}, \quad L \cap f_{=1} = \varnothing.$$

These yield

(21) $$B \subseteq f_{\ge -1} \cap f_{\le 1}$$

and

(22) $$B \cap f_{=1} = K \cap f_{=1} = \big[\big(\alpha,\sqrt{1-\alpha^2},\alpha,0\big),\big(-\alpha,\sqrt{1-\alpha^2},\alpha,0\big)\big].$$

Next note that $\big(\pm\alpha,\pm\sqrt{1-\alpha^2},\pm\alpha,0\big) \in K \subseteq B$ for arbitrary choice of signs. This implies $(\pm\alpha,0,0,0) \in B + \big(0,\pm\sqrt{1-\alpha^2},\pm\alpha,0\big)$, in particular

$$\{(\alpha,0,0,0),(-\alpha,0,0,0)\} \subseteq B\big(\big(0,\sqrt{1-\alpha^2},\alpha,0\big),1\big) \cap B\big(\big(0,-\sqrt{1-\alpha^2},-\alpha,0\big),1\big),$$

and in turn

$$
\begin{aligned}
&\mathrm{bh}_1(\{(\alpha,0,0,0),(-\alpha,0,0,0)\})\\
&\quad\subseteq B\big((0,\sqrt{1-\alpha^2},\alpha,0),1\big)\cap B\big((0,-\sqrt{1-\alpha^2},-\alpha,0),1\big)\\
&\quad\overset{(21)}{=}\big((B\cap f_{\geq -1})+\big(0,\sqrt{1-\alpha^2},\alpha,0\big)\big)\cap\big((B\cap f_{\leq 1})+\big(0,-\sqrt{1-\alpha^2},-\alpha,0\big)\big)\\
&\quad=\big(B+\big(0,\sqrt{1-\alpha^2},\alpha,0\big)\big)\cap f_{\geq 0}\cap\big(B+\big(0,-\sqrt{1-\alpha^2},-\alpha,0\big)\big)\cap f_{\leq 0}\\
&\quad\subseteq\big(B+\big(0,-\sqrt{1-\alpha^2},-\alpha,0\big)\big)\cap f_{=0}\\
&\quad=(B\cap f_{=1})+\big(0,-\sqrt{1-\alpha^2},-\alpha,0\big)\\
&\quad\overset{(22)}{=}[(\alpha,0,0,0),(-\alpha,0,0,0)].
\end{aligned}
$$

This gives the inclusion "$\subseteq$" in (19). The reverse inclusion is obvious, since the ball hull is closed and convex.

<u>Step 2:</u> verification of the equivalence of

$$\{(1,0,0,0),(-1,0,0,0)\}\subseteq B(t,1)=B+t \tag{23}$$

and

$$t=(0,0,\tau_3,0)\quad\text{with }\ \tau_3\in[-1,1]. \tag{24}$$

Suppose that $t=(\tau_1,\tau_2,\tau_3,\tau_4)$ satisfies (23). The inclusion $B=\mathrm{conv}(K\cup L)\subseteq\mathrm{conv}([-1,1]^4)=[-1,1]^4$ together with (23) gives

$$\tau_1=0\quad\text{and}\quad\tau_3\in[-1,1]. \tag{25}$$

Now the assumption $(-1,0,0,0)\in B+t$ amounts to $(-1,-\tau_2,-\tau_3,-\tau_4)\in B$ and, by symmetry, to $(1,\tau_2,\tau_3,\tau_4)\in B$. By (18), this says that there are $\kappa_1,\lambda_1,\kappa_3,\lambda_3\in\mathbb{R}$ such that $\kappa_1+\lambda_1=1,\ \kappa_3+\lambda_3=\tau_3$ and

$$\max\{\|(\kappa_1,\tau_2)\|_2,\|(\tau_2,\kappa_3)\|_2\}+\max\{\|(\lambda_1,\tau_4)\|_2,|\lambda_3|\}\leq 1. \tag{26}$$

From $\kappa_1+\lambda_1=1$ we obtain

$$1\leq|\kappa_1|+|\lambda_1|\leq\|(\kappa_1,\tau_2)\|_2+\|(\lambda_1,\tau_4)\|_2\overset{(26)}{\leq}1.$$

Hence all inequalities in the last formula are identities and

$$\tau_2=\tau_4=0.$$

By (25), the implication "(23)$\Rightarrow$(24)" is proved.

The converse "(24)$\Rightarrow$(23)" amounts to $(\pm1,0,-\tau_3,0)\in B$ for all $\tau_3\in[-1,1]$. This is an obvious consequence of $(\pm1,0,-\tau_3,0)\in L\subseteq B$.

Note that the equivalence "(23)⇔(24)" implies

$$\mathrm{bh}_1\big(\{(1,0,0,0),(-1,0,0,0)\}\big) = \bigcap_{\tau_3\in[-1,1]} (B+(0,0,\tau_3,0)). \tag{27}$$

Step 3: verification of "⊆" from (20). Here the functional

$$g(\xi_1,\xi_2,\xi_3,\xi_4)=\xi_3$$

satisfies $K\cup L\subseteq g_{\geq -1}\cap g_{\leq 1}$. Hence

$$B\subseteq g_{\geq -1}\cap g_{\leq 1} \tag{28}$$

and

$$\begin{aligned} B\cap g_{=1} &= \mathrm{conv}\big((K\cap g_{=1})\cup(L\cap g_{=1})\big)\\ &= \mathrm{conv}\big(\{(\xi_1,0,1,0):|\xi_1|\leq 1\}\cup\{(\xi_1,0,1,\xi_4):\|(\xi_1,\xi_4)\|_2\leq 1\}\big)\\ &= \{(\xi_1,0,1,\xi_4):\|(\xi_1,\xi_4)\|_2\leq 1\}. \end{aligned} \tag{29}$$

Now we obtain the claim "⊆" from (20) by

$$\begin{aligned} &\mathrm{bh}_1(\{(1,0,0,0),(-1,0,0,0)\})\\ &\quad\overset{(27)}{\subseteq} (B+(0,0,1,0))\cap(B+(0,0,-1,0))\\ &\quad\overset{(28)}{=} ((B\cap g_{\geq -1})+(0,0,1,0))\cap((B\cap g_{\leq 1})+(0,0,-1,0))\\ &\quad= (B+(0,0,1,0))\cap g_{\geq 0}\cap(B+(0,0,-1,0))\cap g_{\leq 0}\\ &\quad\subseteq (B+(0,0,-1,0))\cap g_{=0}\\ &\quad= (B\cap g_{=1})+(0,0,-1,0)\\ &\quad\overset{(29)}{=} \{(\xi_1,0,0,\xi_4):\|(\xi_1,\xi_4)\|_2\leq 1\}. \end{aligned}$$

Step 4: verification of "⊇" from (20). Let $\xi_1,\xi_4,\tau_3\in\mathbb{R}$ be such that $\|(\xi_1,\xi_4)\|_2\leq 1$ and $|\tau_3|\leq 1$. Then $(\xi_1,0,-\tau_3,\xi_4)\in L\subseteq B$. Thus

$$(\xi_1,0,0,\xi_4)\in B+(0,0,\tau_3,0) \quad \text{if } \|(\xi_1,\xi_4)\|_2\leq 1,\ |\tau_3|\leq 1.$$

By (27), this implies "⊇" from (20). □

6. Minimal representation of ball convex bodies as ball hulls

In this section we will present, as announced, minimal representations of ball convex bodies in terms of their ball exposed faces.

Theorem 17. *Let K be a b-bounded b-convex body in a Minkowski space $(\mathbb{R}^n,\|\cdot\|)$ and let $S\subseteq K$. Then $\mathrm{bh}_1(S)=K$ if and only if every exposed b-face of K meets $\mathrm{cl}(S)$.*

Proof. For the proof of "$\Rightarrow$", suppose that there is an exposed b-face F of K such that $F \cap \mathrm{cl}(S) = \varnothing$. We have to show that $\mathrm{bh}_1(S) \neq K$. The b-face F has a representation $F = K \cap S(y, 1)$, where $S(y, 1)$ is a supporting sphere of K. By $F \cap \mathrm{cl}(S) = \varnothing$, we obtain $\mathrm{cl}(S) \subseteq \mathrm{int}(B(y, 1))$ and, by Lemma 1(e), $\mathrm{cl}(S) \subseteq B(y, r)$ for some $r < 1$. We fix $x_0 \in F$. Then $\|x_0 - y\| = 1$, because $F \subseteq S(y, 1)$, and $x_0 \notin B(y-(1-r)(x_0-y), 1)$, since $\|x_0-(y-(1-r)(x_0-y))\| = (2-r)\|x_0-y\| > 1$. But $S \subseteq B(y, r) \subseteq B(y - (1 - r)(x_0 - y), 1)$ by the triangle inequality. Thus

$$x_0 \notin B\big(y - (1 - r)(x_0 - y), 1\big) \supseteq \mathrm{bh}_1(S) \quad \text{and} \quad x_0 \in F \subseteq K,$$

showing that $\mathrm{bh}_1(S) \neq K$.

For the converse implication "$\Leftarrow$", we suppose that $\mathrm{bh}_1(S) \neq K$ and will show that $\mathrm{cl}(S)$ misses at least one exposed b-face F_0 of K. Since $\mathrm{bh}_1(S) \neq K$ and $\mathrm{bh}_1(S) \subseteq K$ by Lemma 1, there is $x_0 \in K \setminus \mathrm{bh}_1(S)$. By Proposition 3(b), we can separate x_0 from the b-convex body $\mathrm{bh}_1(S) \subseteq K$ by a sphere $S(y_0, 1)$,

$$\mathrm{cl}(S) \subseteq \mathrm{bh}_1(S) \subseteq B(y_0, 1) \quad \text{and} \quad x_0 \notin B(y_0, 1). \tag{30}$$

The b-boundedness of K gives $y_1 \in \mathbb{R}^n$ such that

$$\mathrm{cl}(S) \subseteq K \subseteq \mathrm{int}(B(y_1, 1)). \tag{31}$$

We consider the balls $B_\lambda := B(y_0 + \lambda(y_1 - y_0), 1)$ for $\lambda \in [0, 1]$. Then $K \nsubseteq B_0$ by (30) and $K \subseteq B_1$ by (31). Consequently, there exists

$$\lambda_0 = \min\{\lambda \in [0, 1] : K \subseteq B_\lambda\} \in (0, 1].$$

(It is a consequence of the continuity of $\|\cdot\|$ that λ_0 is really attained as a minimum.) By the definition of λ_0 and a compactness argument, the set $F_0 = K \cap \mathrm{bd}(B_{\lambda_0})$ is nonempty, so that $S_{\lambda_0} := \mathrm{bd}(B_{\lambda_0})$ is a supporting sphere of K and F_0 is an exposed b-face.

Now it remains to show that $F_0 \cap \mathrm{cl}(S) = \varnothing$. Suppose that this is not the case; i.e., there exists $z_0 \in F_0 \cap \mathrm{cl}(S)$. The inclusions (30) and (31) yield $\|z_0 - y_0\| \leq 1$ and $\|z_0 - y_1\| < 1$. Finally, the inclusion $z_0 \in F_0 \subseteq S_{\lambda_0} = S(y_0 + \lambda_0(y_1 - y_0), 1)$ gives

$$\begin{aligned} 1 &= \|z_0 - (y_0 + \lambda_0(y_1 - y_0))\| \\ &= \|\lambda_0(z_0 - y_1) + (1 - \lambda_0)(z_0 - y_0)\| \\ &\leq \lambda_0\|z_0 - y_1\| + (1 - \lambda_0)\|z_0 - y_0\| \\ &< \lambda_0 + (1 - \lambda_0) = 1. \end{aligned}$$

This contradiction completes the proof. □

Note that the proof of "$\Rightarrow$" did not require b-boundedness of K. However, b-boundedness is essential for "$\Leftarrow$". To see this, consider a closed ball $K = B(y, 1)$ of radius 1. (Proposition 9(i)$\Rightarrow$(vi) says that these are the only b-convex bodies that are not b-bounded, provided that the norm $\|\cdot\|$ is strictly convex.) Then the only supporting sphere of K is $S(y, 1)$, and the only exposed b-face is $F = K \cap S(y, 1) = S(y, 1)$. Then every singleton $S = \{x_0\} \subseteq S(y, 1)$ satisfies the condition from Theorem 17, but $\mathrm{bh}_1(S) = \{x_0\}$ is not K.

Example 18. Consider the space $l_\infty^2 = (\mathbb{R}^2, \|\cdot\|_\infty)$ with unit ball $[-1, 1]^2$. Then all b-convex bodies are of the form $[\alpha_1, \beta_1] \times [\alpha_2, \beta_2]$ with $0 \le \beta_i - \alpha_i \le 2,\ i = 1, 2$. We restrict our consideration to b-bounded b-convex bodies K with nonempty interior. These are rectangles $K = [\alpha_1, \beta_1] \times [\alpha_2, \beta_2]$ with $0 < \beta_i - \alpha_i < 2,\ i = 1, 2$. The exposed b-faces of K are the edges

$$F_1 = [\alpha_1,\beta_1]\times\{\beta_2\},\quad F_2 = \{\alpha_1\}\times[\alpha_2,\beta_2],\quad F_3 = [\alpha_1,\beta_1]\times\{\alpha_2\},\quad F_4 = \{\beta_1\}\times[\alpha_2,\beta_2]$$

and the unions $F_1 \cup F_2,\ F_2 \cup F_3,\ F_3 \cup F_4,\ F_4 \cup F_1$. Theorem 17 says that a set $S \subseteq K$ satisfies $\mathrm{bh}_1(S) = K$ if and only if $\mathrm{cl}(S) \cap F_j \neq \varnothing$ for $j = 1, 2, 3, 4$. Consequently, when searching for minimal sets S (under inclusion) with $\mathrm{bh}_1(S) = K$, we need to find a minimal set S containing at least one point from each of F_1, F_2, F_3, F_4. Such S may consist of 2 (if S is composed of two vertices symmetric with respect to the center of K), 3 or 4 points (if S contains exactly one point from the relative interior of each F_i).

This example can be generalized for boxes in $l_\infty^n,\ n \ge 1$. Corresponding minimal sets must contain a point in every (classical) facet of a box and may consist of $2, \dots, 2n$ elements.

Corollary 19. *If a subset S of a b-bounded b-convex body K in a Minkowski space $(\mathbb{R}^n, \|\cdot\|)$ satisfies $\mathrm{bh}_1(S) = K$, then $\text{b-exp}(K) \subseteq \mathrm{cl}(S)$.*

Proof. If $x \in \text{b-exp}(K)$, then $\{x\}$ is an exposed b-face of K. Now Theorem 17 yields $\{x\} \cap \mathrm{cl}(S) \neq \varnothing$; i.e., $x \in \mathrm{cl}(S)$. □

Theorem 20. *A subset S of a b-bounded b-convex body K in a strictly convex Minkowski space $(\mathbb{R}^n, \|\cdot\|)$ satisfies $\mathrm{bh}_1(S) = K$ if and only if $\text{b-exp}(K) \subseteq \mathrm{cl}(S)$. In particular,*

$$K = \mathrm{bh}_1(\text{b-exp}(K)),$$

and $\mathrm{cl}(\text{b-exp}(K))$ is the unique minimal (under inclusion) closed subset of $\mathbb{R}^n$ whose ball hull is K.

Proof. The implication "⇒" of "$\mathrm{bh}_1(S) = K \Leftrightarrow \text{b-exp}(K) \subseteq \mathrm{cl}(S)$" is given by Corollary 19. To see "⇐", it is enough to show that

$$K \subseteq \mathrm{bh}_1(\text{b-exp}(K)). \tag{32}$$

Indeed, if $\text{b-exp}(K) \subseteq \mathrm{cl}(S)$ and if (32) is verified, parts (b) and then (a) of Lemma 1 give

$$K \subseteq \mathrm{bh}_1(\text{b-exp}(K)) \subseteq \mathrm{bh}_1(\mathrm{cl}(S)) = \mathrm{bh}_1(S) \subseteq \mathrm{bh}_1(K) = K,$$

and "$\Leftarrow$" is proved.

To show (32), we assume there exists $x_0 \in K \setminus \mathrm{bh}_1(\text{b-exp}(K))$. The b-boundedness of K implies b-boundedness of the subset $\widetilde{K} := \mathrm{bh}_1(\text{b-exp}(K)) \subseteq \mathrm{bh}_1(K) = K$. Separation of x_0 from $\widetilde{K}$ by Proposition 3(c) and the b-boundedness of K yield the existence of $y_0, y_1 \in \mathbb{R}^n$ and $r \in (0, 1)$ such that

$$\widetilde{K} \subseteq B(y_0, r) \quad \text{and} \quad x_0 \notin B(y_0, 1), \tag{33}$$

$$\widetilde{K} \subseteq K \subseteq \mathrm{int}(B(y_1, r)). \tag{34}$$

Much as in the proof of Theorem 17, we define $B_\lambda^r := B(y_0 + \lambda(y_1 - y_0), r)$ for $\lambda \in [0, 1]$ and, exploiting $K \nsubseteq B_0^r$ from (33) and $K \subseteq B_1^r$ from (34), find

$$\lambda_0 = \min\{\lambda \in [0, 1] : K \subseteq B_\lambda^r\} \in (0, 1].$$

Then there exists $x_1 \in K \cap S_{\lambda_0}^r$, where $S_{\lambda_0}^r := \mathrm{bd}(B_{\lambda_0}^r)$.

Next we show that

$$x_1 \notin \widetilde{K}. \tag{35}$$

Indeed, if x_1 belongs to $\widetilde{K}$, we have $\|x_1 - y_0\| \le r$ and $\|x_1 - y_1\| < r$. The inclusion $x_1 \in S_{\lambda_0}^r = S(y_0 + \lambda_0(y_1 - y_0), r)$ gives

$$\begin{aligned} r &= \|x_1 - (y_0 + \lambda_0(y_1 - y_0))\| \\ &= \|\lambda_0(x_1 - y_1) + (1 - \lambda_0)(x_1 - y_0)\| \\ &\le \lambda_0\|x_1 - y_1\| + (1 - \lambda_0)\|x_1 - y_0\| \\ &< \lambda_0 r + (1 - \lambda_0)r = r. \end{aligned}$$

This contradiction proves (35).

Since $B_{\lambda_0}^r$ is a b-convex body by Lemma 1(c) and since $x_1 \in S_{\lambda_0}^r = \mathrm{bd}(B_{\lambda_0}^r)$, Proposition 3(a) gives $y_2 \in \mathbb{R}^n$ such that

$$B_{\lambda_0}^r \subseteq B(y_2, 1) \quad \text{and} \quad x_1 \in B_{\lambda_0}^r \cap S(y_2, 1).$$

Proposition 9(i)$\Rightarrow$(xi) tells us that

$$B_{\lambda_0}^r \cap S(y_2, 1) = \{x_1\},$$

because $\|\cdot\|$ is strictly convex. Using the known inclusions $x_1 \in K$ and $K \subseteq B_{\lambda_0}^r$, we get

$$K \subseteq B(y_2, 1) \quad \text{and} \quad K \cap S(y_2, 1) = \{x_1\}.$$

Hence $x_1 \in \text{b-exp}(K)$. But (35) says that $x_1 \notin \mathrm{bh}_1(\text{b-exp}(K))$. This final contradiction establishes (32) and completes the proof. $\square$

Theorem 20 says in particular that every b-bounded b-convex body in a strictly convex Minkowski space gives rise to a unique minimal closed subset whose ball hull is that body. We have seen in Example 18 that this is not necessarily the case if the norm fails to be strictly convex.

Example 21. The set b-exp(K) of all b-exposed points of a b-bounded b-convex body is not necessarily closed. An example of that kind in the Euclidean plane is the convex disc K bounded by the arcs

$$\Gamma_1 = \left\{ \left(\cos\varphi, -\frac{\sqrt{3}}{4} + \sin\varphi \right) : \frac{\pi}{3} \le \varphi \le \frac{2\pi}{3} \right\},$$

$$\Gamma_2 = \left\{ \left(\cos\varphi, \frac{\sqrt{3}}{4} + \sin\varphi \right) : \frac{4\pi}{3} \le \varphi \le \frac{5\pi}{3} \right\},$$

$$\Gamma_3 = \left\{ \left(\frac{1}{4} + \frac{1}{2}\cos\varphi, \frac{1}{2}\sin\varphi \right) : -\frac{\pi}{3} < \varphi < \frac{\pi}{3} \right\},$$

$$\Gamma_4 = \left\{ \left(-\frac{1}{4} + \frac{1}{2}\cos\varphi, \frac{1}{2}\sin\varphi \right) : \frac{2\pi}{3} < \varphi < \frac{4\pi}{3} \right\}.$$

Thus, K is a b-bounded b-convex body with b-exp$(K) = \Gamma_3 \cup \Gamma_4$. Exposed b-faces that are not singletons are Γ_1 and Γ_2.

Corollary 22. *If K is a b-bounded b-convex body in a strictly convex Minkowski space $(\mathbb{R}^n, \|\cdot\|)$, then every exposed b-face of K meets the closure of* b-exp(K).

Proof. This is a consequence of Theorems 17 and 20. □

Corollary 23. *If S is a nonempty b-bounded subset of a strictly convex Minkowski space $(\mathbb{R}^n, \|\cdot\|)$, then* b-exp$(\mathrm{bh}_1(S)) \subseteq \mathrm{cl}(S)$.

Proof. By parts (d) and then (a) of Lemma 1, $K := \mathrm{bh}_1(S)$ is a b-bounded b-convex body and S is a subset of K. Now Theorem 20 says that $\mathrm{cl}(S) \supseteq$ b-exp$(K) =$ b-exp$(\mathrm{bh}_1(S))$. □

Example 18 gives b-bounded b-convex bodies not having any b-exposed points. This shows that Theorem 20 and Corollary 22 fail in general if the underlying norm is not strictly convex.

A similar reason justifies the assumption of b-boundedness in Theorem 20 and Corollary 22:

Proposition 24. *If a b-convex body K in a Minkowski space $(\mathbb{R}^n, \|\cdot\|)$ is not b-bounded, then* b-exp$(K) = \varnothing$.

Proof. We have $\mathrm{rad}(K) = 1$, because K is not b-bounded. Hence every supporting sphere $S(x, 1)$ of K is the boundary of a circumball $B(x, 1)$. By Lemma 2, $|K \cap S(x, 1)| \ge 2$. Hence none of the exposed b-faces of K is a singleton and K has no b-exposed points. □

7. An application to diametrically maximal sets

A bounded nonempty set $C \subseteq \mathbb{R}^n$ is called *complete* (or *diametrically maximal*) if $\operatorname{diam}(C \cup \{x\}) > \operatorname{diam}(C)$ for every $x \in \mathbb{R}^n \setminus C$; see [Meissner 1911; Jessen 1929; Eggleston 1965; Groemer 1986]. Complete sets are necessarily convex bodies, and in the Euclidean case or for $n = 2$, any complete set is of constant width. A complete set C is called a *completion* of a bounded nonempty set S if $S \subseteq C$ and $\operatorname{diam}(C) = \operatorname{diam}(S)$. Zorn's lemma shows that every bounded nonempty subset of $\mathbb{R}^n$ has at least one completion. In n-dimensional Minkowski spaces ($n \geq 3$), the family of complete bodies can form a much richer class than that of bodies of constant width; see [Moreno and Schneider 2012a; 2012b] for recent contributions.

The following problem was posed in [Martini et al. 2014, Section 4]: Given a complete set $C \subseteq \mathbb{R}^n$, find all convex bodies $K_0 \subseteq C$ such that C is the unique completion of K_0 and, moreover, there is no convex body $K \subseteq K_0$, $K \neq K_0$, such that C is the unique completion of K.

Without loss of generality, we can assume that $\operatorname{diam}(C) = 1$. The following lemma summarizes particular relevant statements from the literature (for (a) and (b), see [Eggleston 1965, Section 1(E)]; for (c) and (d), see [Groemer 1986, Theorem 5] and the short proof given there).

Lemma 25. *The following are satisfied in every Minkowski space* $(\mathbb{R}^n, \|\cdot\|)$:

(a) *A set* $C \subseteq \mathbb{R}^n$ *of diameter* 1 *is complete if and only if, for every* $x_1 \in \operatorname{bd}(C)$, *there exists* $x_2 \in \operatorname{bd}(C)$ *such that* $\|x_1 - x_2\| = 1$.

(b) *A set* $C \subseteq \mathbb{R}^n$ *of diameter* 1 *is complete if and only if* $C = \bigcap_{x \in C} B(x, 1)$.

(c) *A set* $S \subseteq \mathbb{R}^n$ *of diameter* 1 *has a unique completion if and only if* $\operatorname{bh}_1(S)$ *is complete.*

(d) *If a set* $S \subseteq \mathbb{R}^n$ *of diameter* 1 *has a unique completion* C, *then* $C = \operatorname{bh}_1(S)$.

Proposition 26. *Let* C *be a complete set of diameter* 1 *in a Minkowski space* $(\mathbb{R}^n, \|\cdot\|)$, *and let* $K \subseteq C$ *be a convex body. The following three conditions are equivalent*:

(I) *C is the unique completion of K.*

(II) $\operatorname{bh}_1(K) = C$.

(III) *K meets every exposed b-face of C.*

If, in addition, $\|\cdot\|$ *is strictly convex, then* (I), (II), *and* (III) *are equivalent to*

(IV) $\operatorname{cl}(\operatorname{conv}(\text{b-exp}(C))) \subseteq K$.

Proof. First note that C is b-bounded by (1), because $\operatorname{diam}(C) = 1$, and Lemma 25(b) shows that C is a b-bounded b-convex body.

(I)⇒(II): Since C is a completion of K, we obtain $\operatorname{diam}(K) = \operatorname{diam}(C) = 1$. Now Lemma 25(d) gives (I)⇒(II).

(II)⇔(III) and (II)⇔(IV) follow from Theorems 17 and 20, respectively.

((II)∧(III))⇒(I): By (III), there exists $x_1 \in K \cap \operatorname{bd}(C)$. Lemma 25(a) gives $x_2' \in \operatorname{bd}(C)$ such that $x_2' \in S(x_1, 1)$. Then $S(x_1, 1)$ is a supporting sphere of C and $F = C \cap S(x_1, 1)$ is an exposed b-face of C. By condition (III), there exists $x_2 \in K \cap F \subseteq K \cap S(x_1, 1)$. We obtain $\operatorname{diam}(K) = 1$, because

$$1 = \|x_1 - x_2\| \leq \operatorname{diam}(K) \leq \operatorname{diam}(C) = 1.$$

Now Lemma 25(c) gives (II)⇒(I), and we are done. □

Criteria (III) and (IV) from Proposition 26 help to characterize minimal convex bodies K_0 in a complete set C such that C is the unique completion of K_0.

Example 27. We consider the space l_∞^n as in Example 18. The only complete sets in that space are closed balls (see [Eggleston 1965, Corollary 2]), so that a complete set C of diameter 1 is necessarily a box (i.e., a square if $n = 2$) with edges of length 1 parallel to the coordinate axes. The equivalence of (I) and (III) in Proposition 26 says that C is the unique completion of a convex body $K \subseteq C$ if and only if K meets each of the $2n$ facets of C. It is easy to find minimal convex bodies K_0 with that property: such a K_0 is the convex hull of a minimal set S consisting of at least one point from each facet of C. If $n = 2$, then such K_0 can be a line segment (a diagonal of C), a triangle or a quadrangle. For arbitrary $n \geq 2$, the number of vertices of such K_0 can be 2 (if K_0 is a diagonal of C passing through the center of C), $3, \dots, 2n$ (e.g., if K_0 is the cross polytope generated by the centers of the $2n$ facets of C).

If the underlying Minkowski space is strictly convex, then Proposition 26 shows that the problem mentioned above has a unique solution.

Corollary 28. *Let C be a complete set of diameter* 1 *in a strictly convex Minkowski space* $(\mathbb{R}^n, \|\cdot\|)$. *Then* $K_0 = \operatorname{cl}(\operatorname{conv}(\text{b-exp}(C)))$ *is the unique minimal* (*under inclusion*) *convex body whose unique completion is C.*

8. Open questions

8.1. ***Spindle convexity in Minkowski spaces.*** In [Lángi et al. 2013], $\operatorname{bh}_1(\{x_1, x_2\})$ is called the *spindle* of $x_1, x_2 \in \mathbb{R}^n$, which generalizes the corresponding notion from Euclidean space (see, e.g., [Bezdek et al. 2007]). A set $S \subseteq \mathbb{R}^n$ is called *spindle convex* if, for all $x_1, x_2 \in S$, S covers the whole spindle of x_1 and x_2. This gives rise to the concept of the *spindle convex hull* of a subset of $\mathbb{R}^n$. Note that spindle convex sets are not necessarily closed, in contrast to b-convex sets. Closed sets turn out to be spindle convex if and only if they are b-convex, provided the underlying Minkowski space is Euclidean or two-dimensional or its unit ball is (an affine image of) a cube (see [Bezdek et al. 2007, Corollary 3.4; Lángi et al. 2013, Corollaries 3.13 and 3.15]). An example in (an affine image of) the space l_1^3 from Example 7 shows that closed spindle convex sets need not be b-convex in general (see [Lángi et al. 2013, Example 3.1]).

We define a related hierarchy of notions of convexity by calling a set $S \subseteq \mathbb{R}^n$ *k-spindle convex*, $k \in \{2, 3, \ldots\}$, if $\mathrm{bh}_1(\{x_1, \ldots, x_k\}) \subseteq S$ for all $x_1, \ldots, x_k \in S$. We call S *$\ast$-spindle convex* if $\mathrm{bh}_1(F) \subseteq S$ for every finite $F \subseteq S$ (i.e., if S is k-spindle convex for all $k = 2, 3, \ldots$).

Are the k-spindle convex hulls and the $\ast$-spindle convex hull of a closed set closed? Clearly, every b-convex set is $\ast$-spindle convex. Is every closed $\ast$-spindle convex set b-convex? Theorem 14 says that in many situations $\mathrm{bh}_1(S)$ is the closure of the $\ast$-spindle convex hull of S. On the other hand, the relatively open segment S from Example 16 is $\ast$-spindle convex, but $\mathrm{cl}(S)$ is not even 2-spindle convex. Given an arbitrary Minkowski space $(\mathbb{R}^n, \|\cdot\|)$, does there exist $k \in \{2, 3, \ldots\}$ such that $\ast$-spindle convexity coincides with k-spindle convexity? Given $k \in \{2, 3, \ldots\}$, does there exist a Minkowski space $(\mathbb{R}^n, \|\cdot\|)$ such that k-spindle convexity differs from $(k+1)$-spindle convexity? These and related questions might be studied to continue naturally our investigations here.

8.2. ***Generalized Minkowski spaces.*** Our results are shown in the framework of a Minkowski space. What remains true if the norm is replaced by a gauge, i.e., if the unit ball is no longer necessarily centered at o?

8.3. ***Möbius geometry.*** One might check whether there are interesting connections (e.g., regarding the used methods and tools) to Möbius geometry where spheres also somehow play the role of hyperplanes; see, e.g., [Volenec 1976; Langevin and Teufel 2009].

References

[Alimov 2012] A. R. Alimov, "Monotone path-connectedness of R-weakly convex sets in spaces with linear ball embedding", *Eurasian Math. J.* **3**:2 (2012), 21–30. MR Zbl

[Amir and Ziegler 1980] D. Amir and Z. Ziegler, "Relative Chebyshev centers in normed linear spaces, I", *J. Approx. Theory* **29**:3 (1980), 235–252. MR Zbl

[Balashov 2002] M. V. Balashov, "An analogue of the Kreĭn–Milman theorem for a strongly convex hull in a Hilbert space", *Math. Notes* **71**:1-2 (2002), 34–38. MR Zbl

[Balashov and Ivanov 2006] M. V. Balashov and G. E. Ivanov, "On the farthest points of sets", *Math. Notes* **80**:1-2 (2006), 159–166. MR Zbl

[Balashov and Polovinkin 2000] M. V. Balashov and E. S. Polovinkin, "M-strongly convex subsets and their generating sets", *Sb. Math.* **191**:1-2 (2000), 25–60. MR Zbl

[Bezdek 2010] K. Bezdek, *Classical topics in discrete geometry*, Springer, New York, 2010. MR Zbl

[Bezdek 2012] K. Bezdek, "Illuminating spindle convex bodies and minimizing the volume of spherical sets of constant width", *Discrete Comput. Geom.* **47**:2 (2012), 275–287. MR Zbl

[Bezdek 2013] K. Bezdek, *Lectures on sphere arrangements: the discrete geometric side*, Fields Institute Monographs **32**, Springer, New York, 2013. MR Zbl

[Bezdek and Naszódi 2006] K. Bezdek and M. Naszódi, "Rigidity of ball-polyhedra in Euclidean 3-space", *European J. Combin.* **27**:2 (2006), 255–268. MR Zbl

[Bezdek and Naszódi 2015] K. Bezdek and M. Naszódi, "Spindle star-shaped sets", *Aequationes Math.* **89**:3 (2015), 803–819. MR Zbl

[Bezdek et al. 2006] K. Bezdek, R. Connelly, and B. Csikós, "On the perimeter of the intersection of congruent disks", *Beiträge Algebra Geom.* **47**:1 (2006), 53–62. MR Zbl

[Bezdek et al. 2007] K. Bezdek, Z. Lángi, M. Naszódi, and P. Papez, "Ball-polyhedra", *Discrete Comput. Geom.* **38**:2 (2007), 201–230. MR Zbl

[Bieberbach 1955] L. Bieberbach, "Zur Euklidischen Geometrie der Kreisbogendreiecke", *Math. Ann.* **130**:1 (1955), 46–86. MR Zbl

[Bieberbach 1970] L. Bieberbach, "Zwei Kongruenzsätze für Kreisbogendreiecke in der Euklidischen Ebene", *Math. Ann.* **190**:2 (1970), 97–118. MR Zbl

[Bohnenblust 1938] F. Bohnenblust, "Convex regions and projections in Minkowski spaces", *Ann. Math.* (2) **39**:2 (1938), 301–308. MR Zbl

[Boltyanski et al. 1997] V. Boltyanski, H. Martini, and P. S. Soltan, *Excursions into combinatorial geometry*, Springer, Berlin, 1997. MR Zbl

[Danzer et al. 1963] L. Danzer, B. Grünbaum, and V. Klee, "Helly's theorem and its relatives", pp. 101–180 in *Convexity* (Seattle, 1961), edited by V. Klee, Proceedings of Symposia in Pure Mathematics **7**, American Mathematical Society, Providence, RI, 1963. MR Zbl

[Eggleston 1965] H. G. Eggleston, "Sets of constant width in finite dimensional Banach spaces", *Israel J. Math.* **3**:3 (1965), 163–172. MR Zbl

[Fodor and Vígh 2012] F. Fodor and V. Vígh, "Disc-polygonal approximations of planar spindle convex sets", *Acta Sci. Math.* (*Szeged*) **78**:1-2 (2012), 331–350. MR Zbl

[Groemer 1986] H. Groemer, "On complete convex bodies", *Geom. Dedicata* **20**:3 (1986), 319–334. MR Zbl

[Grünbaum 1956] B. Grünbaum, "A proof of Vázsonyi's conjecture", *Bull. Res. Council Israel. Sect. A.* **6** (1956), 77–78. MR Zbl

[Heppes 1956] A. Heppes, "Beweis einer Vermutung von A. Vázsonyi", *Acta Math. Acad. Sci. Hungar.* **7**:3 (1956), 463–466. MR Zbl

[Heppes and Révész 1956] A. Heppes and P. Révész, "A splitting problem of Borsuk", *Mat. Lapok* **7** (1956), 108–111. In Hungarian. MR Zbl

[Jessen 1929] B. Jessen, "Über konvexe Punktmengen konstanter Breite", *Math. Z.* **29**:1 (1929), 378–380. MR JFM

[Kupitz et al. 2010] Y. S. Kupitz, H. Martini, and M. A. Perles, "Ball polytopes and the Vázsonyi problem", *Acta Math. Hungar.* **126**:1-2 (2010), 99–163. MR Zbl

[Langevin and Teufel 2009] R. Langevin and E. Teufel, "Conformal width in Möbius geometry", *Beiträge Algebra Geom.* **50**:2 (2009), 369–387. MR Zbl

[Lángi et al. 2013] Z. Lángi, M. Naszódi, and I. Talata, "Ball and spindle convexity with respect to a convex body", *Aequationes Math.* **85**:1-2 (2013), 41–67. MR Zbl

[Lassak 1977] M. Lassak, "On metric B-convexity for which diameters of any set and its hull are equal", *Bull. Acad. Polon. Sci. Sér. Sci. Math. Ast. Phys.* **25**:10 (1977), 969–975. MR Zbl

[Lassak 1979] M. Lassak, "Some properties of B-convexity in Minkowski–Banach spaces", *Bull. Acad. Polon. Sci. Sér. Sci. Math.* **27**:1 (1979), 97–106. MR Zbl

[Lassak 1982] M. Lassak, "Some connections between B-convexity and d-convexity", *Demonstratio Math.* **15**:1 (1982), 261–270. MR Zbl

[Martini and Spirova 2009] H. Martini and M. Spirova, "On the circular hull property in normed planes", *Acta Math. Hungar.* **125**:3 (2009), 275–285. MR Zbl

[Martini and Swanepoel 2004] H. Martini and K. J. Swanepoel, "The geometry of Minkowski spaces: a survey, II", *Expo. Math.* **22**:2 (2004), 93–144. MR Zbl

[Martini et al. 2001] H. Martini, K. J. Swanepoel, and G. Weiß, "The geometry of Minkowski spaces: a survey, I", *Expo. Math.* **19**:2 (2001), 97–142. MR Zbl

[Martini et al. 2002] H. Martini, K. J. Swanepoel, and G. Weiß, "The Fermat–Torricelli problem in normed planes and spaces", *J. Optim. Theory Appl.* **115**:2 (2002), 283–314. MR Zbl

[Martini et al. 2013] H. Martini, C. Richter, and M. Spirova, "Intersections of balls and sets of constant width in finite-dimensional normed spaces", *Mathematika* **59**:2 (2013), 477–492. MR Zbl

[Martini et al. 2014] H. Martini, P. L. Papini, and M. Spirova, "Complete sets and completion of sets in Banach spaces", *Monatsh. Math.* **174**:4 (2014), 587–597. MR Zbl

[Mayer 1935] A. E. Mayer, "Eine Überkonvexität", *Math. Z.* **39**:1 (1935), 511–531. MR Zbl

[Meissner 1911] E. Meissner, "Über Punktmengen konstanter Breite", *Zürich Naturf.-Ges.* **56** (1911), 42–50. JFM

[Montejano 1991] L. Montejano, "A characterization of the Euclidean ball in terms of concurrent sections of constant width", *Geom. Dedicata* **37**:3 (1991), 307–316. MR Zbl

[Moreno and Schneider 2007a] J. P. Moreno and R. Schneider, "Continuity properties of the ball hull mapping", *Nonlinear Anal.* **66**:4 (2007), 914–925. MR Zbl

[Moreno and Schneider 2007b] J. P. Moreno and R. Schneider, "Intersection properties of polyhedral norms", *Adv. Geom.* **7**:3 (2007), 391–402. MR Zbl

[Moreno and Schneider 2012a] J. P. Moreno and R. Schneider, "Diametrically complete sets in Minkowski spaces", *Israel J. Math.* **191**:2 (2012), 701–720. MR Zbl

[Moreno and Schneider 2012b] J. P. Moreno and R. Schneider, "Structure of the space of diametrically complete sets in a Minkowski space", *Discrete Comput. Geom.* **48**:2 (2012), 467–486. MR Zbl

[Papez 2010] P. D. Papez, *Ball-polyhedra, spindle-convexity and their properties*, Ph.D. thesis, University of Calgary, 2010, Available at http://search.proquest.com/docview/734404417. MR

[Polovinkin 1996a] E. S. Polovinkin, "On convex and strongly convex approximations of sets", *Dokl. Akad. Nauk* **350**:3 (1996), 308–311. In Russian; translated in *Dokl. Math.* **54**:2 (1996), 710–713. MR Zbl

[Polovinkin 1996b] E. S. Polovinkin, "Strongly convex analysis", *Sb. Math.* **187**:2 (1996), 259–286. MR Zbl

[Polovinkin 1997] E. S. Polovinkin, "Generalization of Carathéodory and Kreĭn–Milman theorems for strongly convex sets", *Dokl. Akad. Nauk* **355**:2 (1997), 164–166. In Russian; translated in *Dokl. Math.* **56**:1 (1997), 512–514. MR Zbl

[Polovinkin and Balashov 2007] E. S. Polovinkin and M. V. Balashov, Элементы выпуклого и сильно выпуклого анализа, 2nd ed., Fizmatlit, Moscow, 2007. Zbl

[Schneider 1993] R. Schneider, *Convex bodies: the Brunn–Minkowski theory*, Encyclopedia of Mathematics and its Applications **44**, Cambridge Univ. Press, 1993. MR Zbl

[Straszewicz 1957] S. Straszewicz, "Sur un problème géométrique de P. Erdős", *Bull. Acad. Polon. Sci. Cl. III.* **5** (1957), 39–40. MR Zbl

[Valentine 1964] F. A. Valentine, *Convex sets*, McGraw-Hill, New York, 1964. MR Zbl

[Volenec 1976] V. Volenec, "Axiomatic foundation of the n-dimensional Möbius geometry", *Publ. Math. Debrecen* **23**:1-2 (1976), 89–102. MR Zbl

Received March 11, 2016. Revised January 3, 2017.

THOMAS JAHN
FACULTY OF MATHEMATICS
UNIVERSITY OF TECHNOLOGY
D-09107 CHEMNITZ
GERMANY
thomas.jahn@mathematik.tu-chemnitz.de

HORST MARTINI
FACULTY OF MATHEMATICS
UNIVERSITY OF TECHNOLOGY
D-09107 CHEMNITZ
GERMANY
martini@mathematik.tu-chemnitz.de

CHRISTIAN RICHTER
INSTITUTE OF MATHEMATICS
FRIEDRICH SCHILLER UNIVERSITY
D-07737 JENA
GERMANY
christian.richter@uni-jena.de

PACIFIC JOURNAL OF MATHEMATICS
Vol. 289, No. 2, 2017
dx.doi.org/10.2140/pjm.2017.289.317

LOCAL CONSTANCY OF DIMENSION OF SLOPE SUBSPACES OF AUTOMORPHIC FORMS

JOACHIM MAHNKOPF

We prove an analogue of a Gouvêa–Mazur conjecture on local constancy of dimension of slope subspaces of modular forms on the upper half plane for automorphic forms on reductive algebraic groups $\tilde{G}/\mathbb{Q}$ having discrete series. The proof uses a comparison of Bewersdorff's elementary trace formula for pairs of congruent weights and does not make use of methods from p-adic Banach space theory, overconvergent forms or rigid analytic geometry.

We also compare two Goresky–MacPherson trace formulas computing Lefschetz numbers on weighted cohomology for pairs of congruent weights; this has an application to a more explicit version of the Gouvêa–Mazur conjecture for symplectic groups of rank 2.

Introduction

0.1. We fix a prime $p \in \mathbb{N}$ and an integer N not divisible by p. Generalizing Hida's theory [1993; 1988] of ordinary modular forms, Gouvêa and Mazur [1992] conjectured that the dimension $d(\beta, k)$ of the slope β subspace of the space $S_k(\Gamma_0(pN))$ of cuspidal modular forms of level pN and weight k is locally constant in the p-adic topology as a function of k. More precisely, they conjectured that there is a linear polynomial $\boldsymbol{m}(x)$ such that the conditions $k, k' \geq 2\beta + 2$ and $k \equiv k' \pmod{(p-1)p^{m-1}}$ with $m \geq \boldsymbol{m}(\beta)$ imply

$$d(\beta, k) = d(\beta, k'). \tag{1}$$

Using work of Coleman [1997] which is based on rigid analytic geometry and p-adic spectral theory, as well as Katz's theory of p-adic modular forms and results of Gouvêa and Mazur, Wan [1998] proved that there is a quadratic polynomial $\boldsymbol{m}(x)$ such that equation (1) holds. On the other hand, Buzzard and Calegari [2004] showed that in general there is no linear polynomial $\boldsymbol{m}(x)$ such that (1) holds, hence, Wan's result is best possible.

MSC2010: 11F75.
Keywords: Gouvêa–Mazur conjecture, cohomology of arithmetic groups.

0.2. In this article we prove a higher-rank analogue of the Gouvêa–Mazur conjecture. To describe this in more detail, we denote by $\mathbb{A}$ the adeles of $\mathbb{Q}$ and fix a prime $p \in \mathbb{N}$. We let $\tilde{\boldsymbol{G}}/\mathbb{Q}$ be a connected reductive algebraic group which contains a maximal torus $\tilde{\boldsymbol{T}}/\mathbb{Q}$ which splits over $\mathbb{Q}_p$. We select a basis Δ of the root system Φ of $\tilde{\boldsymbol{G}}/F$ where $F/\mathbb{Q}$ is a (minimal) splitting field for $\tilde{\boldsymbol{T}}/\mathbb{Q}$. We let $\tilde{K} \leq \tilde{\boldsymbol{G}}(\mathbb{A}_f)$ be a compact open subgroup with p-component $\tilde{K}_p$ equal to the Iwahori subgroup $\tilde{\mathcal{I}}$ of $\tilde{\boldsymbol{G}}(\mathbb{Q}_p)$. We denote by $\mathbb{T}$ the Hecke operator attached to the double coset $\tilde{K}h^{-1}r\tilde{K}$ where $h \in \tilde{\boldsymbol{T}}(\mathbb{Q})^{++} \leq \tilde{\boldsymbol{T}}(\mathbb{Q}_p)^{++}$ is a strictly dominant element and $r \in \tilde{\boldsymbol{G}}(\mathbb{A}_f)^{(p)}$, i.e., r has trivial p-component. For any dominant weight $\tilde{\lambda} \in X(\tilde{\boldsymbol{T}})$ we understand by $L_{\tilde{\lambda}}$ the irreducible representation of $\tilde{\boldsymbol{G}}/F$ of highest weight $\tilde{\lambda}$. The normalization $\mathbb{T}_{\tilde{\lambda}}$ of $\mathbb{T}$ acts on full cohomology $H^i(S_{\tilde{K}}, L_{\tilde{\lambda}}(\mathbb{Q}_p))$ as well as on cuspidal cohomology $H^i_{\text{cusp}}(S_{\tilde{K}}, L_{\tilde{\lambda}}(\mathbb{C}))$ where $S_{\tilde{K}} = \tilde{\boldsymbol{G}}(\mathbb{Q})\backslash\tilde{\boldsymbol{G}}(\mathbb{A})/\tilde{K}\tilde{K}_\infty$ is the locally symmetric space. We write $H^i(S_{\tilde{K}}, L_{\tilde{\lambda}}(\mathbb{Q}_p))^{\beta}$ (resp. $H^i(S_{\tilde{K}}, L_{\tilde{\lambda}}(\mathbb{Q}_p))^{\leq\beta}$) for the subspace of slope β (resp. slope $\leq \beta$) w.r.t. $\mathbb{T}_{\tilde{\lambda}}$ and we use analogous notation for cuspidal cohomology. Our main results then are as follows.

Theorem A (see 3.10 Corollary, 4.11.4 Theorem). *Let $s = |\Phi^+|$ be the number of positive roots of $\tilde{\boldsymbol{G}}/\mathbb{Q}_p$, $\sigma = \max_{\alpha\in\Phi^+} \operatorname{ht}(\alpha)$ the maximal height of a positive root and $\boldsymbol{g}_i$ the number of i-cells in a finite cell complex $\mathcal{Z}$ which is homotopy equivalent to the Borel Serre compactification $\bar{S}_{\tilde{K}}$ of $S_{\tilde{K}}$. Then for all $\beta \in \mathbb{Q}_{\geq 0}$, $\tilde{\lambda} \in X(\tilde{\boldsymbol{T}})^{\text{dom}}$ and $i \in \mathbb{N}_0$ we obtain*

$$\dim H^i(S_{\tilde{K}}, L_{\tilde{\lambda}}(\mathbb{Q}_p))^{\leq\beta} \leq \mathtt{m}\beta^s + \mathtt{n};$$

here, $\mathtt{m} = 12\boldsymbol{g}_i\sigma^{s+1}/s$ and $\mathtt{n} \in \mathbb{N}$ is an integer which also only depends on $\tilde{K}$ (and, hence, on $\tilde{\boldsymbol{G}}$ and p) and on i.

Theorem B (see 5.2 Theorem). *We assume that $\tilde{\boldsymbol{G}}$ has discrete series and we denote by $d = d_{\tilde{\boldsymbol{G}}}$ the middle degree. There are polynomials $\boldsymbol{m}_1(x), \boldsymbol{m}_2(x) \in \mathbb{Q}[x]$ both of degree $s+1$ and leading term $12\boldsymbol{g}_d\sigma^{s+1}/s$ which only depend on $\tilde{K}$ (hence, on $\tilde{\boldsymbol{G}}$ and p) and $h \in \tilde{\boldsymbol{T}}(\mathbb{Q})^{++}$ with the following property. Let $\beta \in \mathbb{Q}_{\geq 0}$. Suppose the dominant weights $\tilde{\lambda}, \tilde{\lambda}' \in X(\tilde{\boldsymbol{T}})$ satisfy*

- $\langle\tilde{\lambda}, \alpha^\vee\rangle > 2\boldsymbol{m}_1(\beta)$ *and* $\langle\tilde{\lambda}', \alpha^\vee\rangle > 2\boldsymbol{m}_1(\beta)$ *for all* $\alpha \in \Delta_{\tilde{\boldsymbol{G}}}$;
- $\tilde{\lambda} \equiv \tilde{\lambda}' \pmod{(p-1)p^{m-1}X(\tilde{\boldsymbol{T}})}$ *with* $m \geq \boldsymbol{m}_2(\beta)$ ($m \in \mathbb{N}$).

Then

$$\dim H^d_{\text{cusp}}(S_{\tilde{K}}, L_{\tilde{\lambda}}(\mathbb{C}))^{\gamma} = \dim H^d_{\text{cusp}}(S_{\tilde{K}}, L_{\tilde{\lambda}'}(\mathbb{C}))^{\gamma} \quad \textit{for all } 0 \leq \gamma \leq \beta.$$

Remark. In the $\mathbf{GL}_2$-case $\boldsymbol{m}_2(x)$ is a quadratic polynomial; i.e., we obtain the same growth as that of $\boldsymbol{m}(x)$ in [Wan 1998], except that the weights have to satisfy a stronger lower bound (quadratic in β instead of linear as in that paper).

0.3. To prove Theorems A and B we will mostly work in a non-adelic setting; i.e., $\Gamma \leq \tilde{\boldsymbol{G}}(\mathbb{Q})$ denotes an arithmetic subgroup contained in $\tilde{\mathcal{I}}$. Theorem A then is an extension of the main result of [Mahnkopf 2014] (see Section 3.1) and the proof is based on an extension of the notion of truncation of an irreducible representation of $\tilde{\boldsymbol{G}}/\mathbb{Q}_p$ introduced in [Mahnkopf 2013; 2014].

Using the boundedness result of Theorem A the proof of Theorem B reduces to proving certain congruences between traces of powers of the Hecke operator $\mathbb{T}_{\tilde{\lambda}}$ on $H^i_{\mathrm{cusp}}(S_{\tilde{K}}, L_{\tilde{\lambda}}(\mathbb{C}))$ and on $H^i_{\mathrm{cusp}}(S_{\tilde{K}}, L_{\tilde{\lambda}'}(\mathbb{C}))$ for p-adically close weights $\tilde{\lambda}, \tilde{\lambda}'$. We first verify these congruences on full cohomology and our principal tool for this is a comparison of a simple and elementary trace formula of Bewersdorff [1985] for cohomology with coefficients in $L_{\tilde{\lambda}}$ and in $L_{\tilde{\lambda}'}$. The equality of mod p^n reductions of geometric sides essentially follows from p-adic properties of the diagonalization of elements in $\tilde{\mathcal{I}}h^e\tilde{\mathcal{I}} \subseteq \tilde{\boldsymbol{G}}(\mathbb{Q}_p)$, $e \in \mathbb{N}$, which are proved using basic algebra (see 4.3 Lemma and 4.4 Proposition); we note that $\tilde{\mathcal{I}}h^{-e}\tilde{\mathcal{I}}$ is the p-component of $\mathbb{T}^e$. To obtain congruences on cuspidal cohomology we directly prove congruences on the Eisenstein part of full cohomology and subtract from congruences on full cohomology.

Since the Bewersdorff trace formula is elementary we obtain an elementary proof of the congruences on full cohomology and the proofs of Theorems A and B do not make use of methods from p-adic Banach space theory, overconvergent cohomology or rigid analytic geometry (but use the spectral decomposition of full cohomology for regular weight).

0.4. ***Weighted cohomology.*** Goresky and MacPherson [Goresky and MacPherson 2003] proved a trace formula for Lefschetz numbers of Hecke operators on weighted cohomology. Unlike Bewersdorff's formula it contains contributions not only from $\tilde{\boldsymbol{G}}$ but from all $\mathbb{Q}$-parabolic subgroups of $\tilde{\boldsymbol{G}}$. Nevertheless, the same diagonalization of elements in $\tilde{\mathcal{I}}h^e\tilde{\mathcal{I}} \subseteq \tilde{\boldsymbol{G}}(\mathbb{Q}_p)$ as in 0.3 allows to compare two Goresky–MacPherson trace formulas for pairs of congruent weights. This then yields certain congruences on weighted cohomology groups and has an application to a version of the Gouvêa–Mazur conjecture for symplectic groups of rank 2 which is more explicit since we avoid use of the spectral decomposition of full cohomology (see Section 5.8). We note that this depends on properties of the root system C_2 but using instead the Goresky–Kottwitz–MacPherson trace formula [Goresky et al. 1997] together with the calculations of Spallone [2009] it might be possible to extend this to arbitrary reductive groups $\tilde{\boldsymbol{G}}/\mathbb{Q}$.

0.5. Buzzard [Buzzard 2001] gave an elementary proof of boundedness of dimension of slope subspaces in the case $\mathbf{GL}_2/\mathbb{Q}$ also based on an analysis of representations of $\mathbf{GL}_2(\mathbb{Z}_p)$. In the case of quaternion algebras over $\mathbb{Q}$ he also proved in [Buzzard 1998] local constancy of dimension of slope subspaces and his results were generalized to $\mathbf{GL}_2$ over totally real fields by Pande [2009]. Following the method of Ash and Stevens [2008], who introduced overconvergent cohomology, Urban [2011] obtained p-adic families of systems of Hecke eigenvalues; he uses this to also derive a p-adic trace formula on overconvergent cohomology. Andreatta, Iovita and Pilloni also proved existence of p-adic families of eigenforms using rigid analytic geometry; see [Andreatta et al. 2015].

More closely related to our approach is work of Koike [1975; 1976] (and some unpublished work of Clozel); like Buzzard, Koike does not make use of methods from rigid analytic geometry or p-adic Banach space theory. In the case of cuspidal modular forms, i.e., in the case $\mathbf{GL}_2/\mathbb{Q}$ he uses a Selberg trace formula which yields an explicit expression for the trace of Hecke operators to deduce congruences between traces of Hecke operators. Since the Selberg trace formula becomes much more involved this seems difficult to generalize to higher rank. We therefore do not attempt to determine an explicit expression for the trace of Hecke operators but only equate mod p^n reductions of traces for p-adically close weights $\tilde{\lambda}$, $\tilde{\lambda}'$. This can be done even in higher rank by comparing the simple (non-explicit) trace formula of Bewersdorff for weights $\tilde{\lambda}$ and $\tilde{\lambda}'$.

1. Chevalley groups

We recall some basic facts from the theory of Chevalley groups and their representations and we give proofs for some (technical) results for which we do not know a reference.

1.1. ***Complex semisimple Lie algebras.*** Let $\mathfrak{g}$ be a complex semisimple Lie algebra. We denote by $\mathfrak{h}$ a Cartan subalgebra of $\mathfrak{g}$ and by $\Phi = \Phi(\mathfrak{g}, \mathfrak{h})$ the set of roots of $\mathfrak{g}$ w.r.t. to $\mathfrak{h}$. We choose a basis Δ of Φ and we denote by Φ^+ the set of positive roots. For each root $\alpha \in \Phi$ we write $\mathfrak{g}(\alpha)$ for the corresponding root subspace of $\mathfrak{g}$ and we select elements $h_\alpha \in \mathfrak{h}$, $\alpha \in \Delta$, and $x_\alpha \in \mathfrak{g}(\alpha)$, $\alpha \in \Phi$, such that $\{h_\alpha,\ \alpha \in \Delta,\ x_\beta,\ \beta \in \Phi\}$ is a *Chevalley basis* of $\mathfrak{g}$. In particular, h_α is the coroot corresponding to $\alpha \in \Delta$. The Chevalley basis yields $\mathbb{Z}$-forms $\mathfrak{g}(\mathbb{Z}) = \bigoplus_{\beta\in\Phi} \mathbb{Z}x_\beta \oplus \bigoplus_{\alpha\in\Delta} \mathbb{Z}h_\alpha$ (resp. $\mathfrak{h}(\mathbb{Z}) = \bigoplus_{\alpha\in\Delta} \mathbb{Z}h_\alpha$) of $\mathfrak{g}$ (resp. of $\mathfrak{h}$). We denote by $\mathcal{U}_\mathbb{Z}$ the $\mathbb{Z}$-form of the universal enveloping algebra $\mathcal{U}$ of $\mathfrak{g}$ which as a ring is generated by the elements $x_\alpha^n/n!$, $\alpha \in \Phi$, $n \in \mathbb{N}_0$ (see [Humphreys 1972, Theorem 26.4, p. 156]). We set $\mathfrak{g}(R) = \mathfrak{g}(\mathbb{Z}) \otimes R$, $\mathfrak{h}(R) = \mathfrak{h}(\mathbb{Z}) \otimes R$ and $\mathcal{U}_R = \mathcal{U}_\mathbb{Z} \otimes R$, R a $\mathbb{Z}$-algebra. We set $s = |\Phi^+|$ and we fix an ordering $\Phi^+ = \{\alpha_1, \dots, \alpha_s\}$ of the set of positive roots and we set

$$X_\pm^{\boldsymbol{n}} = \frac{x_{\pm\alpha_1}^{n_1}}{n_1!} \cdots \frac{x_{\pm\alpha_s}^{n_s}}{n_s!} \in \mathcal{U}_\mathbb{Z},$$

where $\boldsymbol{n} = (n_1, \dots, n_s) \in \mathbb{N}_0^s$. The $\mathbb{Z}$-span of the elements $X_-^{\boldsymbol{n}}$, where $\boldsymbol{n} \in \mathbb{N}_0^s$, is a

$\mathbb{Z}$-form $\mathcal{U}_{\mathbb{Z}}^-$ of the universal envelopping algebra $\mathcal{U}^-$ of $\mathfrak{n}^- = \bigoplus_{\alpha<0} \mathfrak{g}(\alpha)$. Finally, $\mu \leq \lambda$, for $\lambda, \mu \in \mathfrak{h}^*$, means that $\lambda - \mu$ is a linear combination of positive roots with nonnegative coefficients. For any $\lambda \in \mathfrak{h}^*$ we define a relative height function $\mathrm{ht}_\lambda : \{\mu \in \mathfrak{h}^* : \mu \leq \lambda\} \to \mathbb{N}_0$ by $\mathrm{ht}_\lambda(\mu) = \mathrm{ht}(\lambda - \mu)$ (see [Mahnkopf 2013, 1.3]); here, $\mathrm{ht} = \mathrm{ht}_\Delta$ is the height function corresponding to Δ, i.e., $\mathrm{ht}(\lambda - \mu) = \sum_{\alpha \in \Delta} n_\alpha$ if $\lambda - \mu = \sum_{\alpha \in \Delta} n_\alpha \alpha$. By $\omega_\alpha \in \mathfrak{h}^*$, $\alpha \in \Delta$, we understand the fundamental (dominant) weights, i.e., $\omega_\beta(h_\alpha) = \delta_{\alpha,\beta}$. The fundamental weights span the weight lattice Γ_{sc} of $\mathfrak{g}$ which contains the root lattice Γ_{ad}.

For any integral and dominant weight $\lambda \in \mathfrak{h}^*$ we denote by $(\rho_\lambda, L_\lambda)$ the complex irreducible $\mathfrak{g}$-module of highest weight λ. We denote by Γ_λ the subgroup of the weight lattice Γ_{sc} of $\mathfrak{g}$ which is generated by the (finite) set of weights P_λ of L_λ. The representation ρ_λ is defined over $\mathbb{Z}$, i.e., $L_\lambda = L_\lambda(\mathbb{Z}) \otimes \mathbb{C}$ where $L_\lambda(\mathbb{Z})$ is $\mathcal{U}_{\mathbb{Z}}$-invariant. We select a highest weight vector $v_\lambda \in L_\lambda$; the lattice then is defined as $L_\lambda(\mathbb{Z}) = \mathcal{U}_{\mathbb{Z}} v_\lambda$ (see [Humphreys 1972, proof of Theorem 27.1, p. 158]). Moreover, we set $L_\lambda(\mathbb{Z}, \mu) = L_\lambda(\mathbb{Z}) \cap L_\lambda(\mu)$ where $L_\lambda(\mu) \subseteq L_\lambda$ is the weight μ subspace and obtain $L_\lambda(\mathbb{Z}) = \bigoplus_{\mu \leq \lambda} L_\lambda(\mathbb{Z}, \mu)$ by [Humphreys 1972, Theorem 27.1, p. 158]. More generally, for any $\mathbb{Z}$-algebra R we put $L_\lambda(R) = R \otimes L_\lambda(\mathbb{Z})$ and $L_\lambda(R, \mu) = R \otimes L_\lambda(\mathbb{Z}, \mu)$. The space $L_\lambda(R)$ is a $\mathcal{U}_R$-module and $L_\lambda(R, \mu)$ is a $\mathfrak{h}(R)$-module and the weight decomposition of $L_\lambda(R)$ w.r.t. $\mathfrak{h}(R)$ reads

$$L_\lambda(R) = \bigoplus_{\mu \leq \lambda} L_\lambda(R, \mu).$$

More generally, let (π, L_π) be a faithful complex finite dimensional representation of $\mathfrak{g}$. Since $\pi = \bigoplus_i \rho_{\lambda_i}$ is semisimple (by the theorem just cited) there is a $\mathcal{U}_{\mathbb{Z}}$-invariant lattice $L_\pi(\mathbb{Z})$ in L_π i.e., $L_\pi = L_\pi(\mathbb{Z}) \otimes \mathbb{C}$. Furthermore, for any weight $\mu \in \mathfrak{h}^*$ we set $L_\pi(\mathbb{Z}, \mu) = L_\pi(\mathbb{Z}) \cap L_\pi(\mu)$ and $L_\pi(R, \mu) = R \otimes L_\pi(\mathbb{Z}, \mu)$ (R a $\mathbb{Z}$-algebra) and obtain

$$L_\pi(R) = \bigoplus_{\mu \in P_\pi} L_\pi(R, \mu),$$

where $P_\pi \subseteq \mathfrak{h}^*$ is the set of weights of π. We note that $P_\pi = \bigcup_i P_{\lambda_i}$ and we set $\Gamma_\pi = \langle P_\pi \rangle$.

1.2. ***Chevalley groups $/\mathbb{Z}_p$.*** From now on we fix an algebraic closure $\bar{\mathbb{Q}}_p$ of $\mathbb{Q}_p$ and we denote by $\mathcal{O}_{\bar{\mathbb{Q}}_p}$ the integer ring in $\bar{\mathbb{Q}}_p$. We recall some basic facts from the theory of Chevalley groups. Let (π, L_π) be a finite dimensional complex representation of $\mathfrak{g}$ and let R by a $\mathbb{Z}_p$-algebra. For any $t \in R$ and any root $\alpha \in \Phi$ we define the element $x_\alpha(t) = x_\alpha^\pi(t) = \exp(\pi(t x_\alpha)) \in \mathrm{Aut}(L_\pi(R))$. The subgroup

$$G_{\pi,R} = \langle x_\alpha(t_\alpha), \alpha \in \Phi, t_\alpha \in R \rangle \leq \mathrm{Aut}(L_\pi(R))$$

is called the *Chevalley group* attached to π and R.

The group $G_{\pi,\bar{\mathbb{Q}}_p}$ is a semisimple connected algebraic group, i.e., it is the set of $\bar{\mathbb{Q}}_p$-points of an algebraic group (group scheme) $\boldsymbol{G}_\pi$ which is defined over $\mathbb{Z}_p$. To make this more precise, we denote by $\mathbf{GL}_n$ the general linear group with canonical $\mathbb{Z}_p$-structure $\mathbb{Z}_p[\mathbf{GL}_n] = \mathbb{Z}_p[x_{ij}, \det^{-1}]$, i.e., for any $\mathbb{Z}_p$-algebra R we obtain

$$\mathbf{GL}_n(R) = \mathrm{Mor}_{\mathbb{Z}_p-alg}(\mathbb{Z}_p[\mathbf{GL}_n], R) = \{(x_{ij}) \in R^{n^2} : \det(x_{ij}) \in R^*\}.$$

We select a basis $\mathcal{B}$ of the free $\mathbb{Z}_p$-module $L_\pi(\mathbb{Z}_p)$ and obtain for any $\mathbb{Z}_p$-algebra R an identification

$$\mathrm{Aut}(L_\pi(R)) \overset{\mathcal{B}}{=} \mathbf{GL}_n(R) \quad (n = \dim L_\pi).$$

In particular, $G_{\pi,\bar{\mathbb{Q}}_p}$ is a subset of $\mathrm{Aut}(L_\pi(\bar{\mathbb{Q}}_p)) = \mathbf{GL}_n(\bar{\mathbb{Q}}_p)$ and it is the set of $\bar{\mathbb{Q}}_p$-points $\boldsymbol{G}_\pi(\bar{\mathbb{Q}}_p)$ of a closed algebraic subgroup $\boldsymbol{G}_\pi = \boldsymbol{G}_\pi/\mathbb{Q}_p$ of $\mathbf{GL}_n/\mathbb{Q}_p$ which is defined over $\mathbb{Q}_p$ (see [Borel 1970, 3.3(1), p. 14 and 3.4, p. 18]); in particular, $\mathbb{Q}_p[\boldsymbol{G}_\pi] = \mathbb{Q}_p[\mathbf{GL}_n]/J'$ for some ideal $J' \leq \mathbb{Q}_p[\mathbf{GL}_n]$.

We set $J = J' \cap \mathbb{Z}_p[\mathbf{GL}_n]$ and $\mathbb{Z}_p[\boldsymbol{G}_\pi] := \mathbb{Z}_p[\mathbf{GL}_n]/J$ then is a $\mathbb{Z}_p$-form on $\mathbb{Q}_p[\boldsymbol{G}_\pi]$ which yields a $\mathbb{Z}_p$-structure on $\boldsymbol{G}_\pi$, i.e., which yields a $\mathbb{Z}_p$-group scheme $\boldsymbol{G}_\pi/\mathbb{Z}_p$ whose extension to $\mathbb{Q}_p$ is $\boldsymbol{G}_\pi$ (ibid., 3.4, p. 18). Thus, for any $\mathbb{Z}_p$-algebra R contained in $\bar{\mathbb{Q}}_p$ the group of R-points $\boldsymbol{G}_\pi(R)$ is defined and

$$\boldsymbol{G}_\pi(R) = \boldsymbol{G}_\pi(\bar{\mathbb{Q}}_p) \cap \mathbf{GL}_n(R) = \boldsymbol{G}_{\pi,\bar{\mathbb{Q}}_p} \cap \mathrm{Aut}(L_\pi(R)).$$

In particular, since $x_\alpha(t_\alpha) \in \mathrm{Aut}(L_\pi(R))$, $t_\alpha \in R$, we obtain $x_\alpha(t_\alpha) \in \boldsymbol{G}_\pi(R)$ if $t_\alpha \in R$ which yields

$$G_{\pi,R} \subseteq \boldsymbol{G}_\pi(R). \tag{2}$$

For each $\alpha \in \Phi$ there is a unique morphism $\mu_\alpha = \mu_\alpha^\pi : \mathbf{SL}_2(\bar{\mathbb{Q}}_p) \to \boldsymbol{G}_\pi(\bar{\mathbb{Q}}_p)$ such that $\left(\begin{smallmatrix}1 & t\\ & 1\end{smallmatrix}\right) \mapsto x_\alpha^\pi(t)$ and $\left(\begin{smallmatrix}1 & \\ t & 1\end{smallmatrix}\right) \mapsto x_{-\alpha}^\pi(t)$ ($t \in \bar{\mathbb{Q}}_p$) map (see [Borel 1970, 3.2(1), p. 13]). The μ_α is defined over $\mathbb{Z}_p$ (ibid., 3.3(2), p. 15 and 4.3, p. 22). We denote by $h_\alpha(t) = h_\alpha^\pi(t)$ the image of $\left(\begin{smallmatrix}t & \\ & t^{-1}\end{smallmatrix}\right)$ under μ_α (ibid., 3.2(1), p. 13). The algebraic group $\boldsymbol{G}_\pi/\mathbb{Q}_p$ contains a $\mathbb{Q}_p$-split maximal torus $\boldsymbol{T}/\mathbb{Q}_p = \boldsymbol{T}^\pi/\mathbb{Q}_p$ such that the group of $\bar{\mathbb{Q}}_p$-rational points of $\boldsymbol{T}$ is given as

$$\boldsymbol{T}(\bar{\mathbb{Q}}_p) = \langle h_\alpha(t_\alpha),\ \alpha \in \Delta,\ t_\alpha \in \bar{\mathbb{Q}}_p^*\rangle$$

(ibid., 3.2(1), p. 13 and 3.3(3), p. 15). To any $\lambda \in \Gamma_\pi$ we attach a rational character $\lambda^\circ \in X(\boldsymbol{T})$ by setting

$$\lambda^\circ\Big(\prod_{\alpha\in\Delta} h_\alpha(t_\alpha)\Big) = \prod_{\alpha\in\Delta} t_\alpha^{\lambda(h_\alpha)}$$

for all $t_\alpha \in \bar{\mathbb{Q}}_p^*$ (ibid., 3.3, p. 15). This defines an isomorphism $\Gamma_\pi \to X(\boldsymbol{T})$ (ibid., 3.3(3), p. 15). We note that $d\lambda^\circ = \lambda$ (ibid., 3.3 equation (2), p. 16). The characters α°, $\alpha \in \Phi$, are the roots of $\boldsymbol{G}_\pi/\mathbb{Q}_p$ with respect to $\boldsymbol{T}$ (i.e., the weights of the adjoint action of $\boldsymbol{T}$ on $\mathrm{Lie}(\boldsymbol{G}_\pi/\mathbb{Q}_p)$). To simplify notation we denote the exponential $\alpha^\circ : \boldsymbol{T}(\bar{\mathbb{Q}}_p) \to \bar{\mathbb{Q}}_p^*$ of a root α also by α.

There are closed subgroups $\boldsymbol{N} = \boldsymbol{N}^\pi$ and $\boldsymbol{N}^- = \boldsymbol{N}^{-,\pi}$ of $\boldsymbol{G}_\pi/\mathbb{Q}_p$ such that

$$\begin{aligned} \boldsymbol{N}(\bar{\mathbb{Q}}_p) &= \langle x_\alpha(t),\ t_\alpha \in \bar{\mathbb{Q}}_p, \alpha \in \Phi^+\rangle, \\ \boldsymbol{N}^-(\bar{\mathbb{Q}}_p) &= \langle x_\alpha(t),\ t_\alpha \in \bar{\mathbb{Q}}_p, \alpha \in \Phi^-\rangle. \end{aligned} \tag{3}$$

The subgroups $\boldsymbol{N}$ and $\boldsymbol{N}^-$ are defined over $\mathbb{Q}_p$ and they are maximal unipotent (see [Borel 1970, 3.3(3), p. 15]). In the same way as above the $\mathbb{Z}_p$-structure on $\boldsymbol{G}_\pi$ induces $\mathbb{Z}_p$-structures on closed subgroups H of $\boldsymbol{G}_\pi/\mathbb{Q}_p$ such that

$$H(\mathbb{Z}_p) = H(\bar{\mathbb{Q}}_p) \cap \boldsymbol{G}_\pi(\mathbb{Z}_p) = H(\bar{\mathbb{Q}}_p) \cap \mathrm{Aut}(L_\pi(\mathbb{Z}_p)).$$

For example, this applies to the groups $\boldsymbol{N}$, $\boldsymbol{N}^-$, $\boldsymbol{T}$, which thus have $\mathbb{Z}_p$-structures.

We set $\boldsymbol{B} = \boldsymbol{T}\boldsymbol{N}$, which is a subgroup of $\boldsymbol{G}_\pi$ defined over $\mathbb{Q}_p$. Thus, $\boldsymbol{B}(\bar{\mathbb{Q}}_p)$ is the subgroup of $\boldsymbol{G}_\pi(\bar{\mathbb{Q}}_p)$ which is generated by the root subgroups $\boldsymbol{G}_{\pi,\alpha}(\bar{\mathbb{Q}}_p)$ of $\boldsymbol{G}_\pi(\bar{\mathbb{Q}}_p)$ with $\alpha \in \Phi^+$ together with $\boldsymbol{T}(\bar{\mathbb{Q}}_p)$ and its existence as a closed subgroup defined over $\mathbb{Q}_p$ also follows from [Popov and Vinberg 1994, 5.3.4 Proposition, p. 70]. In particular, $\boldsymbol{B}$ is a minimal parabolic subgroup. We also define the subgroup $\boldsymbol{B}^- = \boldsymbol{T}\boldsymbol{N}^-$ of $\boldsymbol{G}_\pi/\mathbb{Q}_p$. Since $h_\alpha(t) \in \boldsymbol{G}_\pi(\mathbb{Z}_p)$, $t \in \mathbb{Z}_p^*$, because μ_α is defined over $\mathbb{Z}_p$ we obtain

$$h_\alpha(t) \in \boldsymbol{T}(\bar{\mathbb{Q}}_p) \cap \boldsymbol{G}_\pi(\mathbb{Z}_p) = \boldsymbol{T}(\mathbb{Z}_p) \quad (t \in \mathbb{Z}_p^*).$$

Analogously, we obtain for any $t \in \mathbb{Z}_p$ and any positive root α

$$x_\alpha(t) \in \boldsymbol{N}(\bar{\mathbb{Q}}_p) \cap \boldsymbol{G}_\pi(\mathbb{Z}_p) = \boldsymbol{N}(\mathbb{Z}_p)$$

while for a negative root α we get $x_\alpha(t) \in \boldsymbol{N}^-(\mathbb{Z}_p)$.

Notation. If $\pi = \rho_\lambda$ is an irreducible representation then we simplify notation and set $x_\alpha^\lambda(t) = x_\alpha^{\rho_\lambda}(t)$, $\boldsymbol{G}_\lambda = \boldsymbol{G}_{\rho_\lambda}$, $\mu_\alpha^\lambda = \mu_\alpha^{\rho_\lambda}$, $h_\alpha^\lambda(t) = h_\alpha^{\rho_\lambda}(t)$ and $\boldsymbol{T}^\lambda = \boldsymbol{T}^{\rho_\lambda}$; we note that in Section 1.1 we already used the notation L_λ for L_{ρ_λ} and Γ_λ for Γ_{ρ_λ}.

1.3. ***Mod p reduction.*** Let $p \in \mathbb{N}$ be a prime element. We denote by $\boldsymbol{G}_{\pi,(p)}$ or by $\boldsymbol{G}_\pi/\mathbb{F}_p$ the mod p reduction of $\boldsymbol{G}_\pi/\mathbb{Z}_p$; i.e., $\boldsymbol{G}_{\pi,(p)}(\bar{\mathbb{F}}_p)$ is the set of zeros of $\mathbb{F}_p \otimes J \le \mathbb{F}_p \otimes \mathbb{Z}_p[\mathbf{GL}_n]$ in $\mathbf{GL}_n(\bar{\mathbb{F}}_p)$. Hence, $\boldsymbol{G}_{\pi,(p)}$ is an affine variety defined over $\mathbb{F}_p$ (a closed subgroup of $\mathbf{GL}_n/\mathbb{F}_p$). Moreover, since $\boldsymbol{G}_\pi/\mathbb{Z}_p$ has good reduction (see [Borel 1970, 3.4, p. 18]) we know that $\mathbb{F}_p \otimes J$ equals the ideal consisting of all $f \in \mathbb{F}_p \otimes \mathbb{Z}_p[\mathbf{GL}_n]$ which vanish on $\boldsymbol{G}_{\pi,(p)}(\bar{\mathbb{F}}_p)$, hence, $\mathbb{F}_p[\boldsymbol{G}_{\pi,(p)}] = \mathbb{F}_p \otimes \mathbb{Z}_p[\boldsymbol{G}_\pi]$. The mod p reduction $\boldsymbol{G}_{\pi,(p)}$ is a semisimple group defined over $\mathbb{F}_p$ (ibid., 4.3, p. 21/22). Analogously, the mod p reductions $\boldsymbol{N}_{(p)}$, $\boldsymbol{B}_{(p)}$, $\boldsymbol{N}_{(p)}^-$, $\boldsymbol{B}_{(p)}^-$, ... are defined.

We denote by

$$\wp : \mathrm{Aut}(L_\pi(\mathbb{Z}_p)) \overset{\mathcal{B}}{=} \mathbf{GL}_n(\mathbb{Z}_p) \to \mathbf{GL}_n(\mathbb{F}_p) \overset{\bar{\mathcal{B}}}{=} \mathrm{Aut}(L_\pi(\mathbb{F}_p))$$

the mod p reduction map which sends (x_{ij}) to $(x_{ij} \pmod p)$ ($\bar{\mathcal{B}}$ is the basis of $L_\pi(\mathbb{F}_p) = \mathbb{F}_p \otimes L_\pi(\mathbb{Z}_p)$ induced by the basis $\mathcal{B}$ of $L_\pi(\mathbb{Z}_p)$). If $(x_{ij}) \in \boldsymbol{G}_\pi(\mathbb{Z}_p) \leq \mathbf{GL}_n(\mathbb{Z}_p)$ then $\wp(x)$ obviously is contained in $\boldsymbol{G}_\pi(\mathbb{F}_p)$, hence, $\wp$ induces a map $\boldsymbol{G}_\pi(\mathbb{Z}_p) \to \boldsymbol{G}_\pi(\mathbb{F}_p)$.

The Iwahori subgroup $\mathcal{I}$ of $\boldsymbol{G}_\pi(\mathbb{Z}_p)$ then is defined as the set of all $k \in \boldsymbol{G}_\pi(\mathbb{Z}_p)$ such that $\wp(k) \in \boldsymbol{B}^-(\mathbb{F}_p)$, i.e., $\mathcal{I} = \wp^{-1}(\boldsymbol{B}^-(\mathbb{F}_p))$.

1.4. ***Irreducible representations of $\boldsymbol{G_\pi}/\mathbb{Z}_p$.*** The group $\boldsymbol{G}_\pi$ is a semisimple connected $\mathbb{Q}_p$-split group with $\mathbb{Q}_p$-split maximal torus $\boldsymbol{T} = \boldsymbol{T}^\pi$. Let $\lambda^\circ \in X(\boldsymbol{T})$ be a dominant weight. We set $\lambda = \mathrm{d}\,\lambda^\circ \in \mathfrak{h}^*$. As in Section 1.2 the choice of a $\mathbb{Z}_p$-basis $\mathcal{B}$ of the $\mathcal{U}_{\mathbb{Z}_p}$-invariant lattice $L_\lambda(\mathbb{Z}_p)$ $(= L_{\rho_\lambda}(\mathbb{Z}_p))$ yields an identification $\mathrm{Aut}(L_\lambda(\bar{\mathbb{Q}}_p)) \overset{\mathcal{B}}{=} \mathbf{GL}_m(\bar{\mathbb{Q}}_p)$ (where $m = \dim(L_\lambda)$).

If $\Gamma_\pi \supseteq \Gamma_\lambda$ we define a representation of algebraic groups

$$\rho_{\lambda^\circ} : \boldsymbol{G}_\pi(\bar{\mathbb{Q}}_p) \to \mathrm{Aut}(L_\lambda(\bar{\mathbb{Q}}_p)) = \mathbf{GL}_m(\bar{\mathbb{Q}}_p)$$

by mapping $x^\pi_\alpha(t)$ to $x^\lambda_\alpha(t)$, $\alpha \in \Phi$, $t \in \bar{\mathbb{Q}}_p$ (see [Borel 1970, 3.2(4), p. 14 and 3.3(2), p. 15], where ρ_{λ° is denoted by $\lambda_{\rho_\lambda,\pi}$). We note that ρ_{λ° has image $\boldsymbol{G}_\lambda(\bar{\mathbb{Q}}_p)$ (which equals $\boldsymbol{G}_{\rho_\lambda}(\bar{\mathbb{Q}}_p)$).

Lemma. *The map ρ_{λ° induces a map of tori*

$$\rho_{\lambda^\circ} : \boldsymbol{T}^\pi(\bar{\mathbb{Q}}_p) \to \boldsymbol{T}^\lambda(\bar{\mathbb{Q}}_p)$$

which maps $h^\pi_\alpha(t) \mapsto h^\lambda_\alpha(t)$ for all $\alpha \in \Delta$ and $t \in \bar{\mathbb{Q}}^*_p$.

Proof. We first claim that for each $\alpha \in \Phi$ the diagram

$$\begin{tikzcd} \boldsymbol{G}_\pi(\bar{\mathbb{Q}}_p) \arrow[rr, "\rho_{\lambda^\circ}"] & & \boldsymbol{G}_\lambda(\bar{\mathbb{Q}}_p) \\ & \mathbf{SL}_2(\bar{\mathbb{Q}}_p) \arrow[ul, "\mu^\pi_\alpha"] \arrow[ur, "\mu^\lambda_\alpha"'] & \end{tikzcd}$$

commutes. It is sufficient to show commutativity of the diagram for all $\left(\begin{smallmatrix} 1 & t \\ & 1 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 1 & \\ t & 1 \end{smallmatrix}\right)$ with $t \in \bar{\mathbb{Q}}_p$ because these elements generate $\mathbf{SL}_2(\bar{\mathbb{Q}}_p)$. But

$$\rho_{\lambda^\circ}\big(\mu^\pi_\alpha\big(\left(\begin{smallmatrix} 1 & t \\ & 1 \end{smallmatrix}\right)\big)\big) = \rho_{\lambda^\circ}(x^\pi_\alpha(t)) = x^\lambda_\alpha(t) = \mu^\lambda_\alpha\big(\left(\begin{smallmatrix} 1 & t \\ & 1 \end{smallmatrix}\right)\big),$$

and analogously for the lower unipotent matrices. Hence, the diagram commutes and we obtain

$$\rho_{\lambda^\circ}(h^\pi_\alpha(t)) = \rho_{\lambda^\circ}\big(\mu^\pi_\alpha\big(\left(\begin{smallmatrix} t & \\ & t^{-1} \end{smallmatrix}\right)\big)\big) = \mu^\lambda_\alpha\big(\left(\begin{smallmatrix} t & \\ & t^{-1} \end{smallmatrix}\right)\big) = h^\lambda_\alpha(t).$$

Since $\boldsymbol{T}^\pi(\bar{\mathbb{Q}}_p)$ is generated by the $h^\pi_\alpha(t)$, where $\alpha \in \Delta$ and $t \in \bar{\mathbb{Q}}^*_p$, this implies that $\rho_{\lambda^\circ}(\boldsymbol{T}^\pi(\bar{\mathbb{Q}}_p)) \subseteq \boldsymbol{T}^\lambda(\bar{\mathbb{Q}}_p)$ and the lemma is proven. $\square$

The lemma implies that for all $t \in \boldsymbol{T}^\pi(\bar{\mathbb{Q}}_p)$ and any weight vector $v_\mu \in L_\lambda(\bar{\mathbb{Q}}_p, \mu)$, $\mu \in P_\lambda$, (i.e., v_μ has weight μ w.r.t. $\mathfrak{h}$) we have

$$\rho_{\lambda^\circ}(t)(v_\mu) = \mu^\circ(t)v_\mu. \tag{4}$$

In fact since $\boldsymbol{T}^\pi(\bar{\mathbb{Q}}_p)$ is generated by the $h_\alpha^\pi(s)$ we may assume that $t = h_\alpha^\pi(s)$ for some $\alpha \in \Delta$ and $s \in \bar{\mathbb{Q}}_p^*$. Using the lemma and the equation in [Borel 1970, 3.2(1), p. 13], we obtain

$$\rho_{\lambda^\circ}(t)(v_\mu) = \rho_{\lambda^\circ}(h_\alpha^\pi(s))(v_\mu) = h_\alpha^\lambda(s)v_\mu = s^{\mu(h_\alpha)}v_\mu = \mu^\circ(t)v_\mu.$$

The following result seems to be well known. Since we could not find a direct reference we add a proof.

Proposition. 1. *The $\boldsymbol{G}_\pi(\bar{\mathbb{Q}}_p)$-module $L_\lambda(\bar{\mathbb{Q}}_p)$ contains a vector v_{λ° which is invariant under $\boldsymbol{N}^\pi(\bar{\mathbb{Q}}_p)$ and satisfies $tv_{\lambda^\circ} = \lambda^\circ(t)v_{\lambda^\circ}$ for all $t \in \boldsymbol{T}^\pi(\bar{\mathbb{Q}}_p)$.*

2. *The representation ρ_{λ° is the irreducible representation of $\boldsymbol{G}_\pi(\bar{\mathbb{Q}}_p)$ of highest weight λ°.*

Proof. 1. We choose for v_{λ° the highest weight vector $v_\lambda \in L_\lambda(\mathbb{Z})$, which we selected in Section 1.1. Since $x_\alpha v_{\lambda^\circ}$ $(= \rho_\lambda(x_\alpha)v_{\lambda^\circ})$ vanishes for all $\alpha \in \Phi^+$ we obtain

$$\rho_{\lambda^\circ}(x_\alpha^\pi(t))v_{\lambda^\circ} = x_\alpha^\lambda(t)v_{\lambda^\circ} = v_{\lambda^\circ} + t\rho_\lambda(x_\alpha)(v_{\lambda^\circ}) + \cdots = v_{\lambda^\circ}.$$

Since $\boldsymbol{N}^\pi(\bar{\mathbb{Q}}_p)$ is generated by the $x_\alpha^\pi(t)$ with $\alpha \in \Phi^+$, $t \in \bar{\mathbb{Q}}_p$, this yields the first claim about v_{λ°. The second claim is immediate by equation (4) since $v_{\lambda^\circ} = v_\lambda$ has $\mathfrak{h}$-weight λ.

2. For the moment we denote by $(\sigma_{\mu^\circ}, \Sigma_{\mu^\circ})$ the irreducible representation of $\boldsymbol{G}_\pi(\bar{\mathbb{Q}}_p)$ of highest weight $\mu^\circ \in X(\boldsymbol{T}^\pi)$. The derived representation of σ_{μ° is $(\rho_\mu, L_\mu(\bar{\mathbb{Q}}_p))$, where $\mu = d\,\mu^\circ$; hence, $\dim \Sigma_{\mu^\circ} = \dim L_\mu(\bar{\mathbb{Q}}_p)$. Since $\boldsymbol{G}_\pi$ is semisimple and since we are in characteristic 0 any representation of $\boldsymbol{G}_\pi(\bar{\mathbb{Q}}_p)$ is semisimple, hence, we can write

$$\rho_{\lambda^\circ} = \bigoplus_{i=1}^{r} \Sigma_{\mu_i^\circ}.$$

Any representation $\Sigma_{\mu_i^\circ}$ contains a unique (up to scalars) nontrivial vector $v_{\mu_i^\circ}$ invariant under $\boldsymbol{N}^\pi(\bar{\mathbb{Q}}_p)$. This vector $v_{\mu_i^\circ}$ then satisfies $tv_{\mu_i^\circ} = \mu_i^\circ(t)v_{\mu_i^\circ}$, $t \in \boldsymbol{T}^\pi(\bar{\mathbb{Q}}_p)$ (i.e., $\bar{\mathbb{Q}}_p v_{\mu_i^\circ}$ is the unique line which is stable under $\boldsymbol{B}(\bar{\mathbb{Q}}_p)$). The vector v_{λ° decomposes as

$$v_{\lambda^\circ} = \sum_{i=1}^{r} v_i,$$

where $v_i \in \Sigma_{\mu_i^\circ}$ and at least one vector v_j does not vanish. Since v_{λ° is invariant under $\boldsymbol{N}(\bar{\mathbb{Q}}_p)$ by part 1, we obtain $\sum_i nv_i = \sum_i v_i$ for any $n \in \boldsymbol{N}(\bar{\mathbb{Q}}_p)$, hence, $nv_i = v_i$

for all i and all $n \in \boldsymbol{N}(\bar{\mathbb{Q}}_p)$. Thus, $v_i = c_i v_{\mu_i^\circ}$ for some $c_i \in \bar{\mathbb{Q}}_p$ by the uniqueness of $v_{\mu_i^\circ}$; in particular, v_i has weight μ_i° w.r.t. $\boldsymbol{T}^\pi(\bar{\mathbb{Q}}_p)$. On the other hand, since $tv_{\lambda^\circ} = \lambda^\circ(t)v_{\lambda^\circ}$ by part 1, we obtain $\sum_i tv_i = \sum_i \lambda^\circ(t)v_i$; hence, $tv_i = \lambda^\circ(t)v_i$ for all i and $t \in \boldsymbol{T}^\pi(\bar{\mathbb{Q}}_p)$. Since $v_j \neq 0$ we deduce that $\mu_j^\circ = \lambda^\circ$. The representation ρ_{λ° therefore decomposes as a direct sum $\rho_{\lambda^\circ} = \Sigma_{\lambda^\circ} \oplus C$. Since ρ_{λ° is a representation on the space $L_\lambda(\bar{\mathbb{Q}}_p)$ we know that $\dim \rho_{\lambda^\circ} = \dim L_\lambda(\bar{\mathbb{Q}}_p) = \dim \Sigma_{\lambda^\circ}$. This implies that $C = 0$, hence, $\rho_{\lambda^\circ} = \Sigma_{\lambda^\circ}$ is the irreducible representation of $\boldsymbol{G}_\pi(\bar{\mathbb{Q}}_p)$ of highest weight λ°. Thus, the proof is complete. □

From [Borel 1970, 3.5, p. 19], we know that the morphism ρ_{λ° is defined over $\mathbb{Z}_p$, i.e., it is associated to a morphism of $\mathbb{Z}_p$-group schemes $\rho_{\lambda^\circ} : \boldsymbol{G}_\pi/\mathbb{Z}_p \to \boldsymbol{G}_\lambda/\mathbb{Z}_p$. Since $\boldsymbol{G}_\lambda$ is a closed subscheme of $\mathbf{Aut}(L_\lambda) = \mathbf{GL}_m/\mathbb{Z}_p$ we obtain that the representation ρ_{λ° is defined over $\mathbb{Z}_p$, i.e.,

$$\rho_{\lambda^\circ} : \boldsymbol{G}_\pi/\mathbb{Z}_p \to \mathbf{Aut}(L_\lambda) = \mathbf{GL}_m/\mathbb{Z}_p$$

In particular, $L_\lambda(R)$ is a $\boldsymbol{G}_\pi(R)$-module for all $\mathbb{Z}_p$-algebras R. Using equation (4) we deduce that $\boldsymbol{T}^\pi(\mathbb{Z}_p)$ leaves $L_\lambda(\mathbb{Z}_p, \mu) = L_\lambda(\mathbb{Z}_p) \cap L_\lambda(\bar{\mathbb{Q}}_p, \mu)$ invariant and acts via the character μ°.

1.5. ***The level subgroup $K_*(p, \sigma)$.*** From now on we fix a prime element $p \in \mathbb{N}$. For any $\sigma \in \mathbb{N}$ we define the level subgroup

$$K_*(\sigma) = K_*(p, \sigma) = K_*^\pi(p, \sigma) \leq \boldsymbol{G}_\pi(\mathbb{Z}_p)$$

as the subgroup generated by the following elements: all $x_\alpha(t_\alpha)$ with $\alpha \in \Phi^-$ and $t_\alpha \in \mathbb{Z}_p$, all $x_\alpha(t_\alpha)$ with $\alpha \in \Phi^+$ and $t_\alpha \in p^{\lceil \frac{1}{\sigma} \operatorname{ht}(\alpha) \rceil}\mathbb{Z}_p$ and all $h_\alpha^\pi(t_\alpha)$ with $\alpha \in \Delta$ and $t_\alpha \in \mathbb{Z}_p^*$. We note that the equations at the end of Section 1.2 imply that $K_*(\sigma) \leq \boldsymbol{G}_\pi(\mathbb{Z}_p)$ and even that $K_*(\sigma) \leq \mathcal{I}$, because $\wp(x_\alpha(t))$ is the identity in $\operatorname{Aut}(L_\pi(\mathbb{F}_p))$ if $t \in p\mathbb{Z}_p$. If $\sigma \geq \max_{\alpha \in \Phi^+} \operatorname{ht}(\alpha)$ we conclude that $K_*(\sigma)$ equals

$$\bigl\langle x_\alpha(t_\alpha),\ t_\alpha \in p\mathbb{Z}_p \text{ if } \alpha > 0 \text{ and } t_\alpha \in \mathbb{Z}_p \text{ if } \alpha < 0,\ h_\alpha(t_\alpha), \alpha \in \Delta, t_\alpha \in \mathbb{Z}_p^* \bigr\rangle = \mathcal{I};$$

the latter equality follows from [Iwahori and Matsumoto 1965, p. 259] (the Iwahori subgroup is denoted by B there).

We define the subgroups $N_{\mathbb{Z}_p} = N_{\mathbb{Z}_p}^\pi := \langle x_\alpha(t_\alpha),\ \alpha > 0,\ t_\alpha \in \mathbb{Z}_p \rangle \subseteq \boldsymbol{N}(\mathbb{Z}_p)$, $N_{p\mathbb{Z}_p} = N_{p\mathbb{Z}_p}^\pi := \langle x_\alpha(t_\alpha),\ \alpha > 0,\ t_\alpha \in p\mathbb{Z}_p \rangle$, $N_{\mathbb{Z}_p}^- = N_{\mathbb{Z}_p}^{-,\pi} := \langle x_\alpha(t_\alpha),\ \alpha < 0,\ t_\alpha \in \mathbb{Z}_p \rangle$ and $T_{\mathbb{Z}_p} = T_{\mathbb{Z}_p}^\pi := \langle h_\alpha^\pi(t_\alpha),\ \alpha \in \Delta,\ t_\alpha \in \mathbb{Z}_p^* \rangle$ of $\boldsymbol{G}_\pi(\mathbb{Z}_p)$. The Iwahori subgroup then satisfies the decomposition

$$\mathcal{I} = N_{p\mathbb{Z}_p} T_{\mathbb{Z}_p} N_{\mathbb{Z}_p}^-$$

(see [Iwahori and Matsumoto 1965, Theorem 2.5, p. 263]; note that $T_{\mathbb{Z}_p} = \mathfrak{h}_{\mathfrak{O}}$). We note the following consequences.

1. Assume $h \in \boldsymbol{T}(\mathbb{Q}_p)^{++}$; i.e., $v_p(\alpha(h)) > 0$ for all simple roots $\alpha \in \Delta$. Then, for all $e, f \in \mathbb{N}_0$ we have

$$\mathcal{I}h^e\mathcal{I}h^f\mathcal{I} = \mathcal{I}h^{e+f}\mathcal{I}. \tag{5}$$

2. Assume $t, t' \in \boldsymbol{T}(\mathbb{Q}_p)^{++}$. Then

$$\mathcal{I}t\mathcal{I} = \mathcal{I}t'\mathcal{I} \iff T_{\mathbb{Z}_p}tT_{\mathbb{Z}_p} = T_{\mathbb{Z}_p}t'T_{\mathbb{Z}_p}. \tag{6}$$

Proof of equation (6). The leftward implication is trivial. To prove the reverse implication we note that $t \in \mathcal{I}t'\mathcal{I}$ implies that there are $k^+, m^+ \in N_{p\mathbb{Z}_p}$, $k^\circ, m^\circ \in T_{\mathbb{Z}_p}$ and $k^-, m^- \in N^-_{\mathbb{Z}_p}$ such that $tk^+k^\circ k^- = m^+m^\circ m^- t'$. Since $\mathrm{Ad}(t)(x_\alpha(t_\alpha)) = x_\alpha(\alpha(t)t_\alpha)$ the element t normalizes $N_{p\mathbb{Z}_p}$ and $(t')^{-1}$ normalizes $N^-_{\mathbb{Z}_p}$, hence, we obtain $\tilde{k}^+tk^\circ k^- = m^+m^\circ t'\tilde{m}^-$ with $\tilde{k}^+ \in N_{p\mathbb{Z}_p}$, $\tilde{m}^- \in N^-_{\mathbb{Z}_p}$. Equivalently,

$$(m^+)^{-1}\tilde{k}^+tk^\circ = m^\circ t'\tilde{m}^-(k^-)^{-1}. \tag{7}$$

The left-hand side is contained in $\boldsymbol{B}(\mathbb{Q}_p)$ and the right-hand side is contained in $\boldsymbol{B}^-(\mathbb{Q}_p)$, whose intersection is $\boldsymbol{T}(\mathbb{Q}_p)$. Hence, $(m^+)^{-1}\tilde{k}^+ \in \boldsymbol{T}(\mathbb{Q}_p) \cap N_{p\mathbb{Z}_p} = \{1\}$ (note that $N_{p\mathbb{Z}_p} \subseteq N_{\mathbb{Z}_p} \subseteq \boldsymbol{N}(\mathbb{Z}_p)$) and, similarly, $\tilde{m}^-(k^-)^{-1} \in \boldsymbol{T}(\mathbb{Q}_p) \cap N^-_{\mathbb{Z}_p} = \{1\}$. Equation (7) thus implies that $tk^\circ = m^\circ t'$, which proves the claim. □

2. Hecke algebra and cohomology

2.1. ***Reductive algebraic groups.*** From now on, $\tilde{\boldsymbol{G}}$ denotes a connected reductive algebraic group defined over $\mathbb{Q}$. Since $\tilde{\boldsymbol{G}}$ is defined over $\mathbb{Q}$ it contains a maximal torus which is defined over $\mathbb{Q}$ and we assume that $\tilde{\boldsymbol{G}}$ contains a maximal torus $\tilde{\boldsymbol{T}}$ which is defined over $\mathbb{Q}$ and split over $\mathbb{Q}_p$ (hence, $\tilde{\boldsymbol{G}}$ is $\mathbb{Q}_p$-split). This assumption is in particular satisfied if $\tilde{\boldsymbol{G}}$ is $\mathbb{Q}$-split. We denote by $\boldsymbol{G} = \tilde{\boldsymbol{G}}^{\mathrm{der}}$ the derived group and by $\tilde{\boldsymbol{Z}}$ the center of $\tilde{\boldsymbol{G}}$; hence, $\tilde{\boldsymbol{G}} = (\boldsymbol{G} \times \tilde{\boldsymbol{Z}})/\boldsymbol{Z}$ as algebraic groups over $\mathbb{Q}$, where $\boldsymbol{Z}$ is the center of $\boldsymbol{G}$ (embedded via $z \mapsto (z, z^{-1})$). We denote by $\mathrm{Lie}(\boldsymbol{G})$ the Lie algebra of $\boldsymbol{G}$. We use the notations introduced in Section 1.1 for the complex Lie algebra $\mathfrak{g} = \mathrm{Lie}(\boldsymbol{G}) \otimes_{\mathbb{Q}} \mathbb{C}$; e.g., $\mathfrak{h}$ is a Cartan subalgebra in $\mathfrak{g}$, $\Phi = \Phi(\mathfrak{g}, \mathfrak{h})$ the set of roots and Δ a choice of a basis of Φ. Since $\boldsymbol{G}$ is a $\mathbb{Q}_p$-split, semisimple algebraic group, there is a finite dimensional complex representation π of $\mathfrak{g}$ such that

$$\boldsymbol{G}/\mathbb{Q}_p \cong \boldsymbol{G}_\pi/\mathbb{Q}_p$$

as $\mathbb{Q}_p$-groups, where $\boldsymbol{G}_\pi/\mathbb{Z}_p$ is the Chevalley group attached to π; see Section 1.2. In the following we may assume that $\boldsymbol{G}/\mathbb{Q}_p = \boldsymbol{G}_\pi/\mathbb{Q}_p$.

The $\mathbb{Q}_p$-structure on $\tilde{\boldsymbol{G}}$. We denote by $\boldsymbol{T}$ the $\mathbb{Q}_p$-split maximal torus and by $\boldsymbol{N}, \boldsymbol{N}^-$ the maximal unipotent subgroups in $\boldsymbol{G}/\mathbb{Q}_p = \boldsymbol{G}_\pi/\mathbb{Q}_p$ defined in Section 1.2. The subgroups $\boldsymbol{N}$, $\boldsymbol{N}^-$ remain maximal unipotent in $\tilde{\boldsymbol{G}}$ and $\boldsymbol{T}$ is a $\mathbb{Q}_p$-defined and $\mathbb{Q}_p$-split torus in $\tilde{\boldsymbol{G}}$. We denote by $\tilde{\boldsymbol{B}}/\mathbb{Q}_p$ the Borel subgroup in $\tilde{\boldsymbol{G}}$ containing $\tilde{\boldsymbol{T}}$ which corresponds to Δ. The torus $\tilde{\boldsymbol{T}}$ decomposes $\tilde{\boldsymbol{T}} = (\boldsymbol{T}' \times \tilde{\boldsymbol{Z}})/\boldsymbol{Z}$ over $\mathbb{Q}_p$, where $\boldsymbol{T}'/\mathbb{Q}_p$ is a $\mathbb{Q}_p$-split maximal torus in $\boldsymbol{G}/\mathbb{Q}_p$. Since any two $\mathbb{Q}_p$-split maximal tori in $\boldsymbol{G}/\mathbb{Q}_p$ are conjugate by an element $x \in \boldsymbol{G}(\mathbb{Q}_p)$ (see [Springer 1981, 15.2.6 Theorem, p. 256]) we may assume after composing the isomorphism $\boldsymbol{G}/\mathbb{Q}_p \cong \boldsymbol{G}_\pi/\mathbb{Q}_p$ with conjugation by x that $\boldsymbol{T} = \boldsymbol{T}'$, hence, $\tilde{\boldsymbol{T}} = (\boldsymbol{T} \times \tilde{\boldsymbol{Z}})/\boldsymbol{Z}$ as algebraic groups over $\mathbb{Q}_p$. We denote by $X(\tilde{\boldsymbol{T}})$ (resp. $X_*(\tilde{\boldsymbol{T}})$) the (additively written) group of $\mathbb{Q}_p$-characters (resp. $\mathbb{Q}_p$-cocharacters) of $\tilde{\boldsymbol{T}}/\mathbb{Q}_p$ and by $\langle\cdot,\cdot\rangle : X(\tilde{\boldsymbol{T}}) \times X_*(\tilde{\boldsymbol{T}}) \to \mathbb{Z}$ the canonical pairing defined by $\chi \circ \eta(x) = x^{\langle\chi,\eta\rangle}$ for all $x \in \mathbb{G}_m(\bar{\mathbb{Q}}_p)$. We recall that by α we denote a root in $\Phi \subseteq \mathfrak{h}^*$ and also its exponential in $X(\boldsymbol{T})$ (i.e., we write α for α°; see Section 1.2). Any root $\alpha \in X(\boldsymbol{T})$ vanishes on the center $\boldsymbol{Z}$ of $\boldsymbol{G}$, hence, it extends to a character on $\tilde{\boldsymbol{T}} = (\boldsymbol{T} \times \tilde{\boldsymbol{Z}})/\boldsymbol{Z}$ by setting it equal to 1 on $\tilde{\boldsymbol{Z}}$; we denote this extension of the root again by α; hence, $\alpha(t) = \alpha(t^\circ)(=\alpha^\circ(t^\circ))$ if $t = t^\circ z \in \boldsymbol{T}(\bar{\mathbb{Q}}_p)\tilde{\boldsymbol{Z}}(\bar{\mathbb{Q}}_p)$. We denote by $\alpha^\vee \in X_*(\boldsymbol{T}) \subseteq X_*(\tilde{\boldsymbol{T}})$ the coroot corresponding to α; explicitly, $\alpha^\vee(t) = h_\alpha(t)$, $t \in \mathbb{G}_m(\bar{\mathbb{Q}}_p)$. Any character $\tilde{\lambda} \in X(\tilde{\boldsymbol{T}})$ is of the form $\tilde{\lambda} = \lambda^\circ \otimes \kappa$, where $\kappa = \tilde{\lambda}|_{\tilde{\boldsymbol{Z}}} \in X(\tilde{\boldsymbol{Z}})$ and $\lambda^\circ = \tilde{\lambda}|_{\boldsymbol{T}} \in X(\boldsymbol{T})$ satisfy $\lambda^\circ|_{\boldsymbol{Z}} = \kappa|_{\boldsymbol{Z}}$. We note that λ° corresponds to a weight $\lambda \in \Gamma_\pi$, i.e., $\lambda = \mathrm{d}\lambda^\circ$; see Section 1.2. We call $\tilde{\lambda} \in X(\tilde{\boldsymbol{T}})$ dominant if

$$\langle\tilde{\lambda}, \alpha^\vee\rangle = \langle\lambda^\circ, \alpha^\vee\rangle = \lambda(h_\alpha) \geq 0$$

for all $\alpha \in \Delta$. We denote by $\tilde{\boldsymbol{T}}(\mathbb{Q}_p)^+$ (resp. $\tilde{\boldsymbol{T}}(\mathbb{Q}_p)^{++}$, $\tilde{\boldsymbol{T}}(\mathbb{Q}_p)^{--}$) the set of all elements $t \in \tilde{\boldsymbol{T}}(\mathbb{Q}_p)$ such that $v_p(\alpha(t)) \geq 0$ (resp. $v_p(\alpha(t)) > 0$, $v_p(\alpha(t)) < 0$) for all $\alpha \in \Delta$ and by $X(\tilde{\boldsymbol{T}})^{\mathrm{dom}}$ the set of dominant characters.

The $\mathbb{Z}_p$-structure on $\tilde{\boldsymbol{G}}$. We recall that the derived group $\boldsymbol{G}/\mathbb{Q}_p = \boldsymbol{G}_\pi/\mathbb{Q}_p$ has a $\mathbb{Z}_p$-structure; see Section 1.2. In Section 1.5 we defined the level subgroup $K_*(\sigma) = K_*^\pi(p,\sigma)$ which is a subgroup of $\boldsymbol{G}(\mathbb{Z}_p)$. We define a $\mathbb{Z}_p$-structure on $\tilde{\boldsymbol{Z}}$ by selecting as a $\mathbb{Z}_p$-form of $\mathbb{Q}_p[\tilde{\boldsymbol{Z}}]$ the algebra $\mathbb{Z}_p[\mathrm{res}_{\tilde{\boldsymbol{T}}/\tilde{\boldsymbol{Z}}}\ X(\tilde{\boldsymbol{T}})] = \mathbb{Z}_p[X(\tilde{\boldsymbol{T}})/X'(\tilde{\boldsymbol{T}})]$, where $X'(\tilde{\boldsymbol{T}}) = X(\tilde{\boldsymbol{T}}) \cap \sum_{\alpha\in\Phi} \mathbb{Q}\alpha$. It follows that for any $\tilde{\lambda} \in X(\tilde{\boldsymbol{T}})$

$$\tilde{\lambda}|_{\tilde{\boldsymbol{Z}}} : \tilde{\boldsymbol{Z}} \to \mathbb{G}_m \tag{8}$$

is defined over $\mathbb{Z}_p$. The $\mathbb{Z}_p$-structures on $\boldsymbol{G}$ and $\tilde{\boldsymbol{Z}}$ yield a $\mathbb{Z}_p$-structure on $\tilde{\boldsymbol{G}}$.

2.2. From now on, we fix a prime $p \in \mathbb{N}$ and we define the subgroup

$$\tilde{\mathcal{I}} := \langle\mathcal{I}, \tilde{\boldsymbol{Z}}(\mathbb{Z}_p)\rangle \leq \tilde{\boldsymbol{G}}(\mathbb{Z}_p).$$

We select an arithmetic subgroup $\Gamma \leq \tilde{\boldsymbol{G}}(\mathbb{Q})$ satisfying $\Gamma \leq \tilde{\mathcal{I}}$.

2.3. *The Hecke algebra.* Also, from now on, we let h be an element in $\tilde{\boldsymbol{T}}(\mathbb{Q})^{++}$; i.e., $h \in \tilde{\boldsymbol{T}}(\mathbb{Q})$ and $v_p(\alpha(h)) > 0$ for all $\alpha \in \Delta$. We denote by $\mathcal{K} = \mathcal{K}_h = \langle h, \tilde{\mathcal{I}}\rangle_{\mathrm{semigrp}} \leq$

$\tilde{\boldsymbol{G}}(\mathbb{Q}_p)$ the (sub)semigroup of $\tilde{\boldsymbol{G}}(\mathbb{Q}_p)$ which is generated by h and $\tilde{\mathcal{I}}$ and we set

$$\Delta = \Delta_h = \{g \in \tilde{\boldsymbol{G}}(\mathbb{Q}) : g \in \mathcal{K}\}.$$

Thus, $\Delta \leq \tilde{\boldsymbol{G}}(\mathbb{Q})$ is a subsemigroup containing Γ and we denote by

$$\mathcal{H} = \mathcal{H}_h = \mathcal{H}(\Gamma \backslash \Delta / \Gamma)$$

the Hecke algebra attached to the pair (Δ, Γ). Thus, $\mathcal{H}$ is a $\mathbb{Z}$-algebra which is a free $\mathbb{Z}$-module with basis $\{\Gamma\zeta\Gamma,\ \zeta \in \Delta\}$. For any $\mathbb{Z}$-algebra R we set $\mathcal{H}_R = \mathcal{H} \otimes R$ and we put

$$T_\zeta = \Gamma\zeta\Gamma \in \mathcal{H} \quad (\zeta \in \Delta).$$

2.4. *Irreducible representations of $\tilde{\boldsymbol{G}}$.* Let $\tilde{\lambda} \in X(\tilde{\boldsymbol{T}})$ be dominant. We set $\lambda^\circ = \tilde{\lambda}|_{\boldsymbol{T}} \in X(\boldsymbol{T})$ and $\lambda = \mathrm{d}\,\lambda^\circ \in \mathfrak{h}^*$. Since $X(\boldsymbol{T}) \cong \Gamma_\pi$ (see Section 1.2) we know that $\lambda \in \Gamma_\pi$ and we let $\rho_{\lambda^\circ} : \boldsymbol{G}_\pi/\mathbb{Z}_p \to \mathbf{Aut}(L_\lambda)$ be the irreducible representation of $\boldsymbol{G}/\mathbb{Z}_p = \boldsymbol{G}_\pi/\mathbb{Z}_p$ of highest weight λ°; see Section 1.4. The morphism $\rho_{\lambda^\circ} \otimes \tilde{\lambda}|_{\tilde{\boldsymbol{Z}}} : \boldsymbol{G} \times \tilde{\boldsymbol{Z}} \to \mathbf{Aut}(L_\lambda)$, given by sending $(g, z) \in \boldsymbol{G}(R) \times \tilde{\boldsymbol{Z}}(R)$ to $\tilde{\lambda}(z)\rho_{\lambda^\circ}(g) \in \mathbf{Aut}(L_\lambda(R))$, R any $\mathbb{Z}_p$-algebra, is defined over $\mathbb{Z}_p$ (see equation (8)) and factorizes over $\boldsymbol{Z}$, hence, we obtain a representation

$$\rho_{\tilde{\lambda}} : \tilde{\boldsymbol{G}} = (\boldsymbol{G} \times \tilde{\boldsymbol{Z}})/\boldsymbol{Z} \to \mathbf{Aut}(L_\lambda)$$

of $\mathbb{Z}_p$-groups (group schemes). The representation $(\rho_{\tilde{\lambda}}, L_\lambda)$ is irreducible of highest weight $\tilde{\lambda}$ and $\rho_{\lambda^\circ} = \rho_{\tilde{\lambda}}|_{\boldsymbol{G}}$. In particular, for any $\mathbb{Z}_p$-algebra R the $\boldsymbol{G}(R)$-module $L_\lambda(R)$ also is a $\tilde{\boldsymbol{G}}(R)$-module and we write $L_{\tilde{\lambda}}(R)$ for $L_\lambda(R)$ if we view it as $\tilde{\boldsymbol{G}}(R)$-module. Thus, $L_{\tilde{\lambda}}(R)$ and $L_\lambda(R)$ are isomorphic as $\boldsymbol{G}(R)$-modules, but on $L_{\tilde{\lambda}}(R)$ we have an action of $\tilde{\boldsymbol{Z}}(R)$ via $\tilde{\lambda}|_{\tilde{\boldsymbol{Z}}}$ and, hence, an action of $\tilde{\boldsymbol{T}}(R)$. Similarly, we obtain a representation

$$\tilde{\boldsymbol{T}} = (\boldsymbol{T} \times \tilde{\boldsymbol{Z}})/\boldsymbol{Z} \to \mathbf{Aut}(L_\lambda(\mu)), \quad \mu \in P_\lambda,$$

by sending $(t, z) \in \mathbf{T}(R) \times \tilde{\boldsymbol{Z}}(R)$ to $\tilde{\lambda}(z)\rho_{\lambda^\circ}(t) \in \mathbf{Aut}(L_\lambda(\mu, R))$, R any $\mathbb{Z}_p$-algebra (note that $L_\lambda(\mathbb{Z}_p, \mu)$ is a $\mathbb{Z}_p$-module, hence, $\mathbf{Aut}(L_\lambda(\mu))$ is a $\mathbb{Z}_p$-group). If we view the weight space $L_\lambda(R, \mu)$ as $\tilde{\boldsymbol{T}}(R)$-module we write it as $L_{\tilde{\lambda}}(R, \mu)$. Thus, $L_{\tilde{\lambda}}(R, \mu) = L_\lambda(R, \mu)$ as abelian groups and also as $\boldsymbol{T}(R)$-modules but on $L_{\tilde{\lambda}}(R, \mu)$ the torus $\tilde{\boldsymbol{T}}(R)$ acts via the character $\tilde{\mu} := \mu^\circ \otimes \tilde{\lambda}|_{\tilde{\boldsymbol{Z}}}$ of $\tilde{\boldsymbol{T}}$ (see Section 1.4 and equation (4) in particular). The weight decomposition of $L_{\tilde{\lambda}}(R)$ w.r.t. $\tilde{\boldsymbol{T}}(R)$ then reads

$$L_{\tilde{\lambda}}(R) = \bigoplus_{\mu \leq \lambda} L_{\tilde{\lambda}}(R, \mu), \tag{9}$$

where $L_{\tilde{\lambda}}(R, \mu)$ is the weight $\tilde{\mu}$-subspace of $L_{\tilde{\lambda}}(R)$ w.r.t. $\tilde{\boldsymbol{T}}(R)$.

2.5. ***Splitting field.*** Since the maximal torus $\tilde{\boldsymbol{T}}/\mathbb{Q}$ is assumed to be split over $\mathbb{Q}_p$ there is a subfield $F \subseteq \mathbb{Q}_p$ which is a finite extension of $\mathbb{Q}$ such that $\tilde{\boldsymbol{T}}/F$ is split. In particular, $\tilde{\boldsymbol{G}}$ is F-split and $\tilde{\lambda} \in X(\tilde{\boldsymbol{T}})$ and the irreducible highest weight representation $(\rho_{\tilde{\lambda}}, L_{\tilde{\lambda}})$ are defined over F; hence, $L_{\tilde{\lambda}}(F)$ is defined and is a $\tilde{\boldsymbol{G}}(\mathbb{Q})$-module.

We fix an algebraic closure $\bar{\mathbb{Q}}_p$ of $\mathbb{Q}_p$ with valuation v_p normalized by $v_p(p) = 1$. Since $F \subseteq \mathbb{Q}_p \subseteq \bar{\mathbb{Q}}_p$ this induces a p-adic valuation v_p on F and we obtain $F_{v_p} = \mathbb{Q}_p$. We also fix an embedding $F \subseteq \mathbb{C}$. We extend the embeddings of F to embeddings of its algebraic closure $\bar{F} \subseteq \bar{\mathbb{Q}}_p$ and $\bar{F} \subseteq \mathbb{C}$; hence, we may view F as a subfield of $\mathbb{Q}_p$ and of $\mathbb{C}$ and $\bar{F}$ as a subfield of $\mathbb{C}$ and of $\bar{\mathbb{Q}}_p$.

2.6. ***Cohomology with coefficients $L_{\tilde{\lambda}}$.*** We denote by $\Delta^{-1} \leq \tilde{\boldsymbol{G}}(\mathbb{Q})$ the sub semigroup consisting of the inverses of elements in Δ. The representation space $L_{\tilde{\lambda}}(F)$ in particular is a Δ^{-1}-module, hence, the Hecke algebra $\mathcal{H}$ acts on cohomology $H^i(\Gamma, L_{\tilde{\lambda}}(F))$. For later use we recall the definition of this action. Let $T_\zeta = \Gamma\zeta\Gamma \in \mathcal{H}$ ($\zeta \in \Delta$). We select a system of representatives $\gamma_1, \ldots, \gamma_r$ for $(\zeta^{-1}\Gamma\zeta \cap \Gamma)\backslash\Gamma$, hence,

$$T_\zeta = \bigcup_{i=1,\ldots,r} \Gamma\zeta\gamma_i.$$

Thus, for any $\eta \in \Gamma$ and any index i satisfying $1 \leq i \leq r$ there is an index $\eta(i)$ such that

$$\Gamma\zeta\gamma_i\eta = \Gamma\zeta\gamma_{\eta(i)}.$$

In particular, there are $\rho_i(\eta) \in \Gamma$, $i = 1, \ldots, r$, such that $\zeta\gamma_i\eta = \rho_i(\eta)\zeta\gamma_{\eta(i)}$. Let now $c \in C^d(\Gamma, L_{\tilde{\lambda}}(F))$ be any cochain; we then define $T_\zeta(c)$ as the cochain $c' \in C^d(\Gamma, L_{\tilde{\lambda}}(F))$, which is given by

$$c'(\eta_0, \ldots, \eta_d) = \sum_{1 \leq i \leq r} (\zeta\gamma_i)^{-1} c(\rho_i(\eta_0), \ldots, \rho_i(\eta_d)). \tag{10}$$

Since T_ζ commutes with the coboundary operator, T_ζ acts on cohomology with coefficients in $L_{\tilde{\lambda}}(F)$, i.e., T_ζ defines an element in $\mathrm{End}(H^i(\Gamma, L_{\tilde{\lambda}}(F)))$ which does not depend on the choice of the representatives $\gamma_1, \ldots, \gamma_r$ (see [Kuga et al. 1981, p. 227]). We note that this also yields $\mathcal{H}$-module structures on $H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{Q}_p))$ and $H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{C}))$. We denote by

$$H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{Z}_p))_{\mathrm{int}}$$

the image of the canonical mapping $H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{Z}_p)) \to H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{Q}_p))$; this defines a lattice in $H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{Q}_p))$.

Cuspidal cohomology. We select a maximal compact open subgroup $\tilde{K}_\infty \leq \tilde{\boldsymbol{G}}(\mathbb{R})$. We denote by $A_{\tilde{\boldsymbol{G}}}$ the connected component of the real points of a maximal $\mathbb{Q}$-split torus $\boldsymbol{A}_{\tilde{\boldsymbol{G}}}$ in the center of $\tilde{\boldsymbol{G}}$ and we set $X = \tilde{\boldsymbol{G}}(\mathbb{R})/\tilde{K}_\infty A_{\tilde{\boldsymbol{G}}}$. The cuspidal cohomology $H^i_{\mathrm{cusp}}(\Gamma\backslash X, L_{\tilde{\lambda}}(\mathbb{C}))$ is a subspace of full cohomology $H^i(\Gamma\backslash X, L_{\tilde{\lambda}}(\mathbb{C}))$. We note that if the highest weight $\tilde{\lambda}$ is regular and $\tilde{\boldsymbol{G}}$ has discrete series then there are isomorphisms

$$H^i_{(2)}(\Gamma\backslash X, L_{\tilde{\lambda}}(\mathbb{C})) = H^i_!(\Gamma\backslash X, L_{\tilde{\lambda}}(\mathbb{C})) = H^i_{\mathrm{cusp}}(\Gamma\backslash X, L_{\tilde{\lambda}}(\mathbb{C}))$$

and cuspidal cohomology vanishes in all degrees except for the middle degree $d = d_{\tilde{\boldsymbol{G}}} = \frac{1}{2}\dim X$.

Weighted cohomology. For use in Section 4.14 involving the Goresky–MacPherson trace formula we briefly recall the relation between cuspidal cohomology and weighted cohomology. We denote by $W^\nu H^i(\overline{\Gamma\backslash X}, L_{\tilde{\lambda}}(F))$ the weighted cohomology groups of Γ (see [Goresky et al. 1994]). If ν is the middle weight profile and $\tilde{\boldsymbol{G}}$ has discrete series then [Nair 1999, Corollary B, p. 3] (see also Section 5.1 there) implies that there is an isomorphism

$$W^\nu H^i(\overline{\Gamma\backslash X}, L_{\tilde{\lambda}}(\mathbb{C})) = H^i_{(2)}(\Gamma\backslash X, L_{\tilde{\lambda}}(\mathbb{C})).$$

(in the hermitian case this also follows from the Zucker conjecture, which was proven by Saper and Stern, and independently by Looijenga). Thus, if in addition the highest weight $\tilde{\lambda} \in X(\tilde{\boldsymbol{T}})$ is regular then there is a canonical isomorphism of Hecke modules

$$W^\nu H^i(\overline{\Gamma\backslash X}, L_{\tilde{\lambda}}(\mathbb{C})) = H^i_{\mathrm{cusp}}(\Gamma\backslash X, L_{\tilde{\lambda}}(\mathbb{C}))$$

where the cohomology groups are nonvanishing only if $i = d$; in particular, we obtain

$$(-1)^d \mathrm{tr}(T\,|\,H^d_{\mathrm{cusp}}(\Gamma\backslash X, L_{\tilde{\lambda}}(\mathbb{C}))) = \mathrm{Lef}(T\,|\,W^\nu H^\bullet(\overline{\Gamma\backslash X}, L_{\tilde{\lambda}}(\mathbb{C})))$$

where $T \in \mathcal{H}$ is a Hecke operator.

We mention that this implies an F-structure on cuspidal cohomology: we denote by $H^d_{\mathrm{cusp}}(\Gamma\backslash X, L_{\tilde{\lambda}}(F))$ the image of $W^\nu H^d(\overline{\Gamma\backslash X}, L_{\tilde{\lambda}}(F))$ in $H^d_{\mathrm{cusp}}(\Gamma\backslash X, L_{\tilde{\lambda}}(\mathbb{C}))$ and obtain

$$H^d_{\mathrm{cusp}}(\Gamma\backslash X, L_{\tilde{\lambda}}(\mathbb{C})) = H^d_{\mathrm{cusp}}(\Gamma\backslash X, L_{\tilde{\lambda}}(F)) \otimes \mathbb{C}.$$

2.7. *Normalization of Hecke operators.* We want to normalize the Hecke operators so that they act on cohomology with p-adically integral coefficients. We recall the following diagram of inclusions:

$$\begin{array}{ccc} \tilde{\mathcal{I}} & \subseteq & \mathcal{K} \\ \cup & & \cup \\ \Gamma & \subseteq & \Delta. \end{array}$$

Lemma. 1. *$\mathcal{K} = \bigcup_{e \in \mathbb{N}_0} \tilde{\mathcal{I}} h^e \tilde{\mathcal{I}}$, i.e., any element $g \in \mathcal{K}$ can be written $g = k_1 h^e k_2$ with $k_1, k_2 \in \tilde{\mathcal{I}}$ and $e \in \mathbb{N}_0$.*

2. *If $\tilde{\mathcal{I}} h^e \tilde{\mathcal{I}} = \tilde{\mathcal{I}} h^f \tilde{\mathcal{I}}$, $e, f \in \mathbb{N}_0$, then $e = f$.*

3. *Let $\tilde{\lambda} \in X(\tilde{\boldsymbol{T}})$ be a dominant weight. The mapping*

$$\hat{\lambda} : \mathcal{K} \to F^*, \qquad k_1 h^e k_2 \mapsto \tilde{\lambda}(h^e) \quad (k_1, k_2 \in \tilde{\mathcal{I}}, e \in \mathbb{N}_0)$$

is a well defined morphism of semigroups.

Proof. Equations (5) and (6) in Section 1.5 remain valid with the same proof if $\mathcal{I}$ is replaced by $\tilde{\mathcal{I}}$, $T_{\mathbb{Z}_p}$ is replaced by $\tilde{T}_{\mathbb{Z}_p} = T_{\mathbb{Z}_p} \tilde{\boldsymbol{Z}}(\mathbb{Z}_p)$ and if $h, t, t' \in \tilde{\boldsymbol{T}}(\mathbb{Q}_p)$. Conclusion 1 is then immediate by equation (5). As for 2 we note that equation (6) implies $h^e = \delta h^f$, where $\delta \in \tilde{T}_{\mathbb{Z}_p} \subseteq \tilde{\boldsymbol{T}}(\mathbb{Z}_p)$. Applying an arbitrary simple root α and taking p-adic values yields $e v_p(\alpha(h)) = v_p(\alpha(\delta)) + f v_p(\alpha(h))$ and since $\alpha(\delta) \in \mathbb{Z}_p^*$ and $v_p(\alpha(h)) > 0$ we deduce that $e = f$. As for 3 we remark that parts 1 and 2 show that $\hat{\lambda}$ is well defined and equation (5) in Section 1.5 implies that $\hat{\lambda}$ is a morphism of semigroups. Thus, the lemma is proven. □

Let $\tilde{\lambda} \in X(\tilde{\boldsymbol{T}})$ be a dominant weight. By restriction, $\hat{\lambda}$ induces a mapping $\hat{\lambda} : \Delta \to F^*$. For any F-algebra R we define an R-linear mapping

$$\mathcal{H}_R \to \mathcal{H}_R$$

by sending $\Gamma \zeta \Gamma \mapsto \hat{\lambda}(\zeta) \Gamma \zeta \Gamma$, $\zeta \in \Delta$; note that $\{\Gamma \zeta \Gamma,\ \zeta \in \Delta\}$ is a basis for $\mathcal{H}_R$ and that the assignment is well defined since $\hat{\lambda}$ vanishes on $\tilde{\mathcal{I}}$ by definition and, hence, vanishes on $\Gamma \subseteq \tilde{\mathcal{I}}$. We denote the image of $T \in \mathcal{H}_R$ under the above mapping by $T_{\tilde{\lambda}} \in \mathcal{H}_R$ and we call $T_{\tilde{\lambda}}$ the $\tilde{\lambda}$-normalization of T. In particular, if $\zeta \in \Delta$ with $\zeta \in \tilde{\mathcal{I}} h^e \tilde{\mathcal{I}}$ then

$$(T_\zeta)_{\tilde{\lambda}} = \hat{\lambda}(\zeta) T_\zeta = \tilde{\lambda}(h^e) T_\zeta.$$

The normalization $T_{\tilde{\lambda}}$ of any $T \in \mathcal{H}_{\mathbb{Q}_p}$ leaves $H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{Q}_p))$ invariant and we want to show that $T_{\tilde{\lambda}}$ leaves cohomology with integral coefficients $L_{\tilde{\lambda}}(\mathbb{Z}_p)$ invariant. To this end we first show that the "normalization $\hat{\lambda}(g) g^{-1}$" of any $g \in \mathcal{K}$ leaves the lattice $L_{\tilde{\lambda}}(\mathbb{Z}_p)$ invariant.

2.8. *Lemma.* *For all $g \in \mathcal{K}$ and $v \in L_{\tilde{\lambda}}(\mathbb{Z}_p)$ we have $\hat{\lambda}(g) g^{-1} v \in L_{\tilde{\lambda}}(\mathbb{Z}_p)$.*

Proof. Any $g \in \mathcal{K}$ has the form $g = k_1 h^e k_2$ with $k_1, k_2 \in \tilde{\mathcal{I}}$. Since $\hat{\lambda}(g) = \tilde{\lambda}(h^e)$ and since $\tilde{\mathcal{I}} \subseteq \tilde{\boldsymbol{G}}(\mathbb{Z}_p)$ leaves $L_{\tilde{\lambda}}(\mathbb{Z}_p)$ invariant it is sufficient to show that $\tilde{\lambda}(h^e) h^{-e}$ leaves $L_{\tilde{\lambda}}(\mathbb{Z}_p)$ invariant. Since, as we saw in equation (9), we further have

$$L_{\tilde{\lambda}}(\mathbb{Z}_p) = \bigoplus_{\mu \leq \lambda} L_{\tilde{\lambda}}(\mathbb{Z}_p, \mu),$$

it is sufficient to show that $\tilde{\lambda}(h^e) h^{-e} v_\mu \in L_{\tilde{\lambda}}(\mathbb{Z}_p, \mu)$ for any weight vector v_μ

in $L_{\tilde{\lambda}}(\mathbb{Z}_p, \mu)$. Equation (9) implies that $\tilde{\lambda}(h^e)h^{-e}v_\mu = \tilde{\lambda}(h^e)\tilde{\mu}(h^{-e})v_\mu$. We write $\mu = \lambda - \nu$ where $\nu = \sum_{\alpha\in\Delta} n_\alpha \alpha$ with $n_\alpha \in \mathbb{N}_0$ and $h = tz$ with $t \in \boldsymbol{T}(\bar{\mathbb{Q}}_p)$, $z \in \tilde{\boldsymbol{Z}}(\bar{\mathbb{Q}}_p)$ and obtain

$$\begin{aligned}\tilde{\lambda}(h^e)\tilde{\mu}(h^{-e}) &= \lambda^\circ(t^e)\tilde{\lambda}(z^e)\mu^\circ(t^{-e})\tilde{\lambda}(z^{-e}) = \lambda^\circ(t^e)\mu^\circ(t^{-e})\\ &= \lambda^\circ(t^e)\lambda^\circ(t^{-e})\nu^\circ(t^e) = \prod_{\alpha\in\Delta}\alpha(t)^{en_\alpha}.\end{aligned}$$

Since $v_p(\alpha(t)) = v_p(\alpha(h)) \geq 1$ for all simple roots α we deduce that

$$v_p\left(\prod_{\alpha\in\Delta}\alpha(t)^{en_\alpha}\right) \geq e\sum_{\alpha\in\Delta} n_\alpha = e\,\mathrm{ht}(\nu) \geq 0.$$

Thus, taking into account that $\mathrm{ht}(\nu) = \mathrm{ht}_\lambda(\mu)$ we obtain

$$\tilde{\lambda}(h^e)h^{-e}v_\mu \in p^{e\,\mathrm{ht}_\lambda(\mu)}L_{\tilde{\lambda}}(\mathbb{Z}_p, \mu) \subseteq L_{\tilde{\lambda}}(\mathbb{Z}_p, \mu), \tag{11}$$

which implies that $\tilde{\lambda}(h^e)h^{-e}$ leaves $L_{\tilde{\lambda}}(\mathbb{Z}_p)$ invariant. The lemma is proven. □

2.9. Corollary. 1. *For any $T \in \mathcal{H}_{\mathbb{Z}_p}$ the normalized operator $T_{\tilde{\lambda}}$ leaves the group $H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{Z}_p))$ invariant. In particular, $T_{\tilde{\lambda}}$ acts on integral cohomology $H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{Z}_p))_{\mathrm{int}}$.*

2. *For any $T \in \mathcal{H}$ the eigenvalues of $T_{\tilde{\lambda}} \in \mathcal{H}_F$ on $H^i(\Gamma\backslash X, L_{\tilde{\lambda}}(\mathbb{C}))$ and on $H^i_{\mathrm{cusp}}(\Gamma\backslash X, L_{\tilde{\lambda}}(\mathbb{C}))$ are algebraic over F (note that $F \subseteq \mathbb{C}$) and are contained in $\mathcal{O}_{\bar{\mathbb{Q}}_p}$ (note that $\bar{F} \subseteq \bar{\mathbb{Q}}_p$).*

Proof. 1. We may assume that $T = T_\zeta$ for some $\zeta \in \Delta$. The claim then follows directly from the definition of the action of T_ζ on cohomology given in equation (10) and the lemma just proved (note that $\zeta \in \Delta \subseteq \mathcal{K}$ and that $\gamma_i^{-1} \in \Gamma \subseteq \tilde{\boldsymbol{G}}(\mathbb{Z}_p)$ leaves $L_{\tilde{\lambda}}(\mathbb{Z}_p)$ invariant).

2. The cuspidal cohomology $H^i_{\mathrm{cusp}}(\Gamma\backslash X, L_{\tilde{\lambda}}(\mathbb{C})) \subseteq H^i(\Gamma\backslash X, L_{\tilde{\lambda}}(\mathbb{C}))$ is a Hecke submodule of full cohomology, hence, the eigenvectors and (complex) eigenvalues of $T_{\tilde{\lambda}}$ on $H^i_{\mathrm{cusp}}(\Gamma\backslash X, L_{\tilde{\lambda}}(\mathbb{C}))$ are contained in the set of eigenvectors and eigenvalues of $T_{\tilde{\lambda}}$ on $H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{C}))$. Since $H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{C})) = H^i(\Gamma, L_{\tilde{\lambda}}(F)) \otimes \mathbb{C}$ all eigenvalues of $T_{\tilde{\lambda}}$ on $H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{C}))$ are algebraic over F and, hence, they already appear as eigenvalues of $T_{\tilde{\lambda}}$ on $H^i(\Gamma, L_{\tilde{\lambda}}(\bar{F}))$. In particular, they also appear as eigenvalues of $T_{\tilde{\lambda}}$ on $H^i(\Gamma, L_{\tilde{\lambda}}(\bar{\mathbb{Q}}_p))$ where the latter contains the $T_{\tilde{\lambda}}$-invariant $\mathbb{Z}_p$-lattice $H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{Z}_p))_{\mathrm{int}}$. Thus, all eigenvalues of $T_{\tilde{\lambda}}$ on $H^i(\Gamma\backslash X, L_{\tilde{\lambda}}(\mathbb{C}))$ are algebraic over F and contained in the integer ring $\mathcal{O}_{\bar{\mathbb{Q}}_p}$ after embedding $\bar{F}$ in $\bar{\mathbb{Q}}_p$. This completes the proof. □

The diagram of inclusions on the next page recapitulates the objects appearing in the proof above and groups them together for easy lookup as they come up later in the discussion.

$$\begin{array}{ccccc} H^i_{\mathrm{cusp}}(\Gamma\backslash X, L_{\tilde\lambda}(\mathbb{C})) & \subseteq & H^i(\Gamma\backslash X, L_{\tilde\lambda}(\mathbb{C})) & & \\ & & \cup & & \\ & & H^i(\Gamma\backslash X, L_{\tilde\lambda}(\bar F)) & \subseteq & H^i(\Gamma\backslash X, L_{\tilde\lambda}(\bar{\mathbb{Q}}_p)) \\ & & \cup & & \cup \\ & & H^i(\Gamma\backslash X, L_{\tilde\lambda}(F)) & \subseteq & H^i(\Gamma\backslash X, L_{\tilde\lambda}(\mathbb{Q}_p)) \\ & & & & \cup \\ & & & & H^i(\Gamma\backslash X, L_{\tilde\lambda}(\mathbb{Z}_p))_{\mathrm{int}} \end{array}$$

2.10. ***Mod p reduction of irreducible representations.*** We denote by $\boldsymbol{T}_m$ the maximal torus in $\mathbf{GL}_m$ consisting of diagonal matrices, by $\boldsymbol{B}_m^-$ the Borel subgroup in $\mathbf{GL}_m$ consisting of all lower triangular matrices and by $\wp : \mathbf{GL}_m(\mathbb{Z}_p) \to \mathbf{GL}_m(\mathbb{F}_p)$ the mod p reduction map for $\mathbf{GL}_m$. Since $\tilde{\boldsymbol{G}}/\mathbb{Z}_p$ is smooth the reduction map $\wp : \tilde{\boldsymbol{G}}(\mathbb{Z}_p) \to \tilde{\boldsymbol{G}}(\mathbb{F}_p)$ is surjective. We let $\tilde\lambda \in X(\tilde{\boldsymbol{T}})$ be a dominant weight and as before we set $\lambda = d\,(\tilde\lambda|_{\boldsymbol{T}}) \in \mathfrak{h}^*$. We define the mod p reduction

$$\bar\rho_{\tilde\lambda} : \tilde{\boldsymbol{G}}(\mathbb{F}_p) \to \mathbf{GL}_m(\mathbb{F}_p)$$

of the representation $\rho_{\tilde\lambda} : \tilde{\boldsymbol{G}}/\mathbb{Z}_p \to \mathbf{GL}_m/\mathbb{Z}_p$ by $\bar\rho_{\tilde\lambda}(\bar g) = \wp(\rho_{\tilde\lambda}(g))$, where $g \in \tilde{\boldsymbol{G}}(\mathbb{Z}_p)$ satisfies $\wp(g) = \bar g$. We denote by $\mathcal{J}_m \le \mathbf{GL}_m(\mathbb{Z}_p)$ the Iwahori subgroup consisting of all elements $g \in \mathbf{GL}_m(\mathbb{Z}_p)$ such that $\wp(g) \in \boldsymbol{B}_m^-(\mathbb{F}_p)$.

Lemma. $$\rho_{\tilde\lambda}(\tilde{\mathcal{I}}) \subseteq \mathcal{J}_m.$$

Proof. In Section 1.4 we selected a $\mathbb{Z}_p$-basis $\mathcal{B}$ of $L_\lambda(\mathbb{Z}_p)$ to identify $\mathrm{Aut}(L_\lambda(\mathbb{Z}_p)) = \mathbf{GL}_m(\mathbb{Z}_p)$. Since $L_\lambda(\mathbb{Z}_p) = \bigoplus_{\mu\in P_\lambda} L_\lambda(\mathbb{Z}_p,\mu)$ — see Section 1.1 — we may choose a basis $\mathcal{B}$ consisting of weight vectors w.r.t. $\mathfrak{h}$. We order $\mathcal{B}$ so that, if $v_\mu, v_{\mu'} \in \mathcal{B}$ are vectors of respective weights $\mu, \mu' \in \mathfrak{h}^*$, then $\mathrm{ht}_\lambda(\mu) < \mathrm{ht}_\lambda(\mu')$ implies that $v_\mu < v_{\mu'}$. We consider the image $\rho_{\tilde\lambda}(x_\alpha^\pi(t_\alpha)) = \rho_{\lambda^\circ}(x_\alpha^\pi(t_\alpha)) = x_\alpha^\lambda(t_\alpha) \in \mathrm{Aut}(L_\lambda(\bar{\mathbb{Q}}_p))$, where $\alpha \in \Phi^-$ ($t_\alpha \in \bar{\mathbb{Q}}_p$). Let $v_\mu \in \mathcal{B}$ be a basis vector of weight μ. Since $x_\alpha^\lambda(t_\alpha)v_\mu = v_\mu + t_\alpha x_\alpha v_\mu + \frac{1}{2}t_\alpha^2 x_\alpha^2 v_\mu + \cdots$, we see that $x_\alpha^\lambda(t_\alpha)v_\mu$ is a sum of vectors of weights μ, $\mu{+}\alpha$, $\mu{+}2\alpha, \ldots$, which are of strictly increasing relative height (since $\alpha < 0$). Hence, $\rho_{\tilde\lambda}(x_\alpha^\pi(t_\alpha))$ has lower triangular form w.r.t. $\mathcal{B}$, i.e., $\rho_{\tilde\lambda}(x_\alpha^\pi(t_\alpha)) \in \boldsymbol{B}_m^-(\bar{\mathbb{Q}}_p)$. This shows that $\rho_{\tilde\lambda}(\boldsymbol{N}^-(\bar{\mathbb{Q}}_p)) \subseteq \boldsymbol{B}_m^-(\bar{\mathbb{Q}}_p)$. Since $\boldsymbol{T}(\bar{\mathbb{Q}}_p)$ preserves weight spaces by equation (4) in Section 1.4 we find quite analogous that $\rho_{\tilde\lambda}(\boldsymbol{T}(\bar{\mathbb{Q}}_p)) \subseteq \boldsymbol{T}_m(\bar{\mathbb{Q}}_p)$, hence, $\rho_{\tilde\lambda}(\boldsymbol{B}^-(\bar{\mathbb{Q}}_p)) \subseteq \boldsymbol{B}_m^-(\bar{\mathbb{Q}}_p)$. Thus, we obtain $\rho_{\tilde\lambda}(\boldsymbol{B}^-(\mathbb{Z}_p)) \subseteq \boldsymbol{B}_m^-(\bar{\mathbb{Q}}_p) \cap \mathbf{GL}_m(\mathbb{Z}_p) = \boldsymbol{B}_m^-(\mathbb{Z}_p)$ and, hence,

$$\bar\rho_{\tilde\lambda}(\boldsymbol{B}^-(\mathbb{F}_p)) \subseteq \boldsymbol{B}_m^-(\mathbb{F}_p).$$

We obtain $\wp(\rho_{\tilde\lambda}(\mathcal{I})) = \bar\rho_{\tilde\lambda}(\wp(\mathcal{I})) \subseteq \bar\rho_{\tilde\lambda}(\boldsymbol{B}^-(\mathbb{F}_p)) \subseteq \boldsymbol{B}_m^-(\mathbb{F}_p)$ and since $\rho_{\tilde\lambda}(\mathcal{I}) \subseteq \rho_{\tilde\lambda}(\tilde{\boldsymbol{G}}(\mathbb{Z}_p)) \subseteq \mathbf{GL}_m(\mathbb{Z}_p)$ we deduce that $\rho_{\tilde\lambda}(\mathcal{I}) \subseteq \mathcal{J}_m$. Equation (8) implies that $\rho_{\tilde\lambda}(z) = \tilde\lambda(z)1_{\mathbf{GL}_m} \in \mathcal{J}_m$ for all $z \in \tilde{\boldsymbol{Z}}(\mathbb{Z}_p)$, hence, we finally obtain $\rho_{\tilde\lambda}(\tilde{\mathcal{I}}) \subseteq \mathcal{J}_m$ and the lemma is proven. $\square$

3. Boundedness of dimension of slope subspaces

3.1. We keep the assumptions from the previous sections. In particular, $\tilde{\boldsymbol{G}}/\mathbb{Q}$ is a connected reductive group containing a $\mathbb{Q}_p$-split maximal torus $\tilde{\boldsymbol{T}}/\mathbb{Q}$ and $\Gamma \leq \tilde{\boldsymbol{G}}(\mathbb{Q})$ is an arithmetic subgroup such that $\Gamma \subseteq \tilde{\mathcal{I}}$. We will obtain bounds for the dimension of the slope subspaces of $H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{Q}_p))$; see 3.10 Corollary. This extends the main result in [Mahnkopf 2014], since (i) we allow Γ to be an arbitrary subgroup in $\tilde{\mathcal{I}}$ (i.e., we do not assume that Γ is contained in the smaller group $K_*(p) < \tilde{\mathcal{I}}$ defined in [Mahnkopf 2013; 2014]); (ii) we do not assume that $\Gamma \leq \boldsymbol{G}(\mathbb{Q})$ where $\boldsymbol{G}$ is the derived group of $\tilde{\boldsymbol{G}}$; (iii) we obtain stronger bounds for the dimension of the slope subspaces than those in [Mahnkopf 2014]. The proof follows the one in that paper. To deal with arithmetic subgroups Γ which are only contained in $\tilde{\mathcal{I}}$ we have to generalize the notion of truncation of an irreducible representation introduced in [Mahnkopf 2013; 2014] (see Section 3.3).

3.2. We note the following corrections to the works just cited.

1. In [Mahnkopf 2014] we considered a connected reductive group $\tilde{\boldsymbol{G}}$ which is defined over a number field F with $F_{\mathfrak{p}}$-split maximal torus $\tilde{\boldsymbol{T}}$ (Section 1.4 there). As in the present article, we have to assume that $\tilde{\boldsymbol{T}}$ is defined over F (and split over $F_{\mathfrak{p}}$); thus, the F-points $\tilde{\boldsymbol{T}}(F)$ are defined and we may select $h \in \tilde{\boldsymbol{T}}(F)$ as done in Section 1.6 of [Mahnkopf 2014].

2. Let $\boldsymbol{G}$ denote the derived group of the connected reductive group $\tilde{\boldsymbol{G}}$ which we considered in [Mahnkopf 2013; 2014]. Hence, $\boldsymbol{G}$ is a semisimple group and in those two papers we assumed that it is isomorphic over a splitting field to a Chevalley group $\boldsymbol{G}_{\lambda_0}$ for an irreducible representation ρ_{λ_0} of the Lie algebra $\mathfrak{g} = \mathrm{Lie}(\boldsymbol{G}) \otimes \mathbb{C}$. In general, $\boldsymbol{G}$ over its splitting field only is isomorphic to a Chevalley group $\boldsymbol{G}_\pi$ for a semisimple representation π of $\mathfrak{g}$ (if one restricts to irreducible representations π one does not obtain all covering groups of the adjoint group with Lie algebra $\mathfrak{g}$). Since in the cited papers we did not make use of the irreducibility of the representation ρ_{λ_0} the results also hold if we consider a Chevalley group $\boldsymbol{G}_\pi$ which is attached to a semisimple representation π of $\mathfrak{g}$.

3. In the summation formula (7) in Section 2.5.3 of [Mahnkopf 2014], the Bernoulli number $B_s(0)$ has to be replaced by $B_s(1)$ (note that $B_s(1) = B_s(0)$ for all $s > 1$ but $B_1(1) = -B_1(0) = \frac{1}{2}$).

3.3. ***Truncations with slope parameter.*** For the moment we let $\sigma \in \mathbb{N}$ be any natural number and we consider the subgroup

$$\tilde{K}_*(\sigma) = \tilde{K}_*(p, \sigma) = \langle K_*(p, \sigma), \tilde{\boldsymbol{Z}}(\mathbb{Z}_p)\rangle \leq \tilde{\boldsymbol{G}}(\mathbb{Z}_p).$$

Thus, if $\sigma \geq \max_{\alpha\in\Phi^+} \mathrm{ht}(\alpha)$ then $\tilde{K}_*(\sigma) = \tilde{\mathcal{I}}$. For any $r \in \mathbb{N}_0$ we define the $\mathbb{Z}_p$-submodule

$$(12)\quad L_{\tilde{\lambda}}(\mathbb{Z}_p, r, \sigma) := \bigoplus_{\substack{\mu\leq\lambda \\ 0\leq \mathrm{ht}_\lambda(\mu)\leq r\sigma}} p^{r-\lceil \frac{1}{\sigma}\mathrm{ht}_\lambda(\mu)\rceil} L_{\tilde{\lambda}}(\mathbb{Z}_p, \mu) \oplus \bigoplus_{\substack{\mu\leq\lambda \\ \mathrm{ht}_\lambda(\mu)>r\sigma}} L_{\tilde{\lambda}}(\mathbb{Z}_p, \mu)$$

of $L_{\tilde{\lambda}}(\mathbb{Z}_p)$.

Lemma. *The $\mathbb{Z}_p$-module $L_{\tilde{\lambda}}(\mathbb{Z}_p, r, \sigma)$ of $L_{\tilde{\lambda}}(\mathbb{Z}_p)$ is $\tilde{K}_*(\sigma)$-invariant.*

Proof. In view of the definition of $\tilde{K}_*(\sigma)$ we have to show that the three types of generators of $\tilde{K}_*(\sigma)$ map any of the weight subspaces of $L_{\tilde{\lambda}}(\mathbb{Z}_p, r, \sigma)$ (see equation (12)) to $L_{\tilde{\lambda}}(\mathbb{Z}_p, r, \sigma)$. Since $t_\alpha^n x_\alpha^n/n!$, where $t_\alpha \in \mathbb{Z}_p$ and $\alpha \in \Phi$, maps $L_{\tilde{\lambda}}(\mathbb{Z}_p, \mu)$ to $L_{\tilde{\lambda}}(\mathbb{Z}_p, \mu + n\alpha)$, it is immediate that any $t_\alpha^n x_\alpha^n/n!$ with $\alpha < 0$ and $t_\alpha \in \mathbb{Z}_p$ maps any weight subspace $p^m L_{\tilde{\lambda}}(\mathbb{Z}_p, \mu)$ contained in $L_{\tilde{\lambda}}(\mathbb{Z}_p, r, \sigma)$ to $p^m L_{\tilde{\lambda}}(\mathbb{Z}_p, \mu + n\alpha)$, which is contained in $L_{\tilde{\lambda}}(\mathbb{Z}_p, r, \sigma)$ because $\mathrm{ht}_\lambda(\mu + n\alpha) \geq \mathrm{ht}_\lambda(\mu)$. Hence, any generator $x_\alpha(t_\alpha)$, $\alpha \in \Phi^-$, $t_\alpha \in \mathbb{Z}_p$, leaves $L_{\tilde{\lambda}}(\mathbb{Z}_p, r, \sigma)$ invariant. We look at generators $x_\alpha(t_\alpha)$ where $\alpha \in \Phi^+$—hence, $t_\alpha \in p^{\lceil \frac{1}{\sigma}\mathrm{ht}(\alpha)\rceil}\mathbb{Z}_p$. Using the inequalities $\lceil x\rceil - \lceil y\rceil \leq \lceil x - y\rceil$ and $n\lceil x\rceil \geq \lceil nx\rceil$, $x, y \in \mathbb{R}$, $n \in \mathbb{N}$, we find for all $n \in \mathbb{N}$ and all weights $\mu \leq \lambda$ with $\mathrm{ht}_\lambda(\mu) \leq r\sigma$:

$$\begin{aligned} t_\alpha^n \frac{x_\alpha^n}{n!} p^{r-\lceil \frac{1}{\sigma}\mathrm{ht}_\lambda(\mu)\rceil} L_{\tilde{\lambda}}(\mathbb{Z}_p, \mu) &\subseteq p^{n\lceil \frac{1}{\sigma}\mathrm{ht}(\alpha)\rceil} p^{r-\lceil \frac{1}{\sigma}\mathrm{ht}_\lambda(\mu)\rceil} L_{\tilde{\lambda}}(\mathbb{Z}_p, \mu + n\alpha) \\ &\subseteq p^{r-\lceil \frac{1}{\sigma}\mathrm{ht}_\lambda(\mu)\rceil + \lceil n\frac{1}{\sigma}\mathrm{ht}(\alpha)\rceil} L_{\tilde{\lambda}}(\mathbb{Z}_p, \mu + n\alpha) \\ &\subseteq p^{r-(\lceil \frac{1}{\sigma}\mathrm{ht}_\lambda(\mu) - n\frac{1}{\sigma}\mathrm{ht}(\alpha)\rceil)} L_{\tilde{\lambda}}(\mathbb{Z}_p, \mu + n\alpha) \\ &= p^{r-(\lceil \frac{1}{\sigma}\mathrm{ht}_\lambda(\mu+n\alpha)\rceil)} L_{\tilde{\lambda}}(\mathbb{Z}_p, \mu + n\alpha) \\ &\subseteq L_{\tilde{\lambda}}(\mathbb{Z}_p, r, \sigma); \end{aligned}$$

for the last inclusion note that $\mathrm{ht}_\lambda(\mu+n\alpha) \leq \mathrm{ht}_\lambda(\mu) \leq r\sigma$ (we remark that if $\mu+n\alpha$ is not $\leq \lambda$ then $t_\alpha^n(x_\alpha^n/n!)p^{r-\lceil \frac{1}{\sigma}\mathrm{ht}_\lambda(\mu)\rceil} L_{\tilde{\lambda}}(\mathbb{Z}_p, \mu) = 0$). For weights $\mu \leq \lambda$ with $\mathrm{ht}_\lambda(\mu) > r\sigma$ we find

$$t_\alpha^n \frac{x_\alpha^n}{n!} L_{\tilde{\lambda}}(\mathbb{Z}_p, \mu) \subseteq p^{n\lceil \frac{1}{\sigma}\mathrm{ht}(\alpha)\rceil} L_{\tilde{\lambda}}(\mathbb{Z}_p, \mu + n\alpha) \subseteq p^{\lceil \frac{n}{\sigma}\mathrm{ht}(\alpha)\rceil} L_{\tilde{\lambda}}(\mathbb{Z}_p, \mu + n\alpha).$$

Since $\lceil \frac{n}{\sigma}\mathrm{ht}(\alpha)\rceil \geq 0$ this shows that $t_\alpha^n(x_\alpha^n/n!)L_{\tilde{\lambda}}(\mathbb{Z}_p, \mu) \subseteq L_{\tilde{\lambda}}(\mathbb{Z}_p, r, \sigma)$ if $\mathrm{ht}_\lambda(\mu + n\alpha) > r\sigma$. If $\mathrm{ht}_\lambda(\mu + n\alpha) \leq r\sigma$ we note that

$$r - \left\lceil \tfrac{1}{\sigma}\mathrm{ht}_\lambda(\mu + n\alpha)\right\rceil \leq r - \tfrac{1}{\sigma}(\mathrm{ht}_\lambda(\mu) - n\,\mathrm{ht}(\alpha)) \leq \tfrac{n}{\sigma}\mathrm{ht}(\alpha),$$

which shows that again $t_\alpha^n(x_\alpha^n/n!)L_{\tilde{\lambda}}(\mathbb{Z}_p, \mu) \subseteq L_{\tilde{\lambda}}(\mathbb{Z}_p, r, \sigma)$. Hence, the generators $x_\alpha(t_\alpha)$, $\alpha \in \Phi^+$, $t_\alpha \in p^{\lceil \frac{1}{\sigma}\mathrm{ht}(\alpha)\rceil}\mathbb{Z}_p$, also leave $L_{\tilde{\lambda}}(\mathbb{Z}_p, r, \sigma)$ invariant. Finally if $t \in \tilde{\boldsymbol{T}}(\mathbb{Z}_p)$ then t leaves all weight spaces $L_{\tilde{\lambda}}(\mathbb{Z}_p, \mu)$ invariant because the representation

$\rho_{\tilde{\lambda}}$ is defined over $\mathbb{Z}_p$. Hence, $L_{\tilde{\lambda}}(\mathbb{Z}_p, r, \sigma)$ is invariant under all generators of $\tilde{K}_*(\sigma)$ and the lemma is proven. $\square$

Definition. The quotient

$$\boldsymbol{L}_{\tilde{\lambda}}^{[r]}(\mathbb{Z}_p, \sigma) = \frac{L_{\tilde{\lambda}}(\mathbb{Z}_p)}{L_{\tilde{\lambda}}(\mathbb{Z}_p, r, \sigma)} = \bigoplus_{\substack{\mu \le \lambda \\ \mathrm{ht}_\lambda(\mu) \le r\sigma}} \frac{L_{\tilde{\lambda}}(\mathbb{Z}_p, \mu)}{p^{r - \lceil \frac{1}{\sigma} \mathrm{ht}_\lambda(\mu) \rceil} L_{\tilde{\lambda}}(\mathbb{Z}_p, \mu)}$$

is called the truncation of $L_{\tilde{\lambda}}(\mathbb{Z}_p)$ of height $r \in \mathbb{N}_0$ and slope $\frac{1}{\sigma}$, $\sigma \in \mathbb{N}$.

Remark. The lemma just proved implies that $\boldsymbol{L}_{\tilde{\lambda}}^{[r]}(\mathbb{Z}_p, \sigma)$ is a $\tilde{K}_*(\sigma)$-module. Therefore, if $\sigma \ge \max_{\alpha \in \Phi^+} \mathrm{ht}(\alpha)$, i.e., $\tilde{K}_*(\sigma) = \tilde{\mathcal{I}}$, then $\boldsymbol{L}_{\tilde{\lambda}}^{[r]}(\mathbb{Z}_p, \sigma)$ is a Γ-module (recall that $\Gamma \le \tilde{\mathcal{I}}$).

3.4. *Lemma.* *For any dominant and integral weight λ and any $r \in \mathbb{N}$, $\sigma \in \mathbb{N}$ there is an embedding (of $\mathbb{Z}_p$-modules)*

$$\boldsymbol{L}_{\tilde{\lambda}}^{[r]}(\mathbb{Z}_p, \sigma) \le \bigoplus_{h=0}^{r} \left(\frac{\mathbb{Z}_p}{p^{r-h}\mathbb{Z}_p} \right)^{M_{\sigma,h}},$$

where $M_{\sigma,h} = \sigma(\sigma h + 1)^{s-1}$ ($s = |\Phi^+|$).

Proof. We write

$$\boldsymbol{L}_{\tilde{\lambda}}^{[r]}(\mathbb{Z}_p, \sigma) = \bigoplus_{\substack{h \in \mathbb{N}_0 \\ 0 \le h \le r\sigma}} \bigoplus_{\substack{\mu \le \lambda \\ \mathrm{ht}_\lambda(\mu) = h}} \frac{\mathbb{Z}_p}{p^{r - \lceil \frac{1}{\sigma} h \rceil}\mathbb{Z}_p} \otimes L_{\tilde{\lambda}}(\mathbb{Z}_p, \mu).$$

In the proof of 2.2 Lemma in [Mahnkopf 2014] we have seen that

$$\dim_{\mathbb{Z}_p} \bigoplus_{\substack{\mu \le \lambda \\ \mathrm{ht}_\lambda(\mu) = h}} L_\lambda(\mathbb{Z}_p, \mu) \le N_h,$$

where N_h is the number of tuples $\boldsymbol{n} = (n_1, \ldots, n_s) \in \mathbb{N}_0^s$ such that

$$\sum_{i=1}^{s} n_i \, \mathrm{ht}(\alpha_i) = h$$

(recall that $\Phi^+ = \{\alpha_1, \ldots, \alpha_s\}$). Hence,

$$\boldsymbol{L}_{\tilde{\lambda}}^{[r]}(\mathbb{Z}_p, \sigma) \le \bigoplus_{\substack{h \in \mathbb{N}_0 \\ 0 \le h \le r\sigma}} \left(\frac{\mathbb{Z}_p}{p^{r - \lceil \frac{1}{\sigma} h \rceil}\mathbb{Z}_p} \right)^{N_h}. \tag{13}$$

We select an $a \in \mathbb{N}$. The terms of the form $r - \lceil \frac{1}{\sigma} h \rceil$, $h \in \mathbb{N}_0$, which equal $r - a$ are then precisely those with $h = \sigma a - \sigma + 1, \sigma a - \sigma + 2, \ldots, \sigma a$. Thus, the term

$\dfrac{\mathbb{Z}_p}{p^{r-a}\mathbb{Z}_p}$ appears with multiplicity

$$N_{\sigma a-\sigma+1} + N_{\sigma a-\sigma+2} + \cdots + N_{\sigma a}$$

in the right-hand side of equation (13). It is easy to see that $N_h \leq (h+1)^{s-1}$, which implies that

$$N_{\sigma a-\sigma+1} + N_{\sigma a-\sigma+2} + \cdots + N_{\sigma a} \leq \sigma(\sigma a+1)^{s-1}.$$

Thus, we obtain, as desired,

$$\boldsymbol{L}^{[r]}_{\tilde{\lambda}}(\mathbb{Z}_p, \sigma) \leq \bigoplus_{\substack{a\in\mathbb{N}_0 \\ 0\leq a\leq r}} \left(\frac{\mathbb{Z}_p}{p^{r-a}\mathbb{Z}_p}\right)^{M_{\sigma,a}}. \qquad \square$$

3.5. From now on we set

$$\sigma = \max_{\alpha\in\Phi^+} \mathrm{ht}(\alpha).$$

Hence, σ only depends on $\tilde{\boldsymbol{G}}$ and $\tilde{K}_*(\sigma) = \tilde{\mathcal{I}}$. In particular, $L(\mathbb{Z}_p, r, \sigma)$ is a $\tilde{\mathcal{I}}$ and, hence, a Γ-module. The inclusion $i : L_{\tilde{\lambda}}(\mathbb{Z}_p, r, \sigma) \subseteq L_{\tilde{\lambda}}(\mathbb{Z}_p)$ induces a mapping

$$i^* : H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{Z}_p, r, \sigma)) \to H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{Z}_p)).$$

We recall that h is an element in $\tilde{\boldsymbol{T}}(\mathbb{Q})^{++}$ (see Section 2.3).

Lemma. 1. *The mapping i^* induces an injection*

$$i^* : H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{Z}_p, r, \sigma))^{\mathrm{TF}} \hookrightarrow H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{Z}_p))^{\mathrm{TF}},$$

where superscript TF denotes the maximal torsion-free quotient. In particular, we may identify $H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{Z}_p, r, \sigma))^{\mathrm{TF}}$ with its image in $H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{Z}_p))^{\mathrm{TF}}$ under i^*.

2. *Let $\zeta \in \Delta$; hence, $\zeta \in \tilde{\mathcal{I}} h^e \tilde{\mathcal{I}}$ for some $e \in \mathbb{N}_0$ and we assume that $e \in \mathbb{N}$. Then the Hecke operator $(T_\zeta)_{\tilde{\lambda}}$ induces an operator on $H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{Z}_p))^{\mathrm{TF}}$ and we obtain*

$$(T_\zeta)_{\tilde{\lambda}}\big(H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{Z}_p, r, \sigma))^{\mathrm{TF}}\big) \subseteq p^r H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{Z}_p))^{\mathrm{TF}}.$$

Proof. 1. The exact sequence

$$0 \to L_{\tilde{\lambda}}(\mathbb{Z}_p, r, \sigma) \xrightarrow{i} L_{\tilde{\lambda}}(\mathbb{Z}_p) \xrightarrow{\pi} \boldsymbol{L}^{[r]}_{\tilde{\lambda}}(\mathbb{Z}_p, \sigma) \to 0$$

yields an exact sequence

$$H^{i-1}(\Gamma, \boldsymbol{L}^{[r]}_{\tilde{\lambda}}(\mathbb{Z}_p, \sigma)) \to H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{Z}_p, r, \sigma)) \xrightarrow{i^*} H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{Z}_p)) \xrightarrow{\pi^*} H^i(\Gamma, \boldsymbol{L}^{[r]}_{\tilde{\lambda}}(\mathbb{Z}_p, \sigma)).$$

Since $H^i(\Gamma, \boldsymbol{L}^{[r]}_{\tilde\lambda}(\mathbb{Z}_p, \sigma))$ is a finite abelian group we further obtain an exact sequence

$$0 \to H^i(\Gamma, L_{\tilde\lambda}(\mathbb{Z}_p, r, \sigma))^{\mathrm{TF}} \xrightarrow{i^*} H^i(\Gamma, L_{\tilde\lambda}(\mathbb{Z}_p))^{\mathrm{TF}} \xrightarrow{\pi^*} Q \to 0, \tag{14}$$

where Q is a certain subquotient of $H^i(\Gamma, \boldsymbol{L}^{[r]}_{\tilde\lambda}(\mathbb{Z}_p, \sigma))$. Thus, i^* is injective.

2. The first claim follows from 2.9 Corollary. In equation (11) in Section 2.8 we have seen that for any $v_\mu \in L_{\tilde\lambda}(\mathbb{Z}_p, \mu)$

$$\tilde\lambda(h^e)h^{-e}v_\mu \in p^{e\,\mathrm{ht}_\lambda(\mu)} L_{\tilde\lambda}(\mathbb{Z}_p, \mu).$$

Hence, for all weights $\mu \le \lambda$ satisfying $\mathrm{ht}_\lambda(\mu) \le r\sigma$ we obtain

$$\begin{aligned}\tilde\lambda(h^e)h^{-e}p^{r-\lceil\frac{1}{\sigma}\mathrm{ht}_\lambda(\mu)\rceil}L_{\tilde\lambda}(\mathbb{Z}_p, \mu) &\subseteq p^{r-\lceil\frac{1}{\sigma}\mathrm{ht}_\lambda(\mu)\rceil+e\,\mathrm{ht}_\lambda(\mu)}L_{\tilde\lambda}(\mathbb{Z}_p, \mu)\\ &\subseteq p^{r+(e-1)\,\mathrm{ht}_\lambda(\mu)}L_{\tilde\lambda}(\mathbb{Z}_p, \mu) \subseteq p^r L_{\tilde\lambda}(\mathbb{Z}_p, \mu),\end{aligned}$$

and for all weights $\mu \le \lambda$ satisfying $\mathrm{ht}_\lambda(\mu) > r\sigma (\ge r)$ we obtain

$$\tilde\lambda(h^e)h^{-e}L_{\tilde\lambda}(\mathbb{Z}_p, \mu) \subseteq p^{e\,\mathrm{ht}_\lambda(\mu)}L_{\tilde\lambda}(\mathbb{Z}_p, \mu) \subseteq p^r L_{\tilde\lambda}(\mathbb{Z}_p, \mu).$$

Hence, we obtain $\tilde\lambda(h^e)h^{-e}L_{\tilde\lambda}(\mathbb{Z}_p, r, \sigma) \subseteq p^r L_{\tilde\lambda}(\mathbb{Z}_p)$. Since $\zeta \in \tilde{\mathcal{I}}h^e\tilde{\mathcal{I}}$ with $e \ge 1$ and $\tilde{\mathcal{I}}$ leaves $L_{\tilde\lambda}(\mathbb{Z}_p)$ and $L_{\tilde\lambda}(\mathbb{Z}_p, r, \sigma)$ invariant (see 3.3.1 Lemma) we obtain

$$\tilde\lambda(h^e)\zeta^{-1}L_{\tilde\lambda}(\mathbb{Z}_p, r, \sigma) \subseteq p^r L_{\tilde\lambda}(\mathbb{Z}_p)$$

which yields

$$(T_\zeta)_{\tilde\lambda}(C^i(\Gamma, L_{\tilde\lambda}(\mathbb{Z}_p, r, \sigma))) \subseteq p^r C^i(\Gamma, L_{\tilde\lambda}(\mathbb{Z}_p))\ (\subseteq C^i(\Gamma, L_{\tilde\lambda}(\mathbb{Z}_p)))$$

(here, we view $C^i(\Gamma, L_{\tilde\lambda}(\mathbb{Z}_p, r, \sigma))$ as embedded in $C^i(\Gamma, L_{\tilde\lambda}(\mathbb{Z}_p))$ via i^*). The last equation implies the claim. □

We note that $H^i(\Gamma, L_{\tilde\lambda}(\mathbb{Z}_p))^{\mathrm{TF}} \cong H^i(\Gamma, L_{\tilde\lambda}(\mathbb{Z}_p))_{\mathrm{int}}$.

3.6. We select a resolution of the trivial Γ-module $\mathbb{Z}$,

$$0 \to M_d \to \cdots \to M_1 \to M_0 \to \mathbb{Z} \to 0,$$

where M_i is a free $\mathbb{Z}\Gamma$-module of finite rank (see [Brown 1982, p. 199]; note that Γ is of type FL; see p. 218 in the same work). The groups $H^i(\Gamma, \boldsymbol{L}^{[r]}_{\tilde\lambda}(\mathbb{Z}_p, \sigma))$ then may be computed as the cohomology of the complex

$$0 \to \mathrm{Hom}_{\mathbb{Z}_p\Gamma}(M_{0,p}, \boldsymbol{L}^{[r]}_{\tilde\lambda}(\mathbb{Z}_p, \sigma)) \to \cdots \to \mathrm{Hom}_{\mathbb{Z}_p\Gamma}(M_{d,p}, \boldsymbol{L}^{[r]}_{\tilde\lambda}(\mathbb{Z}_p, \sigma)) \to 0$$

where $M_{i,p} = \mathbb{Z}_p \otimes M_i$. We set

$$g_i = g_{i,\Gamma} = \mathrm{rk}_{\mathbb{Z}\Gamma} M_i.$$

Thus, g_i depends on i and the arithmetic group Γ.

3.7. ***Borel–Serre compactification.*** Following [Borel and Serre 1973] (see also [Brown 1982, pp. 14 and 218]) we can construct a finite free resolution $(M_i)_{d\geq i\geq 0}$ of $\mathbb{Z}$ as follows. We denote by $Y=\Gamma\backslash\bar{X}$ the Borel–Serre compactification of the locally symmetric space $\Gamma\backslash X$ attached to $\tilde{\boldsymbol{G}}/\mathbb{Q}$. By the work of Borel and Serre Y is a compact $K(\Gamma,1)$ space; hence, there is a *finite* CW complex $\mathcal{Z}$ having the same homotopy type as Y. The universal cover $\tilde{Y}$ of Y inherits a structure of CW complex $\tilde{\mathcal{Z}}$ from $\mathcal{Z}$ and the cellular complex $\tilde{C}_\bullet=(\tilde{C}_i)_{d\geq i\geq 0}$ ($d=\dim X$) attached to $\tilde{\mathcal{Z}}$ is a complex consisting of free $\mathbb{Z}\Gamma$-modules $\tilde{C}_i$. The module $\tilde{C}_i$ has a natural $\mathbb{Z}\Gamma$-basis which is in bijection with the set of i-cells of $\mathcal{Z}$ (see [Brown 1982, p. 15]); hence, $\tilde{C}_\bullet$ is a finite complex consisting of free, finitely generated $\mathbb{Z}\Gamma$-modules. Since $\tilde{Y}$ is contractible, its cohomology vanishes except in degree 0; hence, the complex

$$0\to\tilde{C}_d\to\cdots\to\tilde{C}_0\to\mathbb{Z}\to 0$$

is exact and thus a resolution of $\mathbb{Z}$. In particular, we may select $M_i=\tilde{C}_i$ and since $\mathrm{rk}_{\mathbb{Z}\Gamma}\,\tilde{C}_i$ equals the number of i-cells of $\mathcal{Z}$ this shows that we can take for g_i the number of i-cells of a CW complex $\mathcal{Z}$ which has the same homotopy type as $\Gamma\backslash\bar{X}$.

3.8. ***Slope subspaces.*** We select a Hecke operator $T\in\mathcal{H}_{\mathbb{Z}_p}$. Let $\tilde{\lambda}\in X(\tilde{\boldsymbol{T}})^{\mathrm{dom}}$ and let $E/\mathbb{Q}_p$ be an extension which is contained in $\bar{\mathbb{Q}}_p$. For any $\beta\in\mathbb{Q}_{\geq 0}$ we denote by

$$H^i(\Gamma,L_{\tilde{\lambda}}(E))^\beta=p_\beta(T_{\tilde{\lambda}})H^i(\Gamma,L_{\tilde{\lambda}}(E))$$

the slope β subspace of $H^i(\Gamma,L_{\tilde{\lambda}}(E))$ w.r.t. to the (normalized) Hecke operator $T_{\tilde{\lambda}}$. Here, $p(X)\in\mathbb{Z}_p[X]$ is the characteristic polynomial of $T_{\tilde{\lambda}}$ acting on $H^i(\Gamma,L_{\tilde{\lambda}}(E))$ and $p_\beta(X)=\prod_{\mu,\,v_p(\mu)\neq\beta}(X-\mu)\in\mathbb{Z}_p[X]$, where μ runs over all roots of $p(X)$ whose p-adic value is different from β. Thus, we obtain $H^i(\Gamma,L_{\tilde{\lambda}}(\bar{\mathbb{Q}}_p))^\beta=\bigoplus_{\mu\in\mathcal{O}_{\bar{\mathbb{Q}}_p},\,v_p(\mu)=\beta}H^i(\Gamma,L_{\tilde{\lambda}}(\bar{\mathbb{Q}}_p))(\mu)$ where $H^i(\Gamma,L_{\tilde{\lambda}}(\bar{\mathbb{Q}}_p))(\mu)$ is the generalized eigenspace attached to the eigenvalue μ. We set

$$H^i(\Gamma,L_{\tilde{\lambda}}(\mathbb{Q}_p))^{\leq\beta}=\bigoplus_{0\leq\gamma\leq\beta}H^i(\Gamma,L_{\tilde{\lambda}}(\mathbb{Q}_p))^\gamma$$

and we denote by $H^i(\Gamma,L_{\tilde{\lambda}}(\mathbb{Q}_p))^{<\infty}=\bigoplus\limits_{0\leq\gamma<\infty}H^i(\Gamma,L_{\tilde{\lambda}}(\mathbb{Q}_p))^\gamma$ the finite slope subspace.

3.9. ***An estimate for the Newton polygon.*** We denote by

$$\mathcal{H}^{\mathrm{reg}}_{\mathbb{Z}_p}\subseteq\mathcal{H}_{\mathbb{Z}_p}$$

the set of all Hecke operators $T=\sum_\zeta c_\zeta T_\zeta\in\mathcal{H}_{\mathbb{Z}_p}$ ($\zeta\in\Delta$, $c_\zeta\in\mathbb{Z}_p$) where $\zeta\in\tilde{\mathcal{I}}h^{e_\zeta}\tilde{\mathcal{I}}$ with $e_\zeta\geq 1$ for all ζ with $c_\zeta\neq 0$. We let $T\in\mathcal{H}^{\mathrm{reg}}_{\mathbb{Z}_p}$. We set $t'=t'(\tilde{\lambda},i)=$

$\dim H^i(\Gamma, L_{\tilde\lambda}(\mathbb{Q}_p))^{<\infty}$ and we denote by $p'(X) = \sum_{i=0}^{t'} a_i X^{t'-i} \in \mathbb{Z}_p[X]$ the characteristic polynomial and by

$$\mathcal{N}^{<\infty} = \mathcal{N}^{<\infty}_{\tilde\lambda,i} : [0, t'] \to \mathbb{R}_{\geq 0}$$

the Newton polygon of $T_{\tilde\lambda}$ acting on $H^i(\Gamma, L_{\tilde\lambda}(\mathbb{Q}_p))^{<\infty}$ which contains the $T_{\tilde\lambda}$-invariant lattice $H^i(\Gamma, L_{\tilde\lambda}(\mathbb{Z}_p))^{<\infty}_{\mathrm{int}} = H^i(\Gamma, L_{\tilde\lambda}(\mathbb{Q}_p))^{<\infty} \cap H^i(\Gamma, L_{\tilde\lambda}(\mathbb{Z}_p))_{\mathrm{int}}$. Thus, $\mathcal{N}^{<\infty}$ is the lower convex hull of the points $(i, v_p(a_i))$, $i = 0, \dots, t'$, where we omit all points with $a_i = 0$ (note that $p'(0) \neq 0$, hence, $a_{t'} \neq 0$). We recall that $g_i = \mathrm{rk}_{\mathbb{Z}\Gamma} M_i$ (see Sections 3.6 and 3.7) and that $B_s \in \mathbb{Q}[X]$ denotes the s-th Bernoulli polynomial.

Theorem. *For all dominant weights $\tilde\lambda \in X(\tilde{\boldsymbol{T}}))$ and all $i \in \mathbb{N}_0$ the Newton polygon $\mathcal{N}^{<\infty} = \mathcal{N}^{<\infty}_{\tilde\lambda,i}$ lies above the restriction to $[0, t']$ of the piecewise linear function $f^*_\infty = f^*_{i,\infty} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ which connects the points $(0, 0)$ and*

$$P_j = \left(g_i \sigma \frac{B_s(\sigma(j+1)+1) - B_s(1)}{s}, \, g_i \sigma^{s+1} \frac{B_{s+1}(j+1) - B_{s+1}(1)}{s+1} \right),$$

where $j = 0, 1, 2, \dots$.

Proof. We proceed in steps.

3.9.1. We let $\tilde\lambda \in X(\tilde{\boldsymbol{T}})^{\mathrm{dom}}$ and set $\lambda = \mathrm{d}\,\tilde\lambda|_{\boldsymbol{T}} \in \mathfrak{h}^*$. Moreover, we select a natural number $r \in \mathbb{N}$. Since $\sigma = \max_{\alpha \in \Phi^+} \mathrm{ht}(\alpha)$ the $\mathbb{Z}_p$-module $\boldsymbol{L}^{[r]}_{\tilde\lambda}(\mathbb{Z}_p, \sigma)$ is a Γ-module and by 3.4 Lemma we know that

$$\boldsymbol{L}^{[r]}_{\tilde\lambda}(\mathbb{Z}_p, \sigma) \leq \bigoplus_{h=0}^{r} (\mathbb{Z}_p/p^{r-h}\mathbb{Z}_p)^{M_{\sigma,h}},$$

which implies that

$$\mathrm{Hom}_{\mathbb{Z}_p\Gamma}(M_{i,p}, \boldsymbol{L}^{[r]}_{\tilde\lambda}(\mathbb{Z}_p, \sigma)) \leq \bigoplus_{h=0}^{r} (\mathbb{Z}_p/p^{r-h}\mathbb{Z}_p)^{g_i M_{\sigma,h}}. \tag{15}$$

We denote by $(p^{a_l})_l$, $a_1 \geq a_2 \geq \cdots \geq a_n > 0$, the sequence of elementary divisors of the right-hand side of equation (15), i.e.,

$$(p^{a_l})_{l=1,\dots,n} = (p^r, \dots, p^r, p^{r-1}, \dots, p^{r-1}, \dots, p, \dots, p), \tag{16}$$

where p^{r-h} appears $g_i M_{\sigma,h}$-times. From 3.5 Lemma it follows that there is a natural embedding of $H^i(\Gamma, L_{\tilde\lambda}(\mathbb{Z}_p, r, \sigma))^{\mathrm{TF}}$ in $H^i(\Gamma, L_{\tilde\lambda}(\mathbb{Z}_p))^{\mathrm{TF}}$ and the exact sequence in equation (14) shows that

$$\frac{H^i(\Gamma, L_{\tilde\lambda}(\mathbb{Z}_p))^{\mathrm{TF}}}{H^i(\Gamma, L_{\tilde\lambda}(\mathbb{Z}_p, r, \sigma))^{\mathrm{TF}}} \quad \text{is a subquotient of } H^i(\Gamma, \boldsymbol{L}^{[r]}_{\tilde\lambda}(\mathbb{Z}_p, \sigma)). \tag{17}$$

We denote by t the rank of $H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{Z}_p))^{\mathrm{TF}}$ and by $(p^{b_l})_l$, $b_1 \geq b_2 \geq \cdots \geq b_m > 0$ ($m \leq t$) the sequence of elementary divisors of the quotient on the left in (17). Equation (17) implies that this quotient is a subquotient of the Hom space on the left-hand side of equation (15); hence, it is a subquotient of the right-hand side of (15) and equation (16) yields $m \leq n$ and

$$b_1 \leq a_1, b_2 \leq a_2, \ldots, b_m \leq a_m. \tag{18}$$

We set $b_l = 0$ for $m < l \leq t$ and $a_l = 0$ for $n < l \leq t$ (if $n < t$); hence, $b_i \leq a_i$ for $i = 1, \ldots, t$.

3.9.2. Using the results so far we can give a lower bound for $\mathcal{N}^{<\infty}$. Equations (16) and (18) imply that the b_l are all smaller than or equal to r. Moreover, since $T = \sum_\zeta c_\zeta T_\zeta$ where $c_\zeta \in \mathbb{Z}_p$ and $\zeta \in \tilde{\mathcal{I}} h^{e_\zeta} \tilde{\mathcal{I}}$ with $e_\zeta \geq 1$ if $c_\zeta \neq 0$ 3.5 Lemma implies that

$$T_{\tilde{\lambda}}(H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{Z}_p, r, \sigma))^{\mathrm{TF}}) \subseteq p^r H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{Z}_p))^{\mathrm{TF}}.$$

Thus, we may apply Lemma 1 in [Buzzard 2001], as recalled in Section 1.7 of [Mahnkopf 2014], to the pair $L = H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{Z}_p))^{\mathrm{TF}}$ and $K = H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{Z}_p, r, \sigma))^{\mathrm{TF}}$ and the operator $\xi = T_{\tilde{\lambda}}$. More precisely, we denote by $f_b = f_{b,r} : [0, t'] \to \mathbb{R}_{\geq 0}$ the piecewise linear function attached to the sequence $(b_1, \ldots, b_t)$; i.e., f_b is the piecewise linear function joining the points $(j, C(j))$, $j = 0, \ldots, t'$, where $C(j) = \sum_{l=1}^{j} (r - b_l)$. Then (1.7) Lemma in [Mahnkopf 2014] states that the Newton polygon $\mathcal{N}^{<\infty}$ of $T_{\tilde{\lambda}}$ acting on

$$\left(H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{Z}_p))^{\mathrm{TF}} \otimes \mathbb{Q}_p\right)^{<\infty} = H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{Q}_p))^{<\infty}$$

is bounded from below by the graph of f_b.

3.9.3. We further estimate the function f_b. Equation (18) implies that f_b lies above the piecewise linear function $f_{a,r} : [0, t'] \to \mathbb{R}_{\geq 0}$ attached to the sequence $(a_1, \ldots, a_{t'})$, i.e., $f_{a,r}$ joins the points $(j, A(j))$, $j = 0, \ldots, t'$, where $A(j) = \sum_{l=1}^{j} r - a_l$ (the $A(j)$'s are equal to or smaller than the $C(j)$'s). Thus, we have

$$\mathcal{N}^{<\infty} \geq f_{b,r} \geq f_{a,r} \tag{19}$$

and this inequality holds for all $r \in \mathbb{N}$ since r was chosen arbitrarily. Using equation (16) it is not difficult to see that the function $f_{a,r}$ is the restriction to $[0, t']$ of the piecewise linear function on $\mathbb{R}_{\geq 0}$ which starts in $(0, 0)$ and has slope j for $g_i \sum_{h=0}^{j-1} M_{\sigma,h} \leq x \leq g_i \sum_{h=0}^{j} M_{\sigma,h}$, $j = 0, \ldots, r-1$, and slope r for $x \geq g_i \sum_{h=0}^{r-1} M_{\sigma,h}$. Since $M_{\sigma,h} = \sigma(\sigma h + 1)^{s-1}$, $h \geq 0$, we see that the function $f_{a,r}$ may be equivalently described as the piecewise linear function $f_{a,r} : \mathbb{R}_{\geq 0} \to \mathbb{R}$

which starts in (0, 0), has slope j for

$$g_i\sigma\sum_{h=0}^{j-1}(\sigma h+1)^{s-1} \le x \le g_i\sigma\sum_{h=0}^{j}(\sigma h+1)^{s-1}, \quad j=0,\dots,r-1, \tag{20}$$

and slope r for $x \ge g_i\sigma\sum_{h=0}^{r-1}(\sigma h+1)^{s-1}$. We set

$$x_s(j) = g_i\sigma\frac{B_s(\sigma(j+1)+1)-B_s(1)}{s}, \quad j=0,1,2,\dots,$$

and we denote by $f_\infty : \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ the piecewise linear function which starts in (0, 0) and has slope 0 in the interval

$$0 \le x \le x_s(0)$$

and slope j in the interval

$$x_s(j-1) \le x \le x_s(j), \quad j=1,2,\dots$$

The function f_∞ like the function $f_{a,r}$ is monotonely increasing. Taking into account that for all $j \in \mathbb{N}_0$

$$\sum_{h=0}^{\sigma j-1}(h+1)^{s-1} = \sum_{h=1}^{\sigma j}h^{s-1} = \begin{cases} \dfrac{B_s(\sigma j+1)-B_s(1)}{s} & \text{if } j \ge 1,\\ 0 & \text{if } j=0, \end{cases} \tag{21}$$

we deduce that

$$x_s(0) = g_i\sigma\frac{B_s(\sigma+1)-B_s(1)}{s} \ge g_i\sigma 1^{s-1} = g_i\sigma$$

and for all $j \in \mathbb{N}$

$$\begin{aligned} x_s(j)-x_s(j-1) &= g_i\sigma\sum_{h=\sigma j}^{\sigma(j+1)-1}(h+1)^{s-1}\\ &\ge g_i\sigma(\sigma j+1)^{s-1} = g_i\sigma\left(\sum_{h=0}^{j}(\sigma h+1)^{s-1} - \sum_{h=0}^{j-1}(\sigma h+1)^{s-1}\right). \end{aligned}$$

Thus, equation (20) implies that the segments of slope $0,\dots,r-1$ of f_∞ are longer than those of $f_{a,r}$, hence, $f_{a,r}(x) \ge f_\infty(x)$ for all $x \in [0, x_s(r-1)]$. This implies by equation (19) that

$$\mathcal{N}^{<\infty}(x) \ge f_\infty(x)$$

for all $x \in [0, x_s(r-1)]$. Since equation (19) holds for arbitrarily large $r \in \mathbb{N}$, and since $x_s(r-1) \to \infty$ for $r \to \infty$ by equation (21), we finally obtain

$$\mathcal{N}^{<\infty}(x) \ge f_\infty(x), \quad x \in [0,\infty). \tag{22}$$

3.9.4. We show that $f_\infty \geq f_\infty^*$. In view of equation (22) this completes the proof. By definition f_∞ is the piecewise linear function joining the points

$$(0,0), \quad Q_j = \left(x_s(j), \sum_{h=1}^{j} h(x_s(h) - x_s(h-1)) \right), \quad j = 0, 1, 2, 3, \dots$$

We obtain the following estimate for the second coordinate $\mathtt{y}_s(j)$ of Q_j, i.e., for the value $f_\infty(x_s(j))$:

$$\mathtt{y}_s(0) = 0 = f_\infty^*(x_s(0))$$

and if $j \geq 1$ then equation (21) implies that

$$\begin{aligned}
\mathtt{y}_s(j) &= \sum_{h=1}^{j} h(x_s(h) - x_s(h-1)) \\
&= \sum_{h=1}^{j} h g_i \sigma \left(\frac{B_s(\sigma(h+1)+1) - B_s(1)}{s} - \frac{B_s(\sigma h+1) - B_s(1)}{s} \right) \\
&\overset{(21)}{=} g_i \sigma \sum_{h=1}^{j} h \sum_{k=\sigma h}^{\sigma(h+1)-1} (k+1)^{s-1} \geq g_i \sigma \sum_{h=1}^{j} h \sum_{k=\sigma h}^{\sigma h+\sigma-1} k^{s-1} \\
&\geq g_i \sigma \sum_{h=1}^{j} h\, \sigma (\sigma h)^{s-1} = g_i \sigma^{s+1} \sum_{h=1}^{j} h^s \\
&\overset{(21)}{=} g_i \sigma^{s+1} \frac{B_{s+1}(j+1) - B_{s+1}(1)}{s+1} = f_\infty^*(x_s(j)).
\end{aligned}$$

Thus, $f_\infty \geq f_\infty^*$ and the theorem is proven. $\square$

3.10. ***A bound for the dimension of slope subspaces.*** We recall that $s = |\Phi^+|$, $\sigma = \max_{\alpha \in \Phi^+} \mathrm{ht}(\alpha)$ and g_i is the number of i-cells in a cell complex $\tilde{\mathcal{Z}}$ which is homotopy equivalent to $\Gamma \backslash \bar{X}$.

Corollary. *For all $\beta \in \mathbb{Q}_{\geq 0}$, all dominant weights $\tilde{\lambda} \in X(\tilde{\boldsymbol{T}})$, all i and all Hecke operators $T \in \mathcal{H}_{\mathbb{Z}_p}^{\mathrm{reg}}$ we have*

$$\dim H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{Q}_p))^{\leq \beta} \leq \mathtt{m} \beta^s + \mathtt{n};$$

here, $\mathtt{m} = \mathtt{m}_\Gamma = 12(g_i/s)\sigma^{s+1} \in \mathbb{Q}_{\geq 0}$ *and* $\mathtt{n} = \mathtt{n}_\Gamma \in \mathbb{N}$ *is an integer which also only depends on* g_i, σ, s (see (26) below); *in particular*, $\mathtt{m}$ *and* $\mathtt{n}$ *only depend on* Γ (*and so on* $\tilde{\boldsymbol{G}}$ *and* p) *and* i, *but not on* $\tilde{\lambda}$, h *and* T.

Proof. Let $h : \mathbb{R}_{\geq 0} \to \mathbb{R}$ be any function such that $f_\infty^*(x) \geq h(x)$ for all $x \geq 0$ and let $(d(\epsilon), y)$ with $d(\epsilon) > 0$ be an intersection point of h and the function

$w_\epsilon : x \mapsto (\beta+\epsilon)x$ $(\epsilon > 0)$. Since

$$(\beta+\epsilon)x > \mathcal{N}^{<\infty}(x) \geq h(x)$$

for all $x \in [0, \dim H^i(\Gamma, L_{\tilde\lambda}(\mathbb{Q}_p))^{\leq\beta}]$ by 3.9 Theorem we deduce that $d(\epsilon) \geq \dim H^i(\Gamma, L_{\tilde\lambda}(\mathbb{Q}_p))^{\leq\beta}$; hence, we obtain an upper bound for the dimension of the slope $\leq \beta$-subspace. We explicitly define a lower bound h for f^*_∞ as follows. Since B_s is a polynomial of degree s and leading coefficient 1 there is a natural number $M = M(\sigma, s) \in \mathbb{N}$ such that

$$x_s(j) = g_i\sigma\frac{B_s(\sigma(j+1)+1) - B_s(1)}{s} \leq 2^{\frac{1}{s+1}}\frac{g_i\sigma^{s+1}}{s}j^s \tag{23}$$

and

$$y_s(j) := g_i\sigma^{s+1}\frac{B_{s+1}(j+1) - B_{s+1}(1)}{s+1} \geq 2^{-\frac{1}{s}}\frac{g_i\sigma^{s+1}}{s+1}j^{s+1} \tag{24}$$

for all $j \geq M$. We define the function

$$h : [x_s(M), \infty) \to \mathbb{R}_{\geq 0}, \quad x \mapsto cx^{\frac{s+1}{s}},$$

where $c = 4^{-\frac{1}{s}}g_i^{-\frac{1}{s}}s^{\frac{s+1}{s}}\frac{1}{(s+1)}\sigma^{-\frac{s+1}{s}}$. We note that $x_s(M) \geq 0$ by equation (21). We then obtain for all $j \geq M$

$$\begin{aligned} h(x_s(j)) = h\left(g_i\sigma\frac{B_s(\sigma(j+1)+1) - B_s(1)}{s}\right) &\overset{(23)}{\leq} c\Big(2^{\frac{1}{s+1}}\frac{g_i\sigma^{s+1}}{s}j^s\Big)^{\frac{s+1}{s}} \\ &= c\,2^{\frac{1}{s}}\Big(\frac{g_i}{s}\Big)^{\frac{s+1}{s}}\sigma^{\frac{(s+1)^2}{s}}j^{s+1} \\ &= 2^{-\frac{1}{s}}\frac{g_i\sigma^{s+1}}{s+1}j^{s+1} \overset{(24)}{\leq} y_s(j). \end{aligned}$$

Since f^*_∞ is the piecewise linear function connecting the points $P_j = (x_s(j), y_s(j))$, $j \in \mathbb{N}_0$ and $(0,0)$, and since h passes below the points P_j, $j \geq M$, and is convex this implies that $h(x) \leq f^*_\infty(x)$ for all $x \geq x_s(M)$. We extend h to a function $h : \mathbb{R}_{\geq 0} \to \mathbb{R}$ by setting $h(x) = f^*_\infty(x)$ for $x \in [0, x_s(M)]$ and $h(x) = cx^{\frac{s+1}{s}}$ if $x > x_s(M)$, hence, $f^*_\infty(x) \geq h(x)$ for all $x \in [0, \infty)$. As in the proof of 3.3 Corollary in [Mahnkopf 2014] we see that for all $\epsilon > 0$ the functions h and $x \mapsto (\beta+\epsilon)x$ always intersect in a point $(d(\epsilon), y)$ with $d(\epsilon) > 0$ and this point satisfies $d(\epsilon) \leq \max((\frac{\beta+\epsilon}{c})^s, x_s(M))$. Since $(\beta/c)^s, x_s(M) \geq 0$ are positive we obtain

$$\dim H^i(\Gamma, L_{\tilde\lambda}(\mathbb{Q}_p))^{\leq\beta} \leq \max((\beta/c)^s, x_s(M)) \leq (\beta/c)^s + x_s(M).$$

Since further

$$c^{-s} = 4g_is^{-(s+1)}(s+1)^s\sigma^{s+1} = 4g_is^{-1}\Big(1+\frac{1}{s}\Big)^s\sigma^{s+1} \leq 4\frac{g_i}{s}\sigma^{s+1}\exp(1) \tag{25}$$

and $\exp(1) \le 3$ the claim of the corollary holds with $\mathtt{m} = 12(g_i/s)\sigma^{s+1}$ and $\mathtt{n} = x_s(M)$. The claim still holds if we replace $\mathtt{n}$ with any larger number and since equation (23) implies that $x_s(M) \le 2^{\frac{1}{s+1}}(g_i\sigma^{s+1}/s)M^s$ the corollary in particular holds with

$$\mathtt{n} = \left\lceil 2^{\frac{1}{s+1}} \frac{g_i\sigma^{s+1}}{s} M^s \right\rceil + 1 \in \mathbb{N}. \tag{26}$$

□

4. Mod p^n reduction of traces of Hecke operators

4.1. In this section we will prove congruences between traces of powers of normalized Hecke operators on cuspidal cohomology for varying weight $\tilde{\lambda}$. Our main tool will be a comparison of Bewersdorff's elementary trace formula for pairs $\tilde{\lambda}$, $\tilde{\lambda}'$ of congruent weights. The equality of mod p^n reductions of geometric sides follows from p-adic properties of the diagonalization of elements in $\tilde{\mathcal{I}}h^e\tilde{\mathcal{I}} \subseteq \tilde{\boldsymbol{G}}(\mathbb{Q}_p)$; see 4.4 Proposition (note that the Hecke operator $\Gamma\zeta\Gamma$, $\zeta \in \Delta$, is contained in $\tilde{\mathcal{I}}h^e\tilde{\mathcal{I}}$ for some $e \in \mathbb{N}_0$). In particular, the comparison is elementary and does not make use of advanced methods such as rigid analytic geometry or p-adic Banach space methods such as overconvergent cohomology. Using an adelic setting we prove analogous congruences on the Eisenstein part of cohomology and subtracting from full cohomology we obtain congrences congruences on cuspidal cohomology (Sections 4.11–4.13).

In Section 4.14 we compare two Goresky–MacPherson trace formulas for two congruent weights. Equality of mod p^n reductions of the geometric sides again follows from the same diagonalization of elements in $\tilde{\mathcal{I}}h^e\tilde{\mathcal{I}} \subseteq \tilde{\boldsymbol{G}}(\mathbb{Q}_p)$ but now applied for all Levi subgroups $\tilde{\boldsymbol{M}}$ of $\mathbb{Q}$-parabolic subgroups of $\tilde{\boldsymbol{G}}$. This yields congruences on weighted cohomology groups and also has an application to a more explicit version of the Gouvêa–Mazur conjecture for symplectic groups of rank 2 (see Section 5.8).

As before, $\tilde{\boldsymbol{G}}/\mathbb{Q}$ is a connected reductive group containing a $\mathbb{Q}_p$-split maximal torus $\tilde{\boldsymbol{T}}/\mathbb{Q}$ and $\Gamma \subseteq \tilde{\boldsymbol{G}}(\mathbb{Q})$ is an arithmetic subgroup satisfying $\Gamma \subseteq \tilde{\mathcal{I}}$.

4.2. ***The fixed point principle of Bewersdorff.*** As in Section 2.6 we denote by $X = \tilde{\boldsymbol{G}}(\mathbb{R})/\tilde{K}_\infty A_{\tilde{\boldsymbol{G}}}$ the symmetric space attached to $\tilde{\boldsymbol{G}}$. The Γ-module $L_{\tilde{\lambda}}(\mathbb{Q}_p)$ defines a locally constant sheaf on the locally symmetric space $\Gamma\backslash X$ and its Borel–Serre compactification $\Gamma\backslash\tilde{X}$, which we will also denote by $L_{\tilde{\lambda}}(\mathbb{Q}_p)$, and the Hecke algebra $\mathcal{H}$ acts on the cohomology groups

$$H^i(\Gamma, L_{\tilde{\lambda}}(\mathbb{Q}_p)) \cong H^i(\Gamma\backslash\bar{X}, L_{\tilde{\lambda}}(\mathbb{Q}_p)).$$

We recall that in Section 2.5 we selected a finite extension $F/\mathbb{Q}$ which splits $\tilde{\boldsymbol{G}}$ and which embeds in $\mathbb{C}$, $\mathbb{Q}_p$; in particular, $(\rho_{\tilde{\lambda}}, L_{\tilde{\lambda}})$ is defined over F. We write

$\Gamma\zeta\Gamma/\sim_\Gamma$ for the set of Γ-conjugacy classes contained in the double coset $\Gamma\zeta\Gamma$, $\zeta \in \Delta$, and $[\xi]_\Gamma$ denotes the Γ-conjugacy class of $\xi \in \tilde{\boldsymbol{G}}(\mathbb{Q})$. We now borrow from [Bewersdorff 1985, Satz 2.6] a simple and elementary formula for the Lefschetz number of Hecke correspondences on full cohomology:

Theorem (Bewersdorff). *Let $\Gamma\zeta\Gamma \in \mathcal{H}$ (i.e., $\zeta \in \Delta$). There are rational integers $c_{[\xi]_\Gamma} \in \mathbb{Z}$, $[\xi]_\Gamma \in \Gamma\zeta\Gamma/\sim_\Gamma$, such that for all dominant weights $\tilde{\lambda} \in X(\tilde{\boldsymbol{T}})$*

$$\operatorname{Lef}(\Gamma\zeta\Gamma | H^\bullet(\Gamma\backslash X, L_{\tilde{\lambda}}(F))) = \sum_{[\xi]_\Gamma \in \Gamma\zeta\Gamma/\sim_\Gamma} c_{[\xi]_\Gamma} \operatorname{tr}(\xi^{-1} | L_{\tilde{\lambda}}(F)) \tag{27}$$

and $c_{[\xi]_\Gamma}$ vanishes if $\xi x \neq x$ for all $x \in \bar{X}$.

Remark. 1. The integers $c_{[\xi]_\Gamma}$ do not depend on the weight $\tilde{\lambda}$.

2. The trace formula (27) is of an elementary nature. Apart from the existence of a nice compactification of the locally symmetric space $\Gamma\backslash X$ (the Borel–Serre compactification) its proof is a direct application of the Lefschetz fixed point principle which is a general and basic principle of algebraic topology.

4.3. The following lemma will be applied in the proof of 4.4 Proposition, where representations $\tilde{\boldsymbol{G}}/\mathbb{Q}_p \hookrightarrow \mathbf{GL}_m/\mathbb{Q}_p$ of $\tilde{\boldsymbol{G}}$ as matrix group are used.

Lemma. *Let $\beta = (\beta_{ij}) \in \mathcal{J}_m$ and let $t = \operatorname{diag}(t_1, \dots, t_m) \in \boldsymbol{T}_m(\mathbb{Q}_p)$ with $v_p(t_1) > v_p(t_i)$ for all $i = 2, \dots, m$. Then the characteristic polynomial $\operatorname{ch}_{\beta t}$ of $\beta t \in \mathbf{GL}_m(\mathbb{Q}_p)$ has m roots $t'_1, t'_2, \dots, t'_m$ in $\bar{\mathbb{Q}}_p$ (roots appearing several times according their multiplicity) such that $v_p(t'_1) > v_p(t'_i)$ for all $i = 2, \dots, m$ (in particular, t'_1 has multiplicity 1) and*

$$v_p(t'_1) = v_p(t_1) \quad \textit{and} \quad t'_1 \equiv \beta_{11} t_1 \pmod{p^{v_p(t_1)+1} \mathcal{O}_{\bar{\mathbb{Q}}_p}}.$$

Proof. We put $[a, b] = \{a, a+1, a+2, \dots, b\}$ ($a, b \in \mathbb{N}$, $a \le b$) and we denote by S_M the symmetric group on the set M. We write the characteristic polynomial of βt as $\operatorname{ch}_{\beta t}(X) = (-1)^m X^m + (-1)^{m-1} c_1 X^{m-1} + \dots - c_{m-1} X + c_m$, $c_i \in \mathbb{Q}_p$ (i.e., $c_0 = 1$). The Leibniz formula

$$\operatorname{ch}_{\beta t}(X) = \det(\beta t - X\mathbf{1}) = \sum_{\pi \in S_{[1,m]}} \operatorname{sgn}(\pi) \prod_{i=1}^m (\beta_{\pi(i),i} t_i - \delta_{\pi(i),i} X)$$

yields

$$c_i = \sum_{\substack{T \subseteq [1,m] \\ |T| = i}} \sum_{\pi \in S_T} c_{T,\pi}$$

for all $i = 1, \dots, m$, where

$$c_{T,\pi} = \operatorname{sgn}(\pi) \prod_{h \in T} \beta_{\pi(h),h} t_h \in \mathbb{Q}_p.$$

Since

$$v_p(\beta_{g,h}t_h)\begin{cases}\geq v_p(t_h)+1 & \text{if } g<h,\\ =v_p(t_h) & \text{if } g=h,\\ \geq v_p(t_h) & \text{if } g>h,\end{cases}\tag{28}$$

we obtain $v_p(c_{T,\pi})\geq v_p(c_{T,\mathrm{id}})+1$ for all $T\subseteq[1,m]$ and all $\pi\in S_T$, $\pi\neq\mathrm{id}$. Hence,

$$c_m=c_{[1,m],\mathrm{id}}+\text{terms with } p\text{-adic value equal to or greater than } v_p(c_{[1,m],\mathrm{id}})+1,$$

which implies that

$$c_m\equiv\prod_{h=1}^m\beta_{hh}t_h \pmod{p^{v_p(c_m)+1}\mathbb{Z}_p}\quad\text{and}\quad v_p(c_m)=v_p\biggl(\prod_{h=1}^m t_h\biggr).\tag{29}$$

Since, moreover, $v_p(t_1)\geq v_p(t_i)+1$ for all $i=2,\dots,m$ we obtain from equation (28) that

- for any subset $T\subset[1,m]$, $T\neq[2,m]$, of cardinality $m-1$ and any $\pi\in S_T$ we have $v_p(c_{T,\pi})\geq v_p(c_{T,\mathrm{id}})\geq v_p(c_{[2,m],\mathrm{id}})+1$
- for any $\pi\in S_{[2,m]}$, $\pi\neq\mathrm{id}$, we have $v_p(c_{[2,m],\pi})\geq v_p(c_{[2,m],\mathrm{id}})+1$.

Hence,

$$c_{m-1}=c_{[2,m],\mathrm{id}}+\text{terms with } p\text{-adic value equal to or greater than } v_p(c_{[2,m],\mathrm{id}})+1,$$

which implies that

$$c_{m-1}\equiv\prod_{h=2}^m\beta_{hh}t_h \pmod{p^{v_p(c_{m-1})+1}\mathbb{Z}_p}\quad\text{and}\quad v_p(c_{m-1})=v_p\biggl(\prod_{h=2}^m t_h\biggr).\tag{30}$$

In particular,

$$v_p(c_m)-v_p(c_{m-1})=v_p(t_1).\tag{31}$$

Finally, for $i=1,\dots,m-1$ we denote by $T_{i,\min}\subset[1,m]$ a subset of cardinality i such that $v_p\bigl(\prod_{h\in T_{i,\min}}t_h\bigr)\leq v_p\bigl(\prod_{h\in T}t_h\bigr)$ for all subsets $T\subseteq[1,m]$ of cardinality i (thus, $T_{m-1,\min}=[2,m]$). As above equation (28) implies

$$v_p(c_i)\geq v_p\biggl(\prod_{h\in T_{i,\min}}t_h\biggr)=\sum_{h\in T_{i,\min}}v_p(t_h)\tag{32}$$

for all $i=1,\dots,m-1$. Since $v_p(t_1)>v_p(t_h)$ for all $h=2,\dots,m$ we obtain $T_{i,\min}\subseteq[2,m]$ and equations (30) and (32) imply that

$$v_p(c_{m-1})-v_p(c_i)\leq\sum_{h\in[2,m]-T_{i,\min}}v_p(t_h)\leq(m-1-i)r$$

for all $i = 1, \ldots, m-1$, where $r = \max_{h=2}^m v_p(t_h)$. Equivalently,

$$v_p(c_i) \geq v_p(c_{m-1}) - (m-1-i)r \tag{33}$$

for all $i = 1, \ldots, m-1$. Since $(m-1)r \geq \sum_{h=2}^m v_p(t_h) = v_p(c_{m-1})$ in view of equation (30), we see that equation (33) also holds for $i = 0$ ($v_p(c_0) = 0$). Thus:

- The line connecting the points $(m-1, v_p(c_{m-1}))$ and $(m, v_p(c_m))$ has slope $v_p(t_1)$; see equation (31).
- All points $(i, v_p(c_i))$ with $0 \leq i \leq m-1$ lie on or above the line g, which has slope r and passes through $(m-1, v_p(c_{m-1}))$; see equation (33).

Since r is strictly smaller than $v_p(t_1)$ this shows that the Newton polygon $\mathcal{N}$ of $\mathrm{ch}_{\beta t}$ has the segment connecting $(m, v_p(c_m))$ and $(m-1, v_p(c_{m-1}))$ as one of its sides while all other segments have slope less than or equal to r. We deduce that there is precisely one root $t_1' \in \bar{\mathbb{Q}}_p$ of $\mathrm{ch}_{\beta t}$ (counted with multiplicity) such that

$$v_p(t_1') = v_p(t_1),$$

while all remaining roots $t_h' \in \bar{\mathbb{Q}}_p$ of $\mathrm{ch}_{\beta t}$, $h = 2, \ldots, m$, have p-adic value smaller than or equal to r (in particular t_1' appears with multiplicity 1). Since $r = \max_{h=2}^m v_p(t_h) < v_p(t_1)$ we obtain $v_p(t_1') \geq v_p(t_h') + 1$ for all $h = 2, \ldots, m$. This implies that

$$c_m = \prod_{h=1}^m t_h' \quad \text{and} \quad c_{m-1} \equiv \prod_{h=2}^m t_h' \pmod{p^{v_p(c_{m-1})+1}\mathcal{O}_{\bar{\mathbb{Q}}_p}}$$

(note that $\mathrm{ch}_{\beta t}(X) = \prod_{h=1}^m (t_h' - X)$ because both sides have leading coefficient $(-1)^m$). Together with equations (29) and (30) we obtain

$$\prod_{h=1}^m \beta_{hh} t_h \equiv \prod_{h=1}^m t_h' \pmod{p^{v_p(c_m)+1}\mathbb{Z}_p}$$

and

$$\prod_{h=2}^m \beta_{hh} t_h \equiv \prod_{h=2}^m t_h' \pmod{p^{v_p(c_{m-1})+1}\mathcal{O}_{\bar{\mathbb{Q}}_p}}.$$

Since $v_p(c_m) - v_p(c_{m-1}) = v_p(t_1)$ the above two equations imply that $\beta_{11} t_1 \equiv t_1' \pmod{p^{v_p(t_1)+1}\mathcal{O}_{\bar{\mathbb{Q}}_p}}$ and the proof of the lemma is complete. □

4.4. The following proposition is an extension of 4.3 Lemma to closed subgroups of $\mathbf{GL}_m$ and is used in the proof of 4.7 Proposition. We denote by $W_{\tilde{\boldsymbol{G}}}$ the Weyl group of $\tilde{\boldsymbol{T}}/F \leq \tilde{\boldsymbol{G}}/F$.

Proposition. *Let $t \in \tilde{\boldsymbol{T}}(\mathbb{Q}_p)^{++}$ and let $x \in \tilde{\mathcal{I}}t\tilde{\mathcal{I}}$. Then the semisimple part x_s of $x \in \tilde{\boldsymbol{G}}(\mathbb{Q}_p)$ is $\tilde{\boldsymbol{G}}(\bar{\mathbb{Q}}_p)$-conjugate to a uniquely determined element $t' = t'_x \in \tilde{\boldsymbol{T}}(\bar{\mathbb{Q}}_p)^{++}$. The element t' satisfies*

$$v_p(\tilde{\lambda}(t')) = v_p(\tilde{\lambda}(t)) \quad \textit{and} \quad \tilde{\lambda}(t') \equiv \epsilon\tilde{\lambda}(t) \pmod{p^{v_p(\tilde{\lambda}(t))+1}\mathcal{O}_{\bar{\mathbb{Q}}_p}}$$

for all $\tilde{\lambda} \in X(\tilde{\boldsymbol{T}})$, where $\epsilon = \epsilon_{\tilde{\lambda},x} \in \mathbb{Z}_p^*$ is a p-adic unit in $\mathbb{Z}_p$.

Proof. Conjugating x by an element in $\tilde{\mathcal{I}} \subseteq \tilde{\boldsymbol{G}}(\mathbb{Q}_p)$ we may assume that $x = \beta t$ with $\beta \in \tilde{\mathcal{I}}$. Since $\mathbb{Q}_p$ is a perfect field we know that the semisimple part $(\beta t)_s$ of $\beta t \in \tilde{\boldsymbol{G}}(\mathbb{Q}_p)$ also is contained in $\tilde{\boldsymbol{G}}(\mathbb{Q}_p)$ (see [Sp 1], 12.1.7 (c), p. 211). By 6.4.5 Theorem (ii) in [Sp 1], p. 109, $(\beta t)_s$ is contained in $\tilde{\boldsymbol{S}}(\bar{\mathbb{Q}}_p)$, where $\tilde{\boldsymbol{S}} = \tilde{\boldsymbol{S}}_{\beta t}$ is a maximal $\bar{\mathbb{Q}}_p$-torus in $\tilde{\boldsymbol{G}}/\bar{\mathbb{Q}}_p$. Since all maximal tori in $\tilde{\boldsymbol{G}}/\bar{\mathbb{Q}}_p$ are conjugate over $\bar{\mathbb{Q}}_p$, there is $g \in \tilde{\boldsymbol{G}}(\bar{\mathbb{Q}}_p)$ such that ${}^g\tilde{\boldsymbol{T}} := g\tilde{\boldsymbol{T}}g^{-1} = \tilde{\boldsymbol{S}}$; in particular,

$$t' := g^{-1}(\beta t)_s g \in \tilde{\boldsymbol{T}}(\bar{\mathbb{Q}}_p).$$

Conjugating further by some $w \in W_{\tilde{\boldsymbol{G}}}$ we may assume that $t' \in \tilde{\boldsymbol{T}}(\bar{\mathbb{Q}}_p)^+$. Thus, t' is $\tilde{\boldsymbol{G}}(\bar{\mathbb{Q}}_p)$-conjugate to $x_s = (\beta t)_s$ and we will show that it satisfies the conditions of the Proposition. To this end we let $\tilde{\lambda} \in X(\tilde{\boldsymbol{T}})$ and we first assume that $\tilde{\lambda}$ is dominant. We denote by $(\rho_{\tilde{\lambda}}, L_{\tilde{\lambda}})$ the irreducible representation of $\tilde{\boldsymbol{G}}/\mathbb{Z}_p$ of highest weight $\tilde{\lambda}$ (see Section 1.4 and 2.4). We select a basis $(v_1, v_2, \ldots, v_m)$ of $L_{\tilde{\lambda}}(\mathbb{Q}_p)$ consisting of weight vectors w.r.t. $\mathfrak{h}$ as in the proof of 2.10 Lemma, i.e., if μ_i denotes the weight of v_i then $\mathrm{ht}_\lambda(\mu_i) > \mathrm{ht}_\lambda(\mu_j)$ implies $i > j$. We note that v_i has weight $\tilde{\mu}_i = \mu_i^\circ \otimes \tilde{\lambda}|_{\tilde{\boldsymbol{Z}}}$ w.r.t. $\tilde{\boldsymbol{T}}(\bar{\mathbb{Q}}_p)$ (see equation (9) in Section 2.4). In particular, v_1 is the highest weight vector, i.e., $\tilde{\mu}_1 = \tilde{\lambda}$. The above choice of a basis of $L_{\tilde{\lambda}}$ yields a matrix representation

$$\rho_{\tilde{\lambda}} : \tilde{\boldsymbol{G}}(\mathbb{Q}_p) \to \mathbf{GL}_m(\mathbb{Q}_p)$$

and 2.10 Lemma implies that

$$\rho_{\tilde{\lambda}}(\tilde{\mathcal{I}}) \subseteq \mathcal{J}_m.$$

Moreover, we obtain

$$\rho_{\tilde{\lambda}}(t) = \mathrm{diag}(\tilde{\lambda}(t), \tilde{\mu}_2(t), \ldots, \tilde{\mu}_m(t)) \in \boldsymbol{T}_m(\mathbb{Q}_p)$$

and

$$\rho_{\tilde{\lambda}}(t') = \mathrm{diag}(\tilde{\lambda}(t'), \tilde{\mu}_2(t'), \ldots, \tilde{\mu}_m(t')) \in \boldsymbol{T}_m(\bar{\mathbb{Q}}_p).$$

Any weight μ_i, $i \geq 2$, has the form $\mu_i = \lambda - \sum_{\alpha\in\Delta} n_\alpha \alpha$ where not all $n_\alpha \in \mathbb{N}_0$ are equal to zero and since $t \in \tilde{\boldsymbol{T}}(\mathbb{Q}_p)^{++}$ (i.e., $v_p(\alpha(t)) > 0$ for all $\alpha \in \Delta$) we obtain

$$v_p(\tilde{\mu}_i(t)) = v_p(\tilde{\lambda}(t)) - \sum_{\alpha\in\Delta} n_\alpha v_p(\alpha(t)) < v_p(\tilde{\lambda}(t)) \tag{34}$$

for all weights $\tilde{\mu}_i$, $i = 2, \ldots, m$. Analogously, since $t' \in \tilde{\boldsymbol{T}}(\bar{\mathbb{Q}}_p)^+$ we obtain

$$v_p(\tilde{\mu}_i(t')) = v_p(\tilde{\lambda}(t')) - \sum_{\alpha\in\Delta} n_\alpha v_p(\alpha(t')) \le v_p(\tilde{\lambda}(t')) \tag{35}$$

for all weights $\tilde{\mu}_i$, $i = 2, \dots, m$. Since $\rho_{\tilde{\lambda}}(\beta) \in \rho_{\tilde{\lambda}}(\tilde{\mathcal{I}}) \subseteq \mathcal{J}_m$ and $\rho_{\tilde{\lambda}}(t) \in \boldsymbol{T}_m(\mathbb{Q}_p)$ where the first entry of $\rho_{\tilde{\lambda}}(t)$ has p-adic value strictly bigger than the remaining entries (see equation (34)) we may apply 4.3 Lemma to $\rho_{\tilde{\lambda}}(\beta t) = \rho_{\tilde{\lambda}}(\beta)\rho_{\tilde{\lambda}}(t) \in \mathbf{GL}_m(\mathbb{Q}_p)$; we obtain that $\rho_{\tilde{\lambda}}(\beta t)$ has an eigenvalue $t_1' \in \bar{\mathbb{Q}}_p$ of multiplicity 1 whose p-adic value is strictly larger than the p-adic values of the $m-1$ remaining eigenvalues of $\rho_{\tilde{\lambda}}(\beta t)$ and which satisfies

$$v_p(\tilde{\lambda}(t)) = v_p(t_1') \quad \text{and} \quad \tilde{\lambda}(t) \equiv \epsilon t_1' \pmod{p^{v_p(\tilde{\lambda}(t))+1}\mathcal{O}_{\bar{\mathbb{Q}}_p}} \tag{36}$$

for some $\epsilon = \epsilon_{\tilde{\lambda},\beta} \in \mathbb{Z}_p^*$. Since the matrix $\rho_{\tilde{\lambda}}(t')$ has the same eigenvalues as $\rho_{\tilde{\lambda}}((\beta t)_s) = (\rho_{\tilde{\lambda}}(\beta t))_s$ it has the same eigenvalues as $\rho_{\tilde{\lambda}}(\beta t)$. In particular, $\rho_{\tilde{\lambda}}(t')$ has the eigenvalue t_1' with multiplicity 1 whose p-adic value is strictly bigger than the p-adic values of the $m-1$ remaining eigenvalues of $\rho_{\tilde{\lambda}}(t')$. Thus, equation (35) shows that $t_1' = \tilde{\lambda}(t')$ and equation (36) yields

$$v_p(\tilde{\lambda}(t)) = v_p(\tilde{\lambda}(t')) \quad \text{and} \quad \tilde{\lambda}(t) \equiv \epsilon\tilde{\lambda}(t') \pmod{p^{v_p(\tilde{\lambda}(t))+1}\mathcal{O}_{\bar{\mathbb{Q}}_p}}. \tag{37}$$

We recall that we have proven (37) under the assumption that $\tilde{\lambda}$ is dominant. For the general case we will need the following consequence of (37). Let $\tilde{\lambda}$ be dominant and write $\tilde{\lambda} = \tilde{\lambda}|_{\tilde{\boldsymbol{Z}}}\lambda^\circ$ where $\lambda^\circ = \tilde{\lambda}|_{\boldsymbol{T}}$. We write $t = t^\circ z$ where $t^\circ \in \boldsymbol{T}(\bar{\mathbb{Q}}_p)$, $z \in \tilde{\boldsymbol{Z}}(\bar{\mathbb{Q}}_p)$, hence, $t' = t'^\circ z$ where $t'^\circ = g^{-1}(\beta t^\circ)_s g$. Equation (37) then implies that

$$1 \equiv \epsilon \frac{\lambda^\circ(t'^\circ)}{\lambda^\circ(t^\circ)} \pmod{p\mathcal{O}_{\bar{\mathbb{Q}}_p}}. \tag{38}$$

Since $\boldsymbol{T} \subseteq \tilde{\boldsymbol{T}}$ is a closed subset the restriction map $X(\tilde{\boldsymbol{T}}) \to X(\boldsymbol{T})$ is surjective, hence, any dominant character $\lambda^\circ \in X(\boldsymbol{T})$ is the restriction of a dominant character $\tilde{\lambda} \in X(\tilde{\boldsymbol{T}})$ and we deduce that equation (38) holds for all dominant $\lambda^\circ \in X(\boldsymbol{T})$.

We now let $\tilde{\lambda} \in X(\tilde{\boldsymbol{T}})$ be arbitrary and we show that equation (37) still holds. We write $\tilde{\lambda} = \tilde{\lambda}|_{\tilde{\boldsymbol{Z}}}\lambda^\circ$ where $\lambda^\circ = \tilde{\lambda}|_{\boldsymbol{T}}$. Applying 4.4.1 Lemma below to $\lambda = \mathrm{d}\,\lambda^\circ \in \Gamma_\pi$ we can write $\lambda^\circ = \prod_i (\mu_i^\circ)^{n_i}$, where $\mu_i^\circ \in X(\boldsymbol{T})$ is dominant and $n_i \in \mathbb{Z}$. Equation (38) implies that

$$1 \equiv \prod_i \epsilon_i^{n_i} \frac{\prod_i \mu_i^\circ(t'^\circ)^{n_i}}{\prod_i \mu_i^\circ(t^\circ)^{n_i}} \pmod{p\mathcal{O}_{\bar{\mathbb{Q}}_p}}$$

with certain $\epsilon_i \in \mathbb{Z}_p^*$. Multiplying this by $\tilde{\lambda}(t) = \tilde{\lambda}(z)\lambda^\circ(t^\circ)$ we finally obtain

$$\tilde{\lambda}(t) \equiv \prod_i \epsilon_i^{n_i}\,\tilde{\lambda}(t') \pmod{p^{v_p(\tilde{\lambda}(t))+1}\mathcal{O}_{\bar{\mathbb{Q}}_p}}.$$

The last equation also implies $v_p(\tilde{\lambda}(t)) = v_p(\tilde{\lambda}(t'))$. Since t was strictly dominant this shows in particular that $t' \in \tilde{\boldsymbol{T}}(\bar{\mathbb{Q}}_p)^{++}$.

It only remains to prove uniqueness of t'. Let $t'' \in \tilde{\boldsymbol{T}}(\bar{\mathbb{Q}}_p)^{++}$ be another element which is $\tilde{\boldsymbol{G}}(\bar{\mathbb{Q}}_p)$-conjugate to $x_s = (\beta t)_s$. Hence, $t'' = gt'g^{-1}$ for some $g \in \tilde{\boldsymbol{G}}(\bar{\mathbb{Q}}_p)$. Since t', t'' are regular we know that $\tilde{\boldsymbol{T}}(\bar{\mathbb{Q}}_p)$ is the centralizer of t' and of t''. This implies $g\tilde{\boldsymbol{T}}(\bar{\mathbb{Q}}_p)g^{-1} = \tilde{\boldsymbol{T}}(\bar{\mathbb{Q}}_p)$, hence, g yields an element in $W_{\tilde{\boldsymbol{G}}}$. Since t', t'' are both strictly dominant g is the unit element in $W_{\tilde{\boldsymbol{G}}}$, i.e., $g \in \mathcal{C}(\tilde{\boldsymbol{T}}(\bar{\mathbb{Q}}_p))$ (centralizer of $\tilde{\boldsymbol{T}}(\bar{\mathbb{Q}}_p)$ in $\tilde{\boldsymbol{G}}(\bar{\mathbb{Q}}_p)$), hence, $t'' = t'$. Thus, the proposition is proven. □

4.4.1. *Lemma. The lattice $\Gamma_\pi \subseteq \mathfrak{h}^*$ is generated by dominant weights.*

Proof. We write the representation π of $\mathfrak{g}$ defining $\boldsymbol{G} = \boldsymbol{G}_\pi$ as $\pi = \bigoplus_{i=1}^n \rho_{\lambda_i}$ with dominant weights $\lambda_i \in \Gamma_\pi$. Since $P_\pi = \bigcup_i P_{\lambda_i}$ we obtain

$$\Gamma_\pi = \langle P_\pi \rangle = \big\langle \textstyle\bigcup_i P_{\lambda_i} \big\rangle \subseteq \big\langle \textstyle\bigcup_i \lambda_i + \Gamma_{\mathrm{ad}} \big\rangle \subseteq \Gamma_\pi;$$

hence, $\Gamma_\pi = \langle \bigcup_i \lambda_i + \Gamma_{\mathrm{ad}} \rangle$ (for the last inclusion note that π is faithful; hence, $\Gamma_{\mathrm{ad}} \subseteq \Gamma_\pi$). We select a weight $\gamma \in \Gamma_{\mathrm{ad}}$ which for the moment is arbitrary and we put $\mu_i = \lambda_i + \gamma$, $i = 1, \dots, n$. Since Γ_{ad} is generated by the simple roots we obtain

$$\Gamma_\pi = \big\langle \textstyle\bigcup_i \mu_i + \Gamma_{\mathrm{ad}} \big\rangle = \langle \mu_i, \mu_i - \alpha,\ i = 1, \dots, n,\ \alpha \in \Delta \rangle.$$

If we choose the weight $\gamma \in \Gamma_{\mathrm{ad}}$ dominant and sufficiently regular, i.e., $\langle \gamma, h_\beta \rangle > 0$ is positive and sufficiently large for all $\beta \in \Delta$ then μ_i and $\mu_i - \alpha$ are dominant for all $i = 1, \dots, n$ and all $\alpha \in \Delta$. This implies that Γ_π is generated by dominant weights. □

4.5. We denote by ${}^w\chi$ or sometimes by $w\chi$, $\chi \in X(\tilde{\boldsymbol{T}})$, $w \in W_{\tilde{\boldsymbol{G}}}$, the character sending t to $\chi(w^{-1}tw)$. We write $\rho = \rho_{\tilde{\boldsymbol{G}}} \in \Gamma_{\mathrm{sc}}$ for the half sum of the positive roots and we put

$$w \cdot \tilde{\lambda} = w(\tilde{\lambda} + \rho^\circ) - \rho^\circ \in X(\tilde{\boldsymbol{T}}) \quad \text{for } \tilde{\lambda} \in X(\tilde{\boldsymbol{T}}),$$

with $\rho^\circ = \frac{1}{2} \sum_{\alpha \in \Phi^+} \alpha \in X(\tilde{\boldsymbol{T}}) \otimes \mathbb{Q}$, where $\alpha = \alpha^\circ \in X(\tilde{\boldsymbol{T}})$ is the exponential of the root α; see Section 1.2.

Lemma. *Let $\tilde{\lambda} \in X(\tilde{\boldsymbol{T}})$ be a dominant weight. For any $w \in W_{\tilde{\boldsymbol{G}}}$, $w \neq 1$, and any $t \in \tilde{\boldsymbol{T}}(\bar{\mathbb{Q}}_p)$ we have $w\tilde{\lambda}(t) = \tilde{\lambda}(t) \left(\sum_{\alpha \in \Delta} -b_\alpha \alpha\right)(t)$, where $b_\alpha \in \mathbb{N}_0$ and*

$$b_{\alpha_0} \geq \frac{\langle \tilde{\lambda}, \alpha_0^\vee \rangle}{2}$$

for at least one root $\alpha_0 \in \Delta$. Also, if $w \neq 1$ we have $w \cdot \tilde{\lambda}(t) = \tilde{\lambda}(t) \left(\sum_{\alpha \in \Delta} -b_\alpha \alpha\right)(t)$, where $b_\alpha \in \mathbb{N}_0$ and

$$b_{\alpha_0} \geq \frac{\langle \tilde{\lambda}, \alpha_0^\vee \rangle}{2}$$

for at least one root $\alpha_0 \in \Delta$.

Proof. We prove the claim about $w \cdot \tilde{\lambda}$. We write $\tilde{\lambda} = \tilde{\lambda}|_{\tilde{\boldsymbol{Z}}}\lambda^\circ$, where $\lambda^\circ = \tilde{\lambda}|_{\boldsymbol{T}}$ and $t = zt^\circ \in \tilde{\boldsymbol{T}}(\bar{\mathbb{Q}}_p)$, where $z \in \tilde{\boldsymbol{Z}}(\bar{\mathbb{Q}}_p)$ and $t^\circ \in \boldsymbol{T}(\bar{\mathbb{Q}}_p)$. Since $w\rho^\circ - \rho^\circ \in \sum_{\alpha\in\Delta}\mathbb{Z}\alpha$ we obtain $w \cdot \tilde{\lambda}(z) = (w\tilde{\lambda} + (w\rho^\circ - \rho^\circ))(z) = \tilde{\lambda}(z)$. Since $\tilde{\lambda} = \lambda^\circ$ on $\boldsymbol{T}$ we obtain $w \cdot \tilde{\lambda}(t^\circ) = w \cdot \lambda^\circ(t^\circ)$. To determine $w \cdot \lambda^\circ(t^\circ)$ we set as before $\lambda = d\,\lambda^\circ \in \Gamma_\pi$, i.e., λ° corresponds to λ under the isomorphism $(\,\cdot\,)^\circ : \Gamma_\pi \to X(\boldsymbol{T})$ (see Section 1.2). The Weyl group $W_{\tilde{\boldsymbol{G}}}$ acts on $\Gamma_\pi \subseteq \mathfrak{h}^*$ via $\lambda \mapsto \lambda \circ \mathrm{Ad}(w^{-1})$, $w \in W_{\tilde{\boldsymbol{G}}}$, and since $(\,\cdot\,)^\circ$ is equivariant w.r.t. the action of $W_{\boldsymbol{G}}$ (see [B], Section 3.3, Remarks (1), p. 16) we obtain $w \cdot \lambda^\circ = (w \cdot \lambda)^\circ$, where $w \cdot \lambda := w(\lambda + \rho) - \rho$. Since $w\lambda$ is a weight of the irreducible $\mathfrak{g}$-module of highest weight λ we know that

$$w\lambda = \lambda - \sum_{\alpha\in\Delta} c_\alpha \alpha$$

for certain $c_\alpha \in \mathbb{N}_0$. Since λ is a dominant element in the weight lattice Γ_{sc} we may write $\lambda = \sum_{\alpha\in\Delta} d_\alpha\omega_\alpha$, where ω_α, $\alpha \in \Delta$, are the fundamental weights and $d_\alpha \in \mathbb{N}_0$. On the other hand, $w \neq 1$ implies that $w\lambda$ is not contained in the Weyl chamber corresponding to the basis Δ, hence, $\langle w\lambda, h_{\alpha_0}\rangle \le 0$ for some root $\alpha_0 \in \Delta$. We obtain

$$0 \ge \langle w\lambda, h_{\alpha_0}\rangle = \left\langle \sum_{\alpha\in\Delta} d_\alpha\omega_\alpha - \sum_{\alpha\in\Delta} c_\alpha\alpha, h_{\alpha_0}\right\rangle = d_{\alpha_0} - \sum_{\alpha\in\Delta} c_\alpha\langle\alpha, h_{\alpha_0}\rangle.$$

Since $\langle\alpha, h_{\alpha_0}\rangle = 2$ if $\alpha = \alpha_0$ and $\langle\alpha, h_{\alpha_0}\rangle \le 0$ if $\alpha \neq \alpha_0$ this yields $0 \ge d_{\alpha_0} - 2c_{\alpha_0}$. Thus,

$$c_{\alpha_0} \ge \tfrac{1}{2}d_{\alpha_0} = \tfrac{1}{2}\langle\lambda, h_{\alpha_0}\rangle = \tfrac{1}{2}\langle\lambda^\circ, \alpha_0^\vee\rangle = \tfrac{1}{2}\langle\tilde{\lambda}, \alpha_0^\vee\rangle.$$

Altogether we obtain

$$\begin{aligned}(w\cdot\tilde{\lambda})(t) &= \tilde{\lambda}(z)\,(w\cdot\lambda^\circ)(t^\circ) = \tilde{\lambda}(z)\,(w\cdot\lambda)^\circ(t^\circ) = \tilde{\lambda}(z)(w\lambda + w\rho - \rho)^\circ(t^\circ)\\ &= \tilde{\lambda}(z)\big(\lambda + \sum_{\alpha\in\Delta} -c_\alpha\alpha + w\rho - \rho\big)^\circ(t^\circ) = \tilde{\lambda}(t)\big(\sum_{\alpha\in\Delta} -c_\alpha\alpha + w\rho - \rho\big)^\circ(t^\circ).\end{aligned}$$

Since $w\rho - \rho \in \mathbb{Z}\Phi$ is a sum of negative roots, this shows that $w \cdot \tilde{\lambda}(t)$ has the claimed form (note that $\alpha(t) = \alpha(t^\circ)$ since $\alpha = \alpha^\circ$ vanishes on $\tilde{\boldsymbol{Z}}(\bar{\mathbb{Q}}_p)$). The claim about $w\tilde{\lambda}$ follows analogously. □

4.6. ***Notation.*** We recall that in Section 2.3 we selected an element $h \in \tilde{\boldsymbol{T}}(\mathbb{Q})^{++}$. We set

$$\kappa_1 = \kappa_{1,\tilde{\boldsymbol{G}}}(h) = \sum_{\alpha\in\Phi^+_{\tilde{G}}} v_p(\alpha(h)) \in \mathbb{N}.$$

Thus, κ_1 depends on $\tilde{\boldsymbol{G}}$ and h. Since $\rho - {}^w\rho$, $w \in W_{\tilde{\boldsymbol{G}}}$, is a sum of certain positive roots all of which occur with multiplicity 1 we obtain

$$v_p((\rho^\circ - {}^w\rho^\circ)(h^e)) \le e\kappa_1.$$

We write $2\rho = 2\rho_{\tilde{G}} = \sum_{\beta\in\Delta_{\tilde{G}}} m_\beta \beta$, where $m_\beta \in \mathbb{N}_0$ for all $\beta \in \Delta_{\tilde{G}}$, and we set

$$\kappa_2 = \kappa_{2,\tilde{G}} = \max_{\beta\in\Delta_{\tilde{G}}} m_\beta \in \mathbb{N}_0;$$

i.e., κ_2 is the maximum multiplicity with which a simple root can occur in $2\rho_{\tilde{G}}$ and only depends on $\tilde{G}$. Since $\rho - {}^w\rho$ is a sum of certain positive roots each of which occurs with multiplicity 1 we can write $\rho - {}^w\rho = \sum_{\beta\in\Delta_{\tilde{G}}} n_\beta\beta$ where $n_\beta \in \mathbb{N}_0$ and $n_\beta \leq m_\beta$ for all $\beta \in \Delta_{\tilde{G}}$. Since $\langle \beta, \alpha^\vee\rangle$, $\alpha, \beta \in \Delta_{\tilde{G}}$, equals 2 if $\alpha = \beta$ and is ≤ 0 otherwise we obtain for all $\alpha \in \Delta_{\tilde{G}}$

$$\langle \rho - {}^w\rho, \alpha^\vee\rangle \leq 2n_\alpha \leq 2m_\alpha \leq 2\kappa_2.$$

Let $\tilde{P}/\mathbb{Q}_p \leq \tilde{G}/\mathbb{Q}_p$ be a standard parabolic subgroup (i.e., $\tilde{P}/\mathbb{Q}_p \supseteq \tilde{B}/\mathbb{Q}_p$) with Levi decomposition $\tilde{P} = \tilde{M}\tilde{N}$. The Levi subgroup $\tilde{M}$ contains the $\mathbb{Q}_p$-split maximal torus $\tilde{T}$ and we denote by $W^{\tilde{P}}$ the set of Kostant representatives for the quotient of Weyl groups $W_{\tilde{M}}\backslash W_{\tilde{G}}$. The intersection $\tilde{B} \cap \tilde{M} \leq \tilde{M}/\mathbb{Q}_p$ is a Borel subgroup in $\tilde{M}/\mathbb{Q}_p$ and for any dominant (w.r.t. to $\tilde{B}\cap\tilde{M}$) weight $\tilde{\lambda} \in X(\tilde{T})$ we then denote by $\rho^{\tilde{M}}_{\tilde{\lambda}} : \tilde{M} \to \mathbf{Aut}(L^{\tilde{M}}_{\tilde{\lambda}})$ the irreducible representation of $\tilde{M}/\mathbb{Q}_p$ of highest weight $\tilde{\lambda}$ (see Section 2.4). Any $p \in \tilde{P}(\bar{\mathbb{Q}}_p)$ can be written $p = \bar{p}u$, $\bar{p} \in \tilde{M}(\bar{\mathbb{Q}}_p)$, $u \in \tilde{N}(\bar{\mathbb{Q}}_p)$ and we denote by $\nu_{\tilde{P}} : \tilde{P}/\mathbb{Q}_p \to \tilde{M}/\mathbb{Q}_p$, $p \mapsto \bar{p}$, the morphism to the Levi subgroup.

In the next proposition we will use the following notation: for $c \in \mathbb{Q}$ and $x, y \in \bar{\mathbb{Q}}_p$ we write $x \equiv y \pmod{p^c\mathcal{O}_{\bar{\mathbb{Q}}_p}}$ to denote that $v_p(x-y) \geq c$. Thus, in case $c \in \mathbb{Z}$ the term "$x \equiv y \pmod{p^c\mathcal{O}_{\bar{\mathbb{Q}}_p}}$" has two meanings that coincide.

4.7. Proposition. *Let $C \in \mathbb{Q}_{>0}$ and assume that $\tilde{\lambda} \in X(\tilde{T})^{\mathrm{dom}}$ satisfies*

$$\langle \tilde{\lambda}, \alpha^\vee\rangle > 2C$$

for all $\alpha \in \Delta_{\tilde{G}}$. Select a standard parabolic subgroup $\tilde{P}/\mathbb{Q}_p \leq \tilde{G}/\mathbb{Q}_p$ with Levi decomposition $\tilde{P} = \tilde{M}\tilde{N}$ and let $\xi \in \tilde{M}(\mathbb{Q}_p)$. Assume that there is $u \in \tilde{N}(\mathbb{Q}_p)$ such that $\xi u \in \tilde{P}(\mathbb{Q}_p)$ is $\tilde{G}(\mathbb{Q}_p)$-conjugate to an element in $\tilde{\mathcal{I}}h^e\tilde{\mathcal{I}}$, $e \in \mathbb{N}$. We denote by $t = t_{\xi u}$ the unique element in $\tilde{T}(\bar{\mathbb{Q}}_p)^{++}$ which is $\tilde{G}(\bar{\mathbb{Q}}_p)$-conjugate to $(\xi u)_s$ (see 4.4 Proposition; note that ξu is conjugate to an element $x \in \tilde{\mathcal{I}}h^e\tilde{\mathcal{I}}$). *Then there is an element $s = s_{\xi u} \in W^{\tilde{P}}$ such that for all $w \in W^{\tilde{P}}$ the following congruence holds*:

$$\tilde{\lambda}(h^e)\,\mathrm{tr}\,(\xi^{-1}|L^{\tilde{M}}_{w\cdot\tilde{\lambda}}(\mathbb{Q}_p)) \equiv \begin{cases} \varepsilon_{\tilde{P}}\,\tilde{\lambda}(h^e t^{-1}) \pmod{p^{(C-\kappa_1-\kappa_2)e}\mathcal{O}_{\bar{\mathbb{Q}}_p}} & \text{if } w = s, \\ 0 \pmod{p^{(C-\kappa_1-\kappa_2)e}\mathcal{O}_{\bar{\mathbb{Q}}_p}} & \text{if } w \neq s. \end{cases}$$

Here, $\varepsilon_{\tilde{P}} = \varepsilon_{\tilde{P},\xi u}$ is an element in $\bar{\mathbb{Q}}_p$ which does not depend on $\tilde{\lambda}$ and satisfies

$$v_p(\varepsilon_{\tilde{P},\xi u}) \geq -e\kappa_1.$$

Proof. By 12.1.7(c) (p. 211) of [Springer 1981], the semisimple part $(\xi u)_s$ is contained in $\tilde{\boldsymbol{P}}(\mathbb{Q}_p)$, and Theorem 6.4.5(ii) (p. 109) of the same reference shows that it is contained in a maximal $\bar{\mathbb{Q}}_p$-torus $\tilde{\boldsymbol{S}} = \tilde{\boldsymbol{S}}_{\xi u}$ in $\tilde{\boldsymbol{P}}/\bar{\mathbb{Q}}_p$. Since all maximal tori in $\tilde{\boldsymbol{P}}/\bar{\mathbb{Q}}_p$ are conjugate there is $y \in \tilde{\boldsymbol{P}}(\bar{\mathbb{Q}}_p)$ such that $\tilde{\boldsymbol{S}} = {}^y\tilde{\boldsymbol{T}}$. In particular there is $t' \in \tilde{\boldsymbol{T}}(\bar{\mathbb{Q}}_p)$ such that

$$(\xi u)_s = {}^y t'.$$

Modifying y by an element in $W_{\tilde{\boldsymbol{M}}}$ we may even assume that t' is dominant w.r.t. $\Delta_{\tilde{\boldsymbol{M}}}$, i.e., $v_p(\alpha(t')) \geq 0$ for all $\alpha \in \Delta_{\tilde{\boldsymbol{M}}}$.

Since ξu is conjugate to an element $x \in \tilde{\mathcal{I}} h^e \tilde{\mathcal{I}}$ the semisimple part $(\xi u)_s$ is $\tilde{\boldsymbol{G}}(\bar{\mathbb{Q}}_p)$-conjugate to an element $t = t_{\xi u} \in \tilde{\boldsymbol{T}}(\bar{\mathbb{Q}}_p)^{++}$ satisfying

$$v_p(\chi(t)) = v_p(\chi(h^e)) \quad \text{and} \quad \chi(t) \equiv \epsilon\chi(h^e) \pmod{p^{v_p(\chi(t))+1}\mathcal{O}_{\bar{\mathbb{Q}}_p}} \tag{39}$$

for all $\chi \in X(\tilde{\boldsymbol{T}})$ where $\epsilon = \epsilon_\chi \in \mathbb{Z}_p^*$ (see 4.4 Proposition). Since t' also is $\tilde{\boldsymbol{G}}(\bar{\mathbb{Q}}_p)$-conjugate to $(\xi u)_s$ we find that $t, t' \in \tilde{\boldsymbol{T}}(\bar{\mathbb{Q}}_p)$ are conjugate by an element $s = s_{\xi u} \in \tilde{\boldsymbol{G}}(\bar{\mathbb{Q}}_p)$:

$$t' = {}^s t.$$

Since t is regular, t' also is regular, and it follows that $\tilde{\boldsymbol{T}} = C(t)^0 = C(t')^0$. This implies ${}^s\tilde{\boldsymbol{T}} = \tilde{\boldsymbol{T}}$; hence, s is contained in the normalizer of $\tilde{\boldsymbol{T}}$ and we therefore can select $s \in W_{\tilde{\boldsymbol{G}}}$ (i.e., s is representative of an element in $W_{\tilde{\boldsymbol{G}}}$). Since t' is dominant w.r.t. $\Delta_{\tilde{\boldsymbol{M}}}$ and t is strictly dominant w.r.t. $\Delta_{\tilde{\boldsymbol{G}}}$ we see that s^{-1} maps $\Phi^+_{\tilde{\boldsymbol{M}}}$ to $\Phi^+_{\tilde{\boldsymbol{G}}}$, which implies $s \in W^{\tilde{P}}$. Denote by $L^{\tilde{P}}_{w\cdot\tilde{\lambda}}$ the extension of the representation $L^{\tilde{M}}_{w\cdot\tilde{\lambda}}$ to $\tilde{\boldsymbol{P}}/\mathbb{Q}_p$ via the morphism $\nu_{\tilde{\boldsymbol{P}}} : \tilde{\boldsymbol{P}} \to \tilde{\boldsymbol{M}}$. We have obtained

$$\begin{aligned}\operatorname{tr}(\xi^{-1}|L^{\tilde{M}}_{w\cdot\tilde{\lambda}}(\mathbb{Q}_p)) &= \operatorname{tr}(\xi^{-1}|L^{\tilde{P}}_{w\cdot\tilde{\lambda}}(\mathbb{Q}_p)) = \operatorname{tr}((\xi u)^{-1}|L^{\tilde{P}}_{w\cdot\tilde{\lambda}}(\mathbb{Q}_p))\\ &= \operatorname{tr}((\xi u)_s^{-1}|L^{\tilde{P}}_{w\cdot\tilde{\lambda}}(\mathbb{Q}_p)) = \operatorname{tr}((t')^{-1})|L^{\tilde{M}}_{w\cdot\tilde{\lambda}}(\mathbb{Q}_p))\\ &= \operatorname{tr}({}^s t^{-1})|L^{\tilde{M}}_{w\cdot\tilde{\lambda}}(\mathbb{Q}_p)).\end{aligned}$$

The Weyl character formula (see [Popov and Vinberg 1994, I.4.6.4 Theorem, p. 45] or [Jantzen 2003, II.5.10 Proposition, p. 223]) then yields

$$\tilde{\lambda}(h^e)\operatorname{tr}(\xi^{-1}|L^{\tilde{M}}_{w\cdot\tilde{\lambda}}(\mathbb{Q}_p)) = \tilde{\lambda}(h^e)\frac{\sum_{v\in W_{\tilde{M}}}(-1)^{\ell(v)}\, v\cdot_{\tilde{M}}(w\cdot\tilde{\lambda})({}^s t^{-1})}{\prod_{\alpha\in\Phi^+_{\tilde{\boldsymbol{M}}}}(1-\alpha^{-1}(\underbrace{{}^s t^{-1}}_{=t'^{-1}}))}. \tag{40}$$

Here, we use the notation $v\cdot_{\tilde{M}}\tilde{\mu} = v(\tilde{\mu}+\rho^\circ_{\tilde{\boldsymbol{M}}}) - \rho^\circ_{\tilde{\boldsymbol{M}}}$. We denote the denominator appearing on the right-hand side of (40) by

$$N(t') = N_{\tilde{\boldsymbol{M}}}(t').$$

For all $\alpha \in \Phi_{\tilde{G}}$ equation (39) implies that $v_p(\alpha(t)) \geq e$ or $v_p(\alpha(t)) \leq -e$. Since $\alpha({}^s t) = (s^{-1}\alpha)(t)$ the same then is true of $t' = {}^s t$, i.e., $v_p(\alpha(t')) \geq e$ or $v_p(\alpha(t')) \leq -e$ for all $\alpha \in \Phi_{\tilde{G}}$. Since $v_p(\alpha(t')) \geq 0$ for all $\alpha \in \Phi^+_{\tilde{M}}$ we obtain

$$v_p(\alpha(t')) \geq e \quad \text{for all } \alpha \in \Phi^+_{\tilde{M}}. \tag{41}$$

Hence,

$$v_p(N(t')) = v_p\Bigg(\prod_{\alpha \in \Phi^+_{\tilde{M}}} (1 - \alpha(t'))\Bigg) = 0,$$

i.e., $N(t')$ is a p-adic unit in $\mathcal{O}_{\bar{\mathbb{Q}}_p}$; in particular, $N(t') \neq 0$.

We look at the individual summands indexed by $v \in W_{\tilde{M}}$ which appear in the numerator in equation (40) and distinguish cases.

4.7.1. We first assume that $v \neq 1$. Using 4.5 Lemma with $\tilde{M}$ in place of $\tilde{G}$ and the definition of $w \cdot \tilde{\lambda}$ we can write

$$\begin{aligned} \tilde{\lambda}(h^e)\, v \cdot_{\tilde{M}} (w \cdot \tilde{\lambda})({}^s t^{-1}) \qquad & \\ = \tilde{\lambda}(h^e)\, (w \cdot \tilde{\lambda})({}^s t^{-1}) &\Bigg(\sum_{\alpha \in \Delta_{\tilde{M}}} -b_{v,\alpha}\alpha\Bigg)({}^s t^{-1}) \\ = \tilde{\lambda}(h^e)(s^{-1}w\tilde{\lambda})(t^{-1})&(w\rho^\circ - \rho^\circ)({}^s t^{-1})\Bigg(\sum_{\alpha \in \Delta_{\tilde{M}}} b_{v,\alpha}\alpha\Bigg)({}^s t) \end{aligned} \tag{42}$$

with $b_{v,\alpha} \in \mathbb{N}_0$ and $b_{v,\alpha_v} \geq \frac{1}{2}\langle w \cdot \tilde{\lambda}, \alpha_v^\vee \rangle$ for (at least) one root $\alpha_v \in \Delta_{\tilde{M}}$ (note that $w \cdot \tilde{\lambda}$ is dominant for $\Delta_{\tilde{M}}$ since $w \in W^{\tilde{P}}$). Since $w \in W^{\tilde{P}}$ we know that ${}^{w^{-1}}\alpha_v \in \Phi^+_{\tilde{G}}$, hence, we obtain

$$b_{v,\alpha_v} \geq \frac{\langle w \cdot \tilde{\lambda}, \alpha_v^\vee \rangle}{2} = \frac{\langle \tilde{\lambda}, {}^{w^{-1}}\alpha_v^\vee \rangle}{2} + \frac{\langle {}^w\rho^\circ - \rho^\circ, \alpha_v^\vee \rangle}{2} \geq C - \kappa_2.$$

Equation (41) implies that

$$v_p\Bigg(\Bigg(\sum_{\alpha \in \Delta_{\tilde{M}}} b_{v,\alpha}\alpha\Bigg)({}^s t)\Bigg) \geq (C - \kappa_2)e. \tag{43}$$

Since $w\rho - \rho$ is a sum of certain negative roots all appearing with multiplicity 1 and since t is strictly dominant we obtain using equation (39)

$$v_p((w\rho^\circ - \rho^\circ)({}^s t^{-1})) \geq -\sum_{\alpha \in \Phi^+_{\tilde{G}}} v_p(\alpha(t)) = -\sum_{\alpha \in \Phi^+_{\tilde{G}}} v_p(\alpha(h^e)) = -e\kappa_1. \tag{44}$$

If $s^{-1}w \neq 1$ then 4.5 Lemma yields

$$\tilde{\lambda}(h^e)(s^{-1}w\tilde{\lambda})(t^{-1}) = \tilde{\lambda}(h^e t^{-1})\Bigg(\sum_{\alpha \in \Delta_{\tilde{G}}} -c_{w,\alpha}\alpha\Bigg)(t^{-1}),$$

where $c_{w,\alpha} \in \mathbb{N}_0$ and $c_{w,\alpha_w} \geq \frac{1}{2}\langle \tilde{\lambda}, \alpha_w^\vee \rangle \geq C$ for (at least) one root $\alpha_w \in \Delta_{\tilde{G}}$. It follows from (39) that $\tilde{\lambda}(h^e)\tilde{\lambda}(t^{-1})$ is a p-adic unit and that $v_p(\alpha(t)) = v_p(\alpha(h^e)) \geq e$ for all $\alpha \in \Delta_{\tilde{\boldsymbol{G}}}$; hence,

$$v_p(\tilde{\lambda}(h^e)(s^{-1}w\tilde{\lambda})(t^{-1})) = v_p\left(\left(\sum_{\alpha\in\Delta_{\tilde{\boldsymbol{G}}}} c_{w,\alpha}\alpha\right)(t)\right) \geq Ce\,(>0). \tag{45}$$

If $s = w$ then $\tilde{\lambda}(h^e)(s^{-1}w\tilde{\lambda})(t^{-1}) = \tilde{\lambda}(h^e t^{-1})$ is a p-adic integer. Thus, if $v \neq 1$ equations (42), (43), (44), (45) yield

$$v_p\big(\tilde{\lambda}(h^e)\, v \cdot_{\tilde{M}} (w\cdot\tilde{\lambda})({}^s t^{-1})\big) \geq (C-\kappa_1-\kappa_2)e \tag{46}$$

for all $w \in W^{\tilde{P}}$. Hence, modulo $p^{(C-\kappa_1-\kappa_2)e}\mathcal{O}_{\bar{\mathbb{Q}}_p}$ we may neglect all summands with $v \neq 1$.

4.7.2. We assume $v = 1$. If $w \neq s$ equations (42), (44), (45) yield

$$\begin{aligned} v_p\big(\tilde{\lambda}(h^e)\, v \cdot_{\tilde{M}} (w\cdot\tilde{\lambda})({}^s t^{-1})\big) &= v_p(\tilde{\lambda}(h^e)(s^{-1}w\tilde{\lambda})(t^{-1})(w\rho^\circ-\rho^\circ)({}^s t^{-1})) \\ &\geq (C-\kappa_1)e. \end{aligned} \tag{47}$$

If $w = s$ we obtain as above

$$\tilde{\lambda}(h^e)\, v \cdot_{\tilde{M}} (w\cdot\tilde{\lambda})({}^s t^{-1}) = \tilde{\lambda}(h^e)\tilde{\lambda}(t^{-1})(s\rho^\circ-\rho^\circ)({}^s t^{-1}). \tag{48}$$

4.7.3. Taking into account that $C-\kappa_1$ is bigger than or equal to $C-\kappa_1-\kappa_2$, equations (40) and (46), (47), (48) now yield (note that $N(t')$ is a p-adic unit)

$$\begin{aligned} &\tilde{\lambda}(h^e)\,\mathrm{tr}\,(\xi^{-1}|L^{\tilde{M}}_{w\cdot\tilde{\lambda}}(\mathbb{Q}_p)) \\ &\qquad \equiv \begin{cases} \dfrac{\tilde{\lambda}(h^e)\tilde{\lambda}(t^{-1})}{N(t')}(s\rho^\circ-\rho^\circ)({}^s t^{-1}) \pmod{p^{(C-\kappa_1-\kappa_2)e}\mathcal{O}_{\bar{\mathbb{Q}}_p}} & \text{if} w = s, \\ 0 \pmod{p^{(C-\kappa_1-\kappa_2)e}\mathcal{O}_{\bar{\mathbb{Q}}_p}} & \text{if} w \neq s. \end{cases} \end{aligned}$$

We put $\varepsilon_{\tilde{\boldsymbol{P}}} = \varepsilon_{\tilde{\boldsymbol{P}},\xi u} = \dfrac{(s\rho^\circ-\rho^\circ)({}^s t^{-1})}{N(t')} \in \bar{\mathbb{Q}}_p$. Since $N(t')$ is a p-adic unit, equation (44) shows that

$$v_p(\varepsilon_{\tilde{\boldsymbol{P}},\xi u}) \geq -e\kappa_1.$$

This completes the proof of the proposition. □

4.8. We look at the special case $\tilde{\boldsymbol{P}} = \tilde{\boldsymbol{G}}$ in 4.7 Proposition which is sufficient for application to Bewersdorff's trace formula. (The general case will be needed in application to the Goresky–MacPherson trace formula which involves contributions from parabolic subgroups of $\tilde{\boldsymbol{G}}$ as well; see Section 4.14). In this case, $\xi \in \tilde{\boldsymbol{G}}(\mathbb{Q}_p)$, $u = 1$, $W^{\tilde{P}} = 1$ and $s = w = 1$; hence, $t = t_\xi = t'$. In particular, $\varepsilon_{\tilde{\boldsymbol{P}}} = \varepsilon_{\tilde{\boldsymbol{G}},1,\xi} = 1/N(t_\xi)$ is a p-adic integer. Moreover, we can choose $\kappa_1 = \kappa_2 = 0$ since then the equations involving ρ and κ_1, κ_2 in Section 4.6 still hold (note that $w = 1$; see also equation (44) and the equation following (42)). We thus obtain:

Corollary. *Let $C \in \mathbb{Q}_{>0}$ and assume that $\tilde{\lambda} \in X(\tilde{\boldsymbol{T}})^{\mathrm{dom}}$ satisfies*

$$\langle \tilde{\lambda}, \alpha^\vee \rangle > 2C$$

for all $\alpha \in \Delta$. Then for any $\xi \in \tilde{\boldsymbol{G}}(\mathbb{Q}_p)$ which is $\tilde{\boldsymbol{G}}(\mathbb{Q}_p)$-conjugate to an element in $\tilde{\mathcal{I}} h^e \tilde{\mathcal{I}}$, $e \in \mathbb{N}$, the following congruence holds:

$$\tilde{\lambda}(h^e) \operatorname{tr}(\xi^{-1} | L_{\tilde{\lambda}}(\mathbb{Q}_p)) \equiv \frac{\tilde{\lambda}(h^e t_\xi^{-1})}{N(t_\xi)} \pmod{p^{Ce} \mathcal{O}_{\bar{\mathbb{Q}}_p}}.$$

Here, $t_\xi \in \tilde{\boldsymbol{T}}(\bar{\mathbb{Q}}_p)^{++}$ denotes the unique element which is $\tilde{\boldsymbol{G}}(\bar{\mathbb{Q}}_p)$-conjugate to ξ_s (see 4.4 Proposition) *and*

$$N(t_\xi) = \prod_{\alpha \in \Phi^+_{\tilde{\boldsymbol{G}}}} (1 - \alpha(t_\xi))$$

is a p-adic unit in $\mathcal{O}_{\bar{\mathbb{Q}}_p}$.

4.9. Proposition. *Let $C \in \mathbb{Q}_{>0}$ and assume that $\tilde{\lambda} \in X(\tilde{\boldsymbol{T}})^{\mathrm{dom}}$ satisfies*

$$\langle \tilde{\lambda}, \alpha^\vee \rangle > 2C$$

for all $\alpha \in \Delta_{\tilde{\boldsymbol{G}}}$. Let ζ be contained in the semigroup Δ, so $\zeta \in \tilde{\mathcal{I}} h^e \tilde{\mathcal{I}}$ for some $e \in \mathbb{N}_0$ (by 2.7 Lemma); *we assume that $e \in \mathbb{N}$. Then the Lefschetz number* $\mathrm{Lef}((\Gamma\zeta\Gamma)_{\tilde{\lambda}} | H^\bullet(\Gamma\backslash X, L_{\tilde{\lambda}}(\mathbb{Q}_p)))$ *is contained in $\mathbb{Z}_p$ and the following congruence holds*:

$$\mathrm{Lef}((\Gamma\zeta\Gamma)_{\tilde{\lambda}} | H^\bullet(\Gamma\backslash X, L_{\tilde{\lambda}}(\mathbb{Q}_p))) \equiv \sum_{[\xi]_\Gamma \in \Gamma\zeta\Gamma/\sim_\Gamma} c_{[\xi]_\Gamma} \frac{\tilde{\lambda}(h^e t_\xi^{-1})}{N(t_\xi)} \pmod{p^{Ce} \mathcal{O}_{\bar{\mathbb{Q}}_p}}.$$

Proof. By 2.9 Corollary, the Lefschetz numbers of normalized Hecke operators are contained in $\mathbb{Z}_p$. Since $\zeta \in \tilde{\mathcal{I}} h^e \tilde{\mathcal{I}}$ we know that $(\Gamma\zeta\Gamma)_{\tilde{\lambda}} = \tilde{\lambda}(h^e)\Gamma\zeta\Gamma$. On the other hand, a representative ξ of a Γ-conjugacy class contained in $\Gamma\zeta\Gamma$ is contained in $\tilde{\boldsymbol{G}}(\mathbb{Q})$ and in $\tilde{\mathcal{I}}\zeta\tilde{\mathcal{I}} = \tilde{\mathcal{I}} h^e \tilde{\mathcal{I}}$. The second claim thus follows from Bewersdorff's trace formula (see 4.2 Theorem) and 4.8 Corollary (note that $F \subseteq \mathbb{Q}_p$). ☐

4.10. Proposition. *Let $C \in \mathbb{Q}_{>0}$ and let the dominant weights $\tilde{\lambda}, \tilde{\lambda}' \in X(\tilde{\boldsymbol{T}})$ satisfy*

- $\langle \tilde{\lambda}, \alpha^\vee \rangle > 2C$ *and* $\langle \tilde{\lambda}', \alpha^\vee \rangle > 2C$ *for all* $\alpha \in \Delta_{\tilde{\boldsymbol{G}}}$;
- $\tilde{\lambda} \equiv \tilde{\lambda}' \pmod{(p-1)p^{m-1} X(\tilde{\boldsymbol{T}})}$ $(m \in \mathbb{N})$.

Let ζ be contained in the semigroup Δ, so $\zeta \in \tilde{\mathcal{I}} h^e \tilde{\mathcal{I}}$ for some $e \in \mathbb{N}_0$ which we assume positive. Then the Lefschetz number $\mathrm{Lef}((\Gamma\zeta\Gamma)_{\tilde{\lambda}} | H^\bullet(\Gamma\backslash X, L_{\tilde{\lambda}}(F)))$ *is contained in $\mathbb{Z}_p$* (*note that $F \subseteq \mathbb{Q}_p$*) *and the following congruence holds*:

$$\begin{aligned} &\mathrm{Lef}((\Gamma\zeta\Gamma)_{\tilde{\lambda}} | H^\bullet(\Gamma\backslash X, L_{\tilde{\lambda}}(F))) \\ &\qquad\equiv \mathrm{Lef}((\Gamma\zeta\Gamma)_{\tilde{\lambda}'} | H^\bullet(\Gamma\backslash X, L_{\tilde{\lambda}'}(F))) \pmod{p^{\lceil \min(m, Ce) \rceil} \mathbb{Z}_p}. \end{aligned}$$

Proof. Since $\mathrm{Lef}((\Gamma\zeta\Gamma)_{\tilde\lambda}|H^\bullet(\Gamma\backslash X, L_{\tilde\lambda}(F))) = \mathrm{Lef}((\Gamma\zeta\Gamma)_{\tilde\lambda}|H^\bullet(\Gamma\backslash X, L_{\tilde\lambda}(\mathbb{Q}_p)))$, in order to prove the proposition we may consider Lefschetz numbers over $\mathbb{Q}_p$. Integrality of the Lefschetz number then follows from 4.9 Proposition. We let $[\xi]_\Gamma$ be any Γ-conjugacy class contained in $\Gamma\zeta\Gamma$. Hence, $\xi \in \tilde{\mathcal{I}} h^e \tilde{\mathcal{I}}$ and we denote by $t_\xi \in \tilde{\boldsymbol{T}}(\bar{\mathbb{Q}}_p)^{++}$ the element which is $\tilde{\boldsymbol{G}}(\bar{\mathbb{Q}}_p)$-conjugate to ξ_s (see 4.4 Proposition). Since $\tilde\lambda \equiv \tilde\lambda'$ (mod $(p-1)p^{m-1}X(\tilde{\boldsymbol{T}})$) there is $\chi \in X(\tilde{\boldsymbol{T}})$ such that $\tilde\lambda - \tilde\lambda' = (p-1)p^{m-1}\chi$. Taking into account that $\chi(h^e t_\xi^{-1}) \equiv \epsilon$ (mod $p\mathcal{O}_{\bar{\mathbb{Q}}_p}$) by 4.4 Proposition, where $\epsilon = \epsilon_\chi \in \mathbb{Z}_p^*$, we therefore obtain

$$\frac{\tilde\lambda(h^e t_\xi^{-1})}{\tilde\lambda'(h^e t_\xi^{-1})} = \chi(h^e t_\xi^{-1})^{(p-1)p^{m-1}} \in 1 + p^m \mathcal{O}_{\bar{\mathbb{Q}}_p}.$$

Since also $\tilde\lambda'(h^e t_\xi^{-1})$ is a p-adic unit by the same proposition, this implies

$$\tilde\lambda(h^e t_\xi^{-1}) \equiv \tilde\lambda'(h^e t_\xi^{-1}) \pmod{p^m \mathcal{O}_{\bar{\mathbb{Q}}_p}}.$$

The claim now follows from 4.9 Proposition taking into account that the Lefschetz numbers are contained in $\mathbb{Q}_p$, hence, their p-adic valuations are integers and that $c_{[\xi]_\Gamma} \in \mathbb{Z}$ and $N(t_\xi)$ is a p-adic unit. Thus, the proof is complete. □

Remark. The proposition also holds trivially for $e = 0$ since both sides of the congruence are integers by 2.9 Corollary.

4.11. ***Adelic formulation.*** Using adelic formulation in Section 4.13 we will prove congruences between traces of Hecke operators on Eisenstein cohomology and, hence, on cuspidal cohomology. In this section we therefore reformulate 3.10 Corollary and 4.10 Proposition in adelic language.

We denote by $\mathbb{A}$ (resp. $\mathbb{A}_f$) the ring of adeles (resp. of finite adeles) of $\mathbb{Q}$. For any compact open subgroup $\tilde K \le \tilde{\boldsymbol{G}}(\mathbb{A}_f)$ we set $S_{\tilde K} = \tilde{\boldsymbol{G}}(\mathbb{Q})\backslash\tilde{\boldsymbol{G}}(\mathbb{A})/\tilde K \tilde K_\infty A_{\tilde{\boldsymbol{G}}}$. We assume that $\boldsymbol{G}/\mathbb{Q}$ satisfies strong approximation; in particular, $\tilde{\boldsymbol{G}}(\mathbb{A})$ is a finite disjoint union $\tilde{\boldsymbol{G}}(\mathbb{A}) = \bigcup_{i=1}^t \tilde{\boldsymbol{G}}(\mathbb{Q}) g_i \tilde{\boldsymbol{G}}(\mathbb{R})\tilde K$, $g_i \in \tilde{\boldsymbol{G}}(\mathbb{A}_f)$, and we obtain

$$S_{\tilde K} = \bigcup_{i=1}^t \Gamma_i\backslash X$$

where

$$\Gamma_i = \tilde{\boldsymbol{G}}(\mathbb{Q}) \cap g_i \tilde K g_i^{-1}.$$

We assume that we can choose a system of double coset representatives g_i as above which is contained in $\tilde{\boldsymbol{G}}(\mathbb{A}_f)^{(p)}$, where $\tilde{\boldsymbol{G}}(\mathbb{A}_f)^{(p)} \le \tilde{\boldsymbol{G}}(\mathbb{A}_f)$ is the subgroup consisting of elements whose p-component equals 1 (e.g., $\tilde{\boldsymbol{G}}$ satisfies weak approximation at p; note that weak approximation holds for almost all primes p).

We fix a compact open subgroup $\tilde{K} = \prod_{\ell \neq \infty} \tilde{K}_\ell \leq \tilde{\boldsymbol{G}}(\mathbb{A}_f)$ such that $\tilde{K}_p = \tilde{\mathcal{I}}$ and we set $\tilde{K}^{(p)} = \prod_{\ell \neq p,\infty} \tilde{K}_\ell$. Since the p-component of g_i is equal to 1 any of the arithmetic subgroups Γ_i is contained in $\tilde{\mathcal{I}}$.

4.11.1. *Hecke algebra.* We fix a Haar measure $dg = \bigotimes_{\ell \neq \infty} dg_\ell$ on $\tilde{\boldsymbol{G}}(\mathbb{A}_f)$ and we denote by $\mathcal{C}_{0,\mathbb{Q}}(\tilde{\boldsymbol{G}}(\mathbb{A}_f))$ the Hecke algebra consisting of compactly supported smooth $\mathbb{Q}$-valued functions on $\tilde{\boldsymbol{G}}(\mathbb{A}_f)$. We have the decomposition $\mathcal{C}_{0,\mathbb{Q}}(\tilde{\boldsymbol{G}}(\mathbb{A}_f)) = \mathcal{C}_{0,\mathbb{Q}}(\tilde{\boldsymbol{G}}(\mathbb{Q}_p)) \otimes \mathcal{C}_{0,\mathbb{Q}}(\tilde{\boldsymbol{G}}(\mathbb{A}_f)^{(p)})$, where the two factors are the Hecke algebras consisting respectively of compactly supported smooth $\mathbb{Q}$-valued functions on $\tilde{\boldsymbol{G}}(\mathbb{Q}_p)$ and on $\tilde{\boldsymbol{G}}(\mathbb{A}_f)^{(p)}$. Let the subalgebra consisting of $\tilde{K}$- (resp. $\tilde{\mathcal{I}}$- or $\tilde{K}^{(p)}$-) bi-invariant functions be denoted by $\mathcal{C}_{0,\mathbb{Q}}(\tilde{\boldsymbol{G}}(\mathbb{A}_f)//\tilde{K}) \leq \mathcal{C}_{0,\mathbb{Q}}(\tilde{\boldsymbol{G}}(\mathbb{A}_f))$ (resp. $\mathcal{C}_{0,\mathbb{Q}}(\tilde{\boldsymbol{G}}(\mathbb{Q}_p)//\tilde{\mathcal{I}}) \leq \mathcal{C}_{0,\mathbb{Q}}(\tilde{\boldsymbol{G}}(\mathbb{Q}_p))$ or $\mathcal{C}_{0,\mathbb{Q}}(\tilde{\boldsymbol{G}}(\mathbb{A}_f)^{(p)}//\tilde{K}^{(p)}) \leq \mathcal{C}_{0,\mathbb{Q}}(\tilde{\boldsymbol{G}}(\mathbb{A}_f)^{(p)})$). Let

$$[\tilde{K}x\tilde{K}] = \frac{1}{\mathrm{vol}(\tilde{K})}\mathbf{1}_{\tilde{K}x\tilde{K}}, \quad x \in \tilde{\boldsymbol{G}}(\mathbb{A}_f),$$

where $\mathbf{1}_X$ is the characteristic function of the set X, and define likewise $[\tilde{\mathcal{I}}x\tilde{\mathcal{I}}]$ for $x \in \tilde{\boldsymbol{G}}(\mathbb{Q}_p)$ and $[\tilde{K}^{(p)}x\tilde{K}^{(p)}]$ for $x \in \tilde{\boldsymbol{G}}(\mathbb{A}_f)^{(p)}$, by replacing $\tilde{K}$ with $\tilde{\mathcal{I}}$ and with $\tilde{K}^{(p)}$. The elements $[\tilde{K}x\tilde{K}]$ (resp. $[\tilde{\mathcal{I}}x\tilde{\mathcal{I}}]$ or $[\tilde{K}^{(p)}x\tilde{K}^{(p)}]$) form a $\mathbb{Q}$-basis of $\mathcal{C}_{0,\mathbb{Q}}(\tilde{\boldsymbol{G}}(\mathbb{A}_f)//\tilde{K})$ (resp. $\mathcal{C}_{0,\mathbb{Q}}(\tilde{\boldsymbol{G}}(\mathbb{Q}_p)//\tilde{\mathcal{I}})$ or $\mathcal{C}_{0,\mathbb{Q}}(\tilde{\boldsymbol{G}}(\mathbb{A}_f)^{(p)}//\tilde{K}^{(p)})$); their $\mathbb{Z}$-spans define $\mathbb{Z}$-structures $\mathcal{C}_0(\tilde{\boldsymbol{G}}(\mathbb{A}_f)//\tilde{K})$ (resp. $\mathcal{C}_0(\tilde{\boldsymbol{G}}(\mathbb{Q}_p)//\tilde{\mathcal{I}})$ or $\mathcal{C}_0(\tilde{\boldsymbol{G}}(\mathbb{A}_f)^{(p)}//\tilde{K}^{(p)})$) in the respective Hecke algebras. Thus, the $\mathbb{Z}$-structure on $\mathcal{C}_{0,\mathbb{Q}}(\tilde{\boldsymbol{G}}(\mathbb{A}_f)//\tilde{K})$ is given as the subspace of $\mathrm{vol}(\tilde{K})^{-1} \cdot \mathbb{Z}$-valued functions and analogously for the other two Hecke algebras.

In Section 2.3 we selected an element $h \in \tilde{\boldsymbol{T}}(\mathbb{Q})^{++}$ and we now denote by

$$\mathcal{C}_0(\tilde{\boldsymbol{G}}(\mathbb{Q}_p)//\tilde{\mathcal{I}})_h$$

the $\mathbb{Z}$-subalgebra of $\mathcal{C}_0(\tilde{\boldsymbol{G}}(\mathbb{Q}_p)//\tilde{\mathcal{I}})$ generated by $[\tilde{\mathcal{I}}h^{-1}\tilde{\mathcal{I}}]$ and we set

$$\mathcal{C}_0(\tilde{\boldsymbol{G}}(\mathbb{A}_f)//\tilde{K})_h = \mathcal{C}_0(\tilde{\boldsymbol{G}}(\mathbb{Q}_p)//\tilde{\mathcal{I}})_h \otimes \mathcal{C}_0(\tilde{\boldsymbol{G}}(\mathbb{A}_f)^{(p)}//\tilde{K}^{(p)}).$$

Since $[\tilde{\mathcal{I}}h^{-1}\tilde{\mathcal{I}}]^e = [\tilde{\mathcal{I}}h^{-e}\tilde{\mathcal{I}}]$ the algebra $\mathcal{C}_0(\tilde{\boldsymbol{G}}(\mathbb{Q}_p)//\tilde{\mathcal{I}})_h$ is the $\mathbb{Z}$-span of $[\tilde{\mathcal{I}}h^{-e}\tilde{\mathcal{I}}]$, $e \in \mathbb{N}_0$, hence, $\mathcal{C}_0(\tilde{\boldsymbol{G}}(\mathbb{A}_f)//\tilde{K})_h$ is the $\mathbb{Z}$-span of

$$[\tilde{K}[h]_p^{-e}r\tilde{K}] = [\tilde{\mathcal{I}}h^{-e}\tilde{\mathcal{I}}] \otimes [\tilde{K}^{(p)}r\tilde{K}^{(p)}], e \in \mathbb{N}_0, r \in \tilde{\boldsymbol{G}}(\mathbb{A}_f)^{(p)};$$

here, $[h]_p \in \tilde{\boldsymbol{G}}(\mathbb{A}_f)$ is the element with h in the p-component and all remaining components equal to 1.

For any $\mathbb{Z}$-algebra R we put $\mathcal{C}_{0,R}(\tilde{\boldsymbol{G}}(\mathbb{A}_f)//\tilde{K})_? = \mathcal{C}_0(\tilde{\boldsymbol{G}}(\mathbb{A}_f)//\tilde{K})_? \otimes R$ where $? = \mathit{blank}, h$. We define the $\tilde{\lambda}$-normalization of $[\tilde{K}[h]_p^{-e}r\tilde{K}]$ as

$$[\tilde{K}[h]_p^{-e}r\tilde{K}]_{\tilde{\lambda}} = \tilde{\lambda}(h)^e[\tilde{K}[h]_p^{-e}r\tilde{K}] \in \mathcal{C}_{0,F}(\tilde{\boldsymbol{G}}(\mathbb{A}_f)//\tilde{K})_h$$

(note that $\tilde\lambda(h)^e \in F$; see Section 2.5) and we extend linearly to $\mathcal{C}_{0,F}(\tilde{\boldsymbol{G}}(\mathbb{A}_f)//\tilde K)_h$. Since $[\tilde{\mathcal{I}}h^{-e}\tilde{\mathcal{I}}][\tilde{\mathcal{I}}h^{-f}\tilde{\mathcal{I}}] = [\tilde{\mathcal{I}}h^{-(e+f)}\tilde{\mathcal{I}}]$ we obtain that the assignment

$$\mathcal{C}_{0,F}(\tilde{\boldsymbol{G}}(\mathbb{A}_f)//\tilde K)_h \to \mathcal{C}_{0,F}(\tilde{\boldsymbol{G}}(\mathbb{A}_f)//\tilde K)_h, \quad \mathbb{T} \mapsto \mathbb{T}_{\tilde\lambda},$$

is a morphism of F-algebras.

4.11.2. *Cohomology.* The algebra $\mathcal{C}_0(\tilde{\boldsymbol{G}}(\mathbb{A}_f)//\tilde K)$, and hence also $\mathcal{C}_0(\tilde{\boldsymbol{G}}(\mathbb{A}_f)//\tilde K)_h$, acts on $H^n(S_{\tilde K}, L_{\tilde\lambda}(F)) = \bigoplus_i H^n(\Gamma_i\backslash X, L_{\tilde\lambda}(F))$. We define the integral cohomology $H^n(S_{\tilde K}, L_{\tilde\lambda}(\mathbb{Z}_p))_{\mathrm{int}}$ as the image of $H^n(S_{\tilde K}, L_{\tilde\lambda}(\mathbb{Z}_p))$ in $H^n(S_{\tilde K}, L_{\tilde\lambda}(\mathbb{Q}_p))$; hence, $H^n(S_{\tilde K}, L_{\tilde\lambda}(\mathbb{Z}_p))_{\mathrm{int}} = \bigoplus_i H^n(\Gamma_i\backslash X, L_{\tilde\lambda}(\mathbb{Z}_p))_{\mathrm{int}}$. Consider $e \in \mathbb{N}_0$ and $r \in \tilde{\boldsymbol{G}}(\mathbb{A}_f)^{(p)}$; we write $g_i[h]_p^{-e}r = \zeta_i g_{j(i)}k$ with $\zeta_i \in \tilde{\boldsymbol{G}}(\mathbb{Q})$, $k = (k_\ell)_\ell \in \tilde K$ and obtain

$$[\tilde K[h]_p^{-e}r\tilde K] = \bigoplus_i \Gamma_i\zeta_i^{-1}\Gamma_{j(i)}.$$

Looking at the p-component and recalling that $g_i \in \tilde{\boldsymbol{G}}(\mathbb{A}_f)$ has trivial p-component we find $h^{-e} = \zeta_i k_p$. Hence, $\zeta_i^{-1} = k_p h^e$ is contained in $\tilde{\mathcal{I}}h^e\tilde{\mathcal{I}}$ and, thus, in $\Delta = \Delta_h$ and we deduce that the normalization $\tilde\lambda(h^e)\Gamma_i\zeta_i^{-1}\Gamma_{j(i)}$ maps $H^n(\Gamma_i\backslash X, L_{\tilde\lambda}(\mathbb{Z}_p))_{\mathrm{int}}$ to $H^n(\Gamma_{j(i)}\backslash X, L_{\tilde\lambda}(\mathbb{Z}_p))_{\mathrm{int}}$. Hence, $[\tilde K[h]_p^{-e}r\tilde K]_{\tilde\lambda}$ and, thus, any $\mathbb{T}_{\tilde\lambda}$ with $\mathbb{T}$ in $\mathcal{C}_{0,\mathbb{Z}_p}(\tilde{\boldsymbol{G}}(\mathbb{A}_f)//\tilde K)_h$ leaves $H^n(S_{\tilde K}, L_{\tilde\lambda}(\mathbb{Z}_p))_{\mathrm{int}}$ invariant. In particular, the Lefschetz number $\mathrm{Lef}(\mathbb{T}_{\tilde\lambda}|H^\bullet(S_{\tilde K}, L_{\tilde\lambda}(F)))$, $\mathbb{T} \in \mathcal{C}_0(\tilde{\boldsymbol{G}}(\mathbb{A}_f)//\tilde K)_h$, is contained in F and in $\mathbb{Z}_p$ (note that $F \subseteq \mathbb{Q}_p$). Moreover, as in the proof of 2.9 Corollary we see that the eigenvalues of $\mathbb{T}_{\tilde\lambda}$, $\mathbb{T} \in \mathcal{C}_0(\tilde{\boldsymbol{G}}(\mathbb{A}_f)//\tilde K)_h$, on $H^n_{\mathrm{cusp}}(S_{\tilde K}, L_{\tilde\lambda}(\mathbb{C})) \subseteq H^n(S_{\tilde K}, L_{\tilde\lambda}(\mathbb{C}))$ are algebraic over F (note that $F \subseteq \mathbb{C}$) and integral over $\mathbb{Z}_p$, hence, they are contained in $\mathcal{O}_{\bar{\mathbb{Q}}_p}$ (note that $\bar F \subseteq \bar{\mathbb{Q}}_p$).

4.11.3. We denote by

$$\mathcal{C}_0(\tilde{\boldsymbol{G}}(\mathbb{A}_f)//\tilde K)_h^{\mathrm{reg}} \subseteq \mathcal{C}_0(\tilde{\boldsymbol{G}}(\mathbb{A}_f)//\tilde K)_h$$

the $\mathbb{Z}$-submodule generated by all Hecke operators $[\tilde K[h]_p^{-e}r\tilde K]$ with $r \in \tilde{\boldsymbol{G}}(\mathbb{A}_f)^{(p)}$ and $e \in \mathbb{N}$ (i.e., $e \geq 1$). Keeping in mind that $[\tilde{\mathcal{I}}h^{-e}\tilde{\mathcal{I}}][\tilde{\mathcal{I}}h^{-f}\tilde{\mathcal{I}}] = [\tilde{\mathcal{I}}h^{-(e+f)}\tilde{\mathcal{I}}]$ we find that $\mathcal{C}_0(\tilde{\boldsymbol{G}}(\mathbb{A}_f)//\tilde K)_h^{\mathrm{reg}}$ is an ideal in $\mathcal{C}_0(\tilde{\boldsymbol{G}}(\mathbb{A}_f)//\tilde K)_h$.

Proposition. *Let $C \in \mathbb{Q}_{>0}$. If the dominant weights $\tilde\lambda, \tilde\lambda' \in X(\tilde{\boldsymbol{T}})$ satisfy*

- $\langle\tilde\lambda, \alpha^\vee\rangle > 2C$ *and* $\langle\tilde\lambda', \alpha^\vee\rangle > 2C$ *for all* $\alpha \in \Delta_{\tilde{\boldsymbol{G}}}$,
- $\tilde\lambda \equiv \tilde\lambda' \pmod{(p-1)p^{m-1}X(\tilde{\boldsymbol{T}})}$ $(m \in \mathbb{N})$,

then for all $e \in \mathbb{N}$ and $r \in \tilde{\boldsymbol{G}}(\mathbb{A}_f)^{(p)}$ the Lefschetz number of $[\tilde K[h]_p^{-e}r\tilde K]_{\tilde\lambda}$ on $H^\bullet(S_{\tilde K}, L_{\tilde\lambda}(F))$ is contained in $\mathbb{Z}_p$ and the following congruence holds:

$$\mathrm{Lef}([\tilde K[h]_p^{-e}r\tilde K]_{\tilde\lambda}|H^\bullet(S_{\tilde K}, L_{\tilde\lambda}(F))) \equiv \mathrm{Lef}([\tilde K[h]_p^{-e}r\tilde K]_{\tilde\lambda'}|H^\bullet(S_{\tilde K}, L_{\tilde\lambda'}(F))) \pmod{p^{\lceil\min(m,Ce)\rceil}\mathbb{Z}_p}.$$

Proof. We have

$$\mathrm{Lef}([\tilde{K}[h]_p^{-e} r \tilde{K}]_{\tilde{\lambda}} | H^\bullet(S_{\tilde{K}}, L_{\tilde{\lambda}}(F))) = \sum_{\substack{i \\ i=j(i)}} \tilde{\lambda}(h^e)\ \mathrm{Lef}(\Gamma_i \zeta_i^{-1} \Gamma_i | H^\bullet(\Gamma_i \backslash X, L_{\tilde{\lambda}}(F)))$$

where $\zeta_i^{-1} \in \Delta_h$ and $\zeta_i^{-1} \in \tilde{\mathcal{I}} h^e \tilde{\mathcal{I}}$. Hence, $\tilde{\lambda}(h^e) = \hat{\lambda}(\zeta_i^{-1})$ and the claim follows from 4.10 Proposition. □

4.11.4. *Slope subspaces.* We select a Hecke operator $\mathbb{T} \in \mathcal{C}_{0,\mathbb{Z}_p}(\tilde{\boldsymbol{G}}(\mathbb{A}_f)/\tilde{K})_h^{\mathrm{reg}}$ and we denote by $H^i(S_{\tilde{K}}, L_{\tilde{\lambda}}(\mathbb{Q}_p))^\beta$ the slope β subspace of $H^i(S_{\tilde{K}}, L_{\tilde{\lambda}}(\mathbb{Q}_p))$ w.r.t. $\mathbb{T}_{\tilde{\lambda}}$. We also denote by $\boldsymbol{g}_i = \boldsymbol{g}_{i,\tilde{K}}$ the number of i-cells in a cell complex which is homotopy equivalent to the Borel–Serre compactification $\bar{S}_{\tilde{K}}$ of $S_{\tilde{K}}$.

Theorem. *For all $\beta \in \mathbb{Q}_{\geq 0}$, all dominant weights $\tilde{\lambda} \in X(\tilde{\boldsymbol{T}})$, all i and all $\mathbb{T} \in \mathcal{C}_{0,\mathbb{Z}_p}(\tilde{\boldsymbol{G}}(\mathbb{A}_f)/\tilde{K})_h^{\mathrm{reg}}$ we have*

$$\dim H^i(S_{\tilde{K}}, L_{\tilde{\lambda}}(\mathbb{Q}_p))^{\leq \beta} \leq \mathtt{m}\beta^s + \mathtt{n};$$

here, $\mathtt{m} = \mathtt{m}_{\tilde{K}} = 12 \frac{\boldsymbol{g}_i}{s} \sigma^{s+1} \in \mathbb{Q}_{\geq 0}$ *and* $\mathtt{n} = \mathtt{n}_{\tilde{K}} \in \mathbb{N}$ *is an integer which also only depends on* $\boldsymbol{g}_i, \sigma, s$ (see (50) below); *in particular*, $\mathtt{m}$ *and* $\mathtt{n}$ *only depend on* $\tilde{K}$ (*and*, *hence*, *on* $\tilde{\boldsymbol{G}}$ *and* p) *and* i, *i.e., they do not depend on* $\tilde{\lambda}$, h *and* $\mathbb{T}$.

The proof follows those of 3.9 Theorem and 3.10 Corollary. More precisely, for any $r \in \mathbb{N}_0$ we define the $\mathbb{Z}_p$-submodule

$$H^i(S_{\tilde{K}}, L_{\tilde{\lambda}}(\mathbb{Z}_p, r, \sigma)) = \bigoplus_{j=1}^{t} H^i(\Gamma_j, L_{\tilde{\lambda}}(\mathbb{Z}_p, r, \sigma))$$

of $H^i(S_{\tilde{K}}, L_{\tilde{\lambda}}(\mathbb{Z}_p)) = \bigoplus_{j=1}^t H^i(\Gamma_j, L_{\tilde{\lambda}}(\mathbb{Z}_p))$ (note that $\Gamma_j \subseteq \tilde{\mathcal{I}}$). Using the decomposition $[\tilde{K}[h]_p^{-e} r \tilde{K}] = \bigoplus_i \Gamma_i \zeta_i^{-1} \Gamma_{j(i)}$ where $\zeta_i^{-1} \in \tilde{\mathcal{I}} h^e \tilde{\mathcal{I}}$ and following the proof in 3.5 Lemma we see that the submodule $H^i(S_{\tilde{K}}, L_{\tilde{\lambda}}(\mathbb{Z}_p, r, \sigma))$ satisfies the following properties.

- $H^i(S_{\tilde{K}}, L_{\tilde{\lambda}}(\mathbb{Z}_p))^{\mathrm{TF}}$ is $\mathbb{T}_{\tilde{\lambda}}$-invariant.
- $\mathbb{T}_{\tilde{\lambda}} H^i(S_{\tilde{K}}, L_{\tilde{\lambda}}(\mathbb{Z}_p, r, \sigma))^{\mathrm{TF}} \subseteq p^r H^i(S_{\tilde{K}}, L_{\tilde{\lambda}}(\mathbb{Z}_p))^{\mathrm{TF}}$.

We denote by $(p^{b_l})_l$ the elementary divisors of the quotient

$$\frac{H^i(S_{\tilde{K}}, L_{\tilde{\lambda}}(\mathbb{Z}_p))^{\mathrm{TF}}}{H^i(S_{\tilde{K}}, L_{\tilde{\lambda}}(\mathbb{Z}_p, r, \sigma))^{\mathrm{TF}}} \tag{49}$$

in decreasing order, i.e., $b_1 \geq b_2 \geq b_3 \geq \cdots$. As in Section 3.9.1 we see that (49) is a subquotient of $\bigoplus_j H^i(\Gamma_j, \boldsymbol{L}_{\tilde{\lambda}}^{[r]}(\mathbb{Z}_p, \sigma))$, which is a subquotient of

$$\bigoplus_j \bigoplus_{h=0}^{r} \left(\frac{\mathbb{Z}_p}{p^{r-h}\mathbb{Z}_p} \right)^{g_{i,\Gamma_j} M_{\sigma,h}} = \bigoplus_{h=0}^{r} \left(\frac{\mathbb{Z}_p}{p^{r-h}\mathbb{Z}_p} \right)^{\boldsymbol{g}_i M_{\sigma,h}}.$$

We denote by $(p^{a_l})_l$ the elementary divisors of the latter sum in decreasing order, i.e., $a_1 \geq a_2 \geq \cdots$ (the elementary divisor p^{r-h} appears $\boldsymbol{g}_i M_{\sigma,h}$-times). We obtain $b_i \leq a_i \leq r$ for all i and following the arguments in Section 3.9.2 - 3.9.4 with $\boldsymbol{g}_i$ in place of g_i we obtain that the piecewise linear function defined in 3.9 Theorem but with g_i replaced by $\boldsymbol{g}_i$ is a lower bound for the Newton polygon of $\mathbb{T}_{\tilde{\lambda}}$ acting on $H^i(S_{\tilde{K}}, L_{\tilde{\lambda}}(\mathbb{Q}_p))^{<\infty}$. Following the proof of 3.10 Corollary we find the bound for the dimension of the slope β subspace as in the statement of the theorem where we can choose

$$\mathtt{n} = \left\lceil 2^{\frac{1}{s+1}} \frac{\boldsymbol{g}_i \sigma^{s+1}}{s} M^s \right\rceil + 1 \in \mathbb{N}; \tag{50}$$

here, M is defined in equations (23) and (24) (note that the definition of M does not depend on g_i). This finishes the proof of the theorem.

4.12. ***Induced representations.*** We look at traces of Hecke operators on induced representations. This will be applied in the next section where we consider Hecke operators on Eisenstein cohomology.

4.12.1. We select a $\mathbb{Q}$-parabolic subgroup $\tilde{\boldsymbol{Q}} \leq \tilde{\boldsymbol{G}}$ with Levi decomposition $\tilde{\boldsymbol{Q}} = \tilde{\boldsymbol{M}}\tilde{\boldsymbol{N}}$ and a representation π of $\tilde{\boldsymbol{M}}(\mathbb{A}_f)$. By $\mathrm{Ind}_{\tilde{\boldsymbol{Q}}(\mathbb{A}_f)}^{\tilde{\boldsymbol{G}}(\mathbb{A}_f)}\pi$ we understand the non-unitarily induced representation. We select a maximal compact open subgroup $\tilde{K}_0 = \prod_{\ell \neq \infty} \tilde{K}_{0,\ell} \leq \tilde{\boldsymbol{G}}(\mathbb{A}_f)$ and we may assume that $\tilde{K} \leq \tilde{K}_0$. We set

$$\tilde{K}^{\tilde{M}} = \prod_{\ell \neq \infty} \tilde{K}_\ell^{\tilde{M}} \quad \text{and} \quad \tilde{K}^{(p),\tilde{M}} = \prod_{\ell \neq p,\infty} \tilde{K}_\ell^{\tilde{M}},$$

where $\tilde{K}_\ell^{\tilde{M}} = \tilde{K}_\ell \cap \tilde{\boldsymbol{M}}(\mathbb{Q}_\ell)$, and we use the same definition with $\tilde{K}$ replaced by $\tilde{K}_0$. We also set $\tilde{K}^{\tilde{N}} = \tilde{K} \cap \tilde{\boldsymbol{N}}(\mathbb{A}_f)$ and $\tilde{K}^{(p),\tilde{N}} = \tilde{K}^{(p)} \cap \tilde{\boldsymbol{N}}(\mathbb{A}_f)^{(p)}$.

Let $f \in \mathcal{C}_{0,\mathbb{Q}}(\tilde{\boldsymbol{G}}(\mathbb{A}_f))$; we define the constant term $f_{\tilde{\boldsymbol{M}}} \in \mathcal{C}_{0,\mathbb{Q}}(\tilde{\boldsymbol{M}}(\mathbb{A}_f))$ by

$$f_{\tilde{\boldsymbol{M}}}(x) = \int_{\tilde{K}_0} \int_{\tilde{\boldsymbol{N}}(\mathbb{A}_f)} f(k^{-1}xnk)\, dn\, dk.$$

Here and below we normalize Haar measures on $\tilde{\boldsymbol{G}}(\mathbb{A}_f)$ (and, hence, on $\tilde{K}_0$), $\tilde{\boldsymbol{M}}(\mathbb{A}_f)$ and $\tilde{\boldsymbol{N}}(\mathbb{A}_f)$ so that $\mathrm{vol}(\tilde{K}) = 1$, $\mathrm{vol}(\tilde{K}^{\tilde{M}}) = 1$, $\mathrm{vol}(\tilde{K}^{\tilde{N}}) = 1$. With these conventions we have

$$\mathrm{tr}(f|\mathrm{Ind}_{\tilde{\boldsymbol{Q}}(\mathbb{A}_f)}^{\tilde{\boldsymbol{G}}(\mathbb{A}_f)}\pi) = \mathrm{tr}(f_{\tilde{\boldsymbol{M}}}|\pi).$$

If $f^{(p)} \in \mathcal{C}_{0,\mathbb{Q}}(\tilde{\boldsymbol{G}}(\mathbb{A}_f)^{(p)})$ we define $f^{(p)}_{\tilde{\boldsymbol{M}}} \in \mathcal{C}_{0,\mathbb{Q}}(\tilde{\boldsymbol{M}}(\mathbb{A}_f)^{(p)})$ by replacing f with $f^{(p)}$, $\tilde{K}_0$ with $\tilde{K}_0^{(p)} = \prod_{\ell \neq p,\infty} \tilde{K}_{0,\ell}$ and $\tilde{\boldsymbol{N}}(\mathbb{A}_f)$ with $\tilde{\boldsymbol{N}}(\mathbb{A}_f)^{(p)}$; an analogous identity for the trace of $f^{(p)}$ on representations induced from $\tilde{\boldsymbol{M}}(\mathbb{A}_f)^{(p)}$ then holds. We will also need the following classical identity which holds for a representation π_p of $\tilde{\boldsymbol{M}}(\mathbb{Q}_p)$. Assume that $\tilde{\boldsymbol{Q}}/F$ contains $\tilde{\boldsymbol{B}}^-/F$, i.e., the unipotent radical of $\tilde{\boldsymbol{Q}}/F$

is generated by root subgroups attached to negative roots (compare the definition of $\mathcal{I}$ in Section 1.3); then

$$\mathrm{tr}([\tilde{\mathcal{I}}h^{-e}\tilde{\mathcal{I}}]|(\mathrm{Ind}_{\tilde{\boldsymbol{Q}}(\mathbb{Q}_p)}^{\tilde{\boldsymbol{G}}(\mathbb{Q}_p)}\pi_p)^{\tilde{\mathcal{I}}}) = \sum_{v\in W^{\tilde{Q}}} c_{v,h^{-e}}\mathrm{tr}([\tilde{\mathcal{I}}^{\tilde{M}}vh^{-e}v^{-1}\tilde{\mathcal{I}}^{\tilde{M}}]|\pi_p^{\tilde{\mathcal{I}}^{\tilde{M}}}), \tag{51}$$

where $c_{v,h^{-e}}\in\mathbb{N}$ and $\tilde{\mathcal{I}}^{\tilde{M}}=\tilde{\mathcal{I}}\cap\tilde{\boldsymbol{M}}(\mathbb{Q}_p)$ (see [Urban 2011, p. 1751], for example). We obtain an analogous formula when $\tilde{\boldsymbol{Q}}$ is standard parabolic, i.e., $\tilde{\boldsymbol{B}}/F\leq\tilde{\boldsymbol{Q}}/F$. To this end we denote by $a\in W_{\tilde{\boldsymbol{G}}}$ and $b\in W_{\tilde{\boldsymbol{M}}}=W(\tilde{\boldsymbol{T}}/\mathbb{Z}_p,\tilde{\boldsymbol{M}}/\mathbb{Z}_p)$ the longest elements; i.e., a maps $\Phi^+_{\tilde{\boldsymbol{G}}}$ to $\Phi^-_{\tilde{\boldsymbol{G}}}$ and b maps $\Phi^+_{\tilde{\boldsymbol{M}}}$ to $\Phi^-_{\tilde{\boldsymbol{M}}}$. Using (51) but with positivity defined by $-\Delta_{\tilde{\boldsymbol{G}}}$ — that is, $\tilde{\boldsymbol{B}}/F\leq\tilde{\boldsymbol{Q}}/F$, $\tilde{\mathcal{I}}\equiv\tilde{\boldsymbol{B}}(\mathbb{F}_p)\pmod p$ and $h\in\tilde{\boldsymbol{T}}(\mathbb{Q})^{--}$ — we obtain

$$\begin{aligned}
\mathrm{tr}([\tilde{\mathcal{I}}h^{-e}\tilde{\mathcal{I}}]|(\mathrm{Ind}_{\tilde{\boldsymbol{Q}}(\mathbb{Q}_p)}^{\tilde{\boldsymbol{G}}(\mathbb{Q}_p)}\pi_p)^{\tilde{\mathcal{I}}}) &= \mathrm{tr}([{}^a\tilde{\mathcal{I}}({}^ah^{-e})\,{}^a\tilde{\mathcal{I}}]|(\mathrm{Ind}_{\tilde{\boldsymbol{Q}}(\mathbb{Q}_p)}^{\tilde{\boldsymbol{G}}(\mathbb{Q}_p)}\pi_p)^{{}^a\tilde{\mathcal{I}}})\\
&=\sum_{v\in W^{\tilde{Q}}} c_{v,h^{-e}}\mathrm{tr}([({}^a\tilde{\mathcal{I}})^{\tilde{M}}v({}^ah^{-e})v^{-1}({}^a\tilde{\mathcal{I}})^{\tilde{M}}]|\pi_p^{({}^a\tilde{\mathcal{I}})^{\tilde{M}}})\\
&=\sum_{v\in W^{\tilde{Q}}} c_{v,h^{-e}}\mathrm{tr}([{}^b({}^a\tilde{\mathcal{I}}^{\tilde{M}})bv({}^ah^{-e})v^{-1}b^{-1}\,{}^b({}^a\tilde{\mathcal{I}}^{\tilde{M}})]|\pi_p^{{}^b({}^a\tilde{\mathcal{I}}^{\tilde{M}})})\\
&=\sum_{v\in W^{\tilde{Q}}} c_{v,h^{-e}}\mathrm{tr}([\tilde{\mathcal{I}}^{\tilde{M}}({}^{bva}h^{-e})\tilde{\mathcal{I}}^{\tilde{M}}]|\pi_p^{\tilde{\mathcal{I}}^{\tilde{M}}}),
\end{aligned}$$

where $c_v=c_{v,h^{-e}}\in\mathbb{Z}$ are certain integers.

4.12.2. We look more closely at the constant term $f^{(p)}_{\tilde{\boldsymbol{M}}}$ of $f^{(p)}\in\mathcal{C}_0(\tilde{\boldsymbol{G}}(\mathbb{A}_f)^{(p)}//\tilde{K}^{(p)})$, i.e., $f^{(p)}$ is $\mathbb{Z}$-valued and $\tilde{K}^{(p)}$ bi-invariant (note that $\mathrm{vol}(\tilde{K}^{(p)})=1$ by our normalizations). For simplicity we assume that $\tilde{K}^{(p)}\leq\tilde{K}^{(p)}_0$ is a normal subgroup. The definition yields

$$\begin{aligned}
f^{(p)}_{\tilde{\boldsymbol{M}}}(x) &= \sum_{k\in\tilde{K}^{(p)}_0/\tilde{K}^{(p)}}\int_{\tilde{\boldsymbol{N}}(\mathbb{A}_f)^{(p)}} f^{(p)}(k^{-1}xnk)\,dn\\
&=\mathrm{vol}(\tilde{K}^{(p),\tilde{N}})\sum_{k\in\tilde{K}^{(p)}_0/\tilde{K}^{(p)}}\int_{\tilde{\boldsymbol{N}}(\mathbb{A}_f)^{(p)}/\tilde{K}^{(p),\tilde{N}}} f^{(p)}(k^{-1}xnk)\,dn,
\end{aligned}$$

The first of these equalities shows that $f^{(p)}_{\tilde{\boldsymbol{M}}}$ is $\tilde{K}^{(p),\tilde{M}}$ bi-invariant (the modulus $\delta_{\tilde{\boldsymbol{Q}}(\mathbb{A}_f)^{(p)}}(m)$ vanishes for $m\in\tilde{K}^{(p),\tilde{M}}$ because $\tilde{K}^{(p),\tilde{M}}$ is contained in the compact group $\tilde{K}^{(p)}\cap\tilde{\boldsymbol{M}}(\mathbb{A}_f)^{(p)}$). Since $\mathrm{vol}(\tilde{K}^{(p),\tilde{N}})=\mathrm{vol}(\tilde{K}^{(p),\tilde{M}})=1$ the second equality implies

$$f^{(p)}_{\tilde{\boldsymbol{M}}}\in\mathcal{C}_0(\tilde{\boldsymbol{M}}(\mathbb{A}_f)^{(p)}//\tilde{K}^{(p),\tilde{M}}). \tag{52}$$

4.13. *Eisenstein cohomology.*

4.13.1. We assume that the highest weight $\tilde{\lambda}\in X(\tilde{\boldsymbol{T}})$ is dominant and regular. The full cohomology then decomposes as

$$H^\bullet(S_{\tilde K}, L_{\tilde\lambda}(\mathbb{C})) = H^\bullet_{\mathrm{cusp}}(S_{\tilde K}, L_{\tilde\lambda}(\mathbb{C})) \oplus H^\bullet_{\mathrm{Eis}}(S_{\tilde K}, L_{\tilde\lambda}(\mathbb{C})),$$

where

$$H^\bullet_{\mathrm{Eis}}(S_{\tilde K}, L_{\tilde\lambda}(\mathbb{C})) = \bigoplus_{\tilde{\boldsymbol{M}}\neq\tilde{\boldsymbol{G}}} \bigoplus_{w} \big(\mathrm{Ind}_{\tilde{\boldsymbol{Q}}(\mathbb{A}_f)}^{\tilde{\boldsymbol{G}}(\mathbb{A}_f)} H^{\bullet-\ell(w)}_{\mathrm{cusp}}(S^{\tilde{\boldsymbol{M}}}, L_{w\cdot\tilde\lambda}(\mathbb{C}))\big)^{\tilde K}; \tag{53}$$

here, $\tilde{\boldsymbol{M}}$ runs over a system of representatives of $\tilde{\boldsymbol{G}}(\mathbb{Q})$-conjugacy classes of proper $\mathbb{Q}$-Levi subgroups of $\tilde{\boldsymbol{G}}$, i.e., there is a (standard) $\mathbb{Q}$-parabolic subgroup $\tilde{\boldsymbol{Q}} \leq \tilde{\boldsymbol{G}}$ with Levi decomposition

$$\tilde{\boldsymbol{Q}} = \tilde{\boldsymbol{M}}\tilde{\boldsymbol{N}}, \quad S^{\tilde{\boldsymbol{M}}} = \tilde{\boldsymbol{M}}(\mathbb{Q})\backslash\tilde{\boldsymbol{M}}(\mathbb{A})/A_{\tilde{\boldsymbol{Q}}}\tilde K^{\tilde M}_\infty,$$

where $\tilde K^{\tilde M}_\infty = \tilde K_\infty \cap \tilde{\boldsymbol{M}}(\mathbb{R})$, is the locally symmetric space attached to $\tilde{\boldsymbol{M}}$, $A_{\tilde{\boldsymbol{Q}}}$ is the connected component of the real points of a maximal $\mathbb{Q}$-split torus $\boldsymbol{A}_{\tilde{\boldsymbol{Q}}}$ in the center of $\tilde{\boldsymbol{M}}$ and w runs over those elements in $W^{\tilde Q}$ which satisfy the condition that $-w(\tilde\lambda+\rho^\circ)|_{A_{\tilde{\boldsymbol{Q}}}}$ is nonnegative, i.e., $\langle -\mathrm{Re}(w(\tilde\lambda+\rho^\circ))|_{A_{\tilde{\boldsymbol{Q}}}}, \alpha^\vee\rangle \geq 0$ for all roots α of $A_{\tilde{\boldsymbol{Q}}}$ acting on $\mathrm{Lie}(\tilde{\boldsymbol{N}}) \otimes \mathbb{R}$ (see [Franke 1998, Theorem 19 II, p. 257] and [Schwermer 1994, Proof of 6.3 Theorem, p. 505]).

4.13.2. Theorem. *Let $C \in \mathbb{Q}_{>0}$ and suppose the dominant weights $\tilde\lambda, \tilde\lambda' \in X(\tilde{\boldsymbol{T}})$ satisfy*

- $\langle\tilde\lambda, \alpha^\vee\rangle > 2C$ *and* $\langle\tilde\lambda', \alpha^\vee\rangle > 2C$ *for all* $\alpha \in \Delta_{\tilde{\boldsymbol{G}}}$,
- $\tilde\lambda \equiv \tilde\lambda' \pmod{(p-1)p^{m-1}X(\tilde{\boldsymbol{T}})}$ $(m \in \mathbb{N})$.

Then, for all $e \in \mathbb{N}$ and $r \in \tilde{\boldsymbol{G}}(\mathbb{A}_f)^{(p)}$, the Lefschetz number of $[\tilde K[h]_p^{-e} r\tilde K]_{\tilde\lambda}$ on $H^\bullet_{\mathrm{cusp}}(S_{\tilde K}, L_{\tilde\lambda}(\mathbb{C}))$ is contained in F and the following congruence holds (*note that* $F \subseteq \mathbb{Q}_p$):

$$\mathrm{Lef}([\tilde K[h]_p^{-e} r\tilde K]_{\tilde\lambda} | H^\bullet_{\mathrm{cusp}}(S_{\tilde K}, L_{\tilde\lambda}(\mathbb{C}))) \equiv \mathrm{Lef}([\tilde K[h]_p^{-e} r\tilde K]_{\tilde\lambda'} | H^\bullet_{\mathrm{cusp}}(S_{\tilde K}, L_{\tilde\lambda'}(\mathbb{C}))) \pmod{p^\dagger\mathbb{Z}_p}.$$

Here, $\dagger = \big\lceil \min\big(m, e(C-\kappa_2\mathrm{rk}(\tilde{\boldsymbol{G}}))\big)\big\rceil - e\kappa_1\mathrm{rk}(\tilde{\boldsymbol{G}})$ *with* $\kappa_i = \kappa_{i,\tilde{\boldsymbol{G}}}$ *and* $\mathrm{rk}(\tilde{\boldsymbol{G}}) = \mathrm{rk}_{\mathbb{Q}}(\tilde{\boldsymbol{G}})$ *is the $\mathbb{Q}$-rank of $\tilde{\boldsymbol{G}}$.*

Proof. We use induction on the $\mathbb{Q}$-rank of $\tilde{\boldsymbol{G}}$. If $\mathrm{rk}(\tilde{\boldsymbol{G}}) = 0$ then

$$H^\bullet_{\mathrm{cusp}}(S_{\tilde K}, L_{\tilde\lambda}(\mathbb{C})) = H^\bullet(S_{\tilde K}, L_{\tilde\lambda}(\mathbb{C}));$$

hence, 4.11.3 Proposition implies the claim. We assume $\mathrm{rk}(\tilde{\boldsymbol{G}}) > 0$. Equation (53) yields

$$\begin{aligned}(54)\quad &\mathrm{Lef}([\tilde K[h]_p^{-e} r \tilde K] | H^\bullet_{\mathrm{Eis}}(S_{\tilde K}, L_{\tilde\lambda}(\mathbb{C})))\\ &= \sum_{\tilde{\boldsymbol M}\neq\tilde{\boldsymbol G}} \sum_w \sum_i (-1)^{i+\ell(w)} \mathrm{tr}\big([\tilde K[h]_p^{-e} r \tilde K] | \big(\mathrm{Ind}_{\tilde{\boldsymbol Q}(\mathbb{A}_f)}^{\tilde G(\mathbb{A}_f)} H^i_{\mathrm{cusp}}(S^{\tilde{\boldsymbol M}}, L_{w\cdot\tilde\lambda}(\mathbb{C}))\big)^{\tilde K}\big)\\ &= \sum_{\tilde{\boldsymbol M}\neq\tilde{\boldsymbol G}} \sum_w (-1)^{\ell(w)} \mathrm{Lef}\big([\tilde K[h]_p^{-e} r \tilde K] | \bigoplus_i (\mathrm{Ind}_{\tilde{\boldsymbol Q}(\mathbb{A}_f)}^{\tilde G(\mathbb{A}_f)} \pi^i_{w\cdot\tilde\lambda})^{\tilde K}\big).\end{aligned}$$

Here, we have set

$$\pi^i_{w\cdot\tilde\lambda} = H^i_{\mathrm{cusp}}(S^{\tilde{\boldsymbol M}}, L_{w\cdot\tilde\lambda}(\mathbb{C}));$$

this is a module under the Hecke algebra attached to $\tilde{\boldsymbol M}(\mathbb{A}_f)$ and we have

$$H^i_{\mathrm{cusp}}(S^{\tilde{\boldsymbol M}}, L_{w\cdot\tilde\lambda}(\mathbb{C}))^{\tilde H} = H^i_{\mathrm{cusp}}(S^{\tilde{\boldsymbol M}}/\tilde H, L_{w\cdot\tilde\lambda}(\mathbb{C})),$$

with $\tilde H \leq \tilde{\boldsymbol M}(\mathbb{A}_f)$ compact open. We select a proper $\mathbb{Q}$-parabolic subgroup $\tilde{\boldsymbol Q} = \tilde{\boldsymbol M}\tilde{\boldsymbol N}$ of $\tilde{\boldsymbol G}$ and an element $w \in W^{\tilde Q}$ as in equation (53). We denote by $\Phi_{\tilde{\boldsymbol M}}$ the set of roots of $\tilde{\boldsymbol T}/F$ acting on $\mathrm{Lie}(\tilde{\boldsymbol M}/F)$ and we set $\Delta_{\tilde{\boldsymbol M}} = \Phi_{\tilde{\boldsymbol M}} \cap \Delta_{\tilde{\boldsymbol G}}$. The set $\Delta_{\tilde{\boldsymbol M}}$ is the basis for the root system $\Phi_{\tilde{\boldsymbol M}}$ corresponding to the Borel subgroup $\tilde{\boldsymbol B}^M/F = \tilde{\boldsymbol B}/F \cap \tilde{\boldsymbol M}/F$ of $\tilde{\boldsymbol M}/F$; in particular, this determines the set of positive roots $\Phi^+_{\tilde{\boldsymbol M}}$. The subgroup $\tilde{\mathcal I}^{\tilde M} = \tilde{\mathcal I} \cap \tilde{\boldsymbol M}(\mathbb{Q}_p)$ is a Iwahori subgroup in $\tilde{\boldsymbol M}(\mathbb{Z}_p)$, i.e., $\tilde{\mathcal I}^{\tilde M} \pmod p$ is contained in the Borel subgroup $(\tilde{\boldsymbol B}^- \cap \tilde{\boldsymbol M})(\mathbb{F}_p) \leq \tilde{\boldsymbol M}(\mathbb{F}_p)$. The identities in Section 4.12.1 for the traces of induced representations and equation (52) yield

$$\begin{aligned}(55)\quad &\mathrm{Lef}\big([\tilde K[h]_p^{-e} r \tilde K] | \bigoplus_i \big(\mathrm{Ind}_{\tilde{\boldsymbol Q}(\mathbb{A}_f)}^{\tilde G(\mathbb{A}_f)} \pi^i_{w\cdot\tilde\lambda}\big)^{\tilde K}\big)\\ &= \mathrm{Lef}\big([\tilde{\mathcal I} h^{-e} \tilde{\mathcal I}] \otimes [\tilde K^{(p)} r \tilde K^{(p)}] | \bigoplus_i \mathrm{Ind}_{\tilde{\boldsymbol Q}(\mathbb{A}_f)}^{\tilde G(\mathbb{A}_f)} \pi^i_{w\cdot\tilde\lambda}\big)\\ &= \sum_{v\in W^{\tilde Q}} c_v\, \mathrm{Lef}\big([\tilde{\mathcal I}^{\tilde M}({}^{bva}h^{-e})\tilde{\mathcal I}^{\tilde M}] \otimes [\tilde K^{(p)} r \tilde K^{(p)}]_{\tilde{\boldsymbol M}} | \bigoplus_i \pi^i_{w\cdot\tilde\lambda}\big)\\ &= \sum_{v\in W^{\tilde Q}} c_v\, \mathrm{Lef}\big([\tilde{\mathcal I}^{\tilde M}({}^{bva}h^{-e})\tilde{\mathcal I}^{\tilde M}] \otimes [\tilde K^{(p)} r \tilde K^{(p)}]_{\tilde{\boldsymbol M}} | \bigoplus_i (\pi^i_{w\cdot\tilde\lambda})^{\tilde K_{\tilde M}}\big),\end{aligned}$$

where $c_v = c_{v,h^{-e}} \in \mathbb{Z}$, a (resp. b) is the longest element in the Weyl group $W_{\tilde{\boldsymbol G}}$ (resp. $W_{\tilde{\boldsymbol M}}$) and $\tilde K^{\tilde M} = \tilde{\mathcal I}^{\tilde M} \times \tilde K^{(p),\tilde M}$. We select an element $v \in W^{\tilde Q}$. Since $v^{-1}\Phi^+_{\tilde{\boldsymbol M}} \subseteq \Phi^+_{\tilde{\boldsymbol G}}$ and ${}^a h \in \tilde{\boldsymbol T}(\mathbb{Q}_p)^{--}$ we obtain for all $\alpha \in \Phi^+_{\tilde{\boldsymbol M}}$ that $v_p(\alpha({}^{bva}h)) = v_p(({}^{v^{-1}b^{-1}}\alpha)({}^a h)) > 0$; hence, ${}^{bva}h$ is regular dominant w.r.t. $\Phi^+_{\tilde{\boldsymbol M}}$. Thus, using equation (52) we obtain that

$$(56)\qquad [\tilde{\mathcal I}^{\tilde M}({}^{bva}h^{-e})\tilde{\mathcal I}^{\tilde M}] \otimes [\tilde K^{(p)} r \tilde K^{(p)}]_{\tilde{\boldsymbol M}} \in \mathcal{C}_0(\tilde{\boldsymbol M}(\mathbb{A}_f)//\tilde K^{\tilde M})_{{}^{bva}h}$$

is contained in the integral Hecke algebra attached to $\tilde{\boldsymbol M}$ and the dominant regular element ${}^{bva}h \in \tilde{\boldsymbol T}(F)$. Now, in Section 4.6 we have seen that $\langle {}^w\rho - \rho, \alpha^\vee\rangle \geq -2\kappa_2$

for all $\alpha \in \Delta_{\tilde{G}}$ ($\rho = \rho_{\tilde{G}}$); hence, we obtain for all $\alpha \in \Delta_{\tilde{M}}$

$$\langle w \cdot \tilde{\lambda}, \alpha^\vee \rangle = \langle \tilde{\lambda}, {}^{w^{-1}}\alpha^\vee \rangle + \langle {}^w\rho^\circ - \rho^\circ, \alpha^\vee \rangle \geq 2C - 2\kappa_2.$$

Since also $w \cdot \tilde{\lambda} \equiv w \cdot \tilde{\lambda}'$ (mod $(p-1)p^m X(\tilde{T})$) the induction hypotheses for the group $\tilde{M}$ which has strictly smaller rank than $\tilde{G}$ (since $\tilde{M} \neq \tilde{G}$) is satisfied and we obtain that the following congruence holds:

$$(57)\quad (w \cdot \tilde{\lambda})({}^{bva}h^e)\ \mathrm{Lef}\big([\tilde{\mathcal{I}}^{\tilde{M}}({}^{bva}h^{-e})\tilde{\mathcal{I}}^{\tilde{M}}] \otimes [\tilde{K}^{(p)} r \tilde{K}^{(p)}]_{\tilde{M}} | \bigoplus_i (\pi^i_{w\cdot\tilde{\lambda}})^{\tilde{K}^{\tilde{M}}}\big)$$
$$\equiv (w \cdot \tilde{\lambda}')({}^{bva}h^e)\ \mathrm{Lef}\big([\tilde{\mathcal{I}}^{\tilde{M}}({}^{bva}h^{-e})\tilde{\mathcal{I}}^{\tilde{M}}] \otimes [\tilde{K}^{(p)} r \tilde{K}^{(p)}]_{\tilde{M}} | \bigoplus_i (\pi^i_{w\cdot\tilde{\lambda}'})^{\tilde{K}^{\tilde{M}}}\big) \pmod{p^\star \mathbb{Z}_p},$$

where $\star = \lceil \min(m, e(C - \kappa_2 - \kappa_{2,\tilde{M}} \mathrm{rk}(\tilde{M}))) \rceil - e\kappa_{1,\tilde{M}} \mathrm{rk}(\tilde{M})$; moreover, the above Lefschetz numbers are contained in F. Equations (54) and (55) thus imply that

$$\tilde{\lambda}(h^e)\ \mathrm{Lef}([\tilde{K}[h]_p^{-e} r \tilde{K}] | H^\bullet_{\mathrm{Eis}}(S_{\tilde{K}}, L_{\tilde{\lambda}}(\mathbb{C}))) \in F$$

(note that $\tilde{\lambda}$ and all $\alpha \in \Phi_{\tilde{G}}$ are defined over F, $h \in \tilde{T}(\mathbb{Q})$ and a, b, v, w normalize $\tilde{T}(F)$) and since full cohomology is defined over F we obtain that

$$\mathrm{Lef}([\tilde{K}[h]_p^{-e} r \tilde{K}]_{\tilde{\lambda}} | H^\bullet_{\mathrm{cusp}}(S_{\tilde{K}}, L_{\tilde{\lambda}}(\mathbb{C}))) \in F.$$

It remains to prove the congruences. Since $\tilde{\lambda} - (bva)^{-1}(w \cdot \tilde{\lambda}) = \tilde{\lambda} - ((bva)^{-1}w) \cdot \tilde{\lambda} + (bva)^{-1}\rho^\circ - \rho^\circ$ we obtain using 4.5 Lemma (note that $h \in \tilde{T}(\mathbb{Q}_p)^{++}$) and $v_p(((bva)^{-1}\rho^\circ - \rho^\circ)(h^e)) \geq -e\kappa_1$ (see Section 4.6) that

$$(58)\qquad \frac{\tilde{\lambda}(h^e)}{(bva)^{-1}(w \cdot \tilde{\lambda})(h^e)} \begin{cases} \text{has } p\text{-adic value } \geq Ce - e\kappa_1 & \text{if } bva \neq w, \\ \text{equals } ((bva)^{-1}\rho^\circ - \rho^\circ)(h^e) & \text{if } bva = w, \end{cases}$$

and the same holds for $\tilde{\lambda}'$. Thus, if $bva = w$, we obtain from equation (57), after multiplying by $((bva)^{-1}\rho^\circ - \rho^\circ)(h^e)$,

$$\tilde{\lambda}(h^e)\,\mathrm{Lef}\big([\tilde{\mathcal{I}}^{\tilde{M}}({}^{bva}h^{-e})\tilde{\mathcal{I}}^{\tilde{M}}] \otimes [\tilde{K}^{(p)} r \tilde{K}^{(p)}]_{\tilde{M}} | \bigoplus_i (\pi^i_{w\cdot\tilde{\lambda}})^{\tilde{K}^{\tilde{M}}}\big)$$
$$\equiv \tilde{\lambda}'(h^e)\,\mathrm{Lef}\big([\tilde{\mathcal{I}}^{\tilde{M}}({}^{bva}h^{-e})\tilde{\mathcal{I}}^{\tilde{M}}] \otimes [\tilde{K}^{(p)} r \tilde{K}^{(p)}]_{\tilde{M}} | \bigoplus_i (\pi^i_{w\cdot\tilde{\lambda}'})^{\tilde{K}^{\tilde{M}}}\big) \pmod{p^{\star - e\kappa_1}\mathbb{Z}_p}.$$

We look at the case $bva \neq w$. Since

$$[\tilde{K}^{(p)} r \tilde{K}^{(p)}]_{\tilde{M}} = \sum_{s \in \tilde{M}(\mathbb{A}_f)^{(p)}} z_s [\tilde{K}^{(p),\tilde{M}} s \tilde{K}^{(p),\tilde{M}}]$$

with $z_s \in \mathbb{Z}$, by equation (52), and since the trace of the $w \cdot \tilde{\lambda}$-normalization of

$$[\tilde{\mathcal{I}}^{\tilde{M}}({}^{bva}h^{-e})\tilde{\mathcal{I}}^{\tilde{M}}] \otimes [\tilde{K}^{(p),\tilde{M}} s \tilde{K}^{(p),\tilde{M}}] \in C_0(\tilde{M}(\mathbb{A}_f)//\tilde{K}^{\tilde{M}})_{{}^{bva}h}$$

on $H^i_{\mathrm{cusp}}(S^{\tilde{M}}, L_{w\cdot\tilde{\lambda}}(\mathbb{C}))$ is contained in $\mathcal{O}_{\bar{\mathbb{Q}}_p}$ (see Section 4.11.2) we obtain that the Lefschetz number of $(w \cdot \tilde{\lambda})({}^{bva}h^e)\, [\tilde{\mathcal{I}}^{\tilde{M}}({}^{bva}h^{-e})\tilde{\mathcal{I}}^{\tilde{M}}] \otimes [\tilde{K}^{(p)} r \tilde{K}^{(p)}]_{\tilde{M}}$ on cuspidal cohomology $\bigoplus_i (\pi^i_{w\cdot\tilde\lambda})^{\tilde K_{\tilde M}}$ is p-adically integral; thus using equation (58) we obtain

$$\begin{aligned}&\tilde\lambda(h^e)\,\mathrm{Lef}\big([\tilde{\mathcal{I}}^{\tilde M}({}^{bva}h^{-e})\tilde{\mathcal{I}}^{\tilde M}]\otimes[\tilde K^{(p)}r\tilde K^{(p)}]_{\tilde{\boldsymbol M}}\,|\,\bigoplus_i(\pi^i_{w\cdot\tilde\lambda})^{\tilde K_{\tilde M}}\big)\\ &\equiv \tilde\lambda'(h^e)\,\mathrm{Lef}\big([\tilde{\mathcal{I}}^{\tilde M}({}^{bva}h^{-e})\tilde{\mathcal{I}}^{\tilde M}]\otimes[\tilde K^{(p)}r\tilde K^{(p)}]_{\tilde{\boldsymbol M}}\,|\,\bigoplus_i(\pi^i_{w\cdot\tilde\lambda'})^{\tilde K_{\tilde M}}\big)\pmod{p^{Ce-e\kappa_1}\mathbb{Z}_p}.\end{aligned}$$

Since $Ce-e\kappa_1\ge\star-e\kappa_1$, in both cases the congruence holds modulo $p^{\star-e\kappa_1}\mathbb{Z}_p$. Using equations (54) and (55) we obtain

$$\begin{aligned}&\mathrm{Lef}([\tilde K[h]_p^{-e}r\tilde K]_{\tilde\lambda}\,|\,H^\bullet_{\mathrm{Eis}}(S_{\tilde K},L_{\tilde\lambda}(\mathbb{C})))\\ &\qquad\equiv\mathrm{Lef}([\tilde K[h]_p^{-e}r\tilde K]_{\tilde\lambda'}\,|\,H^\bullet_{\mathrm{Eis}}(S_{\tilde K},L_{\tilde\lambda'}(\mathbb{C})))\pmod{p^{\star-e\kappa_1}\mathbb{Z}_p}.\end{aligned}$$

The rank of $\tilde{\boldsymbol M}$ is strictly smaller than the rank of $\tilde{\boldsymbol G}$ and $\kappa_{i,\tilde{\boldsymbol M}}\le\kappa_i$, as follows from Section 4.6; hence, $\kappa_i+\kappa_{i,\tilde{\boldsymbol M}}\mathrm{rk}(\tilde{\boldsymbol M})\le\kappa_i\mathrm{rk}(\tilde{\boldsymbol G})$ which yields

$$\star-e\kappa_1\ge\lceil\min(m,e(C-\kappa_2\mathrm{rk}(\tilde{\boldsymbol G})))\rceil-\mathrm{rk}(\tilde{\boldsymbol G})e\kappa_1.$$

Together with 4.11.3 Proposition this implies the claim about congruences for the Lefschetz numbers of $[\tilde K[h]_p^{-e}r\tilde K]_{\tilde\lambda}$ on cuspidal cohomology for $S_{\tilde K}$. Thus, the theorem is proven. □

4.13.3. Remark. Section 4.11.2 implies that the Lefschetz number

$$\mathrm{Lef}(\mathbb{T}_{\tilde\lambda}\,|\,H^\bullet_{\mathrm{cusp}}(S_{\tilde K},L_{\tilde\lambda}(\mathbb{C}))),\qquad \mathbb{T}\in\mathcal{C}_0(\tilde{\boldsymbol G}(\mathbb{A}_f)//\tilde K)_h^{\mathrm{reg}},$$

is contained in $\mathcal{O}_{\bar{\mathbb{Q}}_p}$. Thus, 4.13.2 Theorem implies that $\mathrm{Lef}(\mathbb{T}_{\tilde\lambda}\,|\,H^\bullet_{\mathrm{cusp}}(S_{\tilde K},L_{\tilde\lambda}(\mathbb{C})))$ is contained in $F\cap\mathcal{O}_{\bar{\mathbb{Q}}_p}\subseteq\mathbb{Z}_p$.

4.14. *Weighted cohomology.* In this section we compare two Goresky–MacPherson trace formulas for two different weights $\tilde\lambda$ and $\tilde\lambda'$. This is analogous to the comparison of Bewersdorff's trace formula in previous sections and relies on the same diagonalization of elements in $\tilde{\mathcal{I}}h^e\tilde{\mathcal{I}}$ (see 4.4 Proposition) but now applied to all Levi subgroups $\tilde{\boldsymbol M}$ of parabolic subgroups in $\tilde{\boldsymbol G}/\mathbb{Q}$ (see 4.7 Proposition). As a result we obtain congruences for Hecke operators on weighted cohomology for varying weight $\tilde\lambda$.

Here, we will work again in a classical, non-adelic setting (see Section 2.2 and 2.3); e.g., $\Gamma\le\tilde{\boldsymbol G}(\mathbb{Q})$ is an arithmetic subgroup contained in $\tilde{\mathcal{I}}$, $h\in\tilde{\boldsymbol T}(\mathbb{Q})^{++}$ and we will consider Hecke operators $\Gamma\zeta\Gamma$ where $\zeta\in\Delta=\Delta_h$.

4.14.1. *The trace formula of Goresky–MacPherson.* We select a minimal $\mathbb{Q}$-parabolic subgroup $\tilde{\boldsymbol P}_0$ in $\tilde{\boldsymbol G}/\mathbb{Q}$ with Levi decomposition $\tilde{\boldsymbol P}_0=\tilde{\boldsymbol M}_0\tilde{\boldsymbol N}_0$ and we denote by $\boldsymbol A_0$ a maximal $\mathbb{Q}$-split torus in the center of the Levi subgroup $\tilde{\boldsymbol M}_0$. We may assume that $\tilde{\boldsymbol B}\subseteq\tilde{\boldsymbol P}_0/F$ and $\tilde{\boldsymbol T}\supseteq\boldsymbol A_0/F$. Let $\tilde{\boldsymbol P}=\tilde{\boldsymbol M}\tilde{\boldsymbol N}$ be a $\mathbb{Q}$-parabolic subgroup in $\tilde{\boldsymbol G}$. We denote by $\boldsymbol A_{\tilde{\boldsymbol P}}$ a maximal $\mathbb{Q}$-split torus in the center of $\tilde{\boldsymbol M}$ and we write

$\Delta_{\tilde{P}} = \{\alpha_1, \dots, \alpha_m\} \subset X(A_{\tilde{P}}/A_{\tilde{G}})$ for the set of simple roots of $A_{\tilde{P}}$ occuring in $\mathrm{Lie}(\tilde{N})$ and

$$\{t_{\alpha_1}, \dots, t_{\alpha_m}\} \subset X_*(A_{\tilde{P}}/A_{\tilde{G}}) \otimes \mathbb{Q}$$

for the basis dual to $\Delta_{\tilde{P}}$. We select $h \in \tilde{T}(\mathbb{Q})^{++}$ and we let $\zeta \in \Delta = \Delta_h$. The double coset $\Gamma\zeta\Gamma$ induces an operator on the weighted cohomology groups $W^\nu H^i(\overline{\Gamma\backslash X}, L_{\tilde{\lambda}}(\mathbb{C}))$ with weight profile $\nu \in X(A_0)\otimes\mathbb{Q}$. Goresky and MacPherson computed the Lefschetz number of $\Gamma\zeta\Gamma$ acting on weighted cohomology:

Theorem [Goresky and MacPherson 2003, 1.4 Theorem].

$$\mathrm{Lef}\big(\Gamma\zeta\Gamma | W^\nu H^\bullet(\overline{\Gamma\backslash X}, L_{\tilde{\lambda}}(\mathbb{C}))\big) = \sum_{\{\tilde{P}\}} \sum_i \sum_{\{\xi\}} \Xi_{\tilde{P},\xi} \sum_{\substack{w \in W^{\tilde{P}} \\ I_\nu(w,\tilde{\lambda}) = \Delta^+_{\tilde{P}}(\xi)}} (-1)^{\ell(w)} \operatorname{tr}\big(\xi^{-1} | L^{\tilde{M}}_{w\cdot\tilde{\lambda}}(F)\big).$$

In this formula, $\tilde{P}$ runs over a choice of representatives for the Γ-conjugacy classes of $\mathbb{Q}$-parabolic subgroups of $\tilde{G}$. For each such $\tilde{P}$ we write $\Gamma\zeta\Gamma \cap \tilde{P}(\mathbb{Q}) = \coprod_i \Gamma_{\tilde{P}}\zeta_i\Gamma_{\tilde{P}}$ where $\Gamma_{\tilde{P}} = \Gamma \cap \tilde{P}(\mathbb{Q})$ and $\zeta_i \in \tilde{P}(\mathbb{Q})$. The second sum runs over these finitely many double cosets. We set $\Gamma_{\tilde{M}} = \nu_{\tilde{P}}(\Gamma_{\tilde{P}}) \subseteq \tilde{M}(\mathbb{Q})$ and $\bar{\zeta}_i = \nu_{\tilde{P}}(\zeta_i) \in \tilde{M}(\mathbb{Q})$; here $\tilde{P} = \tilde{M}\tilde{N}$ is the Levi decomposition and $\nu_{\tilde{P}} : \tilde{P} \to \tilde{M}$ is the canonical mapping. The third sum is over a set of representatives ξ of the $\Gamma_{\tilde{M}}$-conjugacy classes of elliptic (modulo $A_{\tilde{P}}(\mathbb{R})$) elements in $\Gamma_{\tilde{M}}\bar{\zeta}_i\Gamma_{\tilde{M}} \subseteq \tilde{M}(\mathbb{Q})$. (The numbers $\Xi_{\tilde{P},\xi}$ are explained in [Goresky and MacPherson 2003, 1.4]; we only need to know that they are contained in $\mathbb{Z}$ and do not depend on the weight $\tilde{\lambda}$.) Moreover,

$$\Delta^+_{\tilde{P}}(\xi) = \{\alpha \in \Delta_{\tilde{P}} : \alpha(a_\xi) < 1\},$$

where a_ξ is the projection of ξ to the identity component $A_{\tilde{P}}$ of $A_{\tilde{P}}(\mathbb{R})$ and $I_\nu(w, \tilde{\lambda})$ is given as

$$I_\nu(w, \tilde{\lambda}) = \{\alpha_i \in \Delta_{\tilde{P}} : \langle w(\tilde{\lambda} + \rho) - \rho - \nu, t_{\alpha_i}\rangle < 0\},$$

where $\rho = \rho_{\tilde{G}}$. Finally, since $\xi \in \tilde{M}(\mathbb{Q})$ and $L^{\tilde{M}}_{w\cdot\tilde{\lambda}}$ is defined over F, the trace may be computed on F-points of $L^{\tilde{M}}_{w\cdot\tilde{\lambda}}$. We note that to make sense of the trace of $\xi^{-1} \in \tilde{M}(\mathbb{Q})$ on $L^{\tilde{M}}_{\tilde{\lambda}}(F)$ ($L^{\tilde{M}}_{\tilde{\lambda}}$ was defined for standard parabolic subgroups) as well as of the definition of $I_\nu(w, \tilde{\lambda})$ ($\tilde{\lambda}$, ρ are characters of $\tilde{T}$) we have to conjugate $\tilde{P}$, i.e., we select $x \in \tilde{G}(F)$ such that ${}^x\tilde{P}/F$ is standard parabolic.

4.14.2. We select a $\mathbb{Q}$-parabolic subgroup $\tilde{P}$ in $\tilde{G}$ with Levi decomposition $\tilde{P} = \tilde{M}\tilde{N}$. For any $w \in W^{\tilde{P}}$ and any $\alpha_i \in \Delta_{\tilde{P}}$ the assignment $L_{w,i} : X(\tilde{T})\otimes\mathbb{Q} \to \mathbb{Q}$ taking $\tilde{\lambda}$ to $\langle w\tilde{\lambda}, t_{\alpha_i}\rangle$ is linear and we denote by $H^+_{w,i}$ (resp. $H^-_{w,i}$) the half space consisting of all $\tilde{\lambda} \in X(\tilde{T})$ such that $\langle w(\tilde{\lambda} + \rho) - \rho - \nu, t_{\alpha_i}\rangle$ is positive (resp. negative). For any $\underline{\epsilon} = (\epsilon_{w,i}) \in \{\pm\}^{W^{\tilde{P}} \times [1,\dots,m]}$ we set

$$C(\underline{\epsilon}) = X(\tilde{\boldsymbol{T}})^{\mathrm{dom}} \cap \bigcap_{w,i} H_{w,i}^{\epsilon(w,i)}.$$

Thus, $C(\underline{\epsilon})$ is an intersection of half spaces which may be empty. For all $w \in W^{\tilde{P}}$ and $i = 1, \dots, m$ the values $\langle w(\tilde{\lambda} + \rho) - \rho - \nu, t_{\alpha_i} \rangle$ and $\langle w(\tilde{\lambda}' + \rho) - \rho - \nu, t_{\alpha_i} \rangle$, where $\tilde{\lambda}, \tilde{\lambda}' \in C(\underline{\epsilon})$ have the same sign, hence, for all $w \in W^{\tilde{P}}$ we obtain $I_\nu(w, \tilde{\lambda}) = I_\nu(w, \tilde{\lambda}')$, i.e., $I_\nu(w, \tilde{\lambda})$ does not depend on $\tilde{\lambda} \in C(\underline{\epsilon})$.

Theorem. *Let $C \in \mathbb{Q}_{>0}$ and $\underline{\epsilon} \in \{\pm\}^{W^{\tilde{P}} \times [1,\dots,m]}$. Let $\tilde{\lambda}, \tilde{\lambda}' \in C(\underline{\epsilon})$ be (dominant) weights satisfying*

- $\langle \tilde{\lambda}, \alpha^\vee \rangle > 2C$ *and* $\langle \tilde{\lambda}', \alpha^\vee \rangle > 2C$ *for all* $\alpha \in \Delta_{\tilde{\boldsymbol{G}}}$,
- $\tilde{\lambda} \equiv \tilde{\lambda}' \pmod{(p-1)p^{m-1} X(\tilde{\boldsymbol{T}})}$ $(m \in \mathbb{N})$.

Let ζ be contained in the semigroup Δ_h, hence, $\zeta \in \tilde{\mathcal{I}} h^e \tilde{\mathcal{I}}$ for some $e \in \mathbb{N}_0$ and we assume that $e \in \mathbb{N}$. Then the Lefschetz number $\mathrm{Lef}\big((\Gamma\zeta\Gamma)_{\tilde{\lambda}} | W^\nu H^\bullet(\overline{\Gamma\backslash X}, L_{\tilde{\lambda}}(F))\big)$ is contained in F and the following congruence holds:

$$\mathrm{Lef}\big((\Gamma\zeta\Gamma)_{\tilde{\lambda}} | W^\nu H^\bullet(\overline{\Gamma\backslash X}, L_{\tilde{\lambda}}(F))\big) \equiv \mathrm{Lef}\big((\Gamma\zeta\Gamma)_{\tilde{\lambda}'} | W^\nu H^\bullet(\overline{\Gamma\backslash X}, L_{\tilde{\lambda}'}(F))\big) \pmod{p^{\lceil \min((C-\kappa_1-\kappa_2)e, m-e\kappa_1)\rceil} \mathbb{Z}_p}.$$

Proof. We look at the Goresky–MacPherson trace formula. Since $\Gamma\zeta\Gamma \cap \tilde{\boldsymbol{P}}(\mathbb{Q}) \supseteq \Gamma_{\tilde{\boldsymbol{P}}} \zeta_i \Gamma_{\tilde{\boldsymbol{P}}}$ we obtain $\zeta_i \in \Gamma\zeta\Gamma \cap \tilde{\boldsymbol{P}}(\mathbb{Q})$, hence, we can write $\zeta_i = \bar{\zeta}_i u$ where $\bar{\zeta}_i \in \tilde{\boldsymbol{M}}(\mathbb{Q})$ and $u \in \tilde{\boldsymbol{N}}(\mathbb{Q})$. Let $\xi \in \Gamma_{\tilde{\boldsymbol{M}}} \bar{\zeta}_i \Gamma_{\tilde{\boldsymbol{M}}}$ be a representative of a $\Gamma_{\tilde{\boldsymbol{M}}}$-conjugacy class. We may assume that $\xi = \gamma_M \bar{\zeta}_i$ for some $\gamma_M \in \Gamma_{\tilde{\boldsymbol{M}}}$, i.e., $\gamma_M = \nu_{\tilde{\boldsymbol{P}}}(\gamma_P)$ for some $\gamma_P \in \Gamma_{\tilde{\boldsymbol{P}}}$. Hence, we can write $\gamma_P = \gamma_N \gamma_M$, where $\gamma_N \in \tilde{\boldsymbol{N}}(\mathbb{Q})$, and obtain

$$\gamma_N \xi u = \gamma_N \gamma_M \bar{\zeta}_i u = \gamma_P \zeta_i \in \Gamma_{\tilde{\boldsymbol{P}}} \zeta_i \subseteq \Gamma\zeta\Gamma \subseteq \tilde{\mathcal{I}} h^e \tilde{\mathcal{I}}.$$

Since $\tilde{\boldsymbol{M}}(\mathbb{Q})$ normalizes $\tilde{\boldsymbol{N}}(\mathbb{Q})$ we can write $\gamma_N \xi u = \xi u'$ with $u' \in \tilde{\boldsymbol{N}}(\mathbb{Q})$. Thus, $\xi u' \in \tilde{\mathcal{I}} h^e \tilde{\mathcal{I}}$ and we may apply 4.7 Proposition to compute $\mathrm{tr}(\xi^{-1} | L^{\tilde{M}}_{w\cdot\tilde{\lambda}}(\mathbb{Q}_p))$. If we put the result in the trace formula of Goresky–MacPherson we obtain

$$\begin{aligned}(59)\quad & \mathrm{Lef}((\Gamma\zeta\Gamma)_{\tilde{\lambda}} | W^\nu H^\bullet(\overline{\Gamma\backslash X}, L_{\tilde{\lambda}}(F))) \\ & \equiv \sum_{\{\tilde{\boldsymbol{P}}\}} \sum_{i\xi} \sum_{\{\xi\}} \Xi_{\tilde{\boldsymbol{P}},\xi} \delta_s (-1)^{\ell(s)} \epsilon_{\tilde{\boldsymbol{P}},\xi u'} \tilde{\lambda}(h^e t^{-1}) \pmod{p^{(C-\kappa_1-\kappa_2)e} \mathcal{O}_{\bar{\mathbb{Q}}_p}},\end{aligned}$$

where $s = s_{\xi u'} \in W^{\tilde{P}}$, $t = t_{\xi u'} \in \tilde{\boldsymbol{T}}(\bar{\mathbb{Q}}_p)^{++}$ is the unique element which is $\tilde{\boldsymbol{G}}(\bar{\mathbb{Q}}_p)$-conjugate to $(\xi u')_s$, $\epsilon_{\tilde{\boldsymbol{P}},\xi u'}$ has p-adic value $\geq -e\kappa_1$ and $\delta_s = 1$ if $I_\nu(s, \tilde{\lambda}) = \Delta^+_{\tilde{P}}(\xi)$ and vanishes otherwise. We note that δ_s does not depend on $\tilde{\lambda}$ since we assume that $\tilde{\lambda}$ is contained in $C(\underline{\epsilon})$. By our assumption there is $\chi \in X(\tilde{\boldsymbol{T}})$ such that $\tilde{\lambda} - \tilde{\lambda}' = (p-1)p^{m-1}\chi$. Since $\chi(h^e t^{-1}) \equiv \epsilon \pmod{p\mathcal{O}_{\bar{\mathbb{Q}}_p}}$ where $\epsilon \in \mathbb{Z}_p^*$ (see 4.4 Proposition) this implies $\tilde{\lambda}(h^e t^{-1}) \equiv \tilde{\lambda}'(h^e t^{-1}) \pmod{p^m \mathcal{O}_{\bar{\mathbb{Q}}_p}}$ (see the proof of

4.10 Proposition) and we find that equation (59) (which also holds for $\tilde{\lambda}'$) implies that the Lefschetz numbers in the Theorem are congruent to each other modulo $p^{\lceil \min((C-\kappa_1-\kappa_2)e, m-e\kappa_1)\rceil}\mathbb{Z}_p$ (note that the Lefschetz numbers are contained in F). Thus the proof is complete. $\square$

4.14.3. *The case C_2.* We assume that $\tilde{\boldsymbol{G}}/\mathbb{Q}$ is connected and $\mathbb{Q}$-split with root system of type C_2, hence, there are two simple roots α_1, α_2. We also assume that $\nu = -\rho$, i.e., ν is the middle weight profile. Thus, if $\tilde{\lambda}$ is regular then $W^\nu H^d(\overline{\Gamma\backslash X}, L_{\tilde{\lambda}}(\mathbb{C})) = H^d_{\rm cusp}(\Gamma\backslash X, L_{\tilde{\lambda}}(\mathbb{C}))$ and both these cohomology groups vanish for all degrees $i \neq d$. Let $\tilde{\lambda} \in X(\tilde{\boldsymbol{T}})^{\rm dom}$. Hence, $\tilde{\lambda} + \rho \in X(\tilde{\boldsymbol{T}})^{\rm dom}$ is strictly dominant and for any $w \in W_{\tilde{\boldsymbol{G}}}$ we write $w(\tilde{\lambda}+\rho) = a_1\alpha_1 + a_2\alpha_2$ with $a_i = a_{i,w,\tilde{\lambda}} \in \mathbb{Q}$. For the root system of type C_2 it happens that the sign of a_1 as well as the sign of a_2 does not depend on $\tilde{\lambda} \in X(\tilde{\boldsymbol{T}})^{\rm dom}$, i.e., $\operatorname{sign}(a_{i,w,\tilde{\lambda}}) = \operatorname{sign}(a_{i,w,\tilde{\lambda}'})$ for any $\tilde{\lambda}, \tilde{\lambda}' \in X(\tilde{\boldsymbol{T}})^{\rm dom}$ and all $w \in W_{\tilde{\boldsymbol{G}}}$ and all $i = 1, 2$. Thus, if $\tilde{\boldsymbol{P}} = \tilde{\boldsymbol{B}}$, hence, $\boldsymbol{A}_{\tilde{\boldsymbol{P}}} = \tilde{\boldsymbol{T}}/\mathbb{Q}$ and if $\{t_{\alpha_1}, t_{\alpha_2}\}$ denotes the basis of $X_*(\tilde{\boldsymbol{T}}) \otimes \mathbb{Q}$ dual to $\{\alpha_1, \alpha_2\}$ then we obtain that the sign of $\langle w(\tilde{\lambda}+\rho), t_{\alpha_i}\rangle = a_i$ does not depend on $\tilde{\lambda} \in X(\tilde{\boldsymbol{T}})^{\rm dom}$. Similarly if $\tilde{\boldsymbol{P}} = \tilde{\boldsymbol{P}}_{\alpha_i}$, hence, $\boldsymbol{A}_{\tilde{\boldsymbol{P}}} = \ker\alpha_j$, $j \neq i$, and if $\{t_{\alpha_i}\}$ denotes the basis of $X_*(\boldsymbol{A}_{\tilde{\boldsymbol{P}}}) \otimes \mathbb{Q}$ dual to $\{\alpha_i\}$ then we obtain that the sign of $\langle w(\tilde{\lambda}+\rho)|_{\boldsymbol{A}_{\tilde{\boldsymbol{P}}}}, t_{\alpha_i}\rangle = a_i$ does not depend on $\tilde{\lambda} \in X(\tilde{\boldsymbol{T}})^{\rm dom}$. This shows that $I_\nu(w, \tilde{\lambda})$ does not depend on $\tilde{\lambda} \in X(\tilde{\boldsymbol{T}})^{\rm dom}$ (if $\tilde{\boldsymbol{P}} = \tilde{\boldsymbol{G}}$ then $I_\nu(w, \tilde{\lambda})$ is empty). 4.14.2 Theorem therefore holds with $X(\tilde{\boldsymbol{T}})^{\rm dom}$ in place of $C(\underline{\epsilon})$:

Corollary. *Assume $\tilde{\boldsymbol{G}}/\mathbb{Q}$ is connected and $\mathbb{Q}$-split with root system C_2. Let $C \in \mathbb{Q}_{>0}$ and let $\tilde{\lambda}, \tilde{\lambda}' \in X(\tilde{\boldsymbol{T}})^{\rm dom}$ be weights satisfying*

- $\langle\tilde{\lambda}, \alpha^\vee\rangle > 2C$ *and* $\langle\tilde{\lambda}', \alpha^\vee\rangle > 2C$ *for all* $\alpha \in \Delta_{\tilde{\boldsymbol{G}}}$, *and*
- $\tilde{\lambda} \equiv \tilde{\lambda}' \pmod{(p-1)p^{m-1}X(\tilde{\boldsymbol{T}})}$ $(m \in \mathbb{N})$.

Then $\operatorname{tr}((\Gamma\zeta\Gamma)_{\tilde{\lambda}}|H^d_{\rm cusp}(\Gamma\backslash X, L_{\tilde{\lambda}}(\mathbb{C})))$ *is contained in F and*

$$\operatorname{tr}\big((\Gamma\zeta\Gamma)_{\tilde{\lambda}}|H^d_{\rm cusp}(\Gamma\backslash X, L_{\tilde{\lambda}}(\mathbb{C}))\big) \equiv \operatorname{tr}\big((\Gamma\zeta\Gamma)_{\tilde{\lambda}'}|H^d_{\rm cusp}(\Gamma\backslash X, L_{\tilde{\lambda}'}(\mathbb{C}))\big)$$
$$(\operatorname{mod} p^{\lceil \min((C-\kappa_1-\kappa_2)e, m-e\kappa_1)\rceil}\mathbb{Z}_p).$$

Proof. Use the equality

$$\operatorname{tr}\big((\Gamma\zeta\Gamma)_{\tilde{\lambda}}|H^d_{\rm cusp}(\Gamma\backslash X, L_{\tilde{\lambda}}(\mathbb{C}))\big) = (-1)^d \operatorname{Lef}\big((\Gamma\zeta\Gamma)_{\tilde{\lambda}}|W^\nu H^\bullet(\overline{\Gamma\backslash X}, L_{\tilde{\lambda}}(\mathbb{C}))\big). \quad \square$$

Remark. The $\mathbb{Q}$-rank of $\tilde{\boldsymbol{G}}$ does not appear in the modulus of these congruences.

5. Local constancy of dimension of slope subspaces

5.1. ***Slope subspaces of cuspidal cohomology.*** As before we let $\tilde{\boldsymbol{G}}/\mathbb{Q}$ be a connected reductive group with $\mathbb{Q}_p$-split maximal torus $\tilde{\boldsymbol{T}}/\mathbb{Q}$ and from now on we assume in addition that $\tilde{\boldsymbol{G}}/\mathbb{Q}$ has discrete series. We denote by ℓ the $\mathbb{Q}$-rank of

$\tilde{\boldsymbol{G}}$ and we keep the notations from Section 4.11. In particular, $\tilde{K} \leq \tilde{\boldsymbol{G}}(\mathbb{A}_f)$ is a compact open subgroup with p-component $\tilde{K}_p = \tilde{\mathcal{I}}$ and $S_{\tilde{K}}$ is the adelic locally symmetric space of level $\tilde{K}$. We select elements $h \in \tilde{\boldsymbol{T}}(\mathbb{Q})^{++}$ and $r \in \tilde{\boldsymbol{G}}(\mathbb{A}_f)^{(p)}$ and we set

$$\mathbb{T} = [\tilde{K}[h]_p^{-1} r \tilde{K}] \in \mathcal{C}_0(\tilde{\boldsymbol{G}}(\mathbb{A}_f)//\tilde{K})_h^{\mathrm{reg}}.$$

As before we denote by $\mathbb{T}_{\tilde{\lambda}}$ the $\tilde{\lambda}$-normalization of $\mathbb{T}$. The (normalized) operator $\mathbb{T}_{\tilde{\lambda}}$ acts on

$$H^d_{\tilde{\lambda},\mathrm{cusp}}(\mathbb{C}) = H^d_{\mathrm{cusp}}(S_{\tilde{K}}, L_{\tilde{\lambda}}(\mathbb{C}))$$

where $d = d_{\tilde{\boldsymbol{G}}}$ is the middle degree. For $\beta \in \mathbb{Q}_{\geq 0}$ we denote by

$$H^d_{\tilde{\lambda},\mathrm{cusp}}(\mathbb{C})^{\beta} = \bigoplus_{\substack{\mu \in \bar{F} \hookrightarrow \bar{\mathbb{Q}}_p \\ v_p(\mu) = \beta}} H^d_{\tilde{\lambda},\mathrm{cusp}}(\mathbb{C})(\mu),$$

the slope β subspace of $H^d_{\tilde{\lambda},\mathrm{cusp}}(\mathbb{C})$ w.r.t. $\mathbb{T}_{\tilde{\lambda}}$; here, $H^d_{\tilde{\lambda},\mathrm{cusp}}(\mathbb{C})(\mu) = H^d_{\tilde{\lambda},\mathrm{cusp},\mathbb{T}}(\mathbb{C})(\mu)$ is the generalized eigenspace attached to $\mathbb{T}_{\tilde{\lambda}}$ and the eigenvalue μ and we remark that the eigenvalues μ of $\mathbb{T}_{\tilde{\lambda}}$ on $H^d_{\tilde{\lambda},\mathrm{cusp}}(\mathbb{C})$ are contained in $\bar{F}(\subseteq \mathbb{C})$ and are p-adically integral, i.e., contained in $\mathcal{O}_{\bar{\mathbb{Q}}_p}$ (note that $\bar{F} \subseteq \bar{\mathbb{Q}}_p$; see Section 4.11.2). We set $H^d_{\tilde{\lambda},\mathrm{cusp}}(\mathbb{C})^{\leq\beta} = \bigoplus_{0 \leq \gamma \leq \beta} H^d_{\tilde{\lambda},\mathrm{cusp}}(\mathbb{C})^{\gamma}$. Since

$$\begin{aligned}
\dim H^d_{\mathrm{cusp}}(S_{\tilde{K}}, L_{\tilde{\lambda}}(\mathbb{C}))^{\leq\beta} &= \dim \bigoplus_{\substack{\mu \in \bar{F} \hookrightarrow \bar{\mathbb{Q}}_p \\ 0 \leq v_p(\mu) \leq \beta}} H^d_{\tilde{\lambda},\mathrm{cusp}}(\mathbb{C})(\mu) \\
&\leq \dim \bigoplus_{\substack{\mu \in \bar{F} \\ 0 \leq v_p(\mu) \leq \beta}} H^d(S_{\tilde{K}}, L_{\tilde{\lambda}}(\mathbb{C}))(\mu) \\
&= \dim \bigoplus_{\substack{\mu \in \bar{F} \\ 0 \leq v_p(\mu) \leq \beta}} H^d(S_{\tilde{K}}, L_{\tilde{\lambda}}(\bar{F}))(\mu) \\
&= \dim \bigoplus_{\substack{\mu \in \bar{F} \hookrightarrow \bar{\mathbb{Q}}_p \\ 0 \leq v_p(\mu) \leq \beta}} H^d(S_{\tilde{K}}, L_{\tilde{\lambda}}(\bar{\mathbb{Q}}_p))(\mu) \\
&= \dim H^d(S_{\tilde{K}}, L_{\tilde{\lambda}}(\bar{\mathbb{Q}}_p))^{\leq\beta} \\
&= \dim H^d(S_{\tilde{K}}, L_{\tilde{\lambda}}(\mathbb{Q}_p))^{\leq\beta},
\end{aligned}$$

we obtain the following bound from 4.11.4 Theorem. Recall that $s = |\Phi^+_{\tilde{\boldsymbol{G}}}|$, $\sigma = \max_{\alpha \in \Phi^+_{\tilde{G}}} \mathrm{ht}(\alpha)$ and we denote by $\boldsymbol{g} = \boldsymbol{g}_{\tilde{K}}$ the number of d cells in a cell complex which is homotopy equivalent to the Borel–Serre compactification $\bar{S}_{\tilde{K}}$. We set

$$\mathtt{m} = \mathtt{m}_{\tilde{K}} = 12\frac{\boldsymbol{g}}{s}\sigma^{s+1} \in \mathbb{Q}_{\geq 0} \quad \text{and} \quad \mathtt{n} = \mathtt{n}_{\tilde{K}} = \left\lceil \frac{1}{2^{s+1}} \frac{\boldsymbol{g}\sigma^{s+1}}{s} M^s \right\rceil + 1 \in \mathbb{N},$$

where $M = M(\sigma, s) \in \mathbb{N}$ is defined in equations (23) and (24); then

$$\dim H^d_{\text{cusp}}(S_{\tilde{K}}, L_{\tilde{\lambda}}(\mathbb{C}))^{\leq \beta} \leq B(\beta) := \mathtt{m}\beta^s + \mathtt{n} \tag{60}$$

for all dominant $\tilde{\lambda} \in X(\tilde{\boldsymbol{T}})$, all $\beta \in \mathbb{Q}_{\geq 0}$ and all $h \in \tilde{\boldsymbol{T}}(\mathbb{Q})^{++}$ and $r \in \tilde{\boldsymbol{G}}(\mathbb{A}_f)^{(p)}$.

5.2. We want to consider the function

$$d(\,\cdot\,,\,\cdot\,) : \mathbb{Q}_{\geq 0} \times X(\tilde{\boldsymbol{T}})^{\text{dom}} \to \mathbb{N}_0, \quad (\beta, \tilde{\lambda}) \mapsto \dim H^d_{\text{cusp}}(S_{\tilde{K}}, L_{\tilde{\lambda}}(\mathbb{C}))^{\beta};$$

i.e., we want to understand how the dimension of the slope subspace varies as a function of the weight. To this end, for any $\beta \in \mathbb{Q}_{\geq 0}$ we set

$$\begin{aligned} \boldsymbol{m}_1(\beta) &= \big(\beta + \tfrac{1}{p-1} + (\kappa_1 + \kappa_2)\ell\big) B(\beta) + 1 \\ &= \mathtt{m}\beta^{s+1} + \mathtt{m}\big(\tfrac{1}{p-1} + (\kappa_1 + \kappa_2)\ell\big)\beta^s + \mathtt{n}\beta + \mathtt{n}\big(\tfrac{1}{p-1} + (\kappa_1 + \kappa_2)\ell\big) + 1 \\ &\qquad \in \mathbb{Q}_{>0}. \end{aligned} \tag{61}$$

and

$$\begin{aligned} \boldsymbol{m}_2(\beta) &= \big(\beta + \tfrac{1}{p-1} + \kappa_1\ell\big) B(\beta) + 1 \\ &= \mathtt{m}\beta^{s+1} + \mathtt{m}\big(\tfrac{1}{p-1} + \kappa_1\ell\big)\beta^s + \mathtt{n}\beta + \mathtt{n}\big(\tfrac{1}{p-1} + \kappa_1\ell\big) + 1 \in \mathbb{Q}_{>0}. \end{aligned} \tag{62}$$

Thus, $\boldsymbol{m}_1(\beta), \boldsymbol{m}_2(\beta) \in \mathbb{Q}[\beta]$ are polynomials in β with positive coefficients, degree $s+1$ and leading term $\mathtt{m} = 12\frac{\boldsymbol{g}}{s}\sigma^{s+1}$ which depend on $\tilde{K}$ (and, hence, on $\tilde{\boldsymbol{G}}$ and p) and on h (since $\kappa_1 = \kappa_1(h)$) but do not depend on $\tilde{\lambda} \in X(\tilde{\boldsymbol{T}})$.

Theorem. *Assume that $\tilde{\boldsymbol{G}}$ has discrete series. Let $\beta \in \mathbb{Q}_{\geq 0}$ and assume the dominant weights $\tilde{\lambda}, \tilde{\lambda}' \in X(\tilde{\boldsymbol{T}})$ satisfy*

- $\langle \tilde{\lambda}, \alpha^\vee \rangle > 2\boldsymbol{m}_1(\beta)$ *and* $\langle \tilde{\lambda}', \alpha^\vee \rangle > 2\boldsymbol{m}_1(\beta)$ *for all* $\alpha \in \Delta_{\tilde{\boldsymbol{G}}}$,
- $\tilde{\lambda} \equiv \tilde{\lambda}' \pmod{(p-1)p^{m-1}X(\tilde{\boldsymbol{T}})}$ *with* $m \geq \boldsymbol{m}_2(\beta)$ ($m \in \mathbb{N}$).

Then

$$\dim H^d_{\text{cusp}}(S_{\tilde{K}}, L_{\tilde{\lambda}}(\mathbb{C}))^{\gamma} = \dim H^d_{\text{cusp}}(S_{\tilde{K}}, L_{\tilde{\lambda}'}(\mathbb{C}))^{\gamma} \quad \textit{for all } 0 \leq \gamma \leq \beta.$$

5.3. ***Characteristic polynomial.*** The proof of the preceding theorem will be given in Section 5.6. To prepare it we consider the characteristic polynomial. We denote by $\text{ch}_{\tilde{\lambda}}(X) = \det(X\mathbf{1} - \mathbb{T}_{\tilde{\lambda}}) = \sum_{i=0}^{m} (-1)^i a_{i,\tilde{\lambda}} X^{m-i} \in \mathbb{C}[X]$, where $m = m_{\tilde{\lambda}} = \dim H^d_{\tilde{\lambda},\text{cusp}}(\mathbb{C})$, the characteristic polynomial of $\mathbb{T}_{\tilde{\lambda}}$ acting on $H^d_{\tilde{\lambda},\text{cusp}}(\mathbb{C})$. We set $a_{i,\tilde{\lambda}} = 0$ if $i > m_{\tilde{\lambda}}$. This definition of the characteristic polynomial differs from the one used in the proof of 4.3 Lemma by a factor $(-1)^m$, but we can refer directly to

[Koecher 1983, 3.4.6 Satz, p. 117], where this definition is used and which yields the following inductive formula for the coefficients of $\mathrm{ch}_{\tilde{\lambda}}(X)$: $a_{0,\tilde{\lambda}}=1$ and

$$ia_{i,\tilde{\lambda}}=\sum_{e=1}^{i}(-1)^{e+1}\mathrm{tr}\,\mathbb{T}^{e}_{\tilde{\lambda}}\,a_{i-e,\tilde{\lambda}},\quad i=1,2,\dots,\tag{63}$$

where we have set

$$\mathrm{tr}\,\mathbb{T}^{e}_{\tilde{\lambda}}=\mathrm{tr}\,((\mathbb{T}_{\tilde{\lambda}})^{e}\,|\,H^{d}_{\tilde{\lambda},\mathrm{cusp}}(\mathbb{C})).$$

Since $[\tilde{\mathcal{I}}h^{-e}\tilde{\mathcal{I}}][\tilde{\mathcal{I}}h^{-f}\tilde{\mathcal{I}}]=[\tilde{\mathcal{I}}h^{-(e+f)}\tilde{\mathcal{I}}]$ we obtain that $\mathbb{T}^{e}=\sum_{s}c_{s}[\tilde{K}[h]_{p}^{-e}s\tilde{K}]$ for all $e\geq 1$, where s runs over $\tilde{\boldsymbol{G}}(\mathbb{A}_{f})^{(p)}$ and $c_{s}\in\mathbb{Z}$. Hence $(\mathbb{T}_{\tilde{\lambda}})^{e}=\sum_{s}c_{s}[\tilde{K}[h]_{p}^{-e}s\tilde{K}]_{\tilde{\lambda}}$. Since $\tilde{\boldsymbol{G}}$ has discrete series we therefore obtain if the highest weight $\tilde{\lambda}$ is regular

$$\begin{aligned}\mathrm{tr}\,\mathbb{T}^{e}_{\tilde{\lambda}}&=\sum_{s}c_{s}\mathrm{tr}\big([\tilde{K}[h]_{p}^{-e}s\tilde{K}]_{\tilde{\lambda}}\,|\,H^{d}_{\tilde{\lambda},\mathrm{cusp}}(\mathbb{C})\big)\\&=\sum_{s}c_{s}(-1)^{d}\,\mathrm{Lef}\big([\tilde{K}[h]_{p}^{-e}s\tilde{K}]_{\tilde{\lambda}}\,|\,H^{\bullet}_{\mathrm{cusp}}(S_{\tilde{K}},L_{\tilde{\lambda}}(\mathbb{C}))\big).\end{aligned}\tag{64}$$

Hence, equation (63) and 4.13.2 Theorem yield $a_{i,\tilde{\lambda}}\in F$ for all i, i.e.,

$$\mathrm{ch}_{\tilde{\lambda}}(X)\in F[X].$$

In particular, the roots of $\mathrm{ch}_{\tilde{\lambda}}$, which are the eigenvalues of $\mathbb{T}_{\tilde{\lambda}}$ on $H^{d}_{\tilde{\lambda},\mathrm{cusp}}(\mathbb{C})$, are algebraic over F and in Section 4.11.2 we have seen that after embedding $\bar{F}\subseteq\bar{\mathbb{Q}}_{p}$ they are contained in $\mathcal{O}_{\bar{\mathbb{Q}}_{p}}$. Hence, $a_{i,\tilde{\lambda}}\in\mathcal{O}_{\bar{\mathbb{Q}}_{p}}$, i.e., the coefficients are p-adically integral which implies that $a_{i,\tilde{\lambda}}\in\mathbb{Z}_{p}$, $i=0,\dots,m$; in particular,

$$\mathrm{ch}_{\tilde{\lambda}}(X)\in\mathbb{Z}_{p}[X].$$

5.4. Proposition. *Let $\beta\in\mathbb{Q}_{\geq 0}$. Assume that the weights $\tilde{\lambda},\tilde{\lambda}'\in X(\tilde{\boldsymbol{T}})^{\mathrm{dom}}$ satisfy the two assumptions of 5.2 Theorem. Then for all $i=0,1,2,3,\dots,B(\beta)$ the following congruence holds*:

$$a_{i,\tilde{\lambda}}\equiv a_{i,\tilde{\lambda}'}\ (\mathrm{mod}\ p^{\lceil\beta B(\beta)+1\rceil}\mathbb{Z}_{p}).$$

Proof. We will prove that for all $i=0,1,\dots,B(\beta)$ the congruence

$$a_{i,\tilde{\lambda}}\equiv a_{i,\tilde{\lambda}'}\ (\mathrm{mod}\ p^{\lceil\beta B(\beta)+\frac{B(\beta)}{p-1}+1\rceil-v_{p}(i!)}\mathbb{Z}_{p})$$

holds. Since $v_{p}(i!)\leq i/(p-1)\leq B(\beta)/(p-1)$ these congruences imply that the congruences of the Proposition hold. The congruences hold trivially for $i=0$ ($a_{0,\tilde{\lambda}}=a_{0,\tilde{\lambda}'}=1$). To prove them for $i=1,2,3,\dots,B(\beta)$ we use equation (63). First, for all $e\in\mathbb{N}$ with $1\leq e\leq B(\beta)$ we have

$$m-e\kappa_{1}\ell\geq\boldsymbol{m}_{2}(\beta)-e\kappa_{1}\ell\geq\beta B(\beta)+\frac{B(\beta)}{p-1}+1$$

and

$$\begin{aligned}\left\lceil e(\boldsymbol{m}_1(\beta)-\kappa_2\ell)\right\rceil - e\kappa_1\ell &\geq \boldsymbol{m}_1(\beta)-e\kappa_2\ell-e\kappa_1\ell\\ &= \big(\beta+\tfrac{1}{p-1}\big)B(\beta)+(\kappa_1+\kappa_2)\ell B(\beta)+1-e\kappa_2\ell-e\kappa_1\ell\\ &\geq \beta B(\beta)+\frac{B(\beta)}{p-1}+1.\end{aligned}$$

Hence, 4.13.2 Theorem yields for all $1\leq e\leq B(\beta)$ and all $s\in\tilde{\boldsymbol{G}}(\mathbb{A}_f)^{(p)}$ that

$$\operatorname{Lef}\big([\tilde K[h]_p^{-e}s\tilde K]_{\tilde\lambda}|H^\bullet_{\mathrm{cusp}}(S_{\tilde K},L_{\tilde\lambda}(\mathbb{C}))\big)\equiv\operatorname{Lef}\big([\tilde K[h]_p^{-e}s\tilde K]_{\tilde\lambda'}|H^\bullet_{\mathrm{cusp}}(S_{\tilde K},L_{\tilde\lambda'}(\mathbb{C}))\big)\\ \pmod{p^{\lceil\beta B(\beta)+\frac{B(\beta)}{p-1}+1\rceil}\mathbb{Z}_p}.$$

Together with equation (64) this implies

$$\operatorname{tr}\mathbb{T}^e_{\tilde\lambda}\equiv\operatorname{tr}\mathbb{T}^e_{\tilde\lambda'}\pmod{p^{\lceil\beta B(\beta)+\frac{B(\beta)}{p-1}+1\rceil}\mathbb{Z}_p}\tag{65}$$

for all $1\leq e\leq B(\beta)$. Since $a_{1,?}=\operatorname{tr}\mathbb{T}_?$, equation (65) implies

$$a_{1,\tilde\lambda}\equiv a_{1,\tilde\lambda'}\pmod{p^{\lceil\beta B(\beta))+\frac{B(\beta)}{p-1}+1\rceil}\mathbb{Z}_p},$$

which is the claim for $i=1$. We now let $i\leq B(\beta)$ be arbitrary and assume that the claim holds for $0,1,2,\dots,i-1$. The recursive relation in equation (63)

$$ia_{i,?}=\sum_{e=1}^{i}(-1)^{e+1}\operatorname{tr}\mathbb{T}^e_?\,a_{i-e,?}$$

together with the induction assumption and equation (65) yields

$$ia_{i,\tilde\lambda}\equiv ia_{i,\tilde\lambda'}\pmod{p^{\lceil\beta B(\beta)+\frac{B(\beta)}{p-1}+1\rceil-v_p((i-1)!)}\mathbb{Z}_p}$$

from which the claim for i is immediate. □

5.5. *Newton polygon.* The *Newton polygon* $\mathcal{N}_{\tilde\lambda}$ of $\mathrm{ch}_{\tilde\lambda}\in F[X]\subseteq\mathbb{Q}_p[X]$ is the lower convex hull of the points $(0,v_p(a_{0,\tilde\lambda})),\dots,(m,v_p(a_{m,\tilde\lambda}))$, where we omit from this list all points $(i,v_p(a_{i,\tilde\lambda}))$ with $v_p(a_{i,\tilde\lambda})=\infty$ (i.e., $a_{i,\tilde\lambda}=0$). Thus, if $a_{m-k+1,\tilde\lambda}=\dots=a_{m,\tilde\lambda}=0$ and $a_{m-k,\tilde\lambda}\neq 0$ (i.e., if 0 occurs in $\mathrm{ch}_{\tilde\lambda}(X)$ with multiplicity $k=\mathrm{ord}_X(\mathrm{ch}_{\tilde\lambda})$), we omit the points $(m-k+1,v_p(a_{m-k+1,\tilde\lambda})),\dots,(m,v_p(a_{m,\tilde\lambda}))$. In particular, $\mathcal{N}_{\tilde\lambda}$ represents a piecewise linear function on the interval $[0,n_{\tilde\lambda}]$ which starts at the point $(0,0)$ corresponding to the leading term $a_{0,\tilde\lambda}=1$ of $\mathrm{ch}_{\tilde\lambda}$; here,

$$n_{\tilde\lambda}=m-k=\dim H^d_{\tilde\lambda,\mathrm{cusp}}(\mathbb{C})-\mathrm{ord}_X(\mathrm{ch}_{\tilde\lambda})=\dim\bigoplus_{\gamma\in\mathbb{Q}_{\geq0}}H^d_{\tilde\lambda,\mathrm{cusp}}(\mathbb{C})^\gamma$$

is the dimension of the finite slope subspace $H^d_{\tilde\lambda,\mathrm{cusp}}(\mathbb{C})^{<\infty}$ of $H^d_{\tilde\lambda,\mathrm{cusp}}(\mathbb{C})$. We have $a_{i,\tilde\lambda}=0$ for all $i>n_{\tilde\lambda}$ (note that $a_{i,\tilde\lambda}=0$ for all $i>m_{\tilde\lambda}$). Since $\mathrm{ch}_{\tilde\lambda}/X^k$, for

$k = \mathrm{ord}_X(\mathrm{ch}_{\tilde{\lambda}})$, also has $\mathcal{N}_{\tilde{\lambda}}$ as its Newton polygon but has nonvanishing constant term we deduce that if the segment S_γ of (necessarily finite) slope $\gamma \in \mathbb{Q}_{\geq 0}$ of $\mathcal{N}_{\tilde{\lambda}}$ has length s_γ (if projected to the x-axis) then $\mathrm{ch}_{\tilde{\lambda}}/X^k$ and, hence, $\mathrm{ch}_{\tilde{\lambda}}$ has precisely s_γ many roots (counted with their multiplicities) in $\bar{F}(\subseteq \bar{\mathbb{Q}}_p)$ of p-adic value equal to γ.

5.6. *Proof of 5.2 Theorem.* We denote by $\mathcal{S}$ the finite set consisting of all $\gamma \in \mathbb{Q}_{\geq 0}$ such that $\gamma \leq \beta$ and the segment S_γ of slope γ of the Newton polygon $\mathcal{N}_{\tilde{\lambda}}$ or the segment S'_γ of slope γ of $\mathcal{N}_{\tilde{\lambda}'}$ has strictly positive length (i.e., $H^d_{\tilde{\lambda},\mathrm{cusp}}(\mathbb{C})^\gamma \neq 0$ or $H^d_{\tilde{\lambda}',\mathrm{cusp}}(\mathbb{C})^\gamma \neq 0$). We have to show that

$$\dim H^d_{\tilde{\lambda},\mathrm{cusp}}(\mathbb{C})^\gamma = \dim H^d_{\tilde{\lambda}',\mathrm{cusp}}(\mathbb{C})^\gamma \quad \text{for all } \gamma \in \mathcal{S}.$$

Since $\dim H^d_{\tilde{\lambda},\mathrm{cusp}}(\mathbb{C})^\gamma$ equals the number of roots $\mu \in \bar{F}$ of $\mathrm{ch}_{\tilde{\lambda}}$ having p-adic value γ this is equivalent to showing that for all $\gamma \in \mathcal{S}$ the corresponding segments S_γ and S'_γ have the same length (length 0 if the slope γ subspace is trivial). We assume this is not the case and we denote by $\gamma \in \mathbb{Q}_{\geq 0}$ the smallest number in $\mathcal{S}$ such that S_γ and S'_γ have different length; without loss of generality we may assume that S'_γ is (strictly) longer than S_γ. For all $\gamma^* \in \mathcal{S}$ with $0 \leq \gamma* < \gamma$ the segments $S_{\gamma*}$ and $S'_{\gamma*}$ have the same length, hence, S_γ and S'_γ have the same initial point (note that $\mathcal{N}_{\tilde{\lambda}}, \mathcal{N}_{\tilde{\lambda}'}$ both start in $(0, 0)$). We denote by $P' = (x', y')$ the endpoint of S'_γ. Hence, (x', y') is a vertex of $\mathcal{N}_{\tilde{\lambda}'}$ which implies that $x' \in \mathbb{N}$ and $y' = v_p(a_{x',\tilde{\lambda}'})$. We also know that $x' \leq B(\beta)$ because $x' = \dim H^d_{\tilde{\lambda}',\mathrm{cusp}}(\mathbb{C})^{\leq\gamma} \leq \dim H^d_{\tilde{\lambda}',\mathrm{cusp}}(\mathbb{C})^{\leq\beta} \leq B(\beta)$. Since $x' \leq B(\beta)$ and since all segments of $\mathcal{N}_{\tilde{\lambda}'}$ which are left to x' have slopes less than or equal to γ we deduce that

$$v_p(a_{x',\tilde{\lambda}'}) = y' \leq \gamma B(\beta) \leq \beta B(\beta). \tag{66}$$

Together with 5.4 Proposition this implies that

$$v_p(a_{x',\tilde{\lambda}}) = v_p(a_{x',\tilde{\lambda}'}). \tag{67}$$

We distinguish cases.

Case 1: $\mathcal{N}_{\tilde{\lambda}}$ is defined at x' (i.e., $x' \leq n_{\tilde{\lambda}}$). Since $\mathcal{N}_{\tilde{\lambda}}$ is convex and S_γ is strictly shorter than S'_γ in this case we know that $\mathcal{N}_{\tilde{\lambda}}(x')$ lies strictly above $\mathcal{N}_{\tilde{\lambda}'}(x') = y'$, hence, we obtain

$$v_p(a_{x',\tilde{\lambda}}) \geq \mathcal{N}_{\tilde{\lambda}}(x') > \mathcal{N}_{\tilde{\lambda}'}(x') = y' = v_p(a_{x',\tilde{\lambda}'}),$$

i.e., we obtain a contradiction to equation (67).

Case 2: $\mathcal{N}_{\tilde{\lambda}}$ is not defined at x' (i.e., $x' > n_{\tilde{\lambda}}$). In this case we know that $a_{x',\tilde{\lambda}} = 0$. Since $v_p(a_{x',\tilde{\lambda}'})$ is finite by equation (66), again, this contradicts equation (67).

Thus, the assumption does not hold and we have shown that all segments of slope $\leq \beta$ in $\mathcal{N}_{\tilde{\lambda}}$ and $\mathcal{N}_{\tilde{\lambda}'}$ have the same length which finishes the proof.

5.7. Remark. The theorem in particular implies that for any $\beta \in \mathbb{Q}_{\geq 0}$ the function $\tilde{\lambda} \mapsto d(\beta, \tilde{\lambda})$ is constant on cosets modulo $(p-1)p^{\lceil \boldsymbol{m}_2(\beta) \rceil - 1} X(\tilde{\boldsymbol{T}})$.

5.8. Example: $C_2/\mathbb{Q}$. We look at a special case and assume that $\tilde{\boldsymbol{G}}/\mathbb{Q}$ is connected and $\mathbb{Q}$-split with root system of type C_2. As in Section 4.14 we use a non-adelic setting, i.e., $\Gamma \leq \tilde{\boldsymbol{G}}(\mathbb{Q})$ is an arithmetic subgroup contained in $\tilde{\mathcal{I}}$, $h \in \tilde{\boldsymbol{T}}(\mathbb{Q})^{++}$ and we set $T = \Gamma\zeta\Gamma \in \mathcal{H}$ where $\zeta \in \Delta_h$. We assume $\zeta \in \tilde{\mathcal{I}}h\tilde{\mathcal{I}}$. All eigenvalues of $T_{\tilde{\lambda}}$ on $H^d_{\text{cusp}}(\Gamma\backslash X, L_{\tilde{\lambda}}(\mathbb{C}))$ are algebraic over F and p-adically integral (see 2.9 Corollary) and we define the slope β subspace $H^d_{\text{cusp}}(\Gamma\backslash X, L_{\tilde{\lambda}}(\mathbb{C}))^{\beta}$ as the sum of the generalized $T_{\tilde{\lambda}}$-eigenspaces attached to eigenvalues of p-adic value β. As in Section 5.1 we obtain that 3.10 Corollary implies

$$\dim H^d_{\text{cusp}}(\Gamma\backslash X, L_{\tilde{\lambda}}(\mathbb{C}))^{\leq\beta} \leq B_\Gamma(\beta),$$

where $B_\Gamma(\beta) = \mathtt{m}_\Gamma \beta^s + \mathtt{n}_\Gamma$. We define the polynomials

$$\begin{aligned}\boldsymbol{m}_1(\beta) &= \beta B_\Gamma(\beta) + \frac{B_\Gamma(\beta)}{p-1} + 1 + B_\Gamma(\beta)(\kappa_1+\kappa_2)\\ &= \mathtt{m}_\Gamma\beta^{s+1} + \Big(\frac{\mathtt{m}_\Gamma}{p-1} + \mathtt{m}_\Gamma(\kappa_1+\kappa_2)\Big)\beta^s + \mathtt{n}_\Gamma\beta + \frac{\mathtt{n}_\Gamma}{p-1} + \mathtt{n}_\Gamma(\kappa_1+\kappa_2) + 1\end{aligned}$$

and

$$\begin{aligned}\boldsymbol{m}_2(\beta) &= \beta B_\Gamma(\beta) + \frac{B_\Gamma(\beta)}{p-1} + 1 + B_\Gamma(\beta)\kappa_1\\ &= \mathtt{m}_\Gamma\beta^{s+1} + \Big(\frac{\mathtt{m}_\Gamma}{p-1} + \mathtt{m}_\Gamma\kappa_1\Big)\beta^s + \mathtt{n}_\Gamma\beta + \frac{\mathtt{n}_\Gamma}{p-1} + \mathtt{n}_\Gamma\kappa_1 + 1.\end{aligned}$$

Following the arguments in the previous subsections but using the congruences for the traces of Hecke operators in 4.14.3 Corollary instead of those in 4.13.2 Theorem we obtain this:

Theorem. *If $\tilde{\boldsymbol{G}}$ is connected and $\mathbb{Q}$-split with root system of type C_2 then* 5.2 *Theorem holds with $\boldsymbol{m}_1, \boldsymbol{m}_2$ as defined above.*

We want to determine the polynomial $\boldsymbol{m}_2$ more explicitly. Since $\tilde{\boldsymbol{G}}/\mathbb{Q}$ is of type C_2 there are two simple roots $\Delta = \{\alpha_1, \alpha_2\}$, the positive roots are $\Phi^+ = \{\alpha_1, \alpha_2, \alpha_1+\alpha_2, 2\alpha_1+\alpha_2\}$, hence, $s = 4$ and $2\rho_{\tilde{\boldsymbol{G}}} = 4\alpha_1 + 3\alpha_2$. We denote by $g = g_d$ the number of d cells in a cell complex which is homotopy equivalent to the Borel–Serre compactification of $\Gamma\backslash X$. We assume that $h \in \tilde{\boldsymbol{T}}(\mathbb{Q})^{++}$ satisfies $v_p(\alpha_1(h)) = v_p(\alpha_2(h)) = 1$. We then obtain

$$\kappa_1 = \sum_{\alpha>0} v_p(\alpha(h)) = 7, \quad \kappa_2 = \max(m_{\alpha_1}, m_{\alpha_2}) = 4, \quad \sigma = \max_{\alpha>0} \text{ht}(\alpha) = 3,$$

and hence

$$\mathtt{m}_\Gamma = \frac{12g\sigma^{s+1}}{s} = \frac{12g3^5}{4} = 729g.$$

To determine $\mathtt{n}_\Gamma$ we have to find $M \in \mathbb{N}$ such that equations (23) and (24) hold. A numerical evaluation shows that we may choose $M = 34$; hence, equation (26) yields

$$\mathtt{n}_\Gamma = \left\lceil 2^{1/(s+1)} \frac{g\sigma^{s+1}M^s}{s} \right\rceil + 1 = \left\lceil \frac{2^{1/5}g3^5 1336336}{4} \right\rceil + 1 \le 93254104g + 1.$$

This yields the bound

$$\begin{aligned}\boldsymbol{m}_2(X) &\le 729gX^5 + 729g\big(\tfrac{1}{p-1} + 7\big)X^4 + (93254104g + 1)X \\ &\qquad\qquad + (93254104g + 1)\big(\tfrac{1}{p-1} + 7\big) + 1 \\ &\le 729gX^5 + 5832gX^4 + (93254104g + 1)X + 746032832g + 9.\end{aligned}$$

We note that since p is in the level of Γ the number $g = g_\Gamma$ depends on the prime p.

Acknowledgement

I am grateful to A. Nair for pointing out an error in an earlier version and for enlightening discussions concerning the work of Goresky–Kottwitz–MacPherson.

References

[Andreatta et al. 2015] F. Andreatta, A. Iovita, and V. Pilloni, "p-adic families of Siegel modular cuspforms", *Ann. of Math.* (2) **181**:2 (2015), 623–697. MR Zbl

[Ash and Stevens 2008] A. Ash and G. Stevens, "p-adic deformations of arithmetic cohomology", preprint, 2008, https://www2.bc.edu/~ashav/Papers/Ash-Stevens-Oct-07-DRAFT-copy.pdf.

[Bewersdorff 1985] J. Bewersdorff, *Eine Lefschetzsche Fixpunktformel für Hecke-Operatoren*, Dissertation (Bonner Mathematische Schriften **164**), Universität Bonn, 1985. MR Zbl

[Borel 1970] A. Borel, "Properties and linear representations of Chevalley groups", pp. 1–55 in *Seminar on algebraic groups and related finite groups* (Princeton, NJ, 1968/69), Lecture Notes in Mathematics **131**, Springer, Berlin, 1970. MR Zbl

[Borel and Serre 1973] A. Borel and J.-P. Serre, "Corners and arithmetic groups", *Comment. Math. Helv.* **48** (1973), 436–491. MR Zbl

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

[Buzzard 1998] K. Buzzard, "p-adic modular forms on definite quaternion algebras", preprint, 1998, http://wwwf.imperial.ac.uk/~buzzard/maths/research/notes/definite.dvi.

[Buzzard 2001] K. Buzzard, "Families of modular forms", *J. Théor. Nombres Bordeaux* **13**:1 (2001), 43–52. MR Zbl

[Buzzard and Calegari 2004] K. Buzzard and F. Calegari, "A counterexample to the Gouvêa–Mazur conjecture", *C. R. Math. Acad. Sci. Paris* **338**:10 (2004), 751–753. MR Zbl

[Coleman 1997] R. F. Coleman, "p-adic Banach spaces and families of modular forms", *Invent. Math.* **127**:3 (1997), 417–479. MR Zbl

[Franke 1998] J. Franke, "Harmonic analysis in weighted L_2-spaces", *Ann. Sci. École Norm. Sup.* (4) **31**:2 (1998), 181–279. MR Zbl

[Goresky and MacPherson 2003] M. Goresky and R. MacPherson, "The topological trace formula", *J. Reine Angew. Math.* **560** (2003), 77–150. MR Zbl

[Goresky et al. 1994] M. Goresky, G. Harder, and R. MacPherson, "Weighted cohomology", *Invent. Math.* **116**:1-3 (1994), 139–213. MR Zbl

[Goresky et al. 1997] M. Goresky, R. Kottwitz, and R. MacPherson, "Discrete series characters and the Lefschetz formula for Hecke operators", *Duke Math. J.* **89**:3 (1997), 477–554. MR Zbl

[Gouvêa and Mazur 1992] F. Gouvêa and B. Mazur, "Families of modular eigenforms", *Math. Comp.* **58**:198 (1992), 793–805. MR Zbl

[Hida 1988] H. Hida, "On p-adic Hecke algebras for GL_2 over totally real fields", *Ann. of Math.* (2) **128**:2 (1988), 295–384. MR Zbl

[Hida 1993] H. Hida, *Elementary theory of L-functions and Eisenstein series*, London Mathematical Society Student Texts **26**, Cambridge University Press, 1993. MR Zbl

[Humphreys 1972] J. E. Humphreys, *Introduction to Lie algebras and representation theory*, Graduate Texts in Mathematics **9**, Springer, New York, 1972. MR Zbl

[Iwahori and Matsumoto 1965] N. Iwahori and H. Matsumoto, "On some Bruhat decomposition and the structure of the Hecke rings of $\mathfrak{p}$-adic Chevalley groups", *Inst. Hautes Études Sci. Publ. Math.* **1965**:25 (1965), 5–48. MR Zbl

[Jantzen 2003] J. C. Jantzen, *Representations of algebraic groups*, 2nd ed., Mathematical Surveys and Monographs **107**, American Mathematical Society, Providence, RI, 2003. MR Zbl

[Koecher 1983] M. Koecher, *Lineare Algebra und analytische Geometrie*, Grundwissen Mathematik **2**, Springer, Berlin, 1983. MR Zbl

[Koike 1975] M. Koike, "On some p-adic properties of the Eichler–Selberg trace formula", *Nagoya Math. J.* **56** (1975), 45–52. MR Zbl

[Koike 1976] M. Koike, "On p-adic properties of the Eichler–Selberg trace formula, II", *Nagoya Math. J.* **64** (1976), 87–96. MR Zbl

[Kuga et al. 1981] M. Kuga, W. Parry, and C. H. Sah, "Group cohomology and Hecke operators", pp. 223–266 in *Manifolds and Lie groups* (Notre Dame, IN, 1980), Progr. Math. **14**, Birkhäuser, Boston, 1981. MR Zbl

[Mahnkopf 2013] J. Mahnkopf, "On truncation of irreducible representations of Chevalley groups", *J. Number Theory* **133**:9 (2013), 3149–3182. MR Zbl

[Mahnkopf 2014] J. Mahnkopf, "On truncation of irreducible representations of Chevalley groups, II", *J. Number Theory* **143** (2014), 46–61. MR Zbl

[Nair 1999] A. Nair, "Weighted cohomology of arithmetic groups", *Ann. of Math.* (2) **150**:1 (1999), 1–31. MR Zbl

[Pande 2009] A. Pande, "Local constancy of dimensions of Hecke eigenspaces of automorphic forms", *J. Number Theory* **129**:1 (2009), 15–27. MR

[Popov and Vinberg 1994] V. L. Popov and E. B. Vinberg, *Invariant theory*, edited by I. R. Shafarevich and A. N. Parshin, Encyclopaedia of Mathematical Sciences **55**, Springer, Berlin, 1994. MR Zbl

[Schwermer 1994] J. Schwermer, "Eisenstein series and cohomology of arithmetic groups: the generic case", *Invent. Math.* **116**:1-3 (1994), 481–511. MR Zbl

[Spallone 2009] S. Spallone, "Stable discrete series characters at singular elements", *Canad. J. Math.* **61**:6 (2009), 1375–1382. MR Zbl

[Springer 1981] T. A. Springer, *Linear algebraic groups*, Birkhäuser, Boston, 1981. MR Zbl

[Urban 2011] E. Urban, "Eigenvarieties for reductive groups", *Ann. of Math.* (2) **174**:3 (2011), 1685–1784. MR Zbl

[Wan 1998] D. Wan, "Dimension variation of classical and p-adic modular forms", *Invent. Math.* **133**:2 (1998), 449–463. MR Zbl

Received March 27, 2015. Revised May 20, 2016.

JOACHIM MAHNKOPF
FAKULTÄT FÜR MATHEMATIK
UNIVERSITÄT WIEN
OSKAR-MORGENSTERN-PLATZ 1
1190 WIEN
AUSTRIA

joachim.mahnkopf@univie.ac.at

PACIFIC JOURNAL OF MATHEMATICS
Vol. 289, No. 2, 2017
dx.doi.org/10.2140/pjm.2017.289.381

WEAKENING IDEMPOTENCY IN K-THEORY

VLADIMIR MANUILOV

We show that the K-theory of C^*-algebras can be defined by pairs of matrices a, b satisfying less strict relations than idempotency, namely $p(a) = p(b)$ for any polynomial p with $p(0) = p(1) = 0$, which means that a and b have to be "projections" only where $a \neq b$.

1. Introduction

The K-theory of a C^*-algebra A is patently defined by pairs (formal differences) of idempotent matrices (projections) over A. Regretfully, projection is a very strict property, and it is usually very hard to find projections in a given C^*-algebra. Many famous conjectures (Kadison, Novikov, Baum–Connes, Bass, etc.) are related to projections and would become more tractable if one could provide enough projections for a given C^*-algebra. Our aim is to show that the K-theory can be defined using less-restrictive relations in the hope that it will be easier to find elements satisfying these relations than the genuine idempotency. We show that K-theory is generated by pairs a, b of matrices over A satisfying $p(a) = p(b)$ for any polynomial p with $p(0) = p(1) = 0$, which means that a and b have to be "projections" only where $a \neq b$.

2. Definitions and some properties

Let A be a C^*-algebra. For $a, b \in A$, consider the relations

$$\|a\| \leq 1, \quad \|b\| \leq 1, \quad a, b \geq 0, \quad (a-a^2)(a-b) = 0, \quad (b-b^2)(a-b) = 0. \tag{1}$$

Definition 2.1. A pair (a, b) of elements in a C^*-algebra is called *balanced* if it satisfies the relations (1).

Two balanced pairs (a_0, b_0) and (a_1, b_1) of elements in A are *homotopy equivalent* if there are paths $a = (a_t)$, $b = (b_t) : [0, 1] \to A$, connecting a_0 with a_1 and b_0 with b_1 respectively, such that the pair (a_t, b_t) is balanced for each $t \in [0, 1]$.

A balanced pair (a, b) is *homotopy trivial* if it is homotopy equivalent to $(0, 0)$.

The author acknowledges partial support by the RFBR Grant No. 16-01-00357.
MSC2010: 19K99, 46L80, 46L05.
Keywords: K-theory, C^*-algebra, projection.

Lemma 2.2. *A balanced pair (a, a) is homotopy trivial for any $a \in A$.*

Proof. The linear homotopy $a_t = t \cdot a$ would do. $\square$

Lemma 2.3. *If a pair (a, b) is balanced then $f(a) = f(b)$ and $f(a)(a-b) = 0$ for any $f \in C_0(0, 1)$.*

Proof. It follows from $(a-a^2)(a-b) = 0$, or, equivalently, from $(a-a^2)a = (a-a^2)b$, that

$$(a-a^2)a^2 = a(a-a^2)a = a(a-a^2)b = (a-a^2)b^2;$$

hence

$$(a-a^2)(a-a^2) = (a-a^2)(b-b^2).$$

Similarly,

$$(b-b^2)(b-b^2) = (a-a^2)(b-b^2);$$

therefore

$$(a-a^2)^2 = (b-b^2)^2. \tag{2}$$

Then (2) and the positivity of $a-a^2$ and $b-b^2$ imply that

$$a-a^2 = b-b^2.$$

Also,

$$(a-a^2)a = (a-a^2)b = (b-b^2)b.$$

Since the two functions g and h given by $g(t) = t-t^2$ and $h(t) = tg(t)$ generate $C_0(0, 1)$, and since $g(a) = g(b)$ and $h(a) = h(b)$, we conclude that the same holds for any $f \in C_0(0, 1)$. Similarly, $g(a)(a-b) = 0$ and $h(a)(a-b) = 0$ imply $f(a)(a-b) = 0$ for any $f \in C_0(0, 1)$. $\square$

Corollary 2.4. *If $\|a\| < 1$, $\|b\| < 1$ and the pair (a, b) is balanced then $a = b$; hence the pair (a, b) is homotopy trivial.*

Proof. Take $f \in C_0(0, 1)$ such that $f(t) = t \in \mathrm{Sp}(a) \cup \mathrm{Sp}(b)$ and $f(1) = 0$. Then $a = f(a)$, $b = f(b)$, and the claim follows from Lemma 2.3. $\square$

Lemma 2.5. *Let $f : [0, 1] \to [0, 1]$ be a continuous map such that $f(0) = 0$ and $f(1) = 1$. If (a, b) is a balanced pair then the pair $(f(a), f(b))$ is also balanced and is homotopy equivalent to (a, b).*

Proof. As the set of all functions with the stated properties is convex, it suffices to show that for any such function f, the pair $(f(a), f(b))$ satisfies the relations (1).

Set $f_0(t) = f(t) - t$. Then $f_0 \in C_0(0, 1)$. As $f_0(a) = f_0(b)$ by Lemma 2.3,

$$f(a) - f(b) = a - b.$$

Set

$$g(t) = t - t^2 + f_0(t) - f_0^2(t) - 2tf_0(t).$$

Then $g \in C_0(0, 1)$ and

$$(f(a) - f^2(a))(f(a) - f(b)) = g(a)(a-b) = 0,$$
$$(f(b) - f^2(b))(f(a) - f(b)) = g(a)(a-b) = 0. \qquad \square$$

Corollary 2.6. *If a pair* (a, b) *is balanced then* $\mathrm{Sp}(a) \setminus \{0, 1\} = \mathrm{Sp}(b) \setminus \{0, 1\}$.

Proof. The inner points of $[0, 1]$ in the two spectra must coincide by Lemma 2.3. $\square$

Let $M_n(A)$ denote the $n \times n$ matrix algebra over A. Two balanced pairs (a_0, b_0) and (a_1, b_1), where $a_0, a_1, b_0, b_1 \in M_n(A)$, are equivalent if there is a homotopy trivial balanced pair (a, b), $a, b \in M_m(A)$ for some integer m, such that the balanced pairs $(a_0 \oplus a, b_0 \oplus b)$ and $(a_1 \oplus a, b_1 \oplus b)$ are homotopy equivalent in $M_{n+m}(A)$. Using the standard inclusion $M_n(A) \subset M_{n+k}(A)$ (as the upper-left corner) we may speak about the equivalence of balanced pairs of different matrix size.

Let $[(a, b)]$ denote the equivalence class of the balanced pair (a, b), $a, b \in M_n(A)$. For two balanced pairs (a, b), $a, b \in M_n(A)$, and (c, d), $c, d \in M_m(A)$, set

$$[(a, b)] + [(c, d)] = [(a \oplus c, b \oplus d)].$$

The result obviously doesn't depend on the choice of representatives. Also $[(a, b)] + [(c, d)] = [(a, b)]$ when (c, d) is homotopy trivial.

Lemma 2.7. *The addition is commutative and associative.*

Proof. If $(u_t)_{t \in [0,1]}$ is a path of unitaries in A with $u_1 = 1$ and $u_0 = u$, then $[(u^*au, u^*bu)] = [(a, b)]$ for any $a, b \in A$, as the relations (1) are not affected by unitary equivalence. The standard argument with a unitary path connecting $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ proves commutativity. A similar argument proves associativity. $\square$

Lemma 2.8. $\qquad [(a, b)] + [(b, a)] = [(0, 0)]$ *for any* a, b.

Proof. Set

$$U_t = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}, \qquad B_t = U_t^* \begin{pmatrix} b & 0 \\ 0 & a \end{pmatrix} U_t.$$

We claim that the pair $\left(\left(\begin{smallmatrix} a & 0 \\ 0 & b \end{smallmatrix}\right), B_t\right)$ is balanced for all t.

One has

$$B_t = \begin{pmatrix} b\cos^2 t + a\sin^2 t & (a-b)\cos t \sin t \\ (a-b)\cos t \sin t & b\sin^2 t + a\cos^2 t \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} + (a-b)C_t, \tag{3}$$

where

$$C_t = \begin{pmatrix} -\cos^2 t & \cos t \sin t \\ \cos t \sin t & \cos^2 t \end{pmatrix}.$$

Then

$$\left(\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} - \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}^2\right)\left(\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} - B_t\right) = \begin{pmatrix} a-a^2 & 0 \\ 0 & b-b^2 \end{pmatrix}(a-b)C_t$$
$$= \begin{pmatrix} (a-a^2)(a-b) & 0 \\ 0 & (b-b^2)(a-b) \end{pmatrix} C_t = 0.$$

It remains to show that

$$A = (B_t - B_t^2)\left(\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} - B_t\right) = 0.$$

Using (3) we have

$$\begin{aligned}
A &= \left(\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} + (a-b)C_t - \left(\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} + (a-b)C_t\right)^2\right)(a-b)C_t \\
&= \left(\begin{pmatrix} a-a^2 & 0 \\ 0 & b-b^2 \end{pmatrix} + (a-b)C_t - \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}(a-b)C_t \right. \\
&\qquad \left. - C_t(a-b)\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} - (a-b)^2 C_t^2\right)(a-b)C_t \\
&= \left((a-b)C_t - \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}(a-b)C_t - C_t(a-b)\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} - (a-b)^2 C_t^2\right)(a-b)C_t \\
&= \left(\begin{pmatrix} a-b-a^2+ab & 0 \\ 0 & a-b-ba+b^2 \end{pmatrix} C_t - C_t(a-b)\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \right. \\
&\qquad \left. - (a-b)^2 \cos^2 t \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)(a-b)C_t \\
&= \left(\begin{pmatrix} -b+ab & 0 \\ 0 & a-ba \end{pmatrix} C_t - C_t\begin{pmatrix} a-ba & 0 \\ 0 & ab-b \end{pmatrix} - (a-b)^2 \cos^2 t \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)(a-b)C_t \\
&= \left(\begin{pmatrix} (ab+ba-a-b)\cos^2 t & 0 \\ 0 & (ab+ba-a-b)\cos^2 t \end{pmatrix} \right. \\
&\qquad \left. - \begin{pmatrix} (a-b)^2\cos^2 t & 0 \\ 0 & (a-b)^2\cos^2 t \end{pmatrix}\right)(a-b)C_t \\
&= 0.
\end{aligned}$$

Thus, the balanced pair $(a \oplus b, b \oplus a)$ is homotopy equivalent to the balanced pair $(a \oplus b, a \oplus b)$, and the latter is homotopy trivial by Lemma 2.2. □

So we see that the equivalence classes of balanced pairs in matrix algebras over A form an abelian group for any C^*-algebra A. Let us denote this group by $L(A)$.

Note that pairs of projections are patently balanced. If A is a unital C^*-algebra then $K_0(A)$ consists of formal differences $[p]-[q]$ with p, q projections in matrices over A. Then

$$\iota([p]-[q]) = [(p,q)]$$

gives rise to a morphism $\iota : K_0(A) \to L(A)$.

In the nonunital case, ι can be defined after unitalization. But, as we shall see, unlike K_0, there is no need to unitalize for L. The following example shows the reason for that in the commutative case.

Example 2.9. Let X be a compact Hausdorff space, $x \in X$, $Y = X \setminus \{x\}$, $A = C_0(Y)$, $A^+ = C(X)$. Let $[p]-[q] \in K_0(A)$, where $p, q \in M_n(A^+)$ are projections. Then $p = p_0 + \alpha$ and $q = p_0 + \beta$, where p_0 is constant on X and $\alpha, \beta \in M_n(A)$. Without loss of generality we may assume that $\alpha, \beta = 0$ not only at the point x, but also in a small neighborhood U of x. Let $h \in C(X)$ satisfy $0 \leq h \leq 1$, $h(x) = 0$ and $h(z) = 1$ for any $z \in X \setminus U$. Set $a = hp_0 + \alpha$, $b = hp_0 + \beta$. Then $a, b \in M_n(A)$ and $[(a,b)] \in L(A)$.

Lemma 2.10. $$L(\mathbb{C}) \cong \mathbb{Z}.$$

Proof. Let $a, b \in M_n$, and let the pair (a,b) be balanced. Let $e_1, \dots, e_n$ (resp. $e'_1, \dots, e'_n$) be an orthonormal basis of eigenvectors for a (resp. for b) with eigenvalues $\lambda_1, \dots, \lambda_n$ (resp. $\lambda'_1, \dots, \lambda'_n$). Let $0 < \lambda_i < 1$. Then e_i is an eigenvector for $a - a^2$ with a nonzero eigenvalue $\lambda_i - \lambda_i^2$. As $(a-a^2)(a-b) = 0$, we have $(a-b)(a-a^2) = 0$; hence

$$(a-b)(a-a^2)(e_i) = (\lambda_i - \lambda_i^2)(a-b)(e_i) = 0.$$

Thus $(a-b)(e_i) = 0$, or, equivalently, $a(e_i) = b(e_i)$. As e_i is an eigenvector for a, it is an eigenvector for b as well: $b(e_i) = \lambda_i e_i$. So the eigenvectors, corresponding to the eigenvalues $\neq 0, 1$, are the same for a and b.

Reorder, if necessary, the eigenvalues so that

$$\lambda_1, \dots, \lambda_k \in (0,1), \quad \lambda_{k+1}, \dots, \lambda_n \in \{0,1\},$$

and denote the linear span of $e_1, \dots, e_k$ by L. Similarly, assume that

$$\lambda'_1, \dots, \lambda'_{k'} \in (0,1), \quad \lambda'_{k'+1}, \dots, \lambda'_n \in \{0,1\},$$

and denote the linear span of $e'_1, \dots, e'_{k'}$ by L'. As $e_1, \dots, e_k \in L'$ and, symmetrically, $e'_1, \dots, e'_{k'} \in L$, we have $\dim L = \dim L'$, $k = k'$, and $\lambda_i = \lambda'_i$ for $i = 1, \dots, k$.

Then $L^\perp$ is an invariant subspace for both a and b, and the restrictions $a|_{L^\perp}$ and $b|_{L^\perp}$ are projections (as their eigenvalues equal 0 or 1). We may write a and b as matrices with respect to the decomposition $L \oplus L^\perp$:

$$a = \begin{pmatrix} c & 0 \\ 0 & p \end{pmatrix}, \quad b = \begin{pmatrix} c & 0 \\ 0 & q \end{pmatrix}, \tag{4}$$

where p, q are projections. The linear homotopy

$$a_t = \begin{pmatrix} tc & 0 \\ 0 & p \end{pmatrix}, \quad b_t = \begin{pmatrix} tc & 0 \\ 0 & q \end{pmatrix}, \qquad t \in [0, 1],$$

connects the pair (a, b) with the pair $(p, q) + (0, 0)$. Therefore, $L(\mathbb{C})$ is a quotient of $\mathbb{Z}$ (which is the set of homotopy classes of pairs of projections modulo stable equivalence). To see that $L(\mathbb{C})$ is exactly $\mathbb{Z}$, note that (4) implies that $\operatorname{tr}(a-b) \in \mathbb{Z}$ for any balanced pair (a, b), so this integer is homotopy invariant. $\square$

Remark 2.11. One may think that the relations (1) imply that balanced pairs (a, b) are something like projections plus a common part and can be reduced to just a pair of projections by cutting out the common part. The following example shows that this is not that simple.

Example 2.12. Let $A = C(X)$, and let Y, Z be closed subsets in X with $Y \cap Z = K$. Let $p, q \in M_n(C(Y))$ be projection-valued functions on Y such that $p|_K = q|_K = r$, where r cannot be extended to a projection-valued function on Z due to a K-theory obstruction, but can be extended to a matrix-valued function $s \in M_n(C(Z))$ on Z (with $0 \leq s \leq 1$). Then set

$$a = \begin{cases} p & \text{on } Y, \\ s & \text{on } Z, \end{cases} \qquad \text{and} \qquad b = \begin{cases} q & \text{on } Y, \\ s & \text{on } Z. \end{cases}$$

3. Universal C^*-algebra for relations (1)

Let (a, b) be a balanced pair in a C^*-algebra A. Denote the C^*-subalgebra generated by a and b by $C^*(a, b)$. The universal C^*-algebra for the relations (1) is a C^*-algebra D generated by elements $\boldsymbol{a}, \boldsymbol{b} \in D$ satisfying the relations (1) such that for any balanced pair (a, b) there is a surjective $*$-homomorphism $\varphi : D \to C^*(a, b)$ with $\varphi(\boldsymbol{a}) = a$ and $\varphi(\boldsymbol{b}) = b$; see [Loring 1997].

Let $I \subset C^*(a, b)$ denote the ideal generated by $a - a^2$, and let $C^*(a, b)/I$ be the quotient C^*-algebra. Then $C^*(a, b)/I$ is generated by $\dot{a} = q(a)$ and $\dot{b} = q(b)$, where q is the quotient map. But since $q(a - a^2) = q(b - b^2) = 0$, $\dot{a}$ and $\dot{b}$ are projections, and $C^*(a, b)/I$ is generated by two projections.

Then the C^*-algebra $C^*(a, b)$ is completely determined by the ideal I, by the quotient $C^*(a, b)/I$, and by the Busby invariant $\tau : C^*(a, b)/I \to Q(I)$ (we denote by $M(I)$ the multiplier algebra of I and by $Q(I) = M(I)/I$ the outer multiplier algebra). The latter is defined by the two projections $\tau(\dot{a}), \tau(\dot{b}) \in Q(C_0(Y))$, where $X = \operatorname{Sp}(a)$, $Y = X \setminus \{0, 1\}$. Let $C_b(Y)$ denote the C^*-algebra of bounded continuous functions on Y and let

$$\pi : C_b(Y) \to C_b(Y)/C_0(Y) = Q(C_0(Y))$$

be the quotient map. Using Gelfand duality, we identify a with the function id on $\mathrm{Sp}(a)$. Let $f \in C_0(Y)$. Then

$$\tau(\dot{a})\pi(f(a)) = \tau(\dot{b})\pi(f(a)) = \pi(af(a)),$$

so we can easily calculate these two projections.

If $1 \notin X$ then $\tau(\dot{a}) = \tau(\dot{b}) = 0$; if $X = \{1\}$ then $I = 0$; if $1 \in X$ and X has at least one more point x then $\tau(\dot{a}) = \tau(\dot{b})$ is the class of functions f on X such that $f(1) = 1$ and $f(t) = 0$ for all $t \leq x$.

Let $M_1 \subset M_2$ denote the upper-left corner in the 2-by-2 matrix algebra. Set

$$D = \big\{ f \in C([-1, 1]; M_2) : f(-1) = 0,\ f(1) \text{ is diagonal},\ f(t) \in M_1 \text{ for } t \in (-1, 0] \big\},$$

and let $\boldsymbol{a}, \boldsymbol{b}$ be functions in $C([-1, 1]; M_2)$ defined by

$$\boldsymbol{a}(t) = \begin{cases} \begin{pmatrix} \cos^2 \frac{\pi}{2} t & 0 \\ 0 & 0 \end{pmatrix} & \text{for } t \in [-1, 0], \\ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} & \text{for } t \in [0, 1], \end{cases} \tag{5}$$

$$\boldsymbol{b}(t) = \begin{cases} \begin{pmatrix} \cos^2 \frac{\pi}{2} t & 0 \\ 0 & 0 \end{pmatrix} & \text{for } t \in [-1, 0], \\ \begin{pmatrix} \cos^2 \frac{\pi}{2} t & \cos \frac{\pi}{2} t \sin \frac{\pi}{2} t \\ \cos \frac{\pi}{2} t \sin \frac{\pi}{2} t & \sin^2 \frac{\pi}{2} t \end{pmatrix} & \text{for } t \in [0, 1]. \end{cases} \tag{6}$$

Then $\boldsymbol{a}, \boldsymbol{b} \in D$, the pair $(\boldsymbol{a}, \boldsymbol{b})$ is balanced, and $D = C^*(\boldsymbol{a}, \boldsymbol{b})$ is generated by these $\boldsymbol{a}$ and $\boldsymbol{b}$.

Like all C^*-algebras of the form $C^*(a, b)$ defined by balanced pairs (a, b), the C^*-algebra D is an extension. It contains the ideal

$$J = \{ f \in D : f(t) = 0 \text{ for } t \in [0, 1] \} \cong C_0(-1, 0),$$

which is generated by $\boldsymbol{a} - \boldsymbol{a}^2$. Note that multiplication by $\boldsymbol{a}$ or by $\boldsymbol{b}$ determines the same multiplier $m_{\boldsymbol{a}} = m_{\boldsymbol{b}} \in M(J)$, and that the C^*-algebra $\bar{J}$ generated by J and by $m_{\boldsymbol{a}}$ is isomorphic to $C_0(-1, 0]$. It is the universal C^*-algebra for the relation $0 \leq a \leq 1$, so there exists a surjective $*$-homomorphism $\bar{\alpha}$ from $\bar{J}$ to the nonunital C^*-algebra generated by a such that $\alpha'(m_{\boldsymbol{a}}) = m_a$, where $m_a \in M(I)$ is the multiplier defined by multiplication by a on A. The restriction $\alpha = \bar{\alpha}|_J$ maps J onto I, and $\alpha(f(\boldsymbol{a})) = f(a)$ for any $f \in C_0(0, 1)$.

The quotient D/J is the universal (nonunital) C^*-algebra

$$D/J = \mathbb{C} * \mathbb{C} = \{ m \in C([0, 1], M_2) : m(1) \text{ is diagonal},\ m(0) \in M_1 \} \tag{7}$$

generated by two projections $\dot{\boldsymbol{a}}$ and $\dot{\boldsymbol{b}}$ [Raeburn and Sinclair 1989]. Therefore, D/J surjects onto any C^*-algebra generated by two projections in a canonical way. Note that D/J is an extension of $\mathbb{C}$ by the C^*-algebra

$$q\mathbb{C} = \{m \in C_0((0, 1], M_2) : m(1) \text{ is diagonal}\}$$

used in the Cuntz picture of K-theory.

Lemma 3.1. *The C^*-algebra D is universal for the relations* (1).

Proof. For any balanced pair (a, b), the universality of $\bar{J}$ and of D/J implies the existence of surjective $*$-homomorphisms $\alpha : J \to I$ and $\gamma : D/J \to C^*(a, b)/I$ such that $\bar{\alpha}(\boldsymbol{a}) = a$ and $\gamma(\dot{\boldsymbol{a}}) = \dot{a}$, $\gamma(\dot{\boldsymbol{b}}) = \dot{b}$. Since α is surjective, it induces $*$-homomorphisms $M(\alpha) : M(J) \to M(I)$ and $Q(\alpha) : Q(J) \to Q(I)$ in a canonical way, and $M(\alpha)|_{\bar{J}} = \bar{\alpha}$. One has

$$D \cong \{(m, f) : m \in M(J),\ f \in D/J,\ q_J(m) = \tau(f)\}, \tag{8}$$

$$C^*(a, b) \cong \{(n, g) : n \in M(I),\ g \in C^*(a, b)/I,\ q_I(n) = \sigma(g)\}, \tag{9}$$

where $q_\bullet : M(\bullet) \to Q(\bullet)$ is the quotient map; hence the map $\varphi : D \to C^*(a, b)$ can be defined by $\varphi(m, f) = (M(\alpha)(m), \gamma(f))$. This map is well defined if the diagram

$$\begin{tikzcd} D/J \arrow[r, "\tau"] \arrow[d, "\gamma"] & Q(J) \arrow[d, "Q(\alpha)"] \\ C^*(a,b)/I \arrow[r, "\sigma"] & Q(I) \end{tikzcd}$$

commutes. It does commute. The case $X = \mathrm{Sp}(a) = \{1\}$ is trivial. For the other cases, notice that the image of τ lies in $C_0(0, 1]/C_0(0, 1) \subset Q(J)$, and the image of σ lies in $C(X)/C_0(X \setminus \{0\})$, which is either $\mathbb{C}$ or 0 (when $1 \in X$ or $1 \notin X$, respectively), and the restriction of $Q(\alpha)$ from the image of τ to the image of σ is induced by the inclusion $X \subset [0, 1]$. So, there is a surjective $*$-homomorphism φ from D to $C^*(a, b)$.

Under the identification (8), $\boldsymbol{a} \in D$ corresponds to the pair $(m_{\boldsymbol{a}}, \dot{\boldsymbol{a}})$; hence $\varphi(\boldsymbol{a}) = (M(\alpha)(m_{\boldsymbol{a}}), \gamma(\dot{\boldsymbol{a}})) = (\alpha'(m_{\boldsymbol{a}}), \dot{a}) = (m_a, \dot{a})$, and the latter corresponds to a under the identification (9). Similarly, one can check that $\varphi(\boldsymbol{b}) = b$. □

The C^*-algebra D allows one more description. Set $A_0 = \mathbb{C}^2$ and $F = \mathbb{C} \oplus M_2$, and define a $*$-homomorphism $\gamma : A_0 \to F \oplus F$ by $\gamma = \gamma_0 \oplus \gamma_1$, where $\gamma_0, \gamma_1 : \mathbb{C}^2 \to \mathbb{C} \oplus M_2$ are given by

$$\gamma_0(\lambda, \mu) = \lambda \oplus \begin{pmatrix} \lambda & 0 \\ 0 & 0 \end{pmatrix}, \quad \gamma_1(\lambda, \mu) = 0 \oplus \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}, \qquad \lambda, \mu \in \mathbb{C}.$$

Let $\partial : C([0,1]; F) \to F \oplus F$ be the boundary map given by $\partial(f) = f(0) \oplus f(1)$, $f \in C([0,1]; F)$. Then D can be identified with the pullback

$$\begin{array}{ccc} D = A_1 & \longrightarrow & A_0 \\ \big\downarrow & & \big\downarrow{\scriptstyle\gamma} \\ C([0,1]; F) & \xrightarrow{\partial} & F \oplus F, \end{array}$$

$$D = \{(f, a) : f \in C([0,1]; F),\ a \in A_0,\ \partial(f) = \gamma(a)\}.$$

Such a pullback is called a 1-dimensional noncommutative CW complex (NCCW complex) in [Eilers et al. 1998]; in this terminology, A_0 is a 0-dimensional NCCW complex.

Recall [Blackadar 1985] that a C^*-algebra B is *semiprojective* if for any C^*-algebra A and increasing chain of ideals $I_n \subset A$, $n \in \mathbb{N}$, with $I = \overline{\bigcup_n I_n}$ and for any $*$-homomorphism $\varphi : B \to A/I$ there exist n and $\hat{\varphi} : B \to A/I_n$ such that $\varphi = q \circ \hat{\varphi}$, where $q : A/I_n \to A/I$ is the quotient map.

Corollary 3.2. *The C^*-algebra D is semiprojective.*

Proof. Essentially, this is Theorem 6.2.2 of [Eilers et al. 1998], where it is proved that all unital 1-dimensional NCCW complexes are semiprojective. The nonunital case is dealt with in Theorem 3.15 of [Thiel 2009], where is it noted that if A_1 is a 1-dimensional NCCW complex then A_1^+ is a 1-dimensional NCCW complex as well, and semiprojectivity of A_1 is equivalent to semiprojectivity of A_1^+. □

One more picture of D can be given in terms of an amalgamated free product: $D = C(0,1] *_{C_0(0,1)} C(0,1]$.

4. Identifying L with K_0

Our definition of $L(A)$ can be reformulated in terms of the universal C^*-algebra D as

$$L(A) = \varinjlim [D, M_n(A)],$$

where $[-,-]$ denotes the set of homotopy classes of $*$-homomorphisms. Recall that semiprojectivity is equivalent to stability of relations that determine D [Loring 1997, Theorem 14.1.4]. The latter means that for any $\varepsilon > 0$ there exists $\delta > 0$ such that whenever $c, d \in A$ satisfy

$$\|c\| \le 1, \quad \|d\| \le 1, \quad c, d \ge 0, \quad \|(c - c^2)(c - d)\| < \delta, \quad \|(d - d^2)(c - d)\| < \delta,$$

there exist $a, b \in A$ such that $\|a - c\| < \varepsilon$, $\|b - d\| < \varepsilon$, and a, b satisfy the relations (1). Stability of the relations (1) implies that

$$L(A) = [D, A \otimes \mathbb{K}] = [\![D, A \otimes \mathbb{K}]\!],$$

where $\mathbb{K}$ denotes the C^*-algebra of compact operators and $[\![\cdot, \cdot]\!]$ is the set of homotopy classes of asymptotic homomorphisms.

Lemma 4.1. *The functor L is half-exact.*

Proof. Let

$$0 \longrightarrow I \xrightarrow{i} B \xrightarrow{p} A \longrightarrow 0$$

be a short exact sequence of C^*-algebras. It is obvious that $p_* \circ i_* = 0$, so it remains to check that $\operatorname{Ker} p_* \subset \operatorname{Im} i_*$. Suppose that $a, b \in M_n(B)$, the pair (a, b) is balanced, and $(p(a), p(b)) = 0$ in $L(A)$. This means that there is a homotopy connecting $(p(a), p(b))$ to $(0, 0)$ in $M_k(A)$ for some $k \geq n$ such that the whole path satisfies (1). This homotopy is given by a $*$-homomorphism $\psi : D \to C([0, 1], M_k(A))$ such that $\mathrm{ev}_1 \circ \psi = 0$, where ev_t denotes the evaluation map at $t \in [0, 1]$.

When D is a semiprojective C^*-algebra, the homotopy lifting theorem [Blackadar 2016, Theorem 5.1] asserts that given a commuting diagram

$$\begin{tikzcd}
D \arrow[drr, "\kappa"] \arrow[dr, dashed, "\varphi"] \arrow[ddr, "\psi"'] & & \\
& C([0, 1]; M_k(B)) \arrow[r, "\mathrm{ev}_0"'] \arrow[d, "\bar{p}_k"] & M_k(B) \arrow[d, "p_k"] \\
& C([0, 1]; M_k(A)) \arrow[r, "\mathrm{ev}_0"] & M_k(A),
\end{tikzcd}$$

where $\bar{p}_k$ and p_k are the $*$-homomorphisms induced by a surjection p, there exists a $*$-homomorphism φ completing the diagram. Replacing A and B by matrices over these C^*-algebras, we get a lifting φ for the given homotopy. It follows from $\mathrm{ev}_1 \circ \psi = 0$ that $\mathrm{ev}_1 \circ \varphi$ maps D to $M_k(I)$. Thus (a, b) lies in the image of i_*. $\square$

In the standard way, set $L_n(A) = L(S^n A)$, where SA denotes the suspension over A. Then, by Theorem 21.4.3 of [Blackadar 1986], $L_n(A)$, being homotopy invariant and half-exact, is a homology theory. Also, by Theorem 22.3.6 of that paper and by Lemma 2.10, it coincides with the K-theory on the bootstrap category of C^*-algebras. We shall show now that it coincides with the K-theory for any C^*-algebra.

Set

$$P = \begin{pmatrix} 1 - \boldsymbol{b} & f(\boldsymbol{a}) \\ f(\boldsymbol{a}) & \boldsymbol{a} \end{pmatrix}, \qquad Q = \begin{pmatrix} 1 - \boldsymbol{b} & f(\boldsymbol{a}) \\ f(\boldsymbol{a}) & \boldsymbol{b} \end{pmatrix},$$

where $\boldsymbol{a}, \boldsymbol{b}$ are generators for D ((5)–(6)), and $f \in C_0(0, 1)$ is given by $f(t) = (t - t^2)^{1/2}$. Then $P, Q \in M_2(D^+)$, where D^+ denotes the unitalization of D.

By Lemma 2.3, $f(\boldsymbol{a}) = f(\boldsymbol{b})$ and $\boldsymbol{a} f(\boldsymbol{a}) = \boldsymbol{b} f(\boldsymbol{a})$, so P and Q are projections. One also has $P - Q \in M_2(D)$; hence

$$x = [P] - [Q] \in K_0(D).$$

Lemma 4.2. *$K_0(D) \cong \mathbb{Z}$ with x as a generator.*

Proof. Consider the short exact sequence

$$0 \longrightarrow J \longrightarrow D \xrightarrow{\pi} \mathbb{C} * \mathbb{C} \longrightarrow 0,$$

where $\mathbb{C}*\mathbb{C}$ is the universal (nonunital) C^*-algebra (7) generated by two projections, p and q [Raeburn and Sinclair 1989], and π is given by restriction to $[0, 1]$, $\pi(\boldsymbol{a}) = p$, $\pi(\boldsymbol{b}) = q$. We have $\pi(P) = (1 - q) \oplus p$ and $\pi(Q) = (1 - q) \oplus q$, so $\pi_*(x) = [p] - [q] \in K_0(\mathbb{C} * \mathbb{C})$. For $t \in [-1, 0]$, one has $P(t) = Q(t)$; hence, for the boundary (exponential) map $\delta : K_0(\mathbb{C}*\mathbb{C}) \to K_1(J)$, we have $\delta(P) = \delta(Q)$. Recall that $J \cong C_0(-1, 0)$. Direct calculation shows that $\delta(P) = \delta(Q) \neq 0$. The claim follows now from the K-theory exact sequence

$$0 = K_0(J) \longrightarrow K_0(D) \xrightarrow{\pi_*} K_0(\mathbb{C} * \mathbb{C}) \xrightarrow{\delta} K_1(J) \cong \mathbb{Z}. \qquad \square$$

Let us define a map $\kappa : L(A) \to K_0(A)$. If $l = [(a, b)] \in L(A)$ then the balanced pair (a, b) determines a $*$-homomorphism $\varphi : D \to M_n(A)$ by $\varphi(\boldsymbol{a}) = a$ and $\varphi(\boldsymbol{b}) = b$. So, $l \in L(A)$ determines a $*$-homomorphism φ up to homotopy (for some n). Put

$$\kappa(l) = \varphi_*(x) \in K_0(A).$$

It is easy to see that the map κ is a well-defined group homomorphism.

Recall that there is also a map $\iota : K_0(A) \to L(A)$ given by $\iota([p]-[q]) = [(p, q)]$, where $[p] - [q] \in K_0(A)$.

Lemma 4.3. *For any unital C^*-algebra A, one has $\kappa \circ \iota = \mathrm{id}_{K_0(A)}$ and $\iota \circ \kappa = \mathrm{id}_{L(A)}$; hence $L(A) = K_0(A)$.*

Proof. To show the first identity, let $z \in K_0(A)$ and let $p, q \in M_n(A)$ be projections such that $z = [p] - [q]$. Let $\varphi : D \to M_n(A)$ be a $*$-homomorphism determined by the pair (p, q). Then, due to the universality of $\mathbb{C} * \mathbb{C}$, φ factorizes through $\mathbb{C} * \mathbb{C}$, $\varphi = \psi \circ \pi$, where $\pi : D \to \mathbb{C} * \mathbb{C}$ is the quotient map and $\psi : \mathbb{C} * \mathbb{C} \to M_n(A)$ is determined by $\psi(i_1(1)) = p$ and $\psi(i_2(1)) = q$, where $i_1, i_2 : \mathbb{C} \to \mathbb{C} * \mathbb{C}$ are inclusions onto the first and the second copy of $\mathbb{C}$. Then

$$\varphi(x) = \psi_*\big([i_1(1)] - [i_2(1)]\big) = [p] - [q];$$

hence $\kappa(\iota(z)) = z$.

Let us show the second identity. For $[(a, b)] \in L(A)$, let $\varphi : D \to M_n(A)$ be a $*$-homomorphism defined by the balanced pair (a, b) (i.e., by $\varphi(\boldsymbol{a}) = a$

and $\varphi(\boldsymbol{b}) = b$), and let $\varphi^+ : D^+ \to M_n(A)$ be its extension, $\varphi^+(1) = 1$. Then $\iota(\kappa([(a, b)])) = [(\varphi_2^+(P), \varphi_2^+(Q))]$, where $\varphi_2^+ = \varphi^+ \otimes \mathrm{id}_{M_2}$.

For $s \in [0, 1]$, set

$$P_s = C_s P C_s, \quad Q_s = C_s Q C_s, \qquad \text{where } C_s = \begin{pmatrix} s \cdot 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Then

$$P_s, Q_s \in M_2(D^+), \quad P_s - Q_s \in M_2(D), \quad 0 \le P_s, Q_s \le 1,$$
$$(P_s - P_s^2)(P_s - Q_s) = 0, \quad (Q_s - Q_s^2)(P_s - Q_s) = 0$$

for all $s \in [0, 1]$, $P_0, Q_0 \in M_2(D)$, and

$$P_1 = P, \quad Q_1 = Q, \qquad P_0 = \begin{pmatrix} 0 & 0 \\ 0 & \boldsymbol{a} \end{pmatrix}, \quad Q_0 = \begin{pmatrix} 0 & 0 \\ 0 & \boldsymbol{b} \end{pmatrix}.$$

Therefore, $(\varphi_2^+(P_s), \varphi_2^+(Q_s))$ provides a homotopy connecting $(\varphi_2^+(P), \varphi_2^+(Q))$ with $(0 \oplus a, 0 \oplus b)$; hence, the balanced pair $(\varphi_2^+(P), \varphi_2^+(Q))$ is equivalent to the balanced pair (a, b). $\square$

Theorem 4.4. *The functors L and K_0 coincide for any C^*-algebra A.*

Proof. Both functors are half-exact and coincide for unital C^*-algebras, so the claim follows. $\square$

Remark 4.5. Similarly to D, one can define a C^*-algebra D_B for any C^*-algebra B as an appropriate extension of $B * B$ by CB, where CB is the cone over B (or by $D_B = CB *_{SB} CB$). Then one gets the group $[D_B, A \otimes \mathbb{K}]$. Regretfully, D_B has no nice presentation (unlike $D = D_{\mathbb{C}}$), so we don't pursue here the bivariant version.

5. Yet another picture for K-theory

Consider the relations

$$a^* = a, \quad b^* = b, \quad a - a^2 = b - b^2, \quad a(a - a^2) = b(b - b^2). \tag{10}$$

This is equivalent to

$$a^* = a, \quad b^* = b, \quad f(a) = f(b)$$

for any polynomial (or, equivalently, for any continuous function) f such that

$$f(0) = f(1) = 0. \tag{11}$$

As before, for a C^*-algebra A we can define a group $L'(A)$ of homotopy classes of pairs (a, b), where a, b are matrices over A satisfying the relations (10) instead of (1). Note that the relations (10) do not impose any bound for norms of a, b;

hence they do not determine a universal C^*-algebra. Nevertheless, the relations (10) give the same functor.

Proposition 5.1. *The group $L'(A)$ is canonically isomorphic to $K_0(A)$.*

Proof. Let us construct maps $i : L(A) \to L'(A)$ and $j : L'(A) \to L(A)$. In the proof of Lemma 2.3 it was shown that if (a, b) is balanced then they satisfy (10) too, so we can define $i([(a, b)]) = [(a, b)]$. For $r \geq 0$, set

$$c_r(t) = \begin{cases} -r & \text{for } t < -r, \\ t & \text{for } -r \leq t \leq r+1, \\ r+1 & \text{for } t > r+1. \end{cases}$$

It is obvious that the pair $(c_r(a), c_r(b))$ satisfies (10) for any $r \geq 0$.

We claim that the pair $(c_0(a), c_0(b))$ is balanced. Indeed, first we obviously have $c_0(a), c_0(b) \geq 0$ and $\|c_0(a)\|, \|c_0(b)\| \leq 1$. Then, $c_0(a) - c_0(a)^2 = f(a)$, where the function

$$f(t) = \begin{cases} t - t^2 & \text{for } t \in [0, 1], \\ 0 & \text{for } t \notin [0, 1] \end{cases}$$

satisfies (11); so $c_0(a) - c_0(a)^2 = c_0(b) - c_0(b)^2$. Similarly, $c_0(a)(c_0(a) - c_0(a)^2) = c_0(b)(c_0(b) - c_0(b)^2)$. Then

$$\begin{aligned}(c_0(a) - c_0(a)^2)(c_0(a) - c_0(b)) &= c_0(a)^2 - c_0(a)^3 - (c_0(a) - c_0(a)^2)c_0(b) \\ &= c_0(b)^2 - c_0(b)^3 - (c_0(b) - c_0(b)^2)c_0(b) = 0.\end{aligned}$$

Therefore, we can set $j([(a, b)]) = [(c_0(a), c_0(b))]$. Obviously, $j \circ i$ is the identity map, so it remains to check that $i \circ j$ is the identity map as well. Set

$$a_s = \begin{cases} a & \text{for } s = 1, \\ c_{\tan \frac{\pi}{2} s}(a) & \text{for } s \in [0, 1). \end{cases}$$

Then (a_s, b_s), $s \in [0, 1]$, is a required continuous homotopy that connects the balanced pairs (a, b) and $(c_0(a), c_0(b))$. $\square$

References

[Blackadar 1985] B. Blackadar, "Shape theory for C^*-algebras", *Math. Scand.* **56**:2 (1985), 249–275. MR Zbl

[Blackadar 1986] B. Blackadar, *K-theory for operator algebras*, Mathematical Sciences Research Institute Publications **5**, Springer, New York, 1986. MR Zbl

[Blackadar 2016] B. Blackadar, "The homotopy lifting theorem for semiprojective C^*-algebras", *Math. Scand.* **118**:2 (2016), 291–302. MR Zbl

[Eilers et al. 1998] S. Eilers, T. A. Loring, and G. K. Pedersen, "Stability of anticommutation relations: an application of noncommutative CW complexes", *J. Reine Angew. Math.* **499** (1998), 101–143. MR Zbl

[Loring 1997] T. A. Loring, *Lifting solutions to perturbing problems in C^*-algebras*, Fields Institute Monographs **8**, American Mathematical Society, Providence, RI, 1997. MR Zbl

[Raeburn and Sinclair 1989] I. Raeburn and A. M. Sinclair, "The C^*-algebra generated by two projections", *Math. Scand.* **65**:2 (1989), 278–290. MR Zbl

[Thiel 2009] H. Thiel, *One-dimensional C^*-algebras and their K-theory*, Diplomarbeit, Westfälische Wilhelms-Universität Münster, 2009, available at http://tinyurl.com/hannesthiel.

Received November 10, 2015. Revised August 20, 2016.

VLADIMIR MANUILOV
DEPARTMENT OF MECHANICS AND MATHEMATICS
MOSCOW STATE UNIVERSITY
LENINSKIE GORY
MOSCOW
119991
RUSSIA
manuilov@mech.math.msu.su

PACIFIC JOURNAL OF MATHEMATICS
Vol. 289, No. 2, 2017
dx.doi.org/10.2140/pjm.2017.289.395

ON LANGLANDS QUOTIENTS OF THE GENERALIZED PRINCIPAL SERIES ISOMORPHIC TO THEIR AUBERT DUALS

IVAN MATIĆ

We determine under which conditions is the Langlands quotient of an induced representation of the form $\delta \rtimes \sigma$, where δ is an irreducible essentially square-integrable representation of a general linear group and σ is a discrete series representation of the classical p-adic group, isomorphic to its Aubert dual.

1. Introduction

Let F denote a nonarchimedean local field and let G_n stand for the symplectic or (full) orthogonal group having split rank n. The involution on the Grothendieck group of the smooth finite-length representations of a reductive group has been intensively studied by many authors, and we use an involution defined for general reductive p-adic groups in [Aubert 1995] and [Schneider and Stuhler 1997]. This involution is known as the Aubert involution and the image of a representation under this involution is called the Aubert dual of a representation. In this paper we regard the Aubert dual of an admissible finite-length representation as a genuine representation, taking the $+$ or $-$ sign in such way that we obtain the positive element in the appropriate Grothendieck group.

The Aubert involution has a number of prominent applications in the representation theory of classical p-adic groups, and one would also like to gain a deeper knowledge on the explicit structure of the Aubert duals of irreducible representations.

In our previous work [Matić 2016a; 2017], we obtained an explicit description of the Aubert duals of certain classes of discrete series representations of G_n, and in this paper we use developed methods to identify certain classes of irreducible representations which are fixed by the Aubert involution, i.e., which are isomorphic to their Aubert duals. We tackle this problem for the Langlands quotients of the generalized principal series of the group G_n. We note that the generalized principal series is an induced representation of the form $\delta \rtimes \sigma$, obtained by the parabolic

MSC2010: primary 22E35; secondary 11F70, 22E50.
Keywords: discrete series, classical p-adic groups, Aubert involution.

induction with respect to the maximal parabolic subgroup, where the inducing representation $\delta \otimes \sigma$ has an irreducible essentially square-integrable representation on the general linear group part and an irreducible square-integrable representation on the classical group part. If $\nu^x \delta$ is unitarizable for $x < 0$, where $\nu = |\det|_F$, then the generalized principal series $\delta \rtimes \sigma$ has a unique irreducible (Langlands) quotient, which is also isomorphic to the unique irreducible subrepresentation of $\tilde{\delta} \rtimes \sigma$. Such irreducible nontempered representations can be observed as the first step in the Langlands classification of the nonunitary dual of G_n.

To obtain the necessary conditions under which the Langlands quotient of the generalized principal series is isomorphic to its Aubert dual, we use the Jacquet modules method and some elementary properties of the Aubert involution, together with descriptions of the Jacquet modules of certain discrete series representations, obtained in [Matić 2013; 2016c].

Afterwards, we explicitly determine the Aubert duals of Langlands quotients satisfying the obtained necessary conditions, using methods introduced in [Matić 2017], and further developed in [Matić 2016a]. Perhaps a bit surprisingly, an important role in such a procedure is, in the considered case, played by the composition factors of the generalized principal series $\delta \rtimes \sigma$ with a strongly positive σ, obtained in [Muić 2004] and [Matić 2016b, Proposition 3.2]. Such a description of the composition factors enables us to control the Jacquet modules of the investigated nontempered representations, similarly to in [Matić 2015].

We summarize our main results in the following theorem.

Theorem 1.1. *The Langlands quotient of the generalized principal series*

$$\delta([\nu^x\rho, \nu^y\rho]) \rtimes \sigma, \quad x + y > 0,$$

is isomorphic to its Aubert dual if and only if one of the following holds:

(1) *The discrete series representation σ is cuspidal*, $x = y$, *and $\nu^x\rho \rtimes \sigma$ is irreducible.*

(2) *The discrete series representation σ is cuspidal*, $x = 0$, $y = 1$, *and $\rho \rtimes \sigma$ reduces.*

(3) *The cuspidal representation ρ is self-contragredient*, *the induced representation $\nu^\alpha\rho \rtimes \sigma_{\text{cusp}}$ reduces for $\alpha > 0$ (here σ_{cusp} stands for the partial cuspidal support of σ)*, $y = \alpha + 1$, *and one of the following holds*:

   (i) *x is a half-integer, $\frac{3}{2} \le x \le \alpha$, and σ is the unique irreducible subrepresentation of the induced representation*

$$\nu^x\rho \times \nu^{x+1}\rho \times \cdots \times \nu^\alpha\rho \rtimes \sigma_{\text{cusp}}.$$

(ii) *x is a positive integer*, $x \leq \alpha$, *and σ is the unique irreducible subrepresentation of the induced representation*

$$\nu^x\rho \times \nu^{x+1}\rho \times \cdots \times \nu^\alpha\rho \rtimes \sigma_{\text{cusp}}.$$

(iii) $x = 0$ *and σ is the unique irreducible subrepresentation of the induced representation*

$$\nu\rho \times \nu^2\rho \times \cdots \times \nu^\alpha\rho \rtimes \sigma_{\text{cusp}}.$$

We now describe the contents of the paper in more detail. In Section 2 we set up the notation and terminology, and prove some technical results which will be helpful in our investigation. In Section 3 we state and prove our main results, using a case-by-case consideration.

2. Notation and preliminaries

Let F denote a nonarchimedean local field of characteristic $\neq 2$.

Let us first recall the definition of the Aubert involution and its basic properties.

For a connected reductive p-adic group G defined over F, let Σ denote the set of roots of G with respect to a fixed minimal parabolic subgroup and let Δ stand for a basis of Σ. For $\Theta \subseteq \Delta$, we let P_Θ be the standard parabolic subgroup of G corresponding to Θ and M_Θ the standard Levi factor of G corresponding to Θ.

For a parabolic subgroup P of G with the Levi factor M and a representation σ of M, we denote by $i_M(\sigma)$ a normalized parabolically induced representation of G_n induced from σ. For an admissible finite-length representation σ of G, the normalized Jacquet module of σ with respect to the standard parabolic subgroup having Levi factor equal to M will be denoted by $r_M(\sigma)$. We recall the following definition and results from [Aubert 1995]:

Theorem 2.1. *Define the operator on the Grothendieck group of admissible representations of G of finite length by*

$$D_G = \sum_{\Theta \subseteq \Delta} (-1)^{|\Theta|} i_{M_\Theta} \circ r_{M_\Theta}.$$

The operator D_G has the following properties:

(1) *D_G is an involution.*

(2) *D_G takes irreducible representations to irreducible ones.*

(3) *If σ is an irreducible cuspidal representation, then $D_G(\sigma) = (-1)^{|\Delta|}\sigma$.*

(4) *For a standard Levi subgroup $M = M_\Theta$, we have*

$$D_G \circ i_M = i_M \circ D_M.$$

(5) *For a standard Levi subgroup $M = M_\Theta$, we have*

$$r_M \circ D_G = \mathrm{Ad}(w) \circ D_{w^{-1}(M)} \circ r_{w^{-1}(M)},$$

where w is the longest element of the set $\{w \in W : w^{-1}(\Theta) > 0\}$.

Let us now describe the groups that we consider. We look at the usual towers of orthogonal or symplectic groups $G_n = G(V_n)$ that are the groups of isometries of F-spaces $(V_n, (\cdot\,,\cdot))$, $n \geq 0$, where the form $(\cdot\,,\cdot)$ is nondegenerate and it is skew-symmetric if the tower is symplectic and symmetric otherwise. The set of standard parabolic subgroups will be fixed in a usual way. Then the Levi factors of standard parabolic subgroups have the form

$$M \cong \mathrm{GL}(n_1, F) \times \mathrm{GL}(n_2, F) \times \cdots \times \mathrm{GL}(n_k, F) \times G_{n'}.$$

If δ_i is a representation of $\mathrm{GL}(n_i, F)$ for $i = 1, 2, \ldots, k$, and τ a representation of $G_{n'}$, the induced representation $i_M(\delta_1 \otimes \delta_2 \otimes \cdots \otimes \delta_k \otimes \tau)$ will be denoted by $\delta_1 \times \delta_2 \times \cdots \times \delta_k \rtimes \tau$. We use a similar notation to denote a parabolically induced representation of $\mathrm{GL}(m, F)$.

If π is an irreducible representation of G_n, we denote by $\hat{\pi}$ the representation $\pm D_{G_n}(\pi)$, taking the sign $+$ or $-$ such that $\hat{\pi}$ is a positive element in the Grothendieck group of finite-length admissible representations of G_n. We call $\hat{\pi}$ the Aubert dual of π.

By $\mathrm{Irr}(G_n)$ we denote the set of all irreducible admissible representations of G_n. Furthermore, let $R(G_n)$ denote the Grothendieck group of admissible representations of finite length of G_n and define $R(G) = \oplus_{n \geq 0} R(G_n)$. Similarly, let $\mathrm{Irr}(\mathrm{GL}(n, F))$ denote the set of all irreducible admissible representations of $\mathrm{GL}(n, F)$, let $R(\mathrm{GL}(n, F))$ denote the Grothendieck group of admissible representations of finite length of $\mathrm{GL}(n, F)$ and define $R(\mathrm{GL}) = \oplus_{n \geq 0} R(\mathrm{GL}(n, F))$.

The generalized principal series are the induced representations of the form $\delta \rtimes \sigma$, where $\delta \in R(\mathrm{GL})$ is an irreducible essentially square-integrable representation and $\sigma \in R(G)$ is a discrete series representation.

There is a unique $e(\delta) \in \mathbb{R}$ such that $\nu^{-e(\delta)}\delta$ is unitarizable, where $\nu = |\det|_F$. If $e(\delta) > 0$, the generalized principal series $\delta \rtimes \sigma$ has a unique irreducible (Langlands) quotient, which is also the unique irreducible subrepresentation of $\tilde{\delta} \rtimes \sigma$, where $\tilde{\delta}$ denotes the contragredient of δ.

By the results of [Zelevinsky 1980], such a representation δ is attached to the segment and we write $\delta = \delta([\nu^a\rho, \nu^b\rho])$, where $a, b \in \mathbb{R}$ are such that $b - a$ is a nonnegative integer and $\rho \in \mathrm{Irr}(\mathrm{GL}(n, F))$ is a unitary cuspidal representation. We recall that $\delta([\nu^a\rho, \nu^b\rho])$ is the unique irreducible subrepresentation of the induced representation $\nu^b\rho \times \nu^{b-1}\rho \times \cdots \times \nu^a\rho$.

For our Jacquet module considerations it is more convenient to use the subrepresentation version of the Langlands classification and write a nontempered irreducible representation π of G_n as the unique irreducible (Langlands) subrepresentation of the induced representation of the form $\delta_1 \times \delta_2 \times \cdots \times \delta_k \rtimes \tau$, where $\tau \in \mathrm{Irr}(G_m)$ is a tempered representation, $\delta_i \in \mathrm{Irr}(\mathrm{GL}(n_i, F))$ is an essentially square-integrable representation attached to the segment $[\nu^{a_i}\rho_i, \nu^{b_i}\rho_i]$ for $i = 1, 2, \ldots, k$, and $a_1 + b_1 \leq a_2 + b_2 \leq \cdots \leq a_k + b_k < 0$. In this case, we write $\pi = L(\delta_1 \times \delta_2 \times \cdots \times \delta_k \rtimes \tau)$.

For $\sigma \in \mathrm{Irr}(G_n)$ and $1 \leq k \leq n$, denote by $r_{(k)}(\sigma)$ the normalized Jacquet module of σ with respect to the parabolic subgroup with Levi factor $\mathrm{GL}(k, F) \times G_{n-k}$. Identify $r_{(k)}(\sigma)$ with its semisimplification in $R(\mathrm{GL}(k, F)) \otimes R(G_{n-k})$ and consider

$$\mu^*(\sigma) = 1 \otimes \sigma + \sum_{k=1}^{n} r_{(k)}(\sigma) \in R(\mathrm{GL}) \otimes R(G).$$

The following result, derived in [Tadić 1995], presents the crucial structural formula for our calculations of Jacquet modules.

Theorem 2.2. *Let ρ be an irreducible cuspidal representation of* $\mathrm{GL}(m, F)$ *and $k, l \in \mathbb{R}$ such that $k + l \in \mathbb{Z}_{\geq 0}$. Let σ be an admissible representation of G_n of finite length. Write $\mu^*(\sigma) = \sum_{\tau,\sigma'} \tau \otimes \sigma'$. Then we have*

$$\begin{aligned}\mu^*(\delta([\nu^{-k}\rho, \nu^{l}\rho]) \rtimes \sigma) \\ = \sum_{i=-k-1}^{l} \sum_{j=i}^{l} \sum_{\tau,\sigma'} \delta([\nu^{-i}\tilde{\rho}, \nu^{k}\tilde{\rho}]) \times \delta([\nu^{j+1}\rho, \nu^{l}\rho]) \times \tau \otimes \delta([\nu^{i+1}\rho, \nu^{j}\rho]) \rtimes \sigma'.\end{aligned}$$

We omit $\delta([\nu^{x}\rho, \nu^{y}\rho])$ if $x > y$.

We note the following direct consequence of the previous theorem and of the Casselman square-integrability criterion:

Corollary 2.3. *Let ρ denote an irreducible self-contragredient cuspidal representation of* $\mathrm{GL}(m, F)$ *and $k, l \in \mathbb{R}$ such that $k + l \in \mathbb{Z}_{\geq 0}$ and $k > 0$. If $\sigma \in \mathrm{Irr}(G_n)$ is a discrete series representation, then $\mu^*(\delta([\nu^{-k}\rho, \nu^{l}\rho]) \rtimes \sigma)$ contains an irreducible constituent of the form $\nu^{r}\rho' \otimes \pi$, where $r \leq 0$ and ρ' is cuspidal, if and only if $l \leq 0$. Furthermore, if $l \leq 0$ and $\mu^*(\delta([\nu^{-k}\rho, \nu^{l}\rho]) \rtimes \sigma)$ contains an irreducible constituent of the form $\nu^{r}\rho' \otimes \pi$, where $r \leq 0$ and ρ' is cuspidal, then $r = l$ and $\rho' \cong \rho$.*

The following technical result will be used several times in the paper:

Lemma 2.4. *Suppose that $\pi \in \mathrm{Irr}(G_n)$ is a subrepresentation of an induced representation of the form $\nu^{a_1}\rho_1 \times \nu^{a_2}\rho_2 \times \cdots \times \nu^{a_k}\rho_k \rtimes \pi_1$, where $\rho_i \in \mathrm{Irr}(\mathrm{GL}(m_i, F))$ is a unitary cuspidal self-contragredient representation for $i = 1, 2, \ldots, k$, and π_1 is an admissible representation of finite length. Then the Jacquet module of $\hat{\pi}$ with* *respect to an appropriate parabolic subgroup contains an irreducible representation of the form* $\nu^{-a_1}\rho_1 \otimes \nu^{-a_2}\rho_2 \otimes \cdots \otimes \nu^{-a_k}\rho_k \otimes \pi_2$.

Proof. Frobenius reciprocity and transitivity of Jacquet modules imply that there is an irreducible cuspidal representation π' such that the Jacquet module of π with respect to an appropriate parabolic subgroup contains the irreducible cuspidal representation $\nu^{a_1}\rho_1 \otimes \nu^{a_2}\rho_2 \otimes \cdots \otimes \nu^{a_k}\rho_k \otimes \pi'$. Using Theorem 2.1, we obtain the claim of the lemma. □

We note that one can also deduce that the representation π_2 in the previous lemma is in fact isomorphic to $\widehat{\pi_1}$. However, we will not use this in the sequel.

We will now recall the Mœglin–Tadić classification of discrete series for groups that we consider. Every discrete series representation of G_n is uniquely determined by three invariants: the partial cuspidal support, the Jordan block and the ϵ function.

The partial cuspidal support of a discrete series $\sigma \in \mathrm{Irr}(G_n)$ is an irreducible cuspidal representation σ_{cusp} of some G_m such that there is an irreducible admissible representation π of $\mathrm{GL}(n-m, F)$ such that σ is a subrepresentation of $\pi \rtimes \sigma_{\mathrm{cusp}}$.

The Jordan block of σ, denoted by $\mathrm{Jord}(\sigma)$, is the set of all pairs (c, ρ) where ρ is an irreducible cuspidal self-contragredient representation of some $\mathrm{GL}(n_\rho, F)$ and $c > 0$ is an integer such that the following two conditions are satisfied:

(1) c is even if and only if $L(s, \rho, r)$ has a pole at $s = 0$. The local L-function $L(s, \rho, r)$ is the one defined by Shahidi [1990; 1992], where $r = \bigwedge^2 \mathbb{C}^{n_\rho}$ is the exterior-square representation of the standard representation on $\mathbb{C}^{n_\rho}$ of $\mathrm{GL}(n_\rho, \mathbb{C})$ if G_n is a symplectic or even-orthogonal group, and $r = \mathrm{Sym}^2\, \mathbb{C}^{n_\rho}$ is the symmetric-square representation of the standard representation on $\mathbb{C}^{n_\rho}$ of $\mathrm{GL}(n_\rho, \mathbb{C})$ if G_n is an odd-orthogonal group.

(2) The induced representation $\delta([\nu^{-(c-1)/2}\rho, \nu^{(c-1)/2}\rho]) \rtimes \sigma$ is irreducible.

To explain the notion of the ϵ function, we will first define Jordan triples. These are triples of the form $(\mathrm{Jord}, \sigma', \epsilon)$, where

- σ' is an irreducible cuspidal representation of some G_n.
- Jord is the finite (possibly empty) set of ordered pairs (c, ρ), where $\rho \in \mathrm{Irr}(\mathrm{GL}(n_\rho, F))$ is a self-contragredient cuspidal representation, and c is a positive integer which is even if and only if $L(s, \rho, r)$ has a pole at $s = 0$ (for the local L function as above). For an irreducible self-contragredient cuspidal representation ρ of $\mathrm{GL}(n_\rho, F)$ we write $\mathrm{Jord}_\rho = \{c : (c, \rho) \in \mathrm{Jord}\}$. If $\mathrm{Jord}_\rho \neq \varnothing$ and $c \in \mathrm{Jord}_\rho$, we put $c\_ = \max\{d \in \mathrm{Jord}_\rho : d < c\}$, if it exists.
- ϵ is the function defined on a subset of $\mathrm{Jord} \cup (\mathrm{Jord} \times \mathrm{Jord})$ and attains the values 1 and -1. If $(c, \rho) \in \mathrm{Jord}$, then $\epsilon(c, \rho)$ is not defined if and only if c is odd and $(c', \rho) \in \mathrm{Jord}(\sigma')$ for some positive integer c'. Next, ϵ is defined on a pair $((c, \rho), (c', \rho')) \in \mathrm{Jord} \times \mathrm{Jord}$ if and only if $\rho \cong \rho'$ and $c \neq c'$.

Suppose that, for the Jordan triple $(\mathrm{Jord}, \sigma', \epsilon)$, there is a $(c, \rho) \in \mathrm{Jord}$ such that $\epsilon((c\_, \rho), (c, \rho)) = 1$. If we put $\mathrm{Jord}' = \mathrm{Jord} \setminus \{(c\_, \rho), (c, \rho)\}$ and consider the restriction ϵ' of ϵ to $\mathrm{Jord}' \cup (\mathrm{Jord}' \times \mathrm{Jord}')$, we obtain a new Jordan triple $(\mathrm{Jord}', \sigma', \epsilon')$, and we say that such Jordan triple is subordinated to $(\mathrm{Jord}, \sigma', \epsilon)$.

We say that the Jordan triple $(\mathrm{Jord}, \sigma', \epsilon)$ is a triple of alternated type if

$$\epsilon((c\_, \rho), (c, \rho)) = -1$$

whenever $c\_$ is defined and there is an increasing bijection $\phi_\rho : \mathrm{Jord}_\rho \to \mathrm{Jord}'_\rho(\sigma')$, where $\mathrm{Jord}'_\rho(\sigma')$ equals $\mathrm{Jord}_\rho(\sigma') \cup \{0\}$ if a is even and $\epsilon(\min \mathrm{Jord}_\rho, \rho) = 1$, and $\mathrm{Jord}'_\rho(\sigma')$ equals $\mathrm{Jord}_\rho(\sigma')$ otherwise.

The Jordan triple $(\mathrm{Jord}, \sigma', \epsilon)$ dominates the Jordan triple $(\mathrm{Jord}', \sigma', \epsilon')$ if there is a sequence of Jordan triples $(\mathrm{Jord}_i, \sigma', \epsilon_i)$, $0 \le i \le k$, such that $(\mathrm{Jord}_0, \sigma', \epsilon_0) = (\mathrm{Jord}, \sigma', \epsilon)$, $(\mathrm{Jord}_k, \sigma', \epsilon_k) = (\mathrm{Jord}', \sigma', \epsilon')$, and $(\mathrm{Jord}_i, \sigma', \epsilon_i)$ is subordinated to $(\mathrm{Jord}_{i-1}, \sigma', \epsilon_{i-1})$ for $i \in \{1, 2, \dots, k\}$. The Jordan triple $(\mathrm{Jord}, \sigma', \epsilon)$ is called admissible if it dominates a triple of alternated type.

The classification given in [Mœglin 2002] and [Mœglin and Tadić 2002] states that there is a one-to-one correspondence between the set of all discrete series in $\mathrm{Irr}(G)$ and the set of all admissible triples $(\mathrm{Jord}, \sigma', \epsilon)$ given by $\sigma = \sigma_{(\mathrm{Jord}, \sigma', \epsilon)}$, such that $\sigma_{\mathrm{cusp}} = \sigma'$ and $\mathrm{Jord}(\sigma) = \mathrm{Jord}$. Furthermore, if $(c, \rho) \in \mathrm{Jord}$ is such that $\epsilon((c\_, \rho), (c, \rho)) = 1$, we set $\mathrm{Jord}' = \mathrm{Jord} \setminus \{(c\_, \rho), (c, \rho)\}$ and consider the restriction ϵ' of ϵ to $\mathrm{Jord}' \cup (\mathrm{Jord}' \times \mathrm{Jord}')$. Then $(\mathrm{Jord}', \sigma', \epsilon')$ is an admissible triple and σ is a subrepresentation of $\delta([\nu^{-(c\_-1)/2}\rho, \nu^{(c-1)/2}\rho]) \rtimes \sigma_{(\mathrm{Jord}', \sigma', \epsilon')}$.

An irreducible representation $\sigma \in R(G)$ is called strongly positive if for every embedding

$$\sigma \hookrightarrow \nu^{s_1}\rho_1 \times \nu^{s_2}\rho_2 \times \cdots \times \nu^{s_k}\rho_k \rtimes \sigma_{\mathrm{cusp}},$$

where $\rho_i \in R$, $i = 1, 2, \dots, k$, are irreducible cuspidal unitary representations and $\sigma_{\mathrm{cusp}} \in R(G)$ is an irreducible cuspidal representation, we have $s_i > 0$ for $i = 1, 2, \dots, k$.

It was shown in [Mœglin 2002, Proposition 5.3] and [Mœglin and Tadić 2002, Proposition 7.1] that triples of alternated type correspond to strongly positive discrete series. Let us recall an inductive description of the noncuspidal strongly positive discrete series, obtained in [Matić 2011, Theorem 5.1], which also holds in the classical group case.

Proposition 2.5. *Suppose that $\sigma_{\mathrm{sp}} \in R(G)$ is an irreducible strongly positive representation and let $\rho \in \mathrm{Irr}(\mathrm{GL}(m, F))$ denote an irreducible cuspidal representation such that some twist of ρ appears in the cuspidal support of σ_{sp}. We denote by σ_{cusp} the partial cuspidal support of σ_{sp}. Then there exist unique $a, b \in \mathbb{R}$ such that $a > 0$, $b > 0$, $b - a$ is a nonnegative integer, and a unique irreducible strongly positive representation $\sigma_{\mathrm{sp}}^{(1)}$ without $\nu^a\rho$ in the cuspidal support, with the property that σ_{sp}*

is the unique irreducible subrepresentation of $\delta([\nu^a\rho, \nu^b\rho]) \rtimes \sigma_{\mathrm{sp}}^{(1)}$. *Furthermore, there is a nonnegative integer* l *such that* $\alpha := a + l > 0$ *and* $\nu^\alpha\rho \rtimes \sigma_{\mathrm{cusp}}$ *reduces. If* $l = 0$ *there are no twists of* ρ *appearing in the cuspidal support of* $\sigma_{\mathrm{sp}}^{(1)}$, *and if* $l > 0$ *there exist a unique* $b' > b$ *and a unique strongly positive discrete series* $\sigma_{\mathrm{sp}}^{(2)}$, *which contains neither* $\nu^a\rho$ *nor* $\nu^{a+1}\rho$ *in its cuspidal support, such that* $\sigma_{\mathrm{sp}}^{(1)}$ *can be written as the unique irreducible subrepresentation of* $\delta([\nu^{a+1}\rho, \nu^{b'}\rho]) \rtimes \sigma_{\mathrm{sp}}^{(2)}$.

We say a representation $\sigma \in \mathrm{Irr}(G_n)$ belongs to the set $D(\rho_1, \dots, \rho_m; \sigma_{\mathrm{cusp}})$ if every element of the cuspidal support of σ belongs to the set $\{\nu^x\rho_1, \dots, \nu^x\rho_m, \sigma_{\mathrm{cusp}} : x \in \mathbb{R}\}$, where $\rho_1, \dots, \rho_m$ are mutually nonisomorphic irreducible cuspidal representations of general linear groups and σ_{cusp} is an irreducible cuspidal representation of $G_{n'}$ for some $n' \le n$.

We note that for a self-contragredient cuspidal $\rho \in \mathrm{Irr}(\mathrm{GL}(m, F))$ and a cuspidal $\sigma_{\mathrm{cusp}} \in \mathrm{Irr}(G_n)$, there is a unique nonnegative α such that the induced representation $\nu^\alpha\rho \rtimes \sigma_{\mathrm{cusp}}$ reduces, and it follows from [Arthur 2013] and [Mœglin 2014, Théorème 3.1.1] that α is a half-integer.

Directly from the previous proposition we obtain

Proposition 2.6. *Let* $\sigma_{\mathrm{sp}} \in \mathrm{Irr}(G_n)$ *denote a strongly positive representation and suppose that* $\sigma_{\mathrm{sp}} \in D(\rho; \sigma_{\mathrm{cusp}})$ *for an irreducible cuspidal self-contragredient representation* ρ. *Let* α *stand for the unique nonnegative half-integer such that* $\nu^\alpha\rho \rtimes \sigma_{\mathrm{cusp}}$ *reduces, and let* $k = \lceil \alpha \rceil$, *the smallest integer which is not smaller than* α. *If* $k = 0$, *then* $\sigma_{\mathrm{sp}} \cong \sigma_{\mathrm{cusp}}$. *Otherwise, there exists a unique* k*-tuple* $(a_1, a_2, \dots, a_k)$ *such that* $a_i - \alpha \in \mathbb{Z}$ *for* $i = 1, 2, \dots, k$, $-1 < a_1 < a_2 < \cdots < a_k$, *and* σ_{sp} *is the unique irreducible subrepresentation of the induced representation*

$$\delta([\nu^{\alpha-k+1}\rho, \nu^{a_1}\rho]) \times \delta([\nu^{\alpha-k+2}\rho, \nu^{a_2}\rho]) \times \cdots \times \delta([\nu^{\alpha}\rho, \nu^{a_k}\rho]) \rtimes \sigma_{\mathrm{cusp}}.$$

3. Langlands quotients fixed by the Aubert involution

In this section we describe all Langlands quotients of the generalized principal series $\delta \rtimes \sigma$ which are isomorphic to their Aubert duals, using case-by-case considerations. We write $\delta = \delta([\nu^x\rho, \nu^y\rho])$, for x, y such that $x + y > 0$. The induced representation $\delta \rtimes \sigma$ then contains a unique irreducible (Langlands) quotient, which is also the unique irreducible subrepresentation of the induced representation $\delta([\nu^{-y}\tilde{\rho}, \nu^{-x}\tilde{\rho}]) \rtimes \sigma$, and in what follows will be denoted by π, i.e., let $\pi = L(\tilde{\delta} \rtimes \sigma)$.

Lemma 3.1. *If* π *is isomorphic to* $\hat{\pi}$, *then* $x \ge 0$.

Proof. Since $x + y > 0$, we obviously have $y > 0$. Suppose that $x < 0$. From the embedding $\pi \hookrightarrow \nu^{-x}\tilde{\rho} \times \delta([\nu^{-y}\tilde{\rho}, \nu^{-x-1}\tilde{\rho}]) \rtimes \sigma$ and the transitivity of Jacquet modules, in the same way as in the proof of Lemma 2.4 we obtain that $\mu^*(\hat{\pi})$ contains an irreducible constituent of the form $\nu^x\rho \otimes \pi'$. Since $y \neq x$, it follows directly from the structural formula that $\mu^*(\delta([\nu^{-y}\tilde{\rho}, \nu^{-x}\tilde{\rho}]) \rtimes \sigma)$ does not contain such an irreducible constituent, so $\hat{\pi}$ is not isomorphic to π, a contradiction. $\square$

Let us first consider the case of cuspidal σ.

Proposition 3.2. *Suppose that $\sigma \in \mathrm{Irr}(G_n)$ is a cuspidal representation. Then π is isomorphic to its Aubert dual if and only if one of the following holds*:

(1) $x = y > 0$ *and the induced representation* $\nu^x\rho \rtimes \sigma$ *is irreducible*,

(2) $(x, y) = (0, 1)$ *and the induced representation* $\rho \rtimes \sigma$ *reduces.*

Proof. We have already seen that if $\pi \cong \hat{\pi}$ then $x \geq 0$. In the same way as in the proof of the previous lemma we deduce that $\mu^*(\hat{\pi}) \geq \nu^{-x}\tilde{\rho} \otimes \pi'$, for some irreducible representation π'. From the structural formula we see that this is possible only if either $x = y$ or $(x, \tilde{\rho}) = (0, \rho)$.

Let us first consider the case $x = y$. Note that then we have $x > 0$. Furthermore, if $\nu^x\rho \rtimes \sigma$ reduces, it follows from [Muić 2004, Proposition 3.1(i)] that $\mu^*(\pi)$ does not contain an irreducible constituent of the form $\nu^{-x}\tilde{\rho} \otimes \pi'$. Consequently, if $\pi \cong \hat{\pi}$ and $x = y$, then $\nu^x\rho \rtimes \sigma$ is irreducible.

Conversely, if the induced representation $\nu^x\rho \rtimes \sigma$ is irreducible, then $\pi \cong \nu^x\rho \rtimes \sigma$ and from part (4) of Theorem 2.1 we have $\hat{\pi} \leq \nu^x\rho \rtimes \sigma \cong \pi$, so π is isomorphic to its Aubert dual.

Let us now assume that $x = 0$ and $\rho \cong \tilde{\rho}$. Let s denote a unique nonnegative half-integer such that $\nu^s\rho \rtimes \sigma$ reduces. We obviously have $y > 0$ and there are two possibilities to consider.

First, suppose that $y = s$ and let σ_{sp} stand for a unique irreducible subrepresentation of the induced representation $\nu\rho \times \nu^2\rho \times \cdots \times \nu^y\rho \rtimes \sigma$. It follows from [Matić 2011, Theorem 4.6] that σ_{sp} is a strongly positive discrete series and one can see directly from [Mœglin and Tadić 2002, Proposition 2.1] that the induced representation $\rho \rtimes \sigma_{\mathrm{sp}}$ reduces. By [Tadić 2013, Section 4], there is a unique irreducible subrepresentation τ of $\rho \rtimes \sigma_{\mathrm{sp}}$ such that $\mu^*(\tau)$ does not contain an irreducible constituent of the form $\nu\rho \otimes \pi'$. We note that τ is a tempered representation. Furthermore, if $\mu^*(\tau)$ contains an irreducible constituent of the form $\nu^a\rho \otimes \pi'_1$, then $a = 0$. Thus, if $\mu^*(\hat{\tau})$ contains an irreducible constituent of the form $\nu^a\rho \otimes \pi'_2$, then $a = 0$. Since τ is a subrepresentation of $\rho \times \nu\rho \times \nu^2\rho \times \cdots \times \nu^y\rho \rtimes \sigma$, using Lemma 2.4 we deduce that $\hat{\tau}$ is a subrepresentation of $\rho \times \nu^{-1}\rho \times \nu^{-2}\rho \times \cdots \times \nu^{-y}\rho \rtimes \sigma$ and, in the same way as in the proof of [Matić 2017, Lemma 3.4], we deduce that $\hat{\tau}$ is a subrepresentation of $\delta([\nu^{-y}\rho, \rho]) \rtimes \sigma$. Consequently, $\hat{\tau} \cong \pi$ and $\hat{\pi} \ncong \pi$.

Now, suppose that $y \neq s$. If $y \geq 2$, we have the following embedding and isomorphism:

$$\pi \hookrightarrow \delta([\nu^{-y+1}\rho, \rho]) \times \nu^{-y}\rho \rtimes \sigma \cong \nu^y\rho \times \delta([\nu^{-y+1}\rho, \rho]) \rtimes \sigma.$$

Lemma 2.4 implies that $\mu^*(\hat{\pi})$ contains an irreducible constituent of the form $\nu^{-y}\rho \otimes \pi'$ and it follows directly from the structural formula that $\hat{\pi} \not\cong \pi$. Thus, we can assume that $y = 1$. If $s > 0$, then $s \neq y$ and [Muić 2004, Theorem 4.1(i)] imply that $\delta([\nu^{-1}\rho, \rho]) \rtimes \sigma$ is irreducible and $\pi \cong \delta([\rho, \nu\rho]) \rtimes \sigma$. Consequently, if $s > 0$ then $\mu^*(\hat{\pi})$ contains an irreducible constituent of the form $\nu^{-1}\rho \otimes \pi'$, and it follows directly from the structural formula that $\hat{\pi} \not\cong \pi$.

It remains to consider the case $s = 0$. According to [Muić 2004, Theorem 2.1], in $R(G)$ we have

$$\delta([\rho, \nu\rho]) \rtimes \sigma = \pi + \sigma_{\mathrm{ds}}^{(1)} + \sigma_{\mathrm{ds}}^{(2)},$$

where $\sigma_{\mathrm{ds}}^{(1)}$ and $\sigma_{\mathrm{ds}}^{(2)}$ are mutually nonisomorphic discrete series subrepresentations of $\delta([\rho, \nu\rho]) \rtimes \sigma$. Frobenius reciprocity implies that both $\mu^*(\sigma_{\mathrm{ds}}^{(1)})$ and $\mu^*(\sigma_{\mathrm{ds}}^{(2)})$ contain irreducible constituents of the form $\nu\rho \otimes \pi'$. It follows from the structural formula that only irreducible constituents of the form $\nu\rho \otimes \pi'$ appearing in $\mu^*(\delta([\rho, \nu\rho]) \rtimes \sigma)$ are $\nu\rho \otimes \tau_1$ and $\nu\rho \otimes \tau_{-1}$, where τ_1 and τ_{-1} are irreducible mutually nonisomorphic tempered representations such that in $R(G)$ we have $\rho \rtimes \sigma = \tau_1 + \tau_{-1}$. Furthermore, both $\nu\rho \otimes \tau_1$ and $\nu\rho \otimes \tau_{-1}$ appear in $\mu^*(\delta([\rho, \nu\rho]) \rtimes \sigma)$ with multiplicity one. Thus, $\mu^*(\pi)$ does not contain an irreducible constituent of the form $\nu\rho \otimes \pi'$ and, consequently, $\mu^*(\hat{\pi})$ does not contain an irreducible constituent of the form $\nu^{-1}\rho \otimes \pi''$. Since π is a subrepresentation of $\rho \times \nu\rho \rtimes \sigma$, using Lemma 2.4 we obtain that $\hat{\pi}$ is a subrepresentation of $\rho \times \nu^{-1}\rho \rtimes \sigma$ and it follows that $\hat{\pi}$ is a unique irreducible subrepresentation of $\delta([\nu^{-1}\rho, \rho]) \rtimes \sigma$, i.e., $\pi \cong \hat{\pi}$. This completes the proof. □

In the rest of this section we assume that σ is a noncuspidal discrete series representation, and let σ_{cusp} denote the partial cuspidal support of σ.

Lemma 3.3. *If π is isomorphic to $\hat{\pi}$, then $\sigma \in D(\rho; \sigma_{\mathrm{cusp}})$. In particular, ρ is self-contragredient.*

Proof. Suppose that $\sigma \notin D(\rho; \sigma_{\mathrm{cusp}})$. Then there is an embedding of the form $\sigma \hookrightarrow \nu^a\rho' \rtimes \sigma'$ such that $a > 0$, ρ' is an irreducible self-contragredient cuspidal representation which is not isomorphic to ρ, and σ' is irreducible. We have

$$\pi \hookrightarrow \tilde{\delta} \times \nu^a\rho' \rtimes \sigma' \cong \nu^a\rho' \times \tilde{\delta} \rtimes \sigma',$$

and Lemma 2.4, together with transitivity of Jacquet modules, implies that Jacquet module of $\hat{\pi}$ with respect to an appropriate parabolic subgroup contains an irreducible representation of the form $\nu^{-a}\rho' \otimes \sigma''$. Since σ is square-integrable, it follows that $\mu^*(\delta \rtimes \sigma)$ does not contain an irreducible constituent of the form $\nu^{-a}\rho' \otimes \sigma''$. Thus, π is not isomorphic to $\hat{\pi}$, a contradiction. □

According to the previous lemma, in what follows we can assume that ρ is a self-contragredient representation and that $\sigma \in D(\rho; \sigma_{\mathrm{cusp}})$. We denote by α a unique nonnegative half-integer s such that $\nu^s\rho \rtimes \sigma_{\mathrm{cusp}}$ reduces.

The following result presents the crucial step towards our description.

Theorem 3.4. *If π is isomorphic to $\hat{\pi}$, then σ is a strongly positive discrete series. In particular, $\alpha > 0$.*

Proof. Suppose on the contrary that σ is not a strongly positive representation and let $(\mathrm{Jord}(\sigma), \sigma_{\mathrm{cusp}}, \epsilon_\sigma)$ denote the corresponding Jordan triple. Since $\sigma \in D(\rho; \sigma_{\mathrm{cusp}})$, there is $c \in \mathrm{Jord}_\rho(\sigma)$ such that $c\_$ is defined and $\epsilon_\sigma((c\_, \rho), (c, \rho)) = 1$. Also, σ is a subrepresentation of an induced representation of the form

$$\delta([\nu^{-(c\_-1)/2}\rho, \nu^{(c-1)/2}\rho]) \rtimes \sigma'$$

for an appropriate discrete series σ'. Using [Mœglin and Tadić 2002, Lemma 3.2], we deduce that σ is a subrepresentation of an induced representation of the form $\nu^{(c-1)/2}\rho \rtimes \pi_1$, for some irreducible π_1. Since $(c-1)/2 \geq 1$ and $-x \leq 0$, if $(c, x) \neq (3, 0)$ we obtain an embedding $\pi \hookrightarrow \nu^{(c-1)/2}\rho \times \delta([\nu^{-y}\rho, \nu^{-x}\rho]) \rtimes \pi_1$. Lemma 2.4 implies that $\mu^*(\hat{\pi})$ contains an irreducible constituent of the form $\nu^{-(c-1)/2}\rho \otimes \pi_2$, and Corollary 2.3 implies that $x = (c-1)/2$.

If $c > 3$, we also have $(c-3)/2 > -(c-1)/2 + 1$, which gives the following embeddings and isomorphisms:

$$\begin{aligned}
\pi &\hookrightarrow \delta([\nu^{-y}\rho, \nu^{-\frac{c-1}{2}}\rho]) \times \delta([\nu^{-\frac{c_-1}{2}}\rho, \nu^{\frac{c-1}{2}}\rho]) \rtimes \sigma' \\
&\hookrightarrow \delta([\nu^{-y}\rho, \nu^{-\frac{c-1}{2}}\rho]) \times \nu^{\frac{c-1}{2}}\rho \times \nu^{\frac{c-3}{2}}\rho \times \delta([\nu^{-\frac{c_-1}{2}}\rho, \nu^{\frac{c-5}{2}}\rho]) \rtimes \sigma' \\
&\cong \nu^{\frac{c-1}{2}}\rho \times \delta([\nu^{-y}\rho, \nu^{-\frac{c-1}{2}}\rho]) \times \nu^{\frac{c-3}{2}}\rho \times \delta([\nu^{-\frac{c_-1}{2}}\rho, \nu^{\frac{c-5}{2}}\rho]) \rtimes \sigma' \\
&\cong \nu^{\frac{c-1}{2}}\rho \times \nu^{\frac{c-3}{2}}\rho \times \delta([\nu^{-y}\rho, \nu^{-\frac{c-1}{2}}\rho]) \times \delta([\nu^{-\frac{c_-1}{2}}\rho, \nu^{\frac{c-5}{2}}\rho]) \rtimes \sigma'.
\end{aligned}$$

Since $\pi \cong \hat{\pi}$, Lemma 2.4 and the transitivity of Jacquet modules imply that the Jacquet module of π with respect to an appropriate parabolic subgroup contains an irreducible representation of the form $\nu^{-(c-1)/2}\rho \otimes \nu^{-(c-3)/2}\rho \otimes \pi_3$.

From $\pi \hookrightarrow \delta([\nu^{-y}\rho, \nu^{-(c-1)/2}\rho]) \rtimes \sigma$, using the structural formula recalled in Theorem 2.2, we obtain that $\nu^{-(c-3)/2}\rho \otimes \pi_3 \leq \mu^*(\delta([\nu^{-y}\rho, \nu^{-(c+1)/2}\rho]) \rtimes \sigma)$, which is impossible.

It remains to consider the case $c = 3$. This directly implies that $c\_ = 1$ and $x \in \{0, 1\}$. In other words, σ is a subrepresentation of an induced representation of the form $\delta([\rho, \nu\rho]) \rtimes \sigma'$, where σ' is a discrete series such that 1 and 3 do not appear in $\mathrm{Jord}_\rho(\sigma')$, and it follows that σ' is a strongly positive representation, since otherwise we can apply the same arguments as before to deduce that π is not isomorphic to $\hat{\pi}$.

Let us first assume that $\mathrm{Jord}_\rho(\sigma') \neq \emptyset$. Then, as in [Matić 2011, Section 4] and [Mœglin and Tadić 2002, Proposition 2.1], we see that there is an $a \geq 2$ such that σ' is a subrepresentation of $\nu^a \rho \rtimes \sigma''$ for an appropriate strongly positive representation σ''.

If $a > 2$, we have $\sigma \hookrightarrow \nu^a \rho \times \delta([\rho, \nu\rho]) \rtimes \sigma''$. Since $x \in \{0, 1\}$, in the same way as before we deduce that $\mu^*(\pi) \geq \nu^{-a}\rho \otimes \pi'$ for an irreducible representation π', which is impossible.

If $a = 2$, the irreducible representation σ' is also a subrepresentation of an induced representation of the form $\nu^2\rho \times \nu\rho \rtimes \sigma_1$. If $x = 0$, we have the following embeddings:

$$\begin{aligned}\pi &\hookrightarrow \delta([\nu^{-y}\rho, \rho]) \times \delta([\rho, \nu\rho]) \times \nu^2\rho \times \nu\rho \rtimes \sigma_1\\ &\hookrightarrow \rho \times \nu\rho \times \nu^2\rho \times \delta([\nu^{-y}\rho, \nu^{-1}\rho]) \times \rho \times \nu\rho \rtimes \sigma_1\\ &\hookrightarrow \rho \times \nu\rho \times \nu^2\rho \times \nu^{-1}\rho \times \rho \times \nu\rho \times \delta([\nu^{-y}\rho, \nu^{-2}\rho]) \rtimes \sigma_1.\end{aligned}$$

Using Lemma 2.4 and the transitivity of Jacquet modules, we obtain that the Jacquet module of π with respect to an appropriate parabolic subgroup contains an irreducible representation of the form $\rho \otimes \nu^{-1}\rho \otimes \nu^{-2}\rho \otimes \nu\rho \otimes \rho \otimes \nu^{-1}\rho \otimes \pi_4$. Since σ is a discrete series representation, applying the structural formula several times, we deduce that $y \geq 2$ and that $\nu\rho \otimes \rho \otimes \nu^{-1}\rho \otimes \pi_4$ is contained in the Jacquet module of $\delta([\nu^{-y}\rho, \nu^{-3}\rho]) \rtimes \sigma$ with respect to an appropriate parabolic subgroup. This directly implies that the Jacquet module of σ with respect to an appropriate parabolic subgroup contains an irreducible representation of the form $\nu\rho \otimes \rho \otimes \nu^{-1}\rho \otimes \pi_5$, contradicting the square-integrability criterion. The case $x = 1$ can be handled in the same way.

Let us now assume that $\mathrm{Jord}_\rho(\sigma') = \emptyset$. This implies that σ' is a cuspidal representation and $\rho \rtimes \sigma'$ reduces. As in the proof of Proposition 3.2, in $R(G)$ we have $\rho \rtimes \sigma' = \tau_1 + \tau_{-1}$ and there is a unique $i \in \{1, -1\}$ such that σ is the unique irreducible subrepresentation of $\nu\rho \rtimes \tau_i$ or, equivalently, such that $\mu^*(\sigma) \geq \nu\rho \otimes \tau_i$. It follows from [Matić 2016a, Theorem 5.1] that $\hat{\sigma} \cong L(\nu^{-1}\rho \rtimes \tau_{-i})$, and if an irreducible constituent of the form $\nu^z\rho \otimes \pi'$ appears in $\mu^*(\hat{\sigma})$, then $z = -1$.

Again, we comment only on the case $x = 0$, since the case $x = 1$ can be handled in the same way.

We have the following embeddings:

$$\begin{aligned}\pi &\hookrightarrow \delta([\nu^{-y}\rho, \rho]) \rtimes \sigma\\ &\hookrightarrow \rho \times \nu^{-1}\rho \times \cdots \times \nu^{-y}\rho \rtimes \sigma\\ &\hookrightarrow \rho \times \nu^{-1}\rho \times \cdots \times \nu^{-y}\rho \times \nu\rho \rtimes \tau_i.\end{aligned} \tag{1}$$

Frobenius reciprocity implies that the Jacquet module of π with respect to an appropriate parabolic subgroup contains the irreducible representation

$$\rho \otimes \nu^{-1}\rho \otimes \cdots \otimes \nu^{-y}\rho \otimes \nu\rho \otimes \tau_i. \tag{2}$$

Since $\pi \cong \hat{\pi}$, applying Theorem 2.1, part (4) to the induced representation appearing in (1), we deduce that π is an irreducible subquotient of

$$\rho \times \nu^{-1}\rho \times \cdots \times \nu^{-y}\rho \rtimes L(\nu^{-1}\rho \rtimes \tau_{-i}). \tag{3}$$

Using a repeated application of the structural formula and $\tau_i \not\cong \tau_{-i}$, one can show that the irreducible representation (2) is not contained in the Jacquet module of the induced representation (3) with respect to an appropriate parabolic subgroup, a contradiction. This completes the proof. $\square$

We denote $\lceil \alpha \rceil$ by k, and let $(a_1, a_2, \ldots, a_k)$ denote a unique ordered k-tuple such that $a_i - \alpha \in \mathbb{Z}$ for $i = 1, 2, \ldots, k$, $-1 < a_1 < a_2 < \cdots < a_k$, and such that σ is the unique irreducible subrepresentation of

$$\delta([\nu^{\alpha-k+1}\rho, \nu^{a_1}\rho]) \times \delta([\nu^{\alpha-k+2}\rho, \nu^{a_2}\rho]) \times \cdots \times \delta([\nu^{\alpha}\rho, \nu^{a_k}\rho]) \rtimes \sigma_{\mathrm{cusp}}.$$

We note that such a k-tuple exists by Proposition 2.6. Since σ is noncuspidal, there is an $i \in \{1,2,\ldots,k\}$ such that $a_i \geq \alpha - k + i$. Denote the minimal such i by $i_{\min}$.

Lemma 3.5. *If π is isomorphic to $\hat{\pi}$, then $a_{i_{\min}} = \alpha - k + i_{\min}$ and $a_{j+1} = a_j + 1$ for $j = i_{\min}, i_{\min}+1, \ldots, k-1$.*

Proof. It follows from [Matić 2013, Theorem 4.6], or from [Matić and Tadić 2015, Section 8], that σ is a subrepresentation of an induced representation of the form $\nu^{a_{i_{\min}}}\rho \rtimes \sigma_{\mathrm{sp}}$, where σ_{sp} is a strongly positive representation. If $-x \neq a_{i_{\min}} - 1$, we have an embedding $\pi \hookrightarrow \nu^{a_{i_{\min}}}\rho \times \delta([\nu^{-y}\rho, \nu^{-x}\rho]) \rtimes \sigma_{\mathrm{sp}}$, and an application of Corollary 2.3 and Lemma 2.4 gives $x = a_{i_{\min}}$.

If $-x = a_{i_{\min}} - 1$, since $x \geq 0$ and $a_{i_{\min}} \geq \frac{1}{2}$, it follows that $x \in \left\{0, \frac{1}{2}\right\}$. Thus, if $-x = a_{i_{\min}} - 1$ then $i_{\min} = 1$ and $a_{i_{\min}} = \alpha - k + 1$.

Let us now assume that $-x \neq a_{i_{\min}} - 1$ and $a_{i_{\min}} \geq \alpha - k + i_{\min} + 1$. It follows that $a_{i_{\min}} \geq \frac{3}{2}$, so $x = a_{i_{\min}}$ and $-x < a_{i_{\min}} - 1$. There is a strongly positive representation σ'_{sp} such that σ is a subrepresentation of $\nu^{a_{i_{\min}}}\rho \times \nu^{a_{i_{\min}}-1}\rho \rtimes \sigma'_{\mathrm{sp}}$, so we have an embedding $\pi \hookrightarrow \nu^{a_{i_{\min}}}\rho \times \nu^{a_{i_{\min}}-1}\rho \times \delta([\nu^{-y}\rho, \nu^{-x}\rho]) \rtimes \sigma'_{\mathrm{sp}}$. Since $\pi \cong \hat{\pi}$, it follows that the Jacquet module of π with respect to an appropriate parabolic subgroup contains an irreducible representation of the form $\nu^{-a_{i_{\min}}}\rho \otimes \nu^{-a_{i_{\min}}+1}\rho \otimes \pi'$. From the structural formula we directly obtain that $\mu^*(\delta([\nu^{-y}\rho, \nu^{-a_{i_{\min}}-1}\rho]) \rtimes \sigma)$ contains $\nu^{-a_{i_{\min}}+1}\rho \otimes \pi'$, a contradiction. Consequently, $a_{i_{\min}} = \alpha - k + i_{\min}$.

Now we assume that there is a $j \in \{i_{\min}, i_{\min}+1, \ldots, k-1\}$ such that $a_{j+1} \neq a_j + 1$. It follows from [Matić 2011, Section 4] that $a_{j+1} \geq a_j + 2$ and π is a subrepresentation of an induced representation of the form $\nu^{a_{j+1}}\rho \rtimes \sigma'_{\mathrm{sp}}$, for a strongly positive representation σ'_{sp}. Since we obviously have that $a_{j+1} > -x+1$, there is an embedding $\pi \hookrightarrow \nu^{a_{j+1}}\rho \times \delta([\nu^{-y}\rho, \nu^{-x}\rho]) \rtimes \sigma'_{\mathrm{sp}}$. This gives $\mu^*(\pi) \geq \nu^{-a_{j+1}}\rho \otimes \pi'$ for some irreducible π', contradicting Corollary 2.3, since $x < a_{j+1}$. $\square$

In the following theorem we state our first main result.

Theorem 3.6. *Suppose that α is a half-integer. Then π is isomorphic to $\hat{\pi}$ if and only if $\alpha \geq \frac{3}{2}$, $a_{i_{\min}} \geq \frac{3}{2}$, $x = a_{i_{\min}}$, and $y = \alpha + 1$.*

Theorem 3.6 follows from the following two propositions:

Proposition 3.7. *Suppose that α is a half-integer and $\pi \cong \hat{\pi}$. Then $\alpha \geq \frac{3}{2}$, $a_{i_{\min}} \geq \frac{3}{2}$, $x = a_{i_{\min}}$, and $y = \alpha + 1$.*

Proof. Let us first show that $x = a_{i_{\min}}$. We have an embedding $\sigma \hookrightarrow \nu^{a_{i_{\min}}}\rho \rtimes \sigma_{\mathrm{sp}}$ for some strongly positive representation σ_{sp}. Since $x \geq 0$ and $a_{i_{\min}} > 0$, if $(x, a_{i_{\min}}) \neq \left(\frac{1}{2}, \frac{1}{2}\right)$, we obtain an embedding $\pi \hookrightarrow \nu^{a_{i_{\min}}}\rho \times \delta([\nu^{-y}\rho, \nu^{-x}\rho]) \rtimes \sigma_{\mathrm{sp}}$, which implies that $\mu^*(\pi)$ contains an irreducible constituent of the form $\nu^{-a_{i_{\min}}}\rho \otimes \pi_1$, since $\pi \cong \hat{\pi}$. This is possible only if $x = a_{i_{\min}}$. Thus, in any case we have $x = a_{i_{\min}}$.

Let us now prove that $a_{i_{\min}} \geq \frac{3}{2}$. Assume on the contrary that $a_{i_{\min}} = \frac{1}{2}$. Using Lemma 3.5 and [Muić 2004, Theorem 5.1], we obtain that $\nu^z\rho \rtimes \sigma$ is irreducible for $z \notin \left\{\frac{1}{2}, \alpha+1\right\}$. If $y \notin \left\{\frac{1}{2}, \alpha+1\right\}$, we have the following embedding and isomorphism:

$$\pi \hookrightarrow \delta([\nu^{-y+1}\rho, \nu^{-1/2}\rho]) \times \nu^{-y}\rho \rtimes \sigma \cong \nu^{y}\rho \times \delta([\nu^{-y+1}\rho, \nu^{-1/2}\rho]) \rtimes \sigma.$$

In the same way as before we conclude that $\mu^*(\pi)$ contains an irreducible constituent of the form $\nu^{-y}\rho \otimes \pi_1$, which is impossible unless $y = \frac{1}{2}$. Thus, $y \in \left\{\frac{1}{2}, \alpha+1\right\}$.

It follows at once that π is a subrepresentation of $\nu^{-1/2}\rho \times \delta([\nu^{-y}\rho, \nu^{-3/2}\rho]) \rtimes \sigma$, and Lemma 2.4, together with transitivity of Jacquet modules, shows that $\mu^*(\pi)$ contains an irreducible constituent of the form $\nu^{1/2}\rho \otimes \pi_1$. We will show that this is impossible, implying $a_{i_{\min}} \geq \frac{3}{2}$. Note that $\epsilon_\sigma(\rho, 2) = 1$, where $(\mathrm{Jord}(\sigma), \sigma_{\mathrm{cusp}}, \epsilon_\sigma)$ stands for the Jordan triple corresponding to σ, so we use [Muić 2004, Theorem 5.1(ii)]. Only the case $y = \alpha + 1$ and $\alpha \geq \frac{3}{2}$ will be described in detail, since other cases can be obtained in the same way and the case $(y, \alpha) = \left(\frac{3}{2}, \frac{1}{2}\right)$ is also, in the split case, discussed in [Jantzen 1996]. The following equality holds in $R(G)$:

$$\begin{aligned}\delta([\nu^{1/2}\rho, \nu^{\alpha+1}\rho]) \rtimes \sigma = \pi &+ L(\delta([\nu^{-\alpha-1}\rho, \nu^{1/2}\rho]) \rtimes \sigma_{\mathrm{sp}}^{(1)}) \\ &+ L(\delta([\nu^{-\alpha}\rho, \nu^{-1/2}\rho]) \rtimes \sigma_{\mathrm{sp}}^{(2)}) \\ &+ L(\delta([\nu^{-\alpha}\rho, \nu^{1/2}\rho]) \rtimes \sigma_{\mathrm{sp}}^{(3)}),\end{aligned}$$

where $\sigma_{\mathrm{sp}}^{(1)}$, $\sigma_{\mathrm{sp}}^{(2)}$, $\sigma_{\mathrm{sp}}^{(3)}$ are the unique irreducible subrepresentations of

$$\nu^{3/2}\rho \times \nu^{5/2}\rho \times \cdots \times \nu^{\alpha}\rho \rtimes \sigma_{\mathrm{cusp}},$$
$$\nu^{1/2}\rho \times \nu^{3/2}\rho \times \cdots \times \nu^{\alpha-1}\rho \times \delta([\nu^{\alpha}\rho, \nu^{\alpha+1}\rho]) \rtimes \sigma_{\mathrm{cusp}},$$
$$\nu^{3/2}\rho \times \nu^{5/2}\rho \times \cdots \times \nu^{\alpha-1}\rho \times \delta([\nu^{\alpha}\rho, \nu^{\alpha+1}\rho]) \rtimes \sigma_{\mathrm{cusp}},$$

respectively. We note that $\sigma_{\mathrm{sp}}^{(i)}$ is strongly positive for $i = 1, 2, 3$.

Using the structural formula, we obtain that if $\nu^{1/2}\rho \otimes \pi_1$ is an irreducible constituent of $\mu^*(\delta([\nu^{1/2}\rho, \nu^{\alpha+1}\rho]) \rtimes \sigma)$, then π_1 is an irreducible subquotient of $\delta([\nu^{1/2}\rho, \nu^{\alpha+1}\rho]) \rtimes \sigma_{\mathrm{sp}}^{(1)}$. By [Muić 2004, Theorem 5.1(i)], in $R(G)$ we have:

$$\delta([\nu^{1/2}\rho, \nu^{\alpha+1}\rho]) \rtimes \sigma_{\mathrm{sp}}^{(1)} = L(\delta([\nu^{-\alpha-1}\rho, \nu^{-1/2}\rho]) \rtimes \sigma_{\mathrm{sp}}^{(1)}) + L(\delta([\nu^{-\alpha}\rho, \nu^{-1/2}\rho]) \rtimes \sigma_{\mathrm{sp}}^{(3)}).$$

Furthermore, both irreducible constituents $\nu^{1/2}\rho \otimes L(\delta([\nu^{-\alpha-1}\rho, \nu^{-1/2}\rho]) \rtimes \sigma_{\mathrm{sp}}^{(1)})$ and $\nu^{1/2}\rho \otimes L(\delta([\nu^{-\alpha}\rho, \nu^{-1/2}\rho]) \rtimes \sigma_{\mathrm{sp}}^{(3)})$ appear in $\mu^*(\delta([\nu^{1/2}\rho, \nu^{\alpha+1}\rho]) \rtimes \sigma)$ with multiplicity one, and Frobenius reciprocity implies that both

$$\mu^*(L(\delta([\nu^{-\alpha-1}\rho, \nu^{1/2}\rho]) \rtimes \sigma_{\mathrm{sp}}^{(1)})) \quad \text{and} \quad \mu^*(L(\delta([\nu^{-\alpha}\rho, \nu^{1/2}\rho]) \rtimes \sigma_{\mathrm{sp}}^{(3)}))$$

contain irreducible constituents of the form $\nu^{1/2}\rho \otimes \pi_1$, so $\mu^*(\pi)$ does not contain such an irreducible constituent.

Since $a_{i_{\min}} \geq \frac{3}{2}$, Lemma 3.5 implies that $\alpha \geq \frac{3}{2}$. From $y \geq a_{i_{\min}}$, using [Muić 2004, Proposition 3.1], we obtain that $\nu^{y}\rho \rtimes \sigma$ is irreducible if $y \neq \alpha + 1$. Suppose that $y \neq \alpha + 1$. Then we have the following embedding and isomorphism:

$$\pi \hookrightarrow \delta([\nu^{-y+1}\rho, \nu^{-x}\rho]) \times \nu^{-y}\rho \rtimes \sigma \cong \nu^{y}\rho \times \delta([\nu^{-y+1}\rho, \nu^{-x}\rho]) \rtimes \sigma.$$

In the same way as before we conclude that $\mu^*(\pi)$ contains an irreducible constituent of the form $\nu^{-y}\rho \otimes \pi_1$, which is impossible unless $x = y$. Thus, $y = a_{i_{\min}}$. It follows from Lemma 3.5 and [Muić 2004, Proposition 3.1] that $\nu^{-a_{i_{\min}}}\rho \rtimes \sigma \cong \nu^{a_{i_{\min}}}\rho \rtimes \sigma$, so π is an irreducible subrepresentation of $\nu^{a_{i_{\min}}}\rho \times \nu^{a_{i_{\min}}}\rho \rtimes \sigma_{\mathrm{sp}}$. Consequently, Lemma 2.4 and the transitivity of Jacquet modules imply that the Jacquet module of π with respect to an appropriate parabolic subgroup contains an irreducible representation of the form $\nu^{-a_{i_{\min}}}\rho \otimes \nu^{-a_{i_{\min}}}\rho \otimes \pi_2$, and an easy application of Theorem 2.2 implies the Jacquet module of $\delta([\nu^{-y}\rho, \nu^{-x}\rho]) \rtimes \sigma$ with respect to an appropriate parabolic subgroup does not contain such a representation. Thus, $y = \alpha + 1$, and the proposition is proved. □

Proposition 3.8. *Suppose that α is a half-integer, $\alpha \geq \frac{3}{2}$, $a_{i_{\min}} \geq \frac{3}{2}$, $x = a_{i_{\min}}$, and $y = \alpha + 1$. Then π is isomorphic to $\hat{\pi}$.*

Proof. Let us first prove that if an irreducible constituent of the form $\nu^{z}\rho \otimes \pi_1$, with $z \geq 0$, appears in $\mu^*(\pi)$, then $z = a_{i_{\min}}$. Using the structural formula and [Matić 2013, Theorem 4.6], we deduce that if an irreducible constituent of the form $\nu^z\rho \otimes \pi_1$, with $z \geq 0$, appears in $\mu^*(\delta([\nu^{-\alpha-1}\rho, \nu^{-a_{i_{\min}}}\rho]) \rtimes \sigma)$, then $z \in \{a_{i_{\min}}, \alpha+1\}$.

By [Muić 2004, Proposition 3.1(i)], in $R(G)$ we have

$$\delta([\nu^{a_{i_{\min}}}\rho, \nu^{\alpha+1}\rho]) \rtimes \sigma = \pi + L(\delta([\nu^{-\alpha}\rho, \nu^{-a_{i_{\min}}}\rho]) \rtimes \sigma_{\mathrm{sp}}),$$

where σ_{sp} denotes the unique irreducible subrepresentation of

$$\nu^{a_{i_{\min}}}\rho \times \nu^{a_{i_{\min}}+1}\rho \times \cdots \times \nu^{\alpha-1}\rho \times \delta([\nu^{\alpha}\rho, \nu^{\alpha+1}\rho]) \rtimes \sigma_{\mathrm{cusp}}.$$

We note that σ_{sp} is a strongly positive representation. Using [Matić 2013, Theorem 4.6], Frobenius reciprocity, and the transitivity of Jacquet modules, we obtain that $\mu^*(L(\delta([\nu^{-\alpha}\rho, \nu^{-a_{i_{\min}}}\rho]) \rtimes \sigma_{\mathrm{sp}}))$ contains an irreducible constituent of the form $\nu^{\alpha+1}\rho \otimes \pi'$. The induced representation $\delta([\nu^{a_{i_{\min}}}\rho, \nu^{\alpha}\rho]) \rtimes \sigma$ is irreducible (by [Muić 2004, Proposition 3.1(ii)]), so the only such irreducible constituent appearing in $\mu^*(\delta([\nu^{a_{i_{\min}}}\rho, \nu^{\alpha+1}\rho]) \rtimes \sigma)$ is $\nu^{\alpha+1}\rho \otimes \delta([\nu^{a_{i_{\min}}}\rho, \nu^{\alpha}\rho]) \rtimes \sigma$, which appears there with multiplicity one. Therefore, there is no irreducible constituent of the form $\nu^{\alpha+1}\rho \otimes \pi_1$ appearing in $\mu^*(\pi)$. Furthermore, Lemma 2.4 implies that if an irreducible constituent of the form $\nu^z\rho \otimes \pi_1$ with $z \leq 0$ appears in $\mu^*(\hat{\pi})$, then $z = -a_{i_{\min}}$.

Since π is a subrepresentation of $\delta([\nu^{-\alpha-1}\rho, \nu^{-a_{i_{\min}}}\rho]) \rtimes \sigma$ and $a_{i_{\min}} \geq \frac{3}{2}$, we have the following embedding and isomorphisms:

$$\begin{aligned} \pi &\hookrightarrow \nu^{a_{i_{\min}}}\rho \times \cdots \times \nu^{\alpha}\rho \times \nu^{-a_{i_{\min}}}\rho \times \cdots \times \nu^{-\alpha}\rho \times \nu^{-\alpha-1}\rho \rtimes \sigma_{\mathrm{cusp}} \\ &\cong \nu^{a_{i_{\min}}}\rho \times \cdots \times \nu^{\alpha}\rho \times \nu^{-a_{i_{\min}}}\rho \times \cdots \times \nu^{-\alpha}\rho \times \nu^{\alpha+1}\rho \rtimes \sigma_{\mathrm{cusp}} \\ &\cong \nu^{a_{i_{\min}}}\rho \times \cdots \times \nu^{\alpha}\rho \times \nu^{\alpha+1}\rho \times \nu^{-a_{i_{\min}}}\rho \times \times \cdots \times \nu^{-\alpha}\rho \rtimes \sigma_{\mathrm{cusp}}. \end{aligned}$$

Using Lemma 2.4, Theorem 2.1, and [Mœglin and Tadić 2002, Lemma 3.1], we obtain that $\hat{\pi}$ is a subrepresentation of the induced representation

$$\nu^{-a_{i_{\min}}}\rho \times \cdots \times \nu^{-\alpha-1}\rho \times \nu^{a_{i_{\min}}}\rho \times \cdots \times \nu^{\alpha}\rho \rtimes \sigma_{\mathrm{cusp}}.$$

It follows from Lemma 3.2 of the same work that there exists an irreducible subquotient π_1 of $\nu^{-a_{i_{\min}}}\rho \times \cdots \times \nu^{-\alpha-1}\rho$ such that $\hat{\pi}$ is a subrepresentation of $\pi_1 \times \nu^{a_{i_{\min}}}\rho \times \cdots \times \nu^{\alpha}\rho \rtimes \sigma_{\mathrm{cusp}}$. Since $\mu^*(\hat{\pi})$ does not contain an irreducible constituent of the form $\nu^z\rho \otimes \pi_1$ for $z \leq 0$ and $z \neq -a_{i_{\min}}$, we deduce that $\pi_1 \cong \delta([\nu^{-\alpha-1}\rho, \nu^{-a_{i_{\min}}}\rho])$.

By the same lemma, there is an irreducible representation π' such that $\hat{\pi}$ is a subrepresentation of $\delta([\nu^{-\alpha-1}\rho, \nu^{-a_{i_{\min}}}\rho]) \rtimes \pi'$ and, obviously, the cuspidal support of π' equals $\{\nu^{a_{i_{\min}}}\rho, \nu^{a_{i_{\min}}+1}\rho, \ldots, \nu^{\alpha}\rho, \sigma_{\mathrm{cusp}}\}$.

Let us first suppose that π' is a nontempered representation and write $\pi' \cong L(\delta_1 \times \delta_2 \times \cdots \times \delta_k \rtimes \tau)$, where $\delta_i \in \mathrm{Irr}(\mathrm{GL}(n_i, F))$ is an irreducible essentially square-integrable representation for $i = 1, 2, \ldots, k$, $e(\delta_i) \leq e(\delta_{i+1}) < 0$ for $i = 1, 2, \ldots, k-1$, and $\tau \in \mathrm{Irr}(G_{n'})$ is an irreducible tempered representation. Write $\delta_i = \delta([\nu^{a_i}\rho, \nu^{b_i}\rho])$. From $a_i + b_i < 0$ and from the description of the cuspidal support of π' follows that $b_i \le -a_{i_{\min}}$, for $i = 1, 2, \ldots, k$. It directly follows that π' is a subrepresentation of an induced representation of the form $\nu^{b_1}\rho \rtimes \pi''$ and, since $-\alpha \le b_1$, we have the following embedding and isomorphism:

$$\hat{\pi} \hookrightarrow \delta([\nu^{-\alpha-1}\rho, \nu^{-a_{i_{\min}}}\rho]) \times \nu^{b_1}\rho \rtimes \pi'' \cong \nu^{b_1}\rho \times \delta([\nu^{-\alpha-1}\rho, \nu^{-a_{i_{\min}}}\rho]) \rtimes \pi'',$$

and it follows that $\mu^*(\hat{\pi})$ contains an irreducible constituent of the form $\nu^{b_1}\rho \otimes \pi_2$, which is impossible unless $b_1 = -a_{i_{\min}}$. If this is the case, we have an embedding

$$\hat{\pi} \hookrightarrow \nu^{-a_{i_{\min}}}\rho \times \nu^{-a_{i_{\min}}}\rho \times \delta([\nu^{-\alpha-1}\rho, \nu^{-a_{i_{\min}}-1}\rho]) \rtimes \pi'',$$

and Lemma 2.4 and the transitivity of Jacquet modules imply that the Jacquet module of π with respect to an appropriate parabolic subgroup contains an irreducible representation of the form $\nu^{a_{i_{\min}}}\rho \otimes \nu^{a_{i_{\min}}}\rho \otimes \pi_2$. Using the structural formula, [Matić 2013, Theorem 4.6], and the fact that $\alpha + 1 > a_{i_{\min}}$, we deduce that a representation of the form $\nu^{a_{i_{\min}}}\rho \otimes \nu^{a_{i_{\min}}}\rho \otimes \pi_2$ does not appear in the Jacquet module of the induced representation $\delta([\nu^{-\alpha-1}\rho, \nu^{-a_{i_{\min}}}\rho]) \rtimes \sigma$ with respect to an appropriate parabolic subgroup, a contradiction.

Consequently, π' is a tempered representation and, using the description of its cuspidal support and [Matić 2012, Theorem 3.5], we conclude that π' is strongly positive. Since the strongly positive representation is completely determined by its cuspidal support ([Matić 2013, Lemma 3.5]), it follows at once that π' is isomorphic to σ. Thus, $\hat{\pi}$ is an irreducible subrepresentation of $\delta([\nu^{-\alpha-1}\rho, \nu^{-a_{i_{\min}}}\rho]) \rtimes \sigma$, leading to $\hat{\pi} \cong \pi$. This completes the proof. □

Now we state our second main result.

Theorem 3.9. *Suppose that α is an integer. Then π is isomorphic to $\hat{\pi}$ if and only if $y = \alpha + 1$ and either $x = a_{i_{\min}}$ or $(a_{i_{\min}}, x) = (1, 0)$.*

Theorem 3.9 follows from the following two propositions:

Proposition 3.10. *Suppose that α is an integer and $\pi \cong \hat{\pi}$. Then $y = \alpha + 1$ and either $x = a_{i_{\min}}$ or $(a_{i_{\min}}, x) = (1, 0)$.*

Proof. If $a_{i_{\min}} \ge 2$, in the same way as in the proof of Proposition 3.7, we deduce that $(x, y) = (a_{i_{\min}}, \alpha + 1)$.

Let us now assume that $a_{i_{\min}} = 1$. Then σ is a subrepresentation of an induced representation of the form $\nu\rho \rtimes \sigma'$ and if $x > 0$ we have an embedding $\pi \hookrightarrow \nu\rho \times \delta([\nu^{-y}\rho, \nu^{-x}\rho]) \rtimes \sigma'$. In the same way as in the proof of Proposition 3.7, we get that $x \in \{0, 1\}$. Note that $y > x$ if $x = 0$. Let us prove that $y = \alpha + 1$. Suppose, contrary to our assumption, that $y \neq \alpha + 1$. Since $y \ge x$, it follows from [Muić

2004, Proposition 3.1] that $\nu^y\rho \rtimes \sigma$ is irreducible. We have

$$\pi \hookrightarrow \delta([\nu^{-y+1}\rho, \nu^{-x}\rho]) \times \nu^{-y}\rho \rtimes \sigma \cong \delta([\nu^{-y+1}\rho, \nu^{-x}\rho]) \times \nu^{y}\rho \rtimes \sigma,$$

and if $y \neq -x+1$ one obtains a contradiction in the same way as in the proof of Proposition 3.7. Since $y \geq x$ and $x \geq 0$, we see that $y = -x+1$ holds only if $(x, y) = (0, 1)$. In that case, we have $\pi \hookrightarrow \rho \times \nu\rho \times \nu\rho \rtimes \sigma'$. Using Lemma 2.4 and the transitivity of Jacquet modules we get that $r_M(\pi)$ contains an irreducible representation of the form $\rho \otimes \nu^{-1}\rho \otimes \nu^{-1}\rho \otimes \sigma''$, where M denotes the Levi factor of an appropriate parabolic subgroup. But, since σ is strongly positive, $r_M(\delta([\rho, \nu\rho]) \rtimes \sigma)$ does not contain an irreducible representation of the form $\rho \otimes \nu^{-1}\rho \otimes \nu^{-1}\rho \otimes \sigma'$, a contradiction. This completes the proof. □

Proposition 3.11. *Suppose that α is an integer. If $y = \alpha+1$ and either $x = a_{i_{\min}}$ or $(a_{i_{\min}}, x) = (1, 0)$, then π is isomorphic to $\hat{\pi}$.*

Proof. First we suppose that $(x, y) = (a_{i_{\min}}, \alpha+1)$. Let us prove that if $\mu^*(\pi)$ contains an irreducible constituent of the form $\nu^z\rho \otimes \pi'$, with $z \geq 0$, then $z = a_{i_{\min}}$. It follows from the structural formula that if $\mu^*(\delta([\nu^{a_{i_{\min}}}\rho, \nu^{\alpha+1}\rho]) \rtimes \sigma)$ contains an irreducible constituent of the form $\nu^z\rho \otimes \pi'$, with $z \geq 0$, then $z \in \{a_{i_{\min}}, \alpha+1\}$. Also, if $z = \alpha+1$, then π' is an irreducible subquotient of $\delta([\nu^{a_{i_{\min}}}\rho, \nu^{\alpha}\rho]) \rtimes \sigma$. We have $\delta([\nu^{a_{i_{\min}}}\rho, \nu^{\alpha}\rho]) \rtimes \sigma$ is irreducible and $\nu^{\alpha+1}\rho \otimes \delta([\nu^{a_{i_{\min}}}\rho, \nu^{\alpha}\rho]) \rtimes \sigma$ is contained in $\mu^*(\delta([\nu^{a_{i_{\min}}}\rho, \nu^{\alpha+1}\rho]) \rtimes \sigma)$ with multiplicity one by [Muić 2004, Proposition 3.1]. Using part (i) of the same proposition, we deduce that in $R(G)$ we have $\delta([\nu^{a_{i_{\min}}}\rho, \nu^{\alpha+1}\rho]) \rtimes \sigma = \pi + L(\delta([\nu^{-\alpha}\rho, \nu^{-a_{i_{\min}}}\rho]) \rtimes \sigma_{\mathrm{sp}})$, where σ_{sp} denotes the unique irreducible subrepresentation of

$$\nu^{a_{i_{\min}}}\rho \times \nu^{a_{i_{\min}}+1}\rho \times \cdots \times \nu^{\alpha-1}\rho \times \delta([\nu^{\alpha}\rho, \nu^{\alpha+1}\rho]) \rtimes \sigma_{\mathrm{cusp}}.$$

We note that σ_{sp} is a strongly positive representation. It is now easy to conclude, using Frobenius reciprocity and the irreducibility of $\nu^x\rho \times \nu^{\alpha+1}\rho$ for $x < \alpha$, that $\mu^*(L(\delta([\nu^{-\alpha}\rho, \nu^{-a_{i_{\min}}}\rho]) \rtimes \sigma_{\mathrm{sp}}))$ contains an irreducible constituent of the form $\nu^{\alpha+1}\rho \otimes \pi'$, so $\mu^*(\pi)$ does not contain such an irreducible constituent. Now, in the same way as in the proof of Proposition 3.8, we obtain that $\hat{\pi}$ is a subrepresentation of $\delta([\nu^{-\alpha-1}\rho, \nu^{-a_{i_{\min}}}\rho]) \rtimes \sigma$, i.e., $\pi \cong \hat{\pi}$.

Now we turn our attention to the case $(a_{i_{\min}}, x, y) = (1, 0, \alpha+1)$. In this case, we have the following embedding and isomorphisms:

$$\begin{aligned}
\pi &\hookrightarrow \rho \times \nu^{-1}\rho \times \cdots \times \nu^{-\alpha-1}\rho \times \nu\rho \times \nu^{2}\rho \times \cdots \times \nu^{\alpha}\rho \rtimes \sigma_{\mathrm{cusp}} \\
&\cong \rho \times \nu\rho \times \nu^{2}\rho \times \cdots \times \nu^{\alpha}\rho \times \nu^{-1}\rho \times \cdots \times \nu^{-\alpha}\rho \times \nu^{-\alpha-1}\rho \rtimes \sigma_{\mathrm{cusp}} \\
&\cong \rho \times \nu\rho \times \nu^{2}\rho \times \cdots \times \nu^{\alpha}\rho \times \nu^{-1}\rho \times \cdots \times \nu^{-\alpha}\rho \times \nu^{\alpha+1}\rho \rtimes \sigma_{\mathrm{cusp}} \\
&\cong \rho \times \nu\rho \times \nu^{2}\rho \times \cdots \times \nu^{\alpha}\rho \times \nu^{\alpha+1}\rho \times \nu^{-1}\rho \times \cdots \times \nu^{-\alpha}\rho \rtimes \sigma_{\mathrm{cusp}}.
\end{aligned}$$

In the same way as before, we obtain that $\hat{\pi}$ is a subrepresentation of the induced representation

$$\rho \times \nu^{-1}\rho \times \nu^{-2}\rho \times \cdots \times \nu^{-\alpha}\rho \times \nu^{-\alpha-1}\rho \times \nu\rho \times \cdots \times \nu^{\alpha}\rho \rtimes \sigma_{\mathrm{cusp}}.$$

We will show that if $\mu^*(\pi)$ contains an irreducible constituent of the form $\nu^z\rho \otimes \pi_1$, with $z \geq 0$, then $z = 0$. The rest of the proof then follows similarly to Proposition 3.8.

Note that if an irreducible constituent of the form $\nu^z\rho \otimes \pi_1$, with $z \geq 0$, appears in $\mu^*(\delta([\rho, \nu^{\alpha+1}\rho]) \rtimes \sigma)$, then $z \in \{0, 1, \alpha+1\}$. We will comment only on the case $\alpha \geq 2$, since the case $\alpha = 1$ can be handled in the same way but more easily, and in the split case it can also be obtained using [Jantzen 1996].

According to [Muić 2004, Theorem 4.1(iv)], in $R(G)$ we have

$$\begin{aligned}\delta([\rho, \nu^{\alpha+1}\rho]) \rtimes \sigma = \pi + L(\delta([\nu^{-\alpha-1}\rho, \nu\rho]) \rtimes \sigma_{\mathrm{sp}}^{(1)}) + L(\delta([\nu^{-\alpha}\rho, \rho]) \rtimes \sigma_{\mathrm{sp}}^{(2)}) \\ + L(\delta([\nu^{-\alpha}\rho, \nu\rho]) \rtimes \sigma_{\mathrm{sp}}^{(3)}),\end{aligned}$$

where $\sigma_{\mathrm{sp}}^{(1)}, \sigma_{\mathrm{sp}}^{(2)}, \sigma_{\mathrm{sp}}^{(3)}$ are the unique irreducible subrepresentations of

$$\begin{aligned}\nu^2\rho \times \cdots \times \nu^{\alpha}\rho \rtimes \sigma_{\mathrm{cusp}}, \quad \nu\rho \times \cdots \times \nu^{\alpha-1}\rho \times \delta([\nu^{\alpha}\rho, \nu^{\alpha+1}\rho]) \rtimes \sigma_{\mathrm{cusp}}, \\ \nu^2\rho \times \cdots \times \nu^{\alpha-1}\rho \times \delta([\nu^{\alpha}\rho, \nu^{\alpha+1}\rho]) \rtimes \sigma_{\mathrm{cusp}},\end{aligned}$$

respectively. We note that $\sigma_{\mathrm{sp}}^{(i)}$ is strongly positive for $i = 1, 2, 3$.

If $\mu^*(\delta([\rho, \nu^{\alpha+1}\rho]) \rtimes \sigma)$ contains an irreducible constituent of the form $\nu\rho \otimes \pi_1$, then π_1 is an irreducible subquotient of $\delta([\rho, \nu^{\alpha+1}\rho]) \rtimes \sigma_{\mathrm{sp}}^{(1)}$. By [Muić 2004, Theorem 4.1(ii)], in $R(G)$ we have

$$\delta([\rho, \nu^{\alpha+1}\rho]) \rtimes \sigma_{\mathrm{sp}}^{(1)} = L(\delta([\nu^{-\alpha-1}\rho, \rho]) \rtimes \sigma_{\mathrm{sp}}^{(1)}) + L(\delta([\nu^{-\alpha}\rho, \rho]) \rtimes \sigma_{\mathrm{sp}}^{(3)}).$$

Also, $\nu\rho \otimes L(\delta([\nu^{-\alpha-1}\rho, \rho]) \rtimes \sigma_{\mathrm{sp}}^{(1)})$ and $\nu\rho \otimes L(\delta([\nu^{-\alpha}\rho, \rho]) \rtimes \sigma_{\mathrm{sp}}^{(3)})$ appear in $\mu^*(\delta([\rho, \nu^{\alpha+1}\rho]) \rtimes \sigma)$ with multiplicity one and are obviously contained in $\mu^*(L(\delta([\nu^{-\alpha-1}\rho, \nu\rho]) \rtimes \sigma_{\mathrm{sp}}^{(1)}))$ and in $\mu^*(L(\delta([\nu^{-\alpha}\rho, \nu\rho]) \rtimes \sigma_{\mathrm{sp}}^{(3)}))$. Thus, there are no irreducible constituents of the form $\nu\rho \otimes \pi_1$ appearing in $\mu^*(\pi)$.

Similarly, if $\mu^*(\delta([\rho, \nu^{\alpha+1}\rho]) \rtimes \sigma)$ contains an irreducible constituent of the form $\nu^{\alpha+1}\rho \otimes \pi_1$, then π_1 is an irreducible subquotient of $\delta([\rho, \nu^{\alpha}\rho]) \rtimes \sigma$. By [Muić 2004, Theorem 4.1(iii)], in $R(G)$ we have

$$\delta([\rho, \nu^{\alpha}\rho]) \rtimes \sigma = L(\delta([\nu^{-\alpha}\rho, \rho]) \rtimes \sigma) + L(\delta([\nu^{-\alpha}\rho, \nu\rho]) \rtimes \sigma_{\mathrm{sp}}^{(1)}).$$

Also, $\nu^{\alpha+1}\rho \otimes L(\delta([\nu^{-\alpha}\rho, \rho]) \rtimes \sigma_{\mathrm{sp}})$ and $\nu^{\alpha+1}\rho \otimes L(\delta([\nu^{-\alpha}\rho, \nu\rho]) \rtimes \sigma_{\mathrm{sp}}^{(1)})$ appear in $\mu^*(\delta([\rho, \nu^{\alpha+1}\rho]) \rtimes \sigma)$ with multiplicity one and obviously appear in $\mu^*(L(\delta([\nu^{-\alpha}\rho, \rho]) \rtimes \sigma_{\mathrm{sp}}^{(2)}))$ and in $\mu^*(L(\delta([\nu^{-\alpha}\rho, \nu\rho]) \rtimes \sigma_{\mathrm{sp}}^{(3)}))$. Thus, $\mu^*(\pi)$ does not contain irreducible constituents of the form $\nu^{\alpha+1}\rho \otimes \pi_1$. Consequently, $\mu^*(\pi)$ does not contain an irreducible constituent of the form $\nu^z\rho \otimes \pi_1$ with $z > 0$, and the proposition is proved. $\square$

Acknowledgements

The author would like to thank the referee for helping to improve the presentation style and for the help with Proposition 3.2. This work has been supported by Croatian Science Foundation under the project 9364.

References

[Arthur 2013] J. Arthur, *The endoscopic classification of representations: orthogonal and symplectic groups*, American Mathematical Society Colloquium Publications **61**, American Mathematical Society, Providence, RI, 2013. MR Zbl

[Aubert 1995] A.-M. Aubert, "Dualité dans le groupe de Grothendieck de la catégorie des représentations lisses de longueur finie d'un groupe réductif p-adique", *Trans. Amer. Math. Soc.* **347**:6 (1995), 2179–2189. Correction in **348**:11 (1996), 4687–4690. MR Zbl

[Jantzen 1996] C. Jantzen, *Degenerate principal series for symplectic and odd-orthogonal groups*, Mem. Amer. Math. Soc. **590**, American Mathematical Society, Providence, RI, 1996. MR Zbl

[Matić 2011] I. Matić, "Strongly positive representations of metaplectic groups", *J. Algebra* **334** (2011), 255–274. MR Zbl

[Matić 2012] I. Matić, "Theta lifts of strongly positive discrete series: the case of $(\widetilde{\mathrm{Sp}}(n), O(V))$", *Pacific J. Math.* **259**:2 (2012), 445–471. MR Zbl

[Matić 2013] I. Matić, "Jacquet modules of strongly positive representations of the metaplectic group $\widetilde{\mathrm{Sp}}(n)$", *Trans. Amer. Math. Soc.* **365**:5 (2013), 2755–2778. MR Zbl

[Matić 2015] I. Matić, "Strongly positive subquotients in a class of induced representations of classical p-adic groups", *J. Algebra* **444** (2015), 504–526. MR Zbl

[Matić 2016a] I. Matić, "Aubert duals of discrete series: the first inductive step", preprint, 2016, Available at http://tinyurl.com/aubertduals.

[Matić 2016b] I. Matić, "On discrete series subrepresentations of the generalized principal series", *Glas. Mat. Ser. III* **51**:1 (2016), 125–152. MR Zbl

[Matić 2016c] I. Matić, "On Jacquet modules of discrete series: the first inductive step", *J. Lie Theory* **26**:1 (2016), 135–168. MR Zbl

[Matić 2017] I. Matić, "Aubert duals of strongly positive discrete series and a class of unitarizable representations", *Proc. Amer. Math. Soc.* **145**:8 (2017), 3561–3570.

[Matić and Tadić 2015] I. Matić and M. Tadić, "On Jacquet modules of representations of segment type", *Manuscripta Math.* **147**:3-4 (2015), 437–476. MR Zbl

[Mœglin 2002] C. Mœglin, "Sur la classification des séries discrètes des groupes classiques p-adiques: paramètres de Langlands et exhaustivité", *J. Eur. Math. Soc.* **4**:2 (2002), 143–200. MR Zbl

[Mœglin 2014] C. Mœglin, "Paquets stables des séries discrètes accessibles par endoscopie tordue; leur paramètre de Langlands", pp. 295–336 in *Automorphic forms and related geometry: assessing the legacy of I. I. Piatetski-Shapiro*, edited by J. W. Cogdell et al., Contemp. Math. **614**, American Mathematical Society, Providence, RI, 2014. MR Zbl

[Mœglin and Tadić 2002] C. Mœglin and M. Tadić, "Construction of discrete series for classical p-adic groups", *J. Amer. Math. Soc.* **15**:3 (2002), 715–786. MR Zbl

[Muić 2004] G. Muić, "Composition series of generalized principal series: the case of strongly positive discrete series", *Israel J. Math.* **140** (2004), 157–202. MR Zbl

[Schneider and Stuhler 1997] P. Schneider and U. Stuhler, "Representation theory and sheaves on the Bruhat–Tits building", *Inst. Hautes Études Sci. Publ. Math.* 85 (1997), 97–191. MR Zbl

[Shahidi 1990] F. Shahidi, "A proof of Langlands' conjecture on Plancherel measures; complementary series for p-adic groups", *Ann. of Math.* (2) **132**:2 (1990), 273–330. MR Zbl

[Shahidi 1992] F. Shahidi, "Twisted endoscopy and reducibility of induced representations for p-adic groups", *Duke Math. J.* **66**:1 (1992), 1–41. MR Zbl

[Tadić 1995] M. Tadić, "Structure arising from induction and Jacquet modules of representations of classical p-adic groups", *J. Algebra* **177**:1 (1995), 1–33. MR Zbl

[Tadić 2013] M. Tadić, "On tempered and square integrable representations of classical p-adic groups", *Sci. China Math.* **56**:11 (2013), 2273–2313. MR Zbl

[Zelevinsky 1980] A. V. Zelevinsky, "Induced representations of reductive p-adic groups, II: On irreducible representations of $\mathrm{GL}(n)$", *Ann. Sci. École Norm. Sup.* (4) **13**:2 (1980), 165–210. MR Zbl

Received October 21, 2016. Revised February 10, 2017.

IVAN MATIĆ
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF OSIJEK
TRG LJUDEVITA GAJA 6
31000 OSIJEK
CROATIA
imatic@mathos.hr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 289, No. 2, 2017
dx.doi.org/10.2140/pjm.2017.289.417

EXACT LAGRANGIAN FILLINGS OF LEGENDRIAN $(2, n)$ TORUS LINKS

YU PAN

Ekholm, Honda, and Kálmán constructed C_n exact Lagrangian fillings for a Legendrian $(2, n)$ torus knot or link with maximal Thurston–Bennequin number, where C_n is the n-th Catalan number. We show that these exact Lagrangian fillings are pairwise nonisotopic through exact Lagrangian isotopy. To do that, we compute the augmentations induced by the exact Lagrangian fillings L to $\mathbb{Z}_2[H_1(L)]$ and distinguish the resulting augmentations.

1. Introduction

A Legendrian submanifold Λ in the standard contact manifold $(\mathbb{R}^3, \xi = \ker\alpha)$, where $\alpha = dz - y\,dx$, is a 1-dimensional closed manifold such that $T\Lambda \subset \xi$ everywhere. An exact Lagrangian filling L of Λ in the symplectization manifold $(\mathbb{R}_t \times \mathbb{R}^3, \omega = d(e^t\alpha))$ is a 2-dimensional surface that is cylindrical over Λ when t is sufficiently large. See Definition 2.5 for more detail, and Figure 1 for a picture.

In this paper, we study oriented exact Lagrangian fillings of the Legendrian $(2, n)$ torus links Λ with maximal Thurston–Bennequin number $(n > 0)$. When n is even, we also require the link to have the right Maslov potential such that Reeb chords $b_1, \dots, b_n$ in Figure 2 are in degree 0 (see Section 2A for detailed definitions). Ekholm, Honda, and Kálmán [Ekholm et al. 2016] gave an algorithm (which we refer to later as the EHK algorithm) to construct exact Lagrangian fillings of the Legendrian $(2, n)$ torus link Λ as follows. Starting with a *Lagrangian projection* (a projection from $\mathbb{R}^3$ to the xy-plane) of Λ as shown in Figure 2, we can successively resolve crossings b_i in any order through pinch moves (see Figure 3), which correspond to saddle cobordisms. As a result, we get two Legendrian unknots, which admit minimum cobordisms as shown in Figure 3. Concatenating the n saddle cobordisms with these two minimum cobordisms, we get an exact Lagrangian filling of Λ.

Different orders of resolving crossings $b_1, \dots, b_n$ may give different exact Lagrangian fillings of Λ up to exact Lagrangian isotopy. Given a permutation $\sigma = (\sigma(1), \sigma(2), \dots, \sigma(n))$ of $\{1, \dots, n\}$, write L_σ for the exact Lagrangian filling achieved by using n successive pinch moves at $b_{\sigma(1)}, b_{\sigma(2)}, \dots, b_{\sigma(n)}$, respectively,

MSC2010: 53D42, 57R17.
Keywords: Exact Lagrangian fillings, (2,n) torus links, augmentation, Legendrian knots.

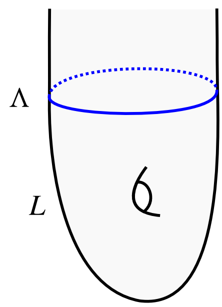


Figure 1. An exact Lagrangian filling.

and then concatenating with the two minimum cobordisms. Observe that two permutations may give isotopic exact Lagrangian fillings. For instance, let Λ be the Legendrian $(2, 3)$ torus knot and consider the exact Lagrangian fillings of Λ that correspond to permutations $(1, 3, 2)$ and $(3, 1, 2)$, respectively. Since the saddles corresponding to the pinch moves at b_1 and b_3 are disjoint when projected to $\mathbb{R}^3$, one can use a Hamiltonian vector field in the t direction to exchange the heights of these two saddles. Therefore, the two fillings $L_{(1,3,2)}$ and $L_{(3,1,2)}$ are Hamiltonian isotopic and thus are exact Lagrangian isotopic. In general, for the Legendrian $(2, n)$ torus link Λ, given any numbers i, j, k such that $i < k < j$, two permutations

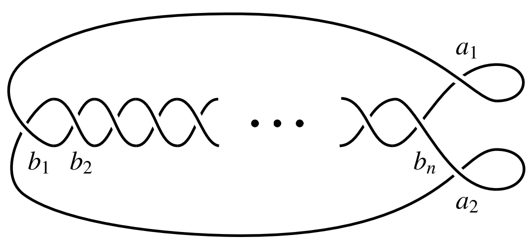


Figure 2. The Lagrangian projection of the Legendrian $(2, n)$ torus knot.

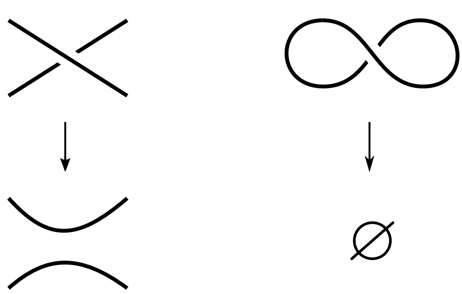


Figure 3. The pinch move (left) and the minimum cobordism (right) between Lagrangian projections of links.

$(\dots, i, j, \dots, k, \dots)$ and $(\dots, j, i, \dots, k, \dots)$, where only i and j are interchanged, give the same exact Lagrangian fillings of Λ up to exact Lagrangian isotopy. Taking all the permutations of $\{1, \dots, n\}$ modded out by this relation, we obtain C_n exact Lagrangian fillings of Λ, where

$$C_n = \frac{1}{n+1}\binom{2n}{n}$$

is the n-th Catalan number. In this paper, we prove the following theorem:

Theorem 1.1 (see Theorem 3.11 and Corollary 3.12). *The C_n exact Lagrangian fillings that come from the EHK algorithm are all of different exact Lagrangian isotopy classes. In other words*, *the Legendrian $(2, n)$ torus link has at least C_n exact Lagrangian fillings up to exact Lagrangian isotopy.*

Shende, Treumann, Williams and Zaslow [Shende et al. 2015] have also constructed C_n exact Lagrangian fillings of the Legendrian $(2, n)$ torus knot using cluster varieties and shown that they are distinct up to Hamiltonian isotopy. They remarked that these are presumably the same as fillings obtained in [Ekholm et al. 2016], but we do not resolve this issue here.

Remark 1.2. We will see from Corollary 3.12 that the conclusion of Theorem 1.1 for the case when n is even can be derived from the result for the case when n is odd. Therefore, for most of the paper, we focus on the case when n is odd, which means Λ is a knot.

Inspired by [Ekholm et al. 2016], we use augmentations to distinguish the C_n exact Lagrangian fillings of the Legendrian $(2, n)$ torus knot Λ. In order to talk about augmentations, we first introduce the Chekanov–Eliashberg differential graded algebra (DGA) of a Legendrian knot Λ, which is a chain complex $(\mathcal{A}(\Lambda), \partial)$. This is an invariant of Legendrian submanifolds introduced by Chekanov [2002] and Eliashberg [1998] in the spirit of symplectic field theory [Eliashberg et al. 2000]. The underlying algebra $\mathcal{A}(\Lambda)$ of the Chekanov–Eliashberg DGA is freely generated by Reeb chords of Λ over a commutative ring $\mathbb{Z}_2[H_1(\Lambda)] = \mathbb{Z}_2[s, s^{-1}]$, where Reeb chords of Λ correspond to double points of the Lagrangian projection of Λ. The differential is defined by a count of rigid holomorphic disks with boundary on Λ, taken with coefficients in $\mathbb{Z}_2[H_1(\Lambda)]$. In general, the Chekanov–Eliashberg DGA of Λ is defined with $\mathbb{Z}[H_1(\Lambda)]$ coefficients. For our purpose, it suffices to consider the DGA with $\mathbb{Z}_2[H_1(\Lambda)]$ coefficients, which means ignoring the orientations of moduli spaces of holomorphic disks. An *augmentation* ϵ of $\mathcal{A}(\Lambda)$ to a commutative ring $\mathbb{F}$ is a DGA map $\epsilon : (\mathcal{A}(\Lambda), \partial) \to (\mathbb{F}, 0)$. As shown in [Ekholm et al. 2016], an exact Lagrangian filling L of Λ gives an augmentation of $\mathcal{A}(\Lambda)$ by counting rigid holomorphic disks with boundary on L. Moreover, by Theorem 1.3 of the same paper, exact Lagrangian isotopic fillings give homotopic augmentations. Therefore, in order to distinguish two fillings, we only need to show their induced augmentations are not chain homotopic.

Ekholm et al. [2016] distinguished all the exact Lagrangian fillings from the EHK algorithm when $n = 3$ by computing all the augmentations of the Legendrian $(2, 3)$ torus knot to $\mathbb{Z}_2$ and finding that they are pairwise non-chain-homotopic. However, when $n \geq 5$, a computation shows that the number of augmentations of the DGA to $\mathbb{Z}_2$ is much less than the Catalan number C_n.

In this paper, for an exact Lagrangian filling L of the Legendrian $(2, n)$ torus knot Λ, we consider its induced augmentation of $\mathcal{A}(\Lambda)$ to $\mathbb{Z}_2[H_1(L)]$, where $H_1(L)$ is the singular homology of L. Note that $H_1(L) \cong H_2(\mathbb{R} \times \mathbb{R}^3, L)$ and thus it is natural to count the rigid holomorphic disks in $\mathbb{R} \times \mathbb{R}^3$ with boundary on L with $\mathbb{Z}_2[H_1(L)]$ coefficients. However, the computation of augmentations is not as easy as for the case with $\mathbb{Z}_2$ coefficients. For each exact Lagrangian filling L from the EHK algorithm, we give a combinatorial formula of the induced augmentation of $\mathcal{A}(\Lambda)$ to $\mathbb{Z}_2[H_1(L)]$. From the formula, we find a combinatorial invariant to show that the augmentations are not pairwise chain homotopic. In this way, we distinguish all of the C_n exact Lagrangian fillings of the Legendrian $(2, n)$ torus knot Λ up to exact Lagrangian isotopy.

Outline. In Section 2, we review the Chekanov–Eliashberg DGA of a Legendrian submanifold and the DGA maps induced by an exact Lagrangian cobordism. In Section 3, we compute all the augmentations of the Legendrian $(2, n)$ torus knot to $\mathbb{Z}_2[H_1(L)]$ induced by the exact Lagrangian fillings L and prove that all the resulting augmentations are distinct up to chain homotopy. In the end, we prove Theorem 1.1 for the case n even as a corollary.

2. Preliminaries

In Section 2A, we review the definition of the Chekanov–Eliashberg DGA of Legendrian submanifolds in $(\mathbb{R}^3, \ker\alpha)$ and its extension to the setting of multiple base points. For the purpose of computing augmentations in Section 3A, the definition of DGA we use here is slightly different from the versions in [Ng 2010] and [Ng et al. 2015], where the underlying algebra is completely noncommutative. In our definition, we allow elements in the coefficient ring to commute with the elements corresponding to Reeb chords. This is a generalization of the definition of Chekanov–Eliashberg DGA from [Etnyre et al. 2002]. See [Ekholm et al. 2013, Section 2.3.2] for further discussions. In Section 2B, we review the DGA map induced by an exact Lagrangian cobordism and revise coefficients of this map for the purpose of computing augmentations in Section 3A.

2A. ***The Chekanov–Eliashberg DGA.*** Let Λ be a Legendrian submanifold in $(\mathbb{R}^3, \ker\alpha)$, where $\alpha = dz - y\,dx$. There are two projection diagrams associated

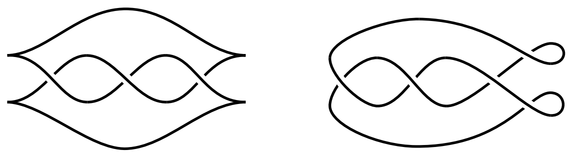


Figure 4. A front projection (left) and a Lagrangian projection (right) of the Legendrian trefoil.

Figure 5. Ng's algorithm to transfer a front projection to a Lagrangian projection by smoothing the left cusp directly and smoothing the right cusp with an additional crossing.

to Λ via the *Lagrangian projection* $\Pi_{xy} : \mathbb{R}^3 \to \mathbb{R}^2,\ (x, y, z) \mapsto (x, y)$ and the *front projection* $\Pi_{xz} : \mathbb{R}^3 \to \mathbb{R}^2,\ (x, y, z) \mapsto (x, z)$, respectively. As an example, a front projection and a Lagrangian projection of the Legendrian trefoil are shown in Figure 4. Moreover, starting from a front projection of Λ, Ng [2003] gave an algorithm to get a Lagrangian projection of Λ by smoothing the cusps of the front projection in a way shown in Figure 5.

Let $\Lambda = \Lambda_1 \cup \Lambda_2 \cup \cdots \cup \Lambda_k$ be an oriented Legendrian link with k connected components. Now let us define the Chekanov–Eliashberg DGA $(\mathcal{A}(\Lambda\,; \mathbb{Z}_2[H_1(\Lambda)]), \partial)$ of Λ. To simplify the definition of grading, we assume throughout the paper that the rotation number of Λ is 0. Note that all the Legendrian $(2, n)$ torus links we consider have rotation number 0.

The underlying algebra. The underlying algebra $\mathcal{A}(\Lambda\,; \mathbb{Z}_2[H_1(\Lambda)])$ is a unital graded algebra freely generated by Reeb chords of Λ over

$$\mathbb{Z}_2[H_1(\Lambda)] = \mathbb{Z}_2[s_1^{\pm 1}, s_2^{\pm 1}, \dots, s_k^{\pm 1}],$$

where $\{s_1, s_2, \dots, s_k\}$ is any basis of $H_1(\Lambda)$. A *Reeb chord* of Λ in $(\mathbb{R}^3, \ker \alpha)$ is a vertical line segment (z direction) with both ends on Λ endowed with an orientation in the positive z direction. Reeb chords of Λ are in one-to-one correspondence to double points of $\Pi_{xy}(\Lambda)$, which by Ng's algorithm correspond to the crossings and right cusps of $\Pi_{xz}(\Lambda)$.

To define the grading of Reeb chords, we work on the front projection $\Pi_{xz}(\Lambda)$. Write $C(\Pi_{xz}(\Lambda))$ for the set of cusps of $\Pi_{xz}(\Lambda)$, which divides $\Pi_{xz}(\Lambda)$ into strands (ignoring double points). The *Maslov potential* is a function μ that assigns an integer to each strand such that around each cusp, the Maslov potential of the lower strand is one less than that of the upper strand. This is well defined up to a global shift on

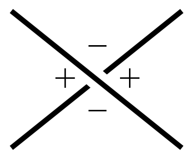


Figure 6. At each crossing, the quadrants labeled with $+$ sign are *positive quadrants* and the ones labeled with $-$ sign are *negative quadrants*.

each component of Λ. Once the Maslov potential is fixed, the grading of a Reeb chord c that corresponds to a crossing of $\Pi_{xz}(\Lambda)$ can be defined by

$$|c| := \mu(u) - \mu(l),$$

where u is the upper strand of the crossing and l is the lower strand of the crossing. The grading of Reeb chords that correspond to right cusps of $\Pi_{xz}(\Lambda)$ are defined to be 1. Extend the definition of grading to $\mathcal{A}(\Lambda\,;\mathbb{Z}_2[H_1(\Lambda)])$ by setting $|s_i| = 0$ for $i = 1, \dots, k$ and using the relation $|ab| = |a| + |b|$.

In the special case of Legendrian $(2, n)$ torus links, when n is odd, the degree is well defined. When n is even, we can choose a Maslov potential of the Legendrian $(2, n)$ torus link such that for any Reeb chord b_i as labeled in Figure 2, the upper strand and the lower strand of b_i have the same Maslov potential. In this setting, for a Legendrian $(2, n)$ torus link (n is either odd or even) whose Lagrangian projection is like Figure 2, we have that $|a_1| = |a_2| = 1$ and $|b_i| = 0$ for $i = 1, \dots, n$.

Differential. The differential ∂ is defined by counting rigid holomorphic disks in $\mathbb{R}^2_{xy}$ with boundary on $\Pi_{xy}(\Lambda)$.

For any Reeb chords $a, b_1, \dots, b_m$ of Λ, define $\mathcal{M}^{\Lambda}(a\,; b_1, \dots, b_m)$ to be the moduli space of holomorphic disks

$$u : (D_{m+1}, \partial D_{m+1}) \to (\mathbb{R}^2, \Pi_{xy}(\Lambda))$$

with the following properties:

- D_{m+1} is a 2-dimensional unit disk with $m+1$ points $s, t_1, \dots, t_m$ removed from the boundary and the points $s, t_1, \dots, t_m$ are labeled in counterclockwise order.
- $\lim_{r\to s} u(r) = a$ and the image of a neighborhood of s under u covers exactly one positive quadrant of the crossing a (see Figure 6).
- $\lim_{r\to t_i} u(r) = b_i$, for $i = 1, \dots, m$, and the image of a neighborhood of t_i under u covers exactly one negative quadrant of the crossing b_i (see Figure 6).

We occasionally abbreviate $(a, b_1, \dots, b_m)$ to $(a\,; \boldsymbol{b})$, where $\boldsymbol{b}$ represents a sequence of Reeb chords, $b_1, \dots, b_m$. According to [Chekanov 2002], we have the

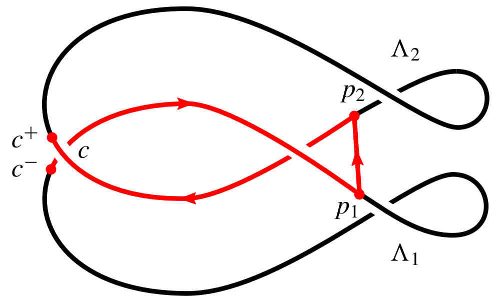


Figure 7. The Legendrian Hopf link $\Lambda_1 \cup \Lambda_2$. For a Reeb chord c from $c^- \in \Lambda_1$ to $c^+ \in \Lambda_2$, the red curve is a capping path γ_c.

following dimension formula:

$$\dim \mathcal{M}^{\Lambda}(a\,; b_1, \ldots, b_m) = |a| - \sum_{i=1}^{m} |b_i| - 1.$$

When $\dim \mathcal{M}^{\Lambda}(a\,; b_1, \ldots, b_m) = 0$, the disk $u \in \mathcal{M}^{\Lambda}(a\,; b_1, \ldots, b_m)$ is called *rigid*. There are finitely many rigid holomorphic disks and hence we can count them.

In order to count with $\mathbb{Z}_2[H_1(\Lambda)]$ coefficients, we want to take the homology class of the boundary of rigid disks in $H_1(\Lambda)$. However, for any rigid holomorphic disk u, the boundary $\Pi_{xy}^{-1}(u(\partial D_{m+1}))$ is not closed. Therefore, we first introduce capping paths. Equip each connected component Λ_i with a reference point p_i, for $i = 1, \ldots, k$. For each $i \neq 1$, pick a path δ_{1i} in $\mathbb{R}^3 \setminus \Lambda$ that goes from p_1 to p_i. For each Reeb chord c of Λ from $c^- \in \Lambda_{i^-}$ to $c^+ \in \Lambda_{i^+}$, the *capping path* γ_c is defined by concatenating

- a path on Λ_{i^-} from c^- to p_{i^-},
- the chosen path $-\delta_{1i^-}$ connecting p_{i^-} to p_1,
- the chosen path δ_{1i^+} connecting p_1 to p_{i^+}, and
- a path on Λ_{i^+} from p_{i^+} to c^+.

See Figure 7 for an example of a capping path.

After associating each Reeb chord with a capping path, for any rigid holomorphic disk $u \in \mathcal{M}^{\Lambda}(a\,; b_1, \ldots, b_m)$, the curve

$$\tilde{u} = \Pi_{xy}^{-1}(u(\partial D_{m+1})) \cup \gamma_a \cup -\gamma_{b_1} \cup \cdots \cup -\gamma_{b_m}$$

is a loop in $\Lambda \cup \delta_{12} \cup \cdots \cup \delta_{1k}$. Notice that $H_1(\Lambda \cup \delta_{12} \cup \cdots \cup \delta_{1k}) \cong H_1(\Lambda)$. Thus we can view the homology class $[\tilde{u}]$ as in $H_1(\Lambda)$.

Now we can define the differential of the Chekanov–Eliashberg DGA of Λ.

Definition 2.1. For any Reeb chord a of Λ, the differential ∂ is defined by:

$$\partial(a) = \sum_{\dim \mathcal{M}^{\Lambda}(a;\boldsymbol{b})=0} \; \sum_{u \in \mathcal{M}^{\Lambda}(a;\boldsymbol{b})} [\tilde{u}]\, b_1 \cdots b_m. \tag{2-1}$$

The definition of differential can be extended to $\mathcal{A}(\Lambda\,;\mathbb{Z}_2[H_1(\Lambda)])$ by setting $\partial(s_i) = 0$ for $i = 1, \dots, k$, and using the Leibniz rule

$$\partial(ab) = \partial(a)b + a\partial(b).$$

According to [Chekanov 2002], the map ∂ is a differential in degree -1, and up to stable tame isomorphism, the Chekanov–Eliashberg DGA $(\mathcal{A}(\Lambda\,;\mathbb{Z}_2[H_1(\Lambda)]), \partial)$ is an invariant of Λ under Legendrian isotopy.

Remark 2.2. In general, for any commutative ring R and a ring homomorphism $\mathbb{Z}_2[H_1(\Lambda)] \to R$, we define the Chekanov–Eliashberg DGA $(\mathcal{A}(\Lambda\,;R), \partial)$ as a tensor product of the DGA $\mathcal{A}(\Lambda\,;\mathbb{Z}_2[H_1(\Lambda)])$ with the ring R:

$$\mathcal{A}(\Lambda\,;R) = \mathcal{A}(\Lambda\,;\mathbb{Z}_2[H_1(\Lambda)]) \otimes_{\mathbb{Z}_2[H_1(\Lambda)]} R,$$

where the ring homomorphism gives R the structure of a module over $\mathbb{Z}_2[H_1(\Lambda)]$.

We give a combinatorial definition of the differential of $(\mathcal{A}(\Lambda\,;\mathbb{Z}_2[H_1(\Lambda)]), \partial)$. Assign Λ an orientation and label each component Λ_i, for $i = 1, \dots, k$, with a base point s_i, which is different from the reference point and ends of Reeb chords. For a union of oriented curves γ in $\Lambda \cup \delta_{12} \cup \cdots \cup \delta_{1k}$, we associate it with a monomial $w(\gamma)$ in $\mathbb{Z}_2[H_1(\Lambda)]$:

$$w(\gamma) = \prod_{i=1}^{k} s_i^{n_i(\gamma)}, \tag{2-2}$$

where $n_i(\gamma)$ is the number of times γ goes through s_i counted with sign. The sign is positive if γ goes through s_i following the link orientation and is negative if γ goes through s_i against the link orientation. In particular, for a rigid holomorphic disk $u \in \mathcal{M}^{\Lambda}(a\,;b_1, \dots, b_m)$, we have

$$[\tilde{u}] = w(\tilde{u}) = w(u)w(\gamma_a) \prod_{i=1}^{m} w(\gamma_{b_i})^{-1}, \tag{2-3}$$

where $w(u)$ is short for $w(\Pi_{xy}^{-1}(u(\partial D_{m+1})))$. Plugging it into the formula (2-1), we get a combinatorial definition of the differential. It seems to depend on the choice of capping paths. However, we have the following well-known proposition.

Proposition 2.3. *Let Λ be a Legendrian link and γ, γ' be two families of capping paths of Reeb chords of Λ. The corresponding DGAs $(\mathcal{A}^{\gamma}(\Lambda), \partial)$ and $(\mathcal{A}^{\gamma'}(\Lambda), \partial')$ are isomorphic.*

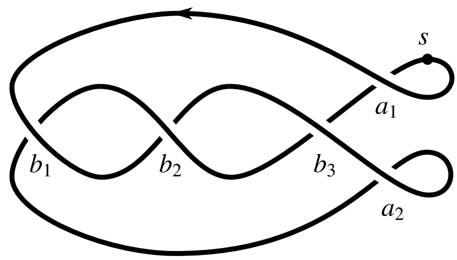


Figure 8. The Lagrangian projection of the Legendrian (2, 3) torus knot with a single base point.

Proof. For a Reeb chord a of Λ, we have

$$\partial(a) = \sum_{\dim \mathcal{M}^{\Lambda}(a;\boldsymbol{b})=0} \sum_{u \in \mathcal{M}^{\Lambda}(a;\boldsymbol{b})} \Big(w(u) w(\gamma_a) \prod_{i=1}^{m} w(\gamma_{b_i})^{-1} \Big) b_1 \cdots b_m,$$

$$\partial'(a) = \sum_{\dim \mathcal{M}^{\Lambda}(a;\boldsymbol{b})=0} \sum_{u \in \mathcal{M}^{\Lambda}(a;\boldsymbol{b})} \Big(w(u) w(\gamma'_a) \prod_{i=1}^{m} w(\gamma'_{b_i})^{-1} \Big) b_1 \cdots b_m.$$

For each Reeb chord c, concatenate $-\gamma'_c$ with γ_c and get a closed curve, denoted by $-\gamma'_c \cup \gamma_c$. It is not hard to check that the map

$$f : (\mathcal{A}^{\gamma}(\Lambda), \partial) \to (\mathcal{A}^{\gamma'}(\Lambda), \partial'), \qquad c \mapsto [-\gamma'_c \cup \gamma_c]\, c = w(\gamma'_c)^{-1} w(\gamma_c)\, c$$

is a chain map and is an isomorphism. □

Note that for an oriented link Λ with minimal base points (i.e., each component has exactly one base point), we can choose a family of capping paths such that none of them pass through any base point. Therefore, we only need to count intersections of the disk boundary and base points. Thanks to Proposition 2.3, we can define the Chekanov–Eliashberg DGA of Λ to be a unital graded algebra over $\mathbb{Z}_2[H_1(\Lambda)] = \mathbb{Z}_2[s_1^{\pm 1}, \dots, s_k^{\pm 1}]$ generated by Reeb chords of Λ endowed with a differential given by

$$\partial(a) = \sum_{\dim \mathcal{M}^{\Lambda}(a;\boldsymbol{b})=0} \sum_{u \in \mathcal{M}^{\Lambda}(a;\boldsymbol{b})} w(u) b_1 \cdots b_m,$$

$$\partial(s_i) = 0, \quad i = 1, \dots, k,$$

with $w(u)$ defined as in (2-3). This DGA is denoted by $(\mathcal{A}(\Lambda, \{s_1, \dots, s_k\}), \partial)$ too.

Example 2.4. Consider the Legendrian (2, 3) torus knot Λ with a single base point s as shown in Figure 8. The underlying algebra $\mathcal{A}(\Lambda, \{s\})$ is generated by Reeb chords a_1, a_2, b_1, b_2, b_3 over $\mathbb{Z}_2[s, s^{-1}]$. Reeb chords a_1 and a_2 are in degree 1 and the rest of the Reeb chords are in degree 0. The differential is given by

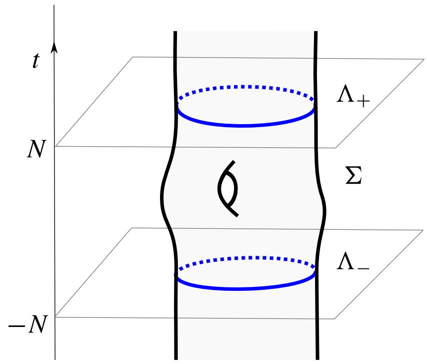


Figure 9. A schematic picture of an exact Lagrangian cobordism.

$$\begin{aligned}\partial(a_1) &= s^{-1} + b_1 + b_3 + b_1 b_2 b_3,\\ \partial(a_2) &= 1 + b_1 + b_3 + b_3 b_2 b_1,\\ \partial(b_i) &= 0, \quad i = 1, 2, 3,\\ \partial(s) &= \partial(s^{-1}) = 0.\end{aligned}$$

The definition of the DGA of a Legendrian link can be generalized to the case where there is more than one base point on some components of the link. Let Λ be an oriented Legendrian link and $\{s_1, \dots, s_l\}$ be a set of points on Λ such that each component of Λ has at least one point in the set and the set does not include any end of any Reeb chord of Λ. For a union of paths γ, associate it with a monomial $w(\gamma) = \prod_{j=1}^{l} s_j^{n_j(\gamma)}$ in $\mathbb{Z}_2[s_1^{\pm 1}, \dots, s_l^{\pm 1}]$, where n_j is defined much as above. The DGA

$$(\mathcal{A}(\Lambda, \{s_1, \dots, s_l\}), \partial)$$

is a unital graded algebra generated by Reeb chords of Λ over $\mathbb{Z}_2[s_1^{\pm 1}, \dots, s_l^{\pm 1}]$ endowed with a differential given by

$$\begin{aligned}\partial(a) &= \sum_{\dim \mathcal{M}^{\Lambda}(a;\boldsymbol{b})=0} \; \sum_{u \in \mathcal{M}^{\Lambda}(a;\boldsymbol{b})} w(u) b_1 \cdots b_m,\\ \partial(s_i) &= 0, \quad i = 1, \dots, l.\end{aligned}$$

2B. ***The DGA map induced by exact Lagrangian cobordisms.*** The Chekanov–Eliashberg DGA acts functorially on exact Lagrangian cobordisms, according to [Ekholm et al. 2016]. We first recall the definition of exact Lagrangian cobordisms.

Definition 2.5. Let Λ_+ and Λ_- be Legendrian submanifolds in $(\mathbb{R}^3, \ker \alpha)$, where $\alpha = dz - y\,dx$. An *exact Lagrangian cobordism* Σ from Λ_- to Λ_+ is a 2-dimensional surface in $(\mathbb{R} \times \mathbb{R}^3, d(e^t \alpha))$ such that there exists $T > 0$ such that Σ is

- cylindrical over Λ_+ on the positive end, i.e., $\Sigma \cap ((T, \infty) \times \mathbb{R}^3) = (T, \infty) \times \Lambda_+$;

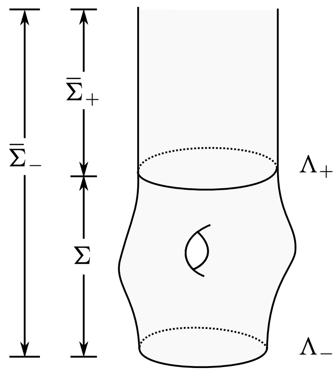


Figure 10. The relation among cobordisms $\overline{\Sigma}_+$, $\overline{\Sigma}_-$, and Σ.

- cylindrical over Λ_- on the negative end, i.e., $\Sigma \cap ((-\infty, -T) \times \mathbb{R}^3) = (-\infty, -T) \times \Lambda_-$;
- compact in $[-T, T] \times \mathbb{R}^3$,

and $e^t \alpha|_{T\Sigma} = df$ for some function $f : \Sigma \to \mathbb{R}$. (See Figure 9.)

When Λ_- is empty, the surface L satisfying the conditions above is called an *exact Lagrangian filling* of Λ_+.

By [Ekholm et al. 2016], an exact Lagrangian cobordism Σ from Λ_- to Λ_+ gives a DGA map from $\mathcal{A}(\Lambda_+)$ to $\mathcal{A}(\Lambda_-)$ with $\mathbb{Z}_2[H_1(\Sigma)]$ coefficients. Thus, an exact Lagrangian filling L of a Legendrian submanifold Λ, which can be viewed as a cobordism from the empty set to Λ, gives a DGA map from $\mathcal{A}(\Lambda)$ to the trivial DGA

$$(\mathbb{Z}_2[H_1(L)], 0),$$

which is an augmentation of $\mathcal{A}(\Lambda)$ to $\mathbb{Z}_2[H_1(L)]$.

For the purpose of computing augmentations of the Legendrian $(2, n)$ torus knots in Section 3A, we revise the coefficient ring of the DGA map induced by exact Lagrangian cobordisms from [Ekholm et al. 2016]. Instead of using $\mathbb{Z}_2[H_1(\Sigma)]$ coefficients, we will show the following proposition:

Proposition 2.6. *Let Λ_+ and Λ_- be Legendrian submanifolds in $(\mathbb{R}^3, \ker\alpha)$ and Σ be a connected exact Lagrangian cobordism from Λ_- to Λ_+. Assume that $\overline{\Sigma}_+$ is a connected exact Lagrangian cobordism from Λ_+ to some other Legendrian link and $\overline{\Sigma}_-$ is the concatenation of $\overline{\Sigma}_+$ and Σ as shown in Figure 10. The exact Lagrangian cobordism Σ induces a DGA map*

$$\Phi : (\mathcal{A}(\Lambda_+ ; \mathbb{Z}_2[H_1(\overline{\Sigma}_+)]), \partial_+) \to (\mathcal{A}(\Lambda_- ; \mathbb{Z}_2[H_1(\overline{\Sigma}_-)]), \partial_-)$$

with $\mathbb{Z}_2[H_1(\overline{\Sigma}_-)]$ coefficients.

Note that when $\overline{\Sigma}_+$ is an exact Lagrangian cylinder over Λ_+, this map agrees with the DGA map introduced in [Ekholm et al. 2016]. The proof of Proposition 2.6 follows Section 3 of that paper. Our revision of the coefficient ring is based on a different choice of capping paths of Λ_+ and Λ_-. Ekholm et al. choose capping paths of Λ_+ and Λ_- on Σ, while we choose capping paths of Λ_+ on $\overline{\Sigma}_+$ and capping paths of Λ_- on $\overline{\Sigma}_-$. For the rest of the section, we will describe this DGA map.

The inclusion map $\Lambda_+ \hookrightarrow \overline{\Sigma}_+$ makes it natural to define the DGA

$$(\mathcal{A}(\Lambda_+\,; \mathbb{Z}_2[H_1(\overline{\Sigma}_+)]), \partial_+).$$

The underlying algebra

$$\mathcal{A}(\Lambda_+\,; \mathbb{Z}_2[H_1(\overline{\Sigma}_+)]) = \mathcal{A}(\Lambda_+\,; \mathbb{Z}_2[H_1(\Lambda_+)]) \otimes_{\mathbb{Z}_2[H_1(\Lambda_+)]} \mathbb{Z}_2[H_1(\overline{\Sigma}_+)]$$

is generated by Reeb chords of Λ_+ over the ring $\mathbb{Z}_2[H_1(\overline{\Sigma}_+)]$. Given that $\overline{\Sigma}_+$ is connected, we can choose a family of capping paths for Λ_+ on $\overline{\Sigma}_+$. Thus, for any rigid holomorphic disk u_+ counted by ∂_+, it is natural to take the homology class of $\tilde{u}_+$ in $H_1(\overline{\Sigma}_+)$. Hence the differential coefficients of ∂_+ are in $\mathbb{Z}_2[H_1(\overline{\Sigma}_+)]$. In addition, the DGA $(\mathcal{A}(\Lambda_+\,; \mathbb{Z}_2[H_1(\overline{\Sigma}_+)]), \partial_+)$ does not depend on the choice of capping paths on $\overline{\Sigma}_+$ for a similar reason as in Proposition 2.3. The DGA $(\mathcal{A}(\Lambda_-\,; \mathbb{Z}_2[H_1(\overline{\Sigma}_-)]), \partial_-)$ is defined similarly.

The DGA map Φ induced by Σ is a composition of two maps. The first map

$$\psi : (\mathcal{A}(\Lambda_+\,; \mathbb{Z}_2[H_1(\overline{\Sigma}_+)]), \partial_+) \to (\mathcal{A}(\Lambda_+\,; \mathbb{Z}_2[H_1(\overline{\Sigma}_-)]), \partial_+)$$

is induced by the inclusion map $\overline{\Sigma}_+ \hookrightarrow \overline{\Sigma}_-$. It is not hard to show ψ is a DGA map. The second map

$$\phi : (\mathcal{A}(\Lambda_+\,; \mathbb{Z}_2[H_1(\overline{\Sigma}_-)]), \partial_+) \to (\mathcal{A}(\Lambda_-\,; \mathbb{Z}_2[H_1(\overline{\Sigma}_-)]), \partial_-)$$

is defined by counting rigid holomorphic disks in $\mathbb{R} \times \mathbb{R}^3$ with boundary on Σ.

Fix an almost complex structure J on $\mathbb{R} \times \mathbb{R}^3$ which is adjusted to the symplectic form ω (see [Ekholm et al. 2016, Section 3.2] for details). For a Reeb chord a of Λ_+ and Reeb chords $b_1, \dots, b_m$ of Λ_-, define $\mathcal{M}^{\Sigma}(a\,; b_1, \dots, b_m)$ to be the moduli space of J-holomorphic disks

$$u : (D_{m+1}, \partial D_{m+1}) \to (\mathbb{R} \times \mathbb{R}^3, \Sigma)$$

with the following properties:

- D_{m+1} is a 2-dimensional unit disk with $m+1$ points $r, s_1, s_2, \dots, s_m$ removed. The points $r, s_1, s_2, \dots, s_m$ are arranged counterclockwise on the boundary of the disk.
- The image of u is asymptotic to a strip $\mathbb{R}_+ \times a$ around r.
- The image of u is asymptotic to a strip $\mathbb{R}_- \times b_i$ around s_i for $i = 1, \dots, m$.

By [Cieliebak et al. 2010], there is a corresponding dimension formula:

$$\dim \mathcal{M}^{\Sigma}(a\,; b_1, \dots, b_m) = |a| - \sum_{i=1}^{m} |b_i|.$$

If $\dim \mathcal{M}^{\Sigma}(a\,; b_1, \dots, b_m) = 0$, the J-holomorphic disk $u \in \mathcal{M}^{\Sigma}(a\,; b_1, \dots, b_m)$ is called *rigid*. For each rigid J-holomorphic disk u, concatenate the image of the disk boundary with the capping paths of corresponding Reeb chords on $\overline{\Sigma}_-$ and get

$$\tilde{u} = u(\partial D_{m+1}) \cup \gamma_a \cup -\gamma_{b_1} \cup \cdots \cup -\gamma_{b_m},$$

which is a loop in $\overline{\Sigma}_-$. Hence we can take the homology class of $\tilde{u}$ in $H_1(\overline{\Sigma}_-)$, denoted by $[\tilde{u}]_{\overline{\Sigma}_-}$. The map

$$\phi : (\mathcal{A}(\Lambda_+\,; \mathbb{Z}_2[H_1(\overline{\Sigma}_-)]), \partial_+) \to (\mathcal{A}(\Lambda_-\,; \mathbb{Z}_2[H_1(\overline{\Sigma}_-)]), \partial_-)$$

is defined as follows. For any Reeb chord a of Λ_+, the map ϕ maps a to

$$\phi(a) = \sum_{\dim \mathcal{M}^{\Sigma}(a;\boldsymbol{b})=0} \; \sum_{u \in \mathcal{M}^{\Sigma}(a;\boldsymbol{b})} [u]_{\overline{\Sigma}_-} b_1 \cdots b_m.$$

The map ϕ is the identity on $\mathbb{Z}_1[H_1(\overline{\Sigma}_-)]$. By [Ekholm et al. 2016, Section 3.5], the map ϕ is a DGA map.

Therefore, the exact Lagrangian cobordism Σ induces a DGA map, $\Phi = \phi \circ \psi$

$$\Phi : (\mathcal{A}(\Lambda_+\,; \mathbb{Z}_2[H_1(\overline{\Sigma}_+)]), \partial_+) \to (\mathcal{A}(\Lambda_-\,; \mathbb{Z}_2[H_1(\overline{\Sigma}_-)]), \partial_-).$$

3. Main results

We consider the exact Lagrangian fillings of the Legendrian $(2, n)$ torus knot constructed from the EHK algorithm. Each filling can be achieved by concatenating n successive saddle cobordisms with two minimum cobordisms. In Section 3A, we combine results in [Ekholm et al. 2016] and Proposition 2.6 to write down combinatorial formulas for the DGA maps induced by a pinch move and a minimum cobordism. Composing all the DGA maps induced by n ordered pinch moves and the two minimum cobordisms, we obtain a combinatorial formula for augmentations of $\mathcal{A}(\Lambda)$ to $\mathbb{Z}_2[H_1(L)]$ induced by exact Lagrangian fillings L. In Section 3B, we find a combinatorial invariant to distinguish these resulting augmentations and hence we show that the C_n exact Lagrangian fillings are distinct up to exact Lagrangian isotopy. As a corollary, we extend the result to the case n is even.

3A. ***Computation of augmentations.*** Consider the Lagrangian projection of the Legendrian $(2, n)$ torus knot Λ with a base point $\tilde{s}_0$ and label the n crossings in degree 0 from left to right by $b_1, \dots, b_n$ as shown in Figure 11.

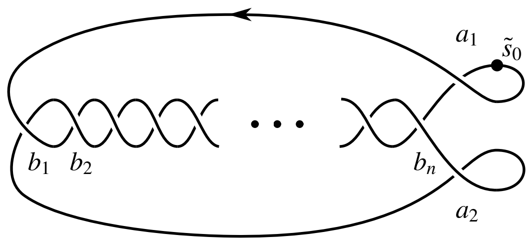


Figure 11. The Lagrangian projection of the Legendrian $(2, n)$ torus knot with a base point.

For each permutation σ of $\{1, \dots, n\}$, the corresponding exact Lagrangian filling L_σ of the Legendrian $(2, n)$ torus knot Λ is achieved in the following way:

- Start with an exact Lagrangian cylinder over Λ, denoted by $\overline{\Sigma}_0$. Label Λ as Λ_0.
- For $i = 1, \dots, n$, concatenate $\overline{\Sigma}_{i-1}$ from the bottom with a saddle cobordism Σ_i corresponding to the pinch move at crossing $b_{\sigma(i)}$ and get a new exact Lagrangian cobordism $\overline{\Sigma}_i$. Label the new Legendrian submanifold after the pinch move as Λ_i.
- Finally, use two minimal cobordisms, denoted by Σ_{n+1}, to close up $\overline{\Sigma}_n$ from the bottom and get the exact Lagrangian filling L_σ. To be consistent, let Λ_{n+1} be the empty set.

By Proposition 2.6, for $i = 1, \dots, n+1$, each exact Lagrangian cobordism Σ_i induces a DGA map:

$$\Phi_i : (\mathcal{A}(\Lambda_{i-1}\,;\mathbb{Z}_2[H_1(\overline{\Sigma}_{i-1})]), \partial_{i-1}) \to (\mathcal{A}(\Lambda_i\,;\mathbb{Z}_2[H_1(\overline{\Sigma}_i)]), \partial_i).$$

The map Φ_{n+1} that is induced by minimum cobordisms is well understood while the maps Φ_i for $i = 1, \dots, n$ that correspond to pinch moves are not. We will first study $H_1(\overline{\Sigma}_n)$ and give a geometric description of the DGA map that corresponds to a pinch move. Combining this with [Ekholm et al. 2016], we will write down an explicit combinatorial formula for each Φ_i, for $i = 1, \dots, n+1$.

To describe $H_1(\overline{\Sigma}_n)$ easily, we chop off the cylindrical top of $\overline{\Sigma}_n$ and view it as a surface with boundary $\Lambda \cup \Lambda_n$, also denoted by $\overline{\Sigma}_n$. By Poincaré duality, we have $H^1(\overline{\Sigma}_n) \cong H_1(\overline{\Sigma}_n, \Lambda \cup \Lambda_n)$. In particular, for each oriented curve α in $\overline{\Sigma}_n$ with ends on $\Lambda \cup \Lambda_n$, which is an element in $H_1(\overline{\Sigma}_n, \Lambda \cup \Lambda_n)$, there exists an element $\theta_\alpha \in H^1(\overline{\Sigma}_n)$ such that for any oriented loop β in $\overline{\Sigma}_n$, the intersection number of α and β is $\theta_\alpha(\beta)$. Thus, in order to know the homology class of a loop β in $H_1(\overline{\Sigma}_n)$, we only need to count the intersection number of each generator curve of $H_1(\overline{\Sigma}_n, \Lambda \cup \Lambda_n)$ with β.

We choose the set of generator curves of $H_1(\overline{\Sigma}_n, \Lambda \cup \Lambda_n)$ as follows. Use the t coordinate to slice $\overline{\Sigma}_n$ into a movie of diagrams (some of them may not be

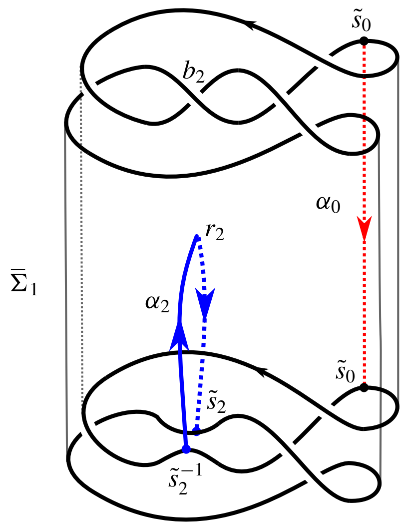


Figure 12. As an example, assume Λ is the Legendrian $(2, 3)$ torus knot and the first pinch move is taken at b_2. The blue curve and the red curve are α_2 and α_0 restricted on $\overline{\Sigma}_1$, respectively.

Legendrian diagrams). We study the trace of points on the diagram when t is decreasing. For $i = 1, \dots, n$, the saddle cobordism Σ_i flows all the points directly downward except ends of the Reeb chord $b_{\sigma(i)}$. According to [Lin 2016], the ends of the Reeb chord $b_{\sigma(i)}$ merge to a point $r_{\sigma(i)}$, and then split into two points, labeled as $\tilde{s}_{\sigma(i)}$ and $\tilde{s}_{\sigma(i)}^{-1}$ respectively. Now for $i = 1, \dots, n$, consider the trace of $\tilde{s}$ in $\overline{\Sigma}_n$, which is a flow line from r_i to the bottom of $\overline{\Sigma}_n$. Concatenating it with the inverse trace of $\tilde{s}_i^{-1}$ in $\overline{\Sigma}_n$, we get a curve α_i in $\overline{\Sigma}_n$ as shown in Figure 12. In addition, denote the trace of the base point $\tilde{s}_0$ in $\overline{\Sigma}_n$ by α_0. In this way, we have that $\alpha = \{\alpha_0, \alpha_1, \dots, \alpha_n\}$ is a set of generator curves of $H_1(\overline{\Sigma}_n, \Lambda \cup \Lambda_n) \cong \mathbb{Z}^{n+1}$.

For each curve α_i, where $i = 0, \dots, n$, Poincaré duality gives an element $\theta_{\alpha_i} \in H^1(\overline{\Sigma}_n)$. Denote its dual in $H_1(\Sigma_n)$ by $\tilde{s}_i$. Therefore, for any union of paths γ in $\overline{\Sigma}_n$, the monomial $w(\gamma)$ associated to γ in $\mathbb{Z}_2[H_1(\overline{\Sigma}_n)]$ is

$$w(\gamma) = \prod_{i=0}^{n} \tilde{s}_i^{n_i(\gamma)},$$

where $n_i(\gamma)$ is the intersection number of α_i and γ counted with signs.

For $i < n$, the map $H_1(\overline{\Sigma}_i) \to H_1(\overline{\Sigma}_n)$ induced by the inclusion map is injective. A similar argument shows that for a union of paths γ in $\overline{\Sigma}_i$, the monomial associated to γ in $\mathbb{Z}_2[H_1(\overline{\Sigma}_i)]$ counts intersections of $\alpha_0, \alpha_{\sigma(1)}, \dots, \alpha_{\sigma(i)}$ with γ. Notice that the curves $\alpha_{\sigma(i+1)}, \dots, \alpha_{\sigma(n)}$ do not intersect $\overline{\Sigma}_i$. Hence the monomial in $\mathbb{Z}_2[H_1(\overline{\Sigma}_i)]$ agrees with $w(\gamma)$ in $\mathbb{Z}_2[H_1(\overline{\Sigma}_n)]$.

Pick a family of capping paths for Λ_i on $\overline{\Sigma}_i$ for $i=0,\dots,n$. By Proposition 2.6, for $i=1,\dots,n+1$, each exact Lagrangian cobordism Σ_i gives a DGA map Φ_i,

$$\Phi_i : (\mathcal{A}(\Lambda_{i-1}\,;\mathbb{Z}_2[H_1(\overline{\Sigma}_{i-1})]),\partial_{i-1}) \to (\mathcal{A}(\Lambda_i\,;\mathbb{Z}_2[H_1(\overline{\Sigma}_i)]),\partial_i),$$

which maps any Reeb chord a of Λ_{i-1} to

$$\sum_{\dim\mathcal{M}^{\Sigma_i}(a;\boldsymbol{b})=0}\ \sum_{u\in\mathcal{M}^{\Sigma_i}(a;\boldsymbol{b})} w(\tilde{u})b_1\cdots b_m = \sum_{\dim\mathcal{M}^{\Sigma_i}(a;\boldsymbol{b})=0}\ \sum_{u\in\mathcal{M}^{\Sigma_i}(a;\boldsymbol{b})} \Bigl(w(\gamma_a)w(u)\prod_{i=1}^{m}w(\gamma_{b_i})^{-1}\Bigr)b_1\cdots b_m.$$

Now we show that the DGA map induced by the exact Lagrangian cobordisms is independent of the choice of capping paths.

Theorem 3.1. *Let γ and γ' be two families of capping paths of Λ_i on $\overline{\Sigma}_i$ for $i=0,\dots,n$. Denote the corresponding DGAs by $(\mathcal{A}^{\gamma}(\Lambda_i\,;\mathbb{Z}_2[H_1(\overline{\Sigma}_i)]),\partial_i^{\gamma})$ and $(\mathcal{A}^{\gamma'}(\Lambda_i\,;\mathbb{Z}_2[H_1(\overline{\Sigma}_i)]),\partial_i^{\gamma'})$. Assume Φ_i^{γ} and $\Phi_i^{\gamma'}$ are the corresponding DGA maps induced by Σ_i. Then the maps*

$$\begin{aligned} f_i : (\mathcal{A}^{\gamma}(\Lambda_i\,;\mathbb{Z}_2[H_1(\overline{\Sigma}_i)]),\partial_i^{\gamma}) &\to (\mathcal{A}^{\gamma'}(\Lambda_i\,;\mathbb{Z}_2[H_1(\overline{\Sigma}_i)]),\partial_i^{\gamma'})\\ c &\mapsto w(\gamma'_c)^{-1}w(\gamma_c)\,c \end{aligned}$$

are DGA isomorphisms for $i=0,\dots,n$. Further, the following diagram commutes:

$$\begin{tikzcd} \bigl(\mathcal{A}^{\gamma}(\Lambda_{i-1}\,;\mathbb{Z}_2[H_1(\overline{\Sigma}_{i-1})]),\partial_{i-1}^{\gamma}\bigr) \arrow[r, "f_{i-1}"] \arrow[d, "\Phi_i^{\gamma}"'] & \bigl(\mathcal{A}^{\gamma'}(\Lambda_{i-1}\,;\mathbb{Z}_2[H_1(\overline{\Sigma}_{i-1})]),\partial_{i-1}^{\gamma'}\bigr) \arrow[d, "\Phi_i^{\gamma'}"] \\ \bigl(\mathcal{A}^{\gamma}(\Lambda_i\,;\mathbb{Z}_2[H_1(\overline{\Sigma}_i)]),\partial_i^{\gamma}\bigr) \arrow[r, "f_i"] & \bigl(\mathcal{A}^{\gamma'}(\Lambda_i\,;\mathbb{Z}_2[H_1(\overline{\Sigma}_i)]),\partial_i^{\gamma'}\bigr) \end{tikzcd}$$

Proof. The maps f_i are DGA isomorphisms for the same reason as in Proposition 2.3. Now we prove the second part. For any Reeb chord a of Λ_{i-1} (and denoting $b_1\cdots b_m$ by $\boldsymbol{b}_*$),

$$\begin{aligned} f_i\circ\Phi_i^{\gamma}(a) &= f_i\Biggl(\sum_{\dim\mathcal{M}^{\Sigma_i}(a;\boldsymbol{b})=0}\ \sum_{u\in\mathcal{M}^{\Sigma_i}(a;\boldsymbol{b})}\Bigl(w(\gamma_a)w(u)\prod_{i=1}^{m}w(\gamma_{b_i})^{-1}\Bigr)\boldsymbol{b}_*\Biggr)\\ &= \sum_{\dim\mathcal{M}^{\Sigma_i}(a;\boldsymbol{b})=0}\ \sum_{u\in\mathcal{M}^{\Sigma_i}(a;\boldsymbol{b})}\Bigl(w(\gamma_a)w(u)\prod_{i=1}^{m}w(\gamma_{b_i})^{-1}w(\gamma'_{b_i})^{-1}w(\gamma_{b_i})\Bigr)\boldsymbol{b}_*\\ &= \sum_{\dim\mathcal{M}^{\Sigma_i}(a;\boldsymbol{b})=0}\ \sum_{u\in\mathcal{M}^{\Sigma_i}(a;\boldsymbol{b})}\Bigl(w(\gamma_a)w(u)\prod_{i=1}^{m}w(\gamma'_{b_i})^{-1}\Bigr)\boldsymbol{b}_*, \end{aligned}$$

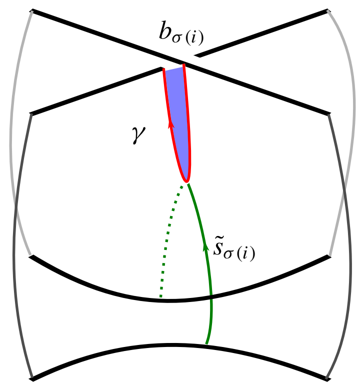


Figure 13. A cobordism corresponding to a pinch move, where the purple disk represents a holomorphic disk with a positive puncture at $b_{\sigma(i)}$.

$$\begin{aligned}\Phi_i^{\gamma'} \circ f_{i-1}(a) &= \Phi_i^{\gamma'}\big(w(\gamma'_a)^{-1} w(\gamma_a) a\big)\\ &= w(\gamma'_a)^{-1} w(\gamma_a) \sum_{\dim \mathcal{M}^{\Sigma_i}(a;\boldsymbol{b})=0} \sum_{u \in \mathcal{M}^{\Sigma_i}(a;\boldsymbol{b})} \Bigg(w(\gamma'_a) w(u) \prod_{i=1}^{m} w(\gamma'_{b_i})^{-1}\Bigg) \boldsymbol{b}_*\\ &= \sum_{\dim \mathcal{M}^{\Sigma_i}(a;\boldsymbol{b})=0} \sum_{u \in \mathcal{M}^{\Sigma_i}(a;\boldsymbol{b})} \Bigg(w(\gamma_a) w(u) \prod_{i=1}^{m} w(\gamma'_{b_i})^{-1}\Bigg) \boldsymbol{b}_*. \qquad \square\end{aligned}$$

Note that, if we cut $\overline{\Sigma}_i$ along the curves $\alpha_0, \alpha_{\sigma(1)}, \dots, \alpha_{\sigma(i)}$, the resulting surface is connected. Therefore, we can choose a family γ of capping paths for Λ_i on $\overline{\Sigma}_i$ such that none of them intersect the curves $\alpha_0, \alpha_{\sigma(1)}, \dots, \alpha_{\sigma(i)}$. Choose families of capping paths for $\Lambda_0, \dots, \Lambda_n$ in a similar way. As a result, for any rigid holomorphic disk u used in differentials of DGAs and DGA maps, we only need to count the intersections of curves in α with the disk boundary, i.e., $w(\tilde{u}) = w(u)$.

With this selection of capping paths, we are able to write down the DGA $(\mathcal{A}(\Lambda_i\,; \mathbb{Z}_2[H_1(\overline{\Sigma}_i)]), \partial_i)$ combinatorially, for $i = 1, \dots, n$. There are $2i+1$ points on Λ_i given by the intersection of α_0 and Λ_i, labeled by $\tilde{s}_0$, along with the two intersections of $\alpha_{\sigma(j)}$ and Λ_i, labeled by $\tilde{s}_{\sigma(j)}$ (positive intersection) and $\tilde{s}'_{\sigma(j)}$ (negative intersection), for $j = 1, \dots, i$. One then takes the DGA of Λ_i with these $2i+1$ base points, which has coefficients $\mathbb{Z}_2[\tilde{s}_0^{\pm 1}, \tilde{s}_{\sigma(1)}^{\pm 1}, \tilde{s}'^{\pm 1}_{\sigma(1)}, \dots, \tilde{s}_{\sigma(i)}^{\pm 1}, \tilde{s}'^{\pm 1}_{\sigma(i)}\,]$, and quotients by the relations $\tilde{s}'_{\sigma(j)} = \tilde{s}_{\sigma(j)}^{-1}$ for $j = 1, \dots, i$, to get the DGA

$$(\mathcal{A}(\Lambda_i\,; \mathbb{Z}_2[H_1(\overline{\Sigma}_i)]), \partial_i),$$

which is a DGA over $\mathbb{Z}_2[\tilde{s}_0^{\pm 1}, \tilde{s}_{\sigma(1)}^{\pm 1}, \dots, \tilde{s}_{\sigma(i)}^{\pm 1}]$, and $\{\tilde{s}_0, \tilde{s}_{\sigma(1)}, \dots, \tilde{s}_{\sigma(i)}\}$ is a basis of $H_1(\overline{\Sigma}_i)$ that corresponds to the curves $\alpha_0, \alpha_{\sigma(1)}, \dots, \alpha_{\sigma(i)}$.

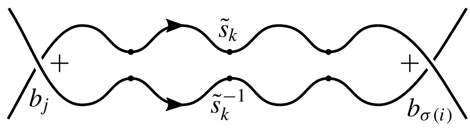


Figure 14. A part of the Lagrangian projection of Λ_{i-1}.

Now we are ready to describe the DGA map Φ_i induced by the exact Lagrangian cobordism Σ_i, for $i = 1, \dots, n$, which corresponds to a pinch move at crossing $b_{\sigma(i)}$. When we combine [Ekholm et al. 2016, Section 6.5] with Proposition 2.6, we find that the DGA map

$$\Phi_i : (\mathcal{A}(\Lambda_{i-1}\,;\mathbb{Z}_2[H_1(\overline{\Sigma}_{i-1})]), \partial_{i-1}) \to (\mathcal{A}(\Lambda_i\,;\mathbb{Z}_2[H_1(\overline{\Sigma}_i)]), \partial_i)$$

maps the Reeb chord $b_{\sigma(i)}$ to $\tilde{s}_{\sigma(i)}$ and any other Reeb chord c to

$$c + \sum_{\dim \mathcal{M}(c, b_{\sigma(i)}\,;\,c_1,\dots,c_m)=1} \; \sum_{u \in \mathcal{M}(c, b_{\sigma(i)}\,;\,c_1,\dots,c_m)} w(u)\tilde{s}_{\sigma(i)}^{-1} c_1 \cdots c_m,$$

where $\mathcal{M}(c, b_{\sigma(i)}\,; c_1, \dots, c_m)$ is the moduli space of holomorphic disks in $\mathbb{R}^2_{xy}$ with boundary on $\Pi_{xy}(\Lambda_{i-1})$ that covers a positive quadrant around c and $b_{\sigma(i)}$ and a negative quadrant around $c_1, \dots, c_m$. Please see [Ekholm et al. 2016, Section 6.5] for a detailed definition.

Here we discuss why the formulas make sense. The pinch move at $b_{\sigma(i)}$ pinches the Reeb chord $b_{\sigma(i)}$ down, which gives a holomorphic disk (as shown in Figure 13) with a positive puncture at $b_{\sigma(i)}$ and intersects $\tilde{s}_{\sigma(i)}$ exactly once. For a holomorphic disk $u \in \mathcal{M}(c, b_{\sigma(i)}\,; c_1, \dots, c_m)$, one can close the puncture of u at $b_{\sigma(i)}$ using the disk in Figure 13, which gives a holomorphic disk that contributes to $\Phi_i(c)$. Note that the boundary of this disk consists of the boundary of u and γ^{-1}. Thus the homology class of the boundary is $w(u)\tilde{s}_{\sigma(i)}^{-1}$, which matches the formula above.

In our case, in order to describe Φ_i combinatorially, we introduce two notations:

Definition 3.2. Let σ be a permutation of $\{1, \dots, n\}$. For $i \in \{1, \dots, n\}$, we define

$$\begin{aligned} T_\sigma^i :=&\{j \in \{1, \dots, n\} \mid \sigma^{-1}(j) > \sigma^{-1}(i) \\ &\qquad \text{and if } i < k < j \text{ or } j < k < i, \text{ then } \sigma^{-1}(k) < \sigma^{-1}(i)\}, \\ S_\sigma^i :=&\{j \in \{1, \dots, n\} \mid i \in T_\sigma^j\} \\ =&\{j \in \{1, \dots, n\} \mid \sigma^{-1}(j) < \sigma^{-1}(i) \\ &\qquad \text{and if } i < k < j \text{ or } j < k < i, \text{ then } \sigma^{-1}(k) < \sigma^{-1}(j)\}. \end{aligned}$$

Here T_σ^i captures all the Reeb chords b_j with the property that, before performing a pinch move at b_i, one can find a holomorphic disk with exactly two positive punctures at b_i and b_j. In other words, it gathers all the Reeb chords on which the

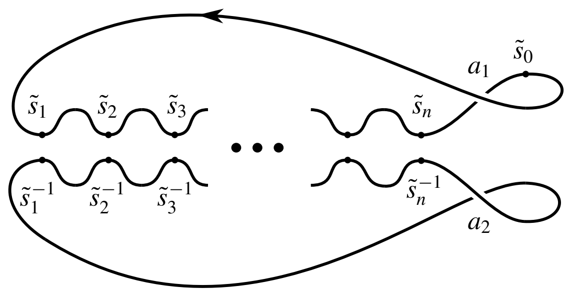


Figure 15. The Lagrangian projection of Λ_n.

DGA map induced by the pinch move at b_i acts nontrivially. The other set S_σ^i, on the other hand, detects all the Reeb chords b_j where a pinch move at b_j gives a DGA map that acts nontrivially on b_i.

If j is in $T_\sigma^{\sigma(i)}$ (an example is shown in Figure 14), the map Φ_i sends b_j to

$$\Phi_i(b_j) = b_j + \tilde{s}_{\sigma(i)}^{-1} \prod_{\substack{j<k<\sigma(i) \text{ or} \\ \sigma(i)<k<j}} \tilde{s}_k^{-2}.$$

For a_1, a_2 and the rest of the b_j where j is not in $T_\sigma^{\sigma(i)}$, the map Φ_i is identity.

Composing all the maps Φ_i for $i = 1, \dots, n$ together, we get a DGA map,

$$\overline{\Phi}_n : (\mathcal{A}(\Lambda\,; \mathbb{Z}_2[H_1(\Lambda)]), \partial) \to (\mathcal{A}(\Lambda_n\,; \mathbb{Z}_2[H_1(\overline{\Sigma}_n)]), \partial_n),$$

that is the identity map on the Reeb chords a_1, a_2. For $i = 1, \dots, n$, in order to know $\overline{\Phi}_n(b_i)$, we consider pinch moves at b_j such that $j \in S_\sigma^i$ together with the pinch move at b_i. These pinch moves correspond to all the DGA maps that contribute to $\overline{\Phi}_n$. Composing all these maps together, we have that

$$\overline{\Phi}_n(b_i) = \Phi_1 \circ \cdots \circ \Phi_{\sigma^{-1}(i)}(b_i) = \tilde{s}_i + \sum_{j \in S_\sigma^i} \Biggl(\tilde{s}_j^{-1} \prod_{\substack{j<k<i \text{ or} \\ i<k<j}} \tilde{s}_k^{-2} \Biggr).$$

Now we describe the last DGA map,

$$\Phi_{n+1} : (\mathcal{A}(\Lambda_n\,; \mathbb{Z}_2[H_1(\overline{\Sigma}_n)]), \partial_n) \to (\mathbb{Z}_2[H_1(L_\sigma)], 0).$$

As shown in Figure 15, the underlying algebra of Λ_n is generated by a_1 and a_2 and the differential is given by

$$\partial_n(a_1) = \tilde{s}_1\tilde{s}_2 \cdots \tilde{s}_n + \tilde{s}_0^{-1}, \qquad \partial_n(a_2) = \tilde{s}_n\tilde{s}_{n-1} \cdots \tilde{s}_1 + 1.$$

Consider the map $\psi : H_1(\overline{\Sigma}_n) \to H_1(L_\sigma)$ induced by the inclusion map $\overline{\Sigma}_n \hookrightarrow L_\sigma$. Since the DGA map

$$\Phi_{n+1} : (\mathcal{A}(\Lambda_n\,; \mathbb{Z}_2[H_1(\overline{\Sigma}_n)]), \partial_n) \to (\mathbb{Z}_2[H_1(L_\sigma)], 0)$$

satisfies $\Phi_{n+1}\circ \partial_n = 0 \circ \Phi_{n+1} = 0$, we have $\psi(\tilde{s}_0) = 1$ and $\psi(\tilde{s}_1)\psi(\tilde{s}_2)\cdots\psi(\tilde{s}_n) = 1$. Given that the map ψ is surjective, we assume a basis of $H_1(L_\sigma)$ is $\{s_1, \dots, s_{n-1}\}$, where $s_i = \tilde{s}_i$, for $i = 1, \dots, n-1$. The DGA map Φ_{n+1} is given by

$$\begin{aligned} &a_1 \mapsto 0, && a_2 \mapsto 0, \\ &\tilde{s}_0 \mapsto 1, && \tilde{s}_i \mapsto s_i, \quad i = 1, \dots, n-1, && \tilde{s}_n \mapsto (s_1 s_2 \cdots s_{n-1})^{-1}. \end{aligned}$$

Composing Φ_{n+1} with $\overline{\Phi}_n$, we get the augmentation ϵ_σ induced by L_σ as follows.

Theorem 3.3. *Given a permutation σ of $\{1, \dots, n\}$, let L_σ be the exact Lagrangian filling of the Legendrian $(2, n)$ torus knot Λ constructed from the EHK algorithm. If we write*

$$\begin{aligned} \mathbb{Z}_2[H_1(\Lambda)] &= \mathbb{Z}_2[\tilde{s}_0, \tilde{s}_0^{-1}], \\ \mathbb{Z}_2[H_1(L_\sigma)] &= \mathbb{Z}_2[s_1^{\pm 1}, \dots, s_{n-1}^{\pm 1}], \end{aligned}$$

and set $s_n = (s_1 s_2 \cdots s_{n-1})^{-1}$, then the augmentation

$$\epsilon_\sigma : \mathcal{A}(\Lambda\,; \mathbb{Z}_2[H_1(\Lambda)]) \to \mathbb{Z}_2[H_1(L_\sigma)]$$

induced by L_σ is given by

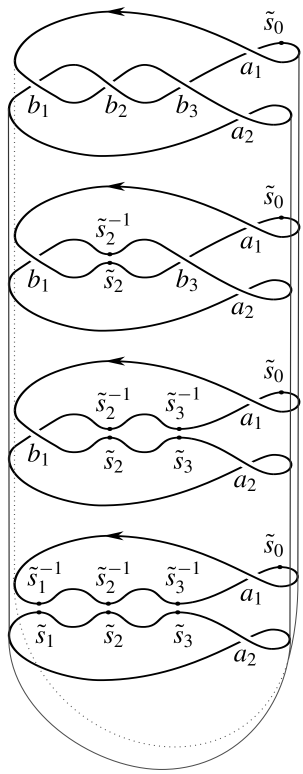
$$\begin{aligned} &\epsilon_\sigma(a_j) = 0, \quad j = 1, 2; \\ &\epsilon_\sigma(b_i) = s_i + \sum_{j \in S_\sigma^i} \Bigl(s_j^{-1} \prod_{\substack{j<k<i \text{ or} \\ i<k<j}} s_k^{-2} \Bigr), \quad i = 1, \dots, n; \\ &\epsilon_\sigma(\tilde{s}_0) = 1. \end{aligned}$$

Example 3.4. In Figure 16, as an example, we compute the augmentation $\epsilon_{(2,3,1)}$ of the Legendrian (2, 3) torus knot induced by the exact Lagrangian filling $L_{(2,3,1)}$.

Similarly, one can compute the augmentation for each permutation of $\{1, 2, 3\}$ and get the following table:

ϵ	$\epsilon(b_1)$	$\epsilon(b_2)$	$\epsilon(b_3)$
$\epsilon_{(1,2,3)}$	s_1	$s_2 + s_1^{-1}$	$s_1^{-1}s_2^{-1} + s_2^{-1}$
$\epsilon_{(1,3,2)} = \epsilon_{(3,1,2)}$	s_1	$s_2 + s_1^{-1} + s_1 s_2$	$s_1^{-1}s_2^{-1}$
$\epsilon_{(2,1,3)}$	$s_1 + s_2^{-1}$	s_2	$s_1^{-1}s_2^{-1} + s_2^{-1} + s_1^{-1}s_2^{-2}$
$\epsilon_{(2,3,1)}$	$s_1 + s_2^{-1} + s_1 s_2^{-1}$	s_2	$s_1^{-1}s_2^{-1} + s_2^{-1}$
$\epsilon_{(3,2,1)}$	$s_1 + s_2^{-1}$	$s_2 + s_1 s_2$	$s_1^{-1}s_2^{-1}$

3B. ***Proof of the main theorem.*** In this section, we use Theorem 3.3 to find an invariant of augmentations induced from the exact Lagrangian fillings obtained



$$
\begin{array}{cccc}
 & b_1 & b_2 & b_3 \\
\Phi_1 & \downarrow & \downarrow & \downarrow \\
 & b_1+\tilde{s}_2^{-1} & \tilde{s}_2 & b_3+\tilde{s}_2^{-1} \\
\Phi_2 & \downarrow & \downarrow & \downarrow \\
 & b_1+\tilde{s}_2^{-1}+\tilde{s}_2^{-2}\tilde{s}_3^{-1} & \tilde{s}_2 & \tilde{s}_3+\tilde{s}_2^{-1} \\
\Phi_3 & \downarrow & \downarrow & \downarrow \\
 & \tilde{s}_1+\tilde{s}_2^{-1}+\tilde{s}_2^{-2}\tilde{s}_3^{-1} & \tilde{s}_2 & \tilde{s}_3+\tilde{s}_2^{-1} \\
\Phi_4 & \downarrow & \downarrow & \downarrow \\
 & s_1+s_2^{-1}+s_1s_2^{-1} & s_2 & s_1^{-1}s_2^{-1}+s_2^{-1}
\end{array}
$$

Figure 16. A computation of the augmentation induced by an exact Lagrangian filling of the Legendrian (2, 3) torus knot. We keep track of the image of b_1, b_2, b_3 under the composition of Φ_1, Φ_2, Φ_3 and Φ_4. The last line is the image of b_1, b_2, b_3 under the augmentation $\epsilon_{(2,3,1)}$.

from the EHK algorithm. As a result, we distinguish all the augmentations in Theorem 3.3 and thus prove Theorem 1.1.

Lemma 3.5. *Let L_1 and L_2 be two exact Lagrangian fillings of the Legendrian $(2, n)$ torus knot Λ constructed from the EHK algorithm. If L_1 and L_2 are exact Lagrangian isotopic, then there exists an invertible map $g : H_1(L_1) \to H_1(L_2)$ such that the following diagram commutes*:

$$
\begin{tikzcd}
(\mathcal{A}(\Lambda), \partial) \arrow[r, "Id"] \arrow[d, "\epsilon_{L_1}"'] & (\mathcal{A}(\Lambda), \partial) \arrow[d, "\epsilon_{L_2}"'] \\
\mathbb{Z}_2[H_1(L_1)] \arrow[r, "g"'] & \mathbb{Z}_2[H_1(L_2)]
\end{tikzcd}
\tag{3-1}
$$

where ϵ_{L_1} and ϵ_{L_2} are augmentations induced by L_1 and L_2 respectively.

Proof. The isotopy between L_1 and L_2 induces an invertible map $g : H_1(L_1) \to H_1(L_2)$. If we identify both $H_1(L_1)$ and $H_1(L_2)$ with $\mathbb{Z}^{n-1}$, then $g \in GL(n-1, \mathbb{Z})$. This map induces a natural map on the corresponding group rings $\mathbb{Z}_2[H_1(L_1)] \to \mathbb{Z}_2[H_1(L_2)]$, also denoted by g. Thus, we have two augmentations of $\mathcal{A}(\Lambda)$ to $\mathbb{Z}_2[H_1(L_2)]$: $\epsilon_1 = g \circ \epsilon_{L_1}$ and $\epsilon_2 = \epsilon_{L_2}$. Since the two fillings L_1 and L_2 are isotopic through a family of exact Lagrangian fillings, according to [Ekholm et al. 2016, Theorem 1.3], we know that ϵ_1 and ϵ_2 are chain homotopic. In other words, there exists a degree 1 map $H : \mathcal{A}(\Lambda) \to \mathbb{Z}_2[H_1(L_2)]$ such that $H \circ \partial = \epsilon_1 - \epsilon_2$ as one can see from following diagram, where C_i denotes the degree i part of $\mathcal{A}(\Lambda)$.

$$\begin{tikzcd}
\cdots & C_{-1} \arrow[l] \arrow[dr, "H"] & C_0 \arrow[l, "\partial"'] \arrow[d, shift right, "\epsilon_1"'] \arrow[d, shift left, "\epsilon_2"] \arrow[dr, "H"] & C_1 \arrow[l, "\partial"'] & \cdots \arrow[l] \\
\cdots & 0 \arrow[l] & \mathbb{Z}_2[H_1(L_2)] \arrow[l] & 0 \arrow[l] & \cdots \arrow[l]
\end{tikzcd}$$

Note that Λ has a Lagrangian projection (as shown in Figure 11) such that no Reeb chords are in negative degree. Hence $C_{-1} = 0$ and $\epsilon_1 - \epsilon_2 = H \circ \partial = 0$. Therefore $\epsilon_1 = \epsilon_2$, i.e., the diagram (3-1) commutes. □

Remark 3.6. For any DGA $\mathcal{A}$ that vanishes on the degree -1 part, by the same argument, we have that two augmentations ϵ_1 and ϵ_2 of $\mathcal{A}$ are chain homotopic if and only if they are identically the same. For a more general criteria of two augmentations to be chain homotopic, check [Ng et al. 2015, Proposition 5.16].

Therefore, in order to distinguish exact Lagrangian fillings, we only need to distinguish their induced augmentations up to a $GL(n-1, \mathbb{Z})$ action. Observing the formula of the augmentation ϵ_σ in Theorem 3.3, we get a combinatorial way to define the number of terms in $\epsilon_\sigma(b_i)$ for $i = 1, \dots, n$ as follows.

Definition 3.7. For each permutation σ of $\{1, \dots, n\}$ and any number $i \in \{1, \dots, n\}$, we define $C_\sigma := (C_\sigma^1, C_\sigma^2, \dots, C_\sigma^n)$, where $C_\sigma^i = |S_\sigma^i| + 1$.

Example 3.8. We compute the vector C_σ for all of the permutations σ of $\{1, 2, 3\}$:

σ	$(1, 2, 3)$	$(1, 3, 2) \sim (3, 1, 2)$	$(2, 1, 3)$	$(2, 3, 1)$	$(3, 2, 1)$
C_σ	$(1, 2, 2)$	$(1, 3, 1)$	$(2, 1, 3)$	$(3, 1, 2)$	$(2, 2, 1)$

Proposition 3.9. *If two exact Lagrangian fillings L_{σ_1} and L_{σ_2} are exact Lagrangian isotopic, then $C_{\sigma_1} = C_{\sigma_2}$. In other words, the vector C_σ is an invariant of the exact Lagrangian filling L_σ up to exact Lagrangian isotopy.*

Proof. Using the formula in Theorem 3.3, we first show that C_σ^i is the number of terms in $\epsilon_\sigma(b_i)$. In order to do that, we need to prove that $\epsilon_\sigma(b_i)$ as a sum of monomials cannot be shorter, i.e, no terms in $\epsilon_\sigma(b_i)$ can be canceled by another term. First, replace s_n with $(s_1 \cdots s_{n-1})^{-1}$. If $i \neq n$, then each term of $\epsilon_\sigma(b_i)$ is one of the following forms:

(1) s_i ,

(2) $s_k^{-1} \prod_{j \in S} s_j^{-2}$ for some $k \neq i \in \{1, \dots, n-1\}$ and a subset $S \subset \{1, \dots, n-1\}$ that does not contain i, k (can be an empty set),

(3) $\prod_{j \in T} s_j^{-1} \prod_{k \notin T} s_k$ for some subset $T \subset \{1, \dots, n-1\}$ that does not contain i (can be an empty set).

If $i = n$, each term of $\epsilon_\sigma(b_n)$ can be either $s_1^{-1} \cdots s_{n-1}^{-1}$ or the form (2). Comparing degrees of $s_1, \dots, s_{n-1}$ of each term, we know that no terms can be canceled.

If L_{σ_1} and L_{σ_2} are exact Lagrangian isotopic, by Lemma 3.5, there is a map $g : \mathbb{Z}_2[H_1(L_1)] \to \mathbb{Z}_2[H_1(L_2)]$ such that $g \circ \epsilon_{L_1} = \epsilon_{L_2}$. Note that the map g on the group rings $\mathbb{Z}_2[H_1(L_1)] \to \mathbb{Z}_2[H_1(L_2)]$ is induced from an invertible map $H_1(L_1) \to H_1(L_2)$ and thus g maps a monomial to a monomial. Therefore $\epsilon_{\sigma_1}(b_i)$ and $\epsilon_{\sigma_2}(b_i)$ have the same number of terms, i.e., $C_{\sigma_1} = C_{\sigma_2}$. □

We say that two permutations σ_1 and σ_2 of $\{1, \dots, n\}$ are *isotopy equivalent* if they are equivalent via a sequence of relations of the form

$$(\dots, i, j, \dots, k, \dots) \sim (\dots, j, i, \dots, k, \dots), \quad \text{where } i < k < j. \tag{3-2}$$

By [Ekholm et al. 2016], if σ_1 and σ_2 are isotopy equivalent, the corresponding exact Lagrangian fillings L_{ϵ_1} and L_{ϵ_2} are exact Lagrangian isotopic and hence $C_{\sigma_1} = C_{\sigma_2}$. Conversely, we have the following:

Lemma 3.10. *If $C_{\sigma_1} = C_{\sigma_2}$, then σ_1 and σ_2 are isotopy equivalent.*

Proof. If $\sigma_1(1) = k$, then $C_{\sigma_1}^k = 1$. So $C_{\sigma_2}^k = 1$, i.e., we have that $S_{\sigma_2}^k = \varnothing$. If $\sigma_2(1) \neq k$, assume the element in σ_2 right before k is l, i.e., $\sigma_2(\sigma_2^{-1}(k) - 1) = l$. Note that $l \notin S_{\sigma_2}^k$, i.e., there exists i such that $l < i < k$ or $k < i < l$ and $\sigma_2^{-1}(i) > \sigma_2^{-1}(l)$. Note that $i \neq k$ and hence $\sigma_2^{-1}(i) > \sigma_2^{-1}(k) = \sigma_2^{-1}(l) + 1$. Thus we can use the relation (3-2) to switch l and k. In this way we can switch k to the first position in σ_2, i.e., $\sigma_2(1) = k = \sigma_1(1)$.

By induction, assume $\sigma_2(i) = \sigma_1(i)$ for $i < l$ and $\sigma_1(l) = k$. Then $S_{\sigma_1}^k \subset S_{\sigma_2}^k$. The assumption $C_{\sigma_2}^k = C_{\sigma_1}^k$ implies that $|S_{\sigma_1}^k| = |S_{\sigma_2}^k|$ and thus $S_{\sigma_1}^k = S_{\sigma_2}^k$. If $\sigma_2(l) \neq k$, for a similar reason to above, one can switch k to the l-th position and get $\sigma_2(l) = \sigma_1(l)$. Therefore, σ_1 and σ_2 are isotopy equivalent. □

Theorem 3.11. *If n is odd, the C_n exact Lagrangian fillings of the Legendrian $(2, n)$ torus knot Λ from the EHK algorithm are all of different exact Lagrangian isotopy classes.*

Proof. If two augmentations σ_1 and σ_2 are not isotopy equivalent, by Lemma 3.10, we have $C_{\sigma_1} \neq C_{\sigma_2}$. According to Proposition 3.9, the corresponding exact Lagrangian fillings L_{σ_1} and L_{σ_2} are not exact Lagrangian isotopic. Therefore, the

Legendrian $(2, n)$ torus knot has at least C_n exact Lagrangian fillings up to exact Lagrangian isotopy. □

Corollary 3.12. *When n is even, the Legendrian* $(2, n)$ *torus link* Λ *has at least* C_n *exact Lagrangian fillings.*

Proof. Start with the Legendrian $(2, n+1)$-knot Λ_0 and label its degree 0 Reeb chords from left to right by $b_1, \dots, b_{n+1}$ as usual. Let Σ be the exact Lagrangian cobordism from Λ to Λ_0 that corresponds to a pinch move of Λ_0 at b_{n+1}. For any permutation σ of $\{1, \dots, n\}$, the exact Lagrangian filling L_σ of Λ gives an exact Lagrangian filling of Λ_0 by concatenating with Σ on the top. This new exact Lagrangian filling of Λ_0 corresponds to the permutation $\tilde{\sigma} = (n+1, \sigma(1), \dots, \sigma(n))$ of $\{1, 2, \dots, n+1\}$, i.e., it is the filling $L_{\tilde{\sigma}}$ of Λ_0. Note that $C_{\tilde{\sigma}}^{n+1} = 1$. Moreover, we have that $C_{\tilde{\sigma}}^i = C_\sigma^i$ for $i = 1, \dots, n-1$ and $C_{\tilde{\sigma}}^n = C_\sigma^n + 1$. Thus $C_{\tilde{\sigma}}$ is determined by C_σ. Therefore, by Proposition 3.9 and Lemma 3.10, if two permutations σ_1 and σ_2 of $\{1, \dots, n\}$ are not isotopy equivalent, their induced permutations $\tilde{\sigma}_1$ and $\tilde{\sigma}_2$ of $\{1, \dots, n+1\}$ are not isotopy equivalent. According to Theorem 3.11, the corresponding exact Lagrangian fillings $L_{\tilde{\sigma}_1}$ and $L_{\tilde{\sigma}_2}$ of Λ_0 are not exact Lagrangian isotopic. Hence L_{σ_1} and L_{σ_2} are not exact Lagrangian isotopic. □

Acknowledgements

The author would like to thank Lenhard Ng for introducing her to the problem, and for many enlightening discussions. This work was partially supported by NSF grants DMS-0846346 and DMS-1406371.

References

[Chekanov 2002] Y. Chekanov, "Differential algebra of Legendrian links", *Invent. Math.* **150**:3 (2002), 441–483. MR Zbl

[Cieliebak et al. 2010] K. Cieliebak, T. Ekholm, and J. Latschev, "Compactness for holomorphic curves with switching Lagrangian boundary conditions", *J. Symplectic Geom.* **8**:3 (2010), 267–298. MR Zbl

[Ekholm et al. 2013] T. Ekholm, J. B. Etnyre, L. Ng, and M. G. Sullivan, "Knot contact homology", *Geom. Topol.* **17**:2 (2013), 975–1112. MR Zbl

[Ekholm et al. 2016] T. Ekholm, K. Honda, and T. Kálmán, "Legendrian knots and exact Lagrangian cobordisms", *J. Eur. Math. Soc.* (*JEMS*) **18**:11 (2016), 2627–2689. MR Zbl

[Eliashberg 1998] Y. Eliashberg, "Invariants in contact topology", *Doc. Math.* Extra Vol. 2 (1998), 327–338. MR Zbl

[Eliashberg et al. 2000] Y. Eliashberg, A. Givental, and H. Hofer, "Introduction to symplectic field theory", *Geom. Funct. Anal.* Special Volume, Part 2 (2000), 560–673. MR Zbl

[Etnyre et al. 2002] J. B. Etnyre, L. L. Ng, and J. M. Sabloff, "Invariants of Legendrian knots and coherent orientations", *J. Symplectic Geom.* **1**:2 (2002), 321–367. MR Zbl

[Lin 2016] F. Lin, "Exact Lagrangian caps of Legendrian knots", *J. Symplectic Geom.* **14**:1 (2016), 269–295. MR Zbl

[Ng 2003] L. L. Ng, "Computable Legendrian invariants", *Topology* **42**:1 (2003), 55–82. MR Zbl

[Ng 2010] L. Ng, "Rational symplectic field theory for Legendrian knots", *Invent. Math.* **182**:3 (2010), 451–512. MR Zbl

[Ng et al. 2015] L. Ng, D. Rutherford, V. Shende, S. Sivek, and E. Zaslow, "Augmentations are sheaves", preprint, 2015. arXiv

[Shende et al. 2015] V. Shende, D. Treumann, H. Williams, and E. Zaslow, "Cluster varieties from Legendrian knots", preprint, 2015. arXiv

Received August 1, 2016. Revised January 13, 2017.

YU PAN
MATHEMATICS DEPARTMENT
DUKE UNIVERSITY
BOX 90320
DURHAM, NC 27708
UNITED STATES
yp37@math.duke.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 289, No. 2, 2017
dx.doi.org/10.2140/pjm.2017.289.443

ELEMENTARY CALCULATION OF THE COHOMOLOGY RINGS OF REAL GRASSMANN MANIFOLDS

RUSTAM SADYKOV

We give elementary proofs of the Takeuchi and He theorems on the real cohomology rings and real equivariant cohomology rings of real Grassmann manifolds.

1. Introduction

In an influential paper, Borel [1953] developed a general technique of computing cohomology rings of compact symmetric spaces. However, there are some exceptional cases including those of real Grassmann manifolds of odd dimension that do not immediately fit the Borel theory. In these cases the cohomology rings with real coefficients were determined by Takeuchi [1962].

Let $\tilde{\mathrm{G}}(m,n)$ denote the Grassmann manifold of oriented planes of dimension m in $\mathbb{R}^{m+n}$. Its tautological m- and n-vector bundles support the total Pontrjagin classes

$$p = 1 + p_1 + \cdots + p_{\lfloor m/2 \rfloor}, \quad \bar{p} = 1 + \bar{p}_1 + \cdots + \cdots \bar{p}_{\lfloor n/2 \rfloor},$$

as well as the Euler classes e_m and $\bar{e}_n$. If mn is odd, there is also a cohomology class r in $\tilde{\mathrm{G}}(m,n)$ of degree $m+n-1$. Let $\mathbb{P} = \mathbb{P}(m,n)$ denote the symmetric algebra over $\mathbb{R}$ on the Pontrjagin classes p_i, $\bar{p}_j$ subject to the relation $p \cdot \bar{p} = 1$.

Theorem 1 [Takeuchi 1962]. *For $m, n > 1$, the cohomology algebra $H^*\tilde{\mathrm{G}}(m,n)$ over $\mathbb{R}$ is isomorphic to*

- $\mathbb{P} \otimes \Lambda(r)$ *if mn is odd,*
- $\mathbb{P}[e_m]$ *subject to $e_m^2 = p_{m/2}$ if m is even and n is odd,*
- $\mathbb{P}[\bar{e}_n]$ *subject to $\bar{e}_n^2 = \bar{p}_{n/2}$ if m is odd and n is even,*
- $\mathbb{P}[e_m, \bar{e}_n]$ *subject to $e_m\bar{e}_n = 0$, $e_m^2 = p_{m/2}$ and $\bar{e}_n^2 = \bar{p}_{n/2}$, if m and n are even.*

The original proof of Theorem 1 by Takeuchi [1962] relies on the Borel theory [Borel 1953] as well as the Borel–Hirzebruch theory [Borel and Hirzebruch 1958]. The algebra $H^*\tilde{\mathrm{G}}(m,n)$ as well as its equivariant version were also recently

MSC2010: primary 14M15, 57T15; secondary 22C05.
Keywords: Grassmann manifolds, equivariant cohomology.

computed by means of the GKM theory by He [2016]. Furthermore, there is an elegant computation of these algebras by means of pure Sullivan models by Carlson [2016] whose work relies on a model constructed by Kapovitch [2002]. We give a short elementary proof of Theorem 1 based on an observation that the total spaces of the tautological $(m-1)$-sphere bundle over $\tilde{\mathrm{G}}(m,n)$ and n-sphere bundle over $\tilde{\mathrm{G}}(n+1,m-1)$ are isomorphic. We also deduce from Theorem 1 its equivariant version, see Theorem 6.

2. Proof of Theorem 1

2.1. ***The case of even mn.*** The calculations in these three cases can be carried out directly as in the case where both m and n are odd, see Section 2.2. Alternatively, it suffices to observe that if mn is even, then the Lie groups $\mathrm{SO}_m \times \mathrm{SO}_n$ and SO_{m+n} are of the same rank, and therefore, Theorem 1 follows from the Borel Theorem [Borel 1953, §26].

2.2. ***The case of odd mn.*** To simplify notation, we fix the dimension $m+n$ of the ambient space and write $\tilde{\mathrm{G}}_m$ for $\tilde{\mathrm{G}}(m,n)$. Let $\mathrm{S}\tilde{\mathrm{G}}_m$ denote the total space of the (*tautological*) sphere bundle associated with the tautological m-vector bundle $E\tilde{\mathrm{G}}_m$ over $\tilde{\mathrm{G}}_m$. There are isomorphisms

$$\tilde{\mathrm{G}}_{m-1} = \tilde{\mathrm{G}}_{n+1}, \qquad \mathrm{S}\tilde{\mathrm{G}}_m = \mathrm{S}\tilde{\mathrm{G}}_{n+1}. \tag{1}$$

Remark 2. Since $\tilde{\mathrm{G}}_{n+m}$ consists of two points, it is reasonable to define $\tilde{\mathrm{G}}_0$ to be a two-point set.

The multiplication by $\bar{e}_{n+1}$ defines an endomorphism of $H^*\tilde{\mathrm{G}}_{n+1}$. By the result in Section 2.1, its cokernel I is the quotient of the algebra $\mathbb{P}[e_{m-1}]$ by the ideal generated by $e_{m-1}^2 - p_{\lfloor m/2\rfloor}$, while its kernel K is the ideal $e_{m-1}I$. Let r be a cohomology class in $\mathrm{S}\tilde{\mathrm{G}}_m$ such that $\delta(r) = e_{m-1}$, where δ is the coboundary homomorphism in the Gysin exact sequence

$$\cdots \xrightarrow{\smile \bar{e}_{n+1}} H^*\tilde{\mathrm{G}}_{n+1} \xrightarrow{i^*} H^*\,\mathrm{S}\tilde{\mathrm{G}}_m \xrightarrow{\delta} H^{*-n}\tilde{\mathrm{G}}_{n+1} \xrightarrow{\smile \bar{e}_{n+1}} \cdots \tag{*}$$

of the tautological sphere bundle over $\tilde{\mathrm{G}}_{n+1}$ with total space $\mathrm{S}\tilde{\mathrm{G}}_{n+1} = \mathrm{S}\tilde{\mathrm{G}}_m$.

Proposition 3. *The cohomology algebra of* $\mathrm{S}\tilde{\mathrm{G}}_m$ *is isomorphic to* $I \otimes \Lambda(r)$.

Proof. Since the restriction of i^* to I is injective, we will identify its image with I. By the Leibniz formula [Dold 1972, VII.8.10], the restriction of δ to the vector space rI is an isomorphism onto K. It follows now from (*) that the vector space $H^*\mathrm{S}\tilde{\mathrm{G}}_m$ is isomorphic to $I \oplus rI$. Finally, the class $r \smile r$ is trivial since r is of odd degree. □

Proof in the case of odd mn. In the Gysin exact sequence of the tautological sphere bundle over $\tilde{\mathrm{G}}_m$,

$$\cdots \xrightarrow{0} H^*\tilde{\mathrm{G}}_m \xrightarrow{i^*} H^*\,\mathrm{S}\tilde{\mathrm{G}}_m \xrightarrow{\bar{\delta}} H^{*-m+1}\tilde{\mathrm{G}}_m \xrightarrow{0} \cdots \tag{**}$$

the (surjective) coboundary homomorphism $\bar{\delta}$ restricted to $I = \mathbb{P} \oplus e_{m-1}\mathbb{P}/_\sim$ is trivial on $\mathbb{P}$ and takes $e_{m-1}p$ to p for all $p \in \mathbb{P}$; compare $\bar{\delta}$ with the coboundary homomorphism in the Gysin exact sequence of the tautological sphere bundle over BSO_m, see [Milnor and Stasheff 1974, p.180]. Since the Euler class e_m of the tautological bundle over $\tilde{\mathrm{G}}_m$ is trivial, the group $H^n\tilde{\mathrm{G}}_m$ is a subgroup of the trivial group $H^n\,\mathrm{S}\tilde{\mathrm{G}}_m$. Hence $\bar{\delta}(r) \in H^n\tilde{\mathrm{G}}_m$ is trivial, and therefore r extends to a class in $H^*\tilde{\mathrm{G}}_m$. This completes the proof of Theorem 1 in the case of odd mn. □

Remark 4. There is a free involution σ on $\tilde{\mathrm{G}}(m,n)$ whose orbit space is the Grassmann manifold $\mathrm{G}(m,n)$ of nonoriented planes. Hence $H^*\,\mathrm{G}(m,n)$ is isomorphic to the subring of $H^*\tilde{\mathrm{G}}(m,n)$ of σ-invariant classes. Casian and Kodama [2013, Theorem 3.2] gave a description of the adjacencies of Schubert cells in $\mathrm{G}(m,n)$, from which it follows that the class r corresponds to the Schubert cell with the Young diagram $(n) \times 1^{m-1}$; alternatively, it also follows from the Ehresmann's adjacency formulas.

Remark 5. From Giambelli's formula, the mod 2 reduction of r is $\bar{w}_n w_{m-1} = \bar{w}_{n-1} w_m$.

3. Equivariant case

Recall that $\tilde{\mathrm{G}} = \tilde{\mathrm{G}}(m,n)$ can be identified with the quotient of $\mathrm{SO}(m+n)$ by $\mathrm{SO}(m) \times \mathrm{SO}(n)$. Let T denote the maximal torus of the latter group. There is a left action of $T < \mathrm{SO}(m+n)$ on the Grassmann manifold $\tilde{\mathrm{G}}$. Let k be the dimension of T; it equals $\lfloor (m+n)/2 \rfloor$ if mn is even, and $\lfloor (m+n-1)/2 \rfloor$ if mn is odd. In this section we give a short computation of the equivariant cohomology ring $H_T^*\tilde{\mathrm{G}}$ which was earlier computed by He [2016] and Carlson [2016].

Recall that the equivariant cohomology ring $H_T^*\tilde{\mathrm{G}}$ is defined to be the cohomology ring of $\tilde{\mathrm{G}}_T = ET \times_T \tilde{\mathrm{G}}$, where ET is the total space of the principle T-bundle $ET \to BT$. In the cohomology ring of $\tilde{\mathrm{G}}_T$ there are total Pontrjagin classes p^T and $\bar{p}^T$ and Euler classes e_m^T and $\bar{e}_n^T$ of the tautological vector bundles over $\tilde{\mathrm{G}}_T$, as well as the first Chern classes $t_1, \dots, t_k$ of the k complex line bundles $L_1, \dots, L_k$ that are pulled back from the tautological complex line bundles over $BT = \mathbb{C}P^\infty \times \cdots \times \mathbb{C}P^\infty$. Since the sum of the two tautological vector bundles over $\tilde{\mathrm{G}}_T$ is stably equivalent to $L_1 \oplus \cdots \oplus L_k$, we have a relation $p^T\,\bar{p}^T = \prod(1+t_i^2)$. Similarly, $e_m^T \bar{e}_n^T = \prod t_i$ if m and n are even and $m+n = 2k$, and $e_{m-1}^T \bar{e}_{n+1}^T = 0$ if m and n are odd and $m+n = 2k+2$. Let $\mathbb{P}^T$ denote the symmetric algebra over $\mathbb{R}$ generated by the Pontrjagin classes p_i^T, $\bar{p}_i^T$ as well as the Chern classes t_i subject to the relation $p^T\,\bar{p}^T = \prod(1+t_i^2)$. When mn is odd, the equivariant version of the Gysin exact sequence (*) defines a cohomology class $\tilde{r}$ in $\mathrm{S}\tilde{\mathrm{G}}_m$, while from the equivariant version of the Gysin exact sequence (**) it follows that $\tilde{r}$ extends to a class in $\tilde{\mathrm{G}}_m$.

Theorem 6 [He 2016; Carlson 2016]. *For $m, n > 1$, the algebra $H^*_T\tilde{\mathrm{G}}(m,n)$ is isomorphic to*

- $\mathbb{P}^T \otimes \Lambda(\tilde{r})$ *if mn is odd,*
- $\mathbb{P}^T[e^T_m]$ *subject to $(e^T_m)^2 = p^T_{m/2}$ if m is even and n is odd,*
- $\mathbb{P}^T[\bar{e}^T_n]$ *subject to $(\bar{e}^T_n)^2 = \bar{p}_{n/2}$ if m is odd and n is even,*
- $\mathbb{P}^T[e_m, \bar{e}^T_n]$ *subject to $e^T_m\bar{e}^T_n = \prod t_i$, $(e^T_m)^2 = p_{m/2}$ and $(\bar{e}^T_n)^2 = \bar{p}_{n/2}$ if m and n are even.*

Proof. We have seen that all cohomology classes of the fiber $\tilde{\mathrm{G}}$ of the fiber bundle $\tilde{\mathrm{G}}_T \to BT$ extend over the total space. Thus, $H^*_T\tilde{\mathrm{G}}$ is a free H^*BT-module on the set of generators given by a basis of the vector space $H^*\tilde{\mathrm{G}}$. In particular, in $H^*_T\tilde{\mathrm{G}}$ there are no relations besides those listed in Theorem 6. Indeed, assume to the contrary that there is a trivial algebraic combination y of classes $\tilde{r}$, e^T_m, $\bar{e}^T_n$, p_i, $\bar{p}_i$ and t_i not in the ideal $\mathcal{I}$ generated by the relations in Theorem 6. Using relations in Theorem 6 we can reduce y to an H^*BT-linear combination of basis vectors of $H^*\tilde{\mathrm{G}}$. Since y is trivial, all coefficients in the reduced linear combination are zero. Hence $y \in \mathcal{I}$ contrary to the assumption. □

Acknowledgment

I am grateful to Toru Ohmoto; this project accidentally came out of our conversations at a Sapporo onsen during my visiting Hokkaido University. Ohmoto could not be convinced to coauthor the note.

I am thankful to the referee who informed me about the papers [Takeuchi 1962] and [He 2016], made many helpful comments and shared with me a copy of [Takeuchi 1962]. I am grateful to Osamu Saeki for many references, and to László Fehér for letting me know about his unpublished work. Namely, Fehér [2013] observed that the Schubert calculus of the Grassmann manifold $\mathrm{G}(2m, 2n)$ is analogous to the Schubert calculus of the Grassmann manifold of complex m-planes in $\mathbb{C}^n$.

References

[Borel 1953] A. Borel, "Sur la cohomologie des espaces fibrés principaux et des espaces homogènes de groupes de Lie compacts", *Ann. of Math.* (2) **57** (1953), 115–207. MR Zbl

[Borel and Hirzebruch 1958] A. Borel and F. Hirzebruch, "Characteristic classes and homogeneous spaces, I", *Amer. J. Math.* **80** (1958), 458–538. MR Zbl

[Carlson 2016] J. D. Carlson, "The Borel equivariant cohomology of real Grassmannians", preprint, 2016. arXiv

[Casian and Kodama 2013] L. Casian and Y. Kodama, "On the cohomology of real Grassmann manifolds", preprint, 2013. arXiv

[Dold 1972] A. Dold, *Lectures on algebraic topology*, Grundlehren der math. Wissenschaften **200**, Springer, 1972. MR Zbl

[Fehér 2013] L. Fehér, "Real Schubert calculus", web page, Math Overflow, 2013, available at http://mathoverflow.net/questions/119273/real-schubert-calculus.

[He 2016] C. He, "GKM theory, characteristic classes and the equivariant cohomology ring of real Grassmannian", preprint, 2016. arXiv

[Kapovitch 2002] V. Kapovitch, "A note on rational homotopy of biquotients", preprint, 2002, available at http://www.math.toronto.edu/vtk/biquotient.pdf.

[Milnor and Stasheff 1974] J. W. Milnor and J. D. Stasheff, *Characteristic classes*, Annals of Mathematics Studies **76**, Princeton Univ. Press, 1974. MR Zbl

[Takeuchi 1962] M. Takeuchi, "On Pontrjagin classes of compact symmetric spaces", *J. Fac. Sci. Univ. Tokyo Sect. I* **9**:1962 (1962), 313–328. MR Zbl

Received August 30, 2015. Revised February 5, 2017.

RUSTAM SADYKOV
MATHEMATICS DEPARTMENT
KANSAS STATE UNIVERSITY
138 CARDWELL HALL
MANHATTAN, KS 66503
UNITED STATES

rstsdk@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 289, No. 2, 2017
dx.doi.org/10.2140/pjm.2017.289.449

CLUSTER TILTING MODULES AND NONCOMMUTATIVE PROJECTIVE SCHEMES

KENTA UEYAMA

We study the relationship between equivalences of noncommutative projective schemes and cluster tilting modules. In particular, we prove the following result. Let A be an AS-Gorenstein algebra of dimension $d \geq 2$ and $\mathsf{tails}\, A$ the noncommutative projective scheme associated to A. If $\mathrm{gldim}(\mathsf{tails}\, A) < \infty$ and A has a $(d-1)$-cluster tilting module X with the property that its graded endomorphism algebra is $\mathbb{N}$-graded, then the graded endomorphism algebra B of a basic $(d-1)$-cluster tilting submodule of X is a two-sided noetherian $\mathbb{N}$-graded AS-regular algebra over B_0 of global dimension d such that $\mathsf{tails}\, B$ is equivalent to $\mathsf{tails}\, A$.

1. Introduction

Artin and Zhang [1994] introduced the notion of a noncommutative projective scheme, and established a fundamental and comprehensive theory of noncommutative projective schemes. Since the study of the categories of coherent sheaves on commutative projective schemes (or their derived categories) is of increasing importance in algebraic geometry, the study of noncommutative projective schemes has been a major project in noncommutative projective geometry.

Let A, A' be right noetherian graded algebras, and $\mathsf{tails}\, A$, $\mathsf{tails}\, A'$ the noncommutative projective schemes associated to A and A' respectively. Clearly, if $A \cong A'$ as graded algebras, then $\mathsf{tails}\, A \cong \mathsf{tails}\, A'$. It is well known that the converse does not hold, so the following question is a natural one to ask.

Question 1.1. Given a right noetherian graded algebra A, can we find a better homogeneous coordinate ring of $\mathsf{tails}\, A$? That is, can we find a better graded algebra B (e.g., $\mathrm{gldim}\, B < \infty$) such that $\mathsf{tails}\, B \cong \mathsf{tails}\, A$?

For example, take the commutative graded algebra $A = k[x, y, z, w]/(xw - yz)$. Then $\mathsf{tails}\, A$ ($\cong \mathsf{coh}\, \mathbb{P}^1 \times \mathbb{P}^1$) is not equivalent to $\mathsf{tails}\, k[x_1, \dots, x_n]$ ($\cong \mathsf{coh}\, \mathbb{P}^{n-1}$), but we can find the noncommutative graded algebra $B = k\langle x, y\rangle/(x^2y - yx^2, y^2x - xy^2)$

The author was supported by JSPS Grant-in-Aid for Young Scientists (B) 15K17503.
MSC2010: primary 16S38; secondary 14A22, 16G50, 16W50.
Keywords: noncommutative projective scheme, cluster tilting module, AS-Gorenstein algebra, AS-regular algebra, ASF-regular algebra.

of global dimension 3 such that $\mathsf{tails}\, A \cong \mathsf{tails}\, B$, so the above question has a significant meaning in noncommutative projective geometry.

The purpose of this paper is to give an answer to Question 1.1 by investigating cluster tilting modules. Cluster tilting modules are crucial in the study of higher-dimensional analogues of Auslander–Reiten theory, and also attract attention from the viewpoint of Van den Bergh's noncommutative crepant resolutions. In particular, cluster tilting modules have been extensively studied for a certain class of algebras, called orders, including commutative Cohen–Macaulay rings and finite-dimensional algebras (see [Iyama 2008]). One of the goals of this paper is to develop cluster tilting theory for nonorders in terms of noncommutative projective geometry.

The main result of this paper is as follows. Let A be a two-sided noetherian connected graded algebra satisfying χ and let X be a finitely generated graded right A-module. If A is an AS-Gorenstein algebra of dimension $d \geq 2$ and X is a $(d-1)$-cluster tilting module satisfying some additional conditions, then the graded endomorphism algebra $B = \underline{\mathrm{End}}_A(X)$ is a two-sided noetherian ASF-regular algebra of global dimension d such that the functors

$$\mathsf{tails}\, B \to \mathsf{tails}\, A \ \text{ induced by } \ - \otimes_B X$$

and

$$\mathsf{tails}\, B^{\mathrm{op}} \to \mathsf{tails}\, A^{\mathrm{op}} \ \text{ induced by } \ \underline{\mathrm{Hom}}_A(X, A) \otimes_B -$$

are equivalences (Theorem 3.10 (1)). Moreover, a certain converse statement also holds (Theorem 3.10 (2)). As a corollary of this result, we can answer Question 1.1.

Theorem 1.2 (Corollary 3.12). *Let A be an AS-Gorenstein algebra of dimension $d \geq 2$. If $\mathrm{gldim}(\mathsf{tails}\, A) < \infty$ and A has a $(d-1)$-cluster tilting module X with the property that its graded endomorphism algebra is $\mathbb{N}$-graded, then the graded endomorphism algebra B of a basic $(d-1)$-cluster tilting submodule of X is a two-sided noetherian $\mathbb{N}$-graded AS-regular algebra over B_0 of global dimension d such that $\mathsf{tails}\, B \cong \mathsf{tails}\, A$.*

We note that the notions of ASF-regular and AS-regular over R were recently introduced by Minamoto and Mori [2011], and these are natural generalizations of AS-regular algebras for $\mathbb{N}$-graded (not necessarily connected graded) algebras. A comparison theorem for these algebras is described in Theorem 2.10 (see also Corollary 2.11).

2. Preliminaries

Throughout, let k be a field. A graded k-vector space $V = \bigoplus_{i \in \mathbb{Z}} V_i$ is called locally finite if $\dim_k V_i < \infty$ for all $i \in \mathbb{Z}$, and it is called left (resp. right) bounded if $V_i = 0$ for all $i \ll 0$ (resp. $i \gg 0$). We denote by $DV = \underline{\mathrm{Hom}}_k(V, k)$ the graded vector space dual of a locally finite graded k-vector space V.

In this paper, a graded algebra means a $\mathbb{Z}$-graded algebra over k unless otherwise stated. For a graded algebra A, we denote by $\mathsf{GrMod}\,A$ the category of graded right A-modules with A-module homomorphisms of degree 0, and by $\mathsf{grmod}\,A$ the full subcategory consisting of finitely generated graded A-modules. Note that if A is right noetherian, then $\mathsf{grmod}\,A$ is an abelian category. We denote by A^{op} the opposite algebra of A, and by $A^e = A^{\mathrm{op}} \otimes_k A$ the enveloping algebra. The category of graded left A-modules is identified with $\mathsf{GrMod}\,A^{\mathrm{op}}$, and the category of graded A-A bimodules is identified with $\mathsf{GrMod}\,A^e$.

For a graded module $M \in \mathsf{GrMod}\,A$ and an integer $n \in \mathbb{Z}$, we define the shift $M(n) \in \mathsf{GrMod}\,A$ by $M(n)_i := M_{n+i}$ for $i \in \mathbb{Z}$. Note that the rule $M \mapsto M(n)$ is a k-linear autoequivalence for $\mathsf{GrMod}\,A$ and $\mathsf{grmod}\,A$, called the shift functor. For $M, N \in \mathsf{GrMod}\,A$, we write $\mathrm{Ext}^i_{\mathsf{GrMod}\,A}(M, N)$ for the extension group in $\mathsf{GrMod}\,A$, and define

$$\underline{\mathrm{Ext}}^i_A(M, N) := \bigoplus_{i \in \mathbb{Z}} \mathrm{Ext}^i_{\mathsf{GrMod}\,A}(M, N(i)).$$

Let A be a right noetherian locally finite $\mathbb{N}$-graded algebra. For $M \in \mathsf{GrMod}\,A$ and an integer $n \in \mathbb{Z}$, we define the truncated submodule $M_{\geq n} \in \mathsf{GrMod}\,A$ by $M_{\geq n} := \bigoplus_{i \geq n} M_i$. We say that an element x of a graded module $M \in \mathsf{GrMod}\,A$ is torsion if there exists a positive integer n such that $xA_{\geq n} = 0$. We denote by $t(M)$ the submodule of M consisting of all torsion elements. A graded module $M \in \mathsf{GrMod}\,A$ is called torsion if $M = t(M)$, and torsion-free if $t(M) = 0$. We denote by $\mathsf{Tors}\,A$ (resp. $\mathsf{tors}\,A$) the full subcategory of $\mathsf{GrMod}\,A$ (resp. $\mathsf{grmod}\,A$) consisting of torsion modules. One can define the Serre quotient categories

$$\mathsf{Tails}\,A = \mathsf{GrMod}\,A/\mathsf{Tors}\,A \quad \text{and} \quad \mathsf{tails}\,A = \mathsf{grmod}\,A/\mathsf{tors}\,A.$$

Note that $\mathsf{tails}\,A$ is the full subcategory of noetherian objects of $\mathsf{Tails}\,A$. The quotient functor is denoted by $\pi : \mathsf{GrMod}\,A \to \mathsf{Tails}\,A$. We often denote by $\mathcal{M} = \pi M \in \mathsf{Tails}\,A$ the image of $M \in \mathsf{GrMod}\,A$. Note that the shift functor preserves torsion modules, so it induces a k-linear autoequivalence $\mathcal{M} \mapsto \mathcal{M}(n)$ for $\mathsf{Tails}\,A$ and $\mathsf{tails}\,A$, again called the shift functor. For $\mathcal{M}, \mathcal{N} \in \mathsf{Tails}\,A$, we write $\mathrm{Ext}^i_{\mathsf{Tails}\,A}(\mathcal{M}, \mathcal{N})$ for the extension group in $\mathsf{Tails}\,A$, and define

$$\underline{\mathrm{Ext}}^i_{\mathcal{A}}(\mathcal{M}, \mathcal{N}) := \bigoplus_{i \in \mathbb{Z}} \mathrm{Ext}^i_{\mathsf{Tails}\,A}(\mathcal{M}, \mathcal{N}(i)).$$

See [Artin and Zhang 1994, Section 7] for details on Ext groups in $\mathsf{tails}\,A$. Following Serre's theorem and the Gabriel–Rosenberg reconstruction theorem, $\mathsf{tails}\,A$ is called the noncommutative projective scheme associated to A (see [Artin and Zhang 1994] for details). We define the global dimension of $\mathsf{tails}\,A$ by

$$\mathrm{gldim}(\mathsf{tails}\,A) = \sup\{i \mid \mathrm{Ext}^i_{\mathsf{tails}\,A}(\mathcal{M}, \mathcal{N}) \neq 0 \text{ for some } \mathcal{M}, \mathcal{N} \in \mathsf{tails}\,A\}.$$

If $\operatorname{gldim} A < \infty$, it is clear that $\operatorname{gldim}(\mathsf{tails}\, A) < \infty$. The condition $\operatorname{gldim}(\mathsf{tails}\, A) < \infty$ is considered as a noncommutative graded isolated singularity property (see [Ueyama 2013; 2015; Mori and Ueyama 2016a; 2016b]).

Recall that we say that χ_i holds for A if $\underline{\operatorname{Ext}}_A^j(A/A_{\geq 1}, M)$ is finite-dimensional over k for every $M \in \mathsf{grmod}\, A$ and every $j \leq i$, and we say that χ holds for A if χ_i holds for every i. The condition χ_i plays an essential role in the study of the noncommutative projective scheme $\mathsf{tails}\, A$ (see [Artin and Zhang 1994; Yekutieli and Zhang 1997] for details).

We call $(\mathsf{C}, \mathcal{O}, s)$ an algebraic triple if it consists of a k-linear abelian category C, an object $\mathcal{O} \in \mathsf{C}$, and a k-linear autoequivalence $s \in \operatorname{Aut}_k \mathsf{C}$.

Definition 2.1 [Artin and Zhang 1994]. Let $(\mathsf{C}, \mathcal{O}, s)$ be an algebraic triple. We say that the pair $(\mathcal{O}, s)$ is ample for C if

(Am1) for every object $\mathcal{M} \in \mathsf{C}$, there are positive integers $r_1, \ldots, r_p \in \mathbb{N}^+$ and an epimorphism $\bigoplus_{i=1}^p s^{-r_i}\mathcal{O} \to \mathcal{M}$ in C, and

(Am2) for every epimorphism $\mathcal{M} \to \mathcal{N}$ in C, there is an integer n_0 such that the induced map $\operatorname{Hom}_{\mathsf{C}}(s^{-n}\mathcal{O}, \mathcal{M}) \to \operatorname{Hom}_{\mathsf{C}}(s^{-n}\mathcal{O}, \mathcal{N})$ is surjective for every $n \geq n_0$.

We define the graded algebra associated to an algebraic triple $(\mathsf{C}, \mathcal{O}, s)$ by $B(\mathsf{C}, \mathcal{O}, s) := \bigoplus_{i \in \mathbb{Z}} \operatorname{Hom}_{\mathsf{C}}(\mathcal{O}, s^i\mathcal{O})$. Moreover, for any object $\mathcal{M} \in \mathsf{C}$, it is known that $\bigoplus_{i \in \mathbb{Z}} \operatorname{Hom}_{\mathsf{C}}(\mathcal{O}, s^i\mathcal{M})$ has a natural graded right $B(\mathsf{C}, \mathcal{O}, s)$-module structure.

Theorem 2.2 [Artin and Zhang 1994, Corollary 4.6 (1)]. *Let $(\mathsf{C}, \mathcal{O}, s)$ be an algebraic triple. If $\mathcal{O} \in \mathsf{C}$ is a noetherian object, $\dim_k \operatorname{Hom}_{\mathsf{C}}(\mathcal{O}, \mathcal{M}) < \infty$ for all $\mathcal{M} \in \mathsf{C}$, and $(\mathcal{O}, s)$ is ample for C, then $B = B(\mathsf{C}, \mathcal{O}, s)_{\geq 0}$ is a right noetherian locally finite $\mathbb{N}$-graded algebra satisfying χ_1, and the functor*

$$\mathsf{C} \to \mathsf{tails}\, B, \quad \mathcal{F} \mapsto \pi\Big(\bigoplus_{i \in \mathbb{N}} \operatorname{Hom}_{\mathsf{C}}(\mathcal{O}, s^i\mathcal{F})\Big)$$

induces an equivalence of algebraic triples $(\mathsf{C}, \mathcal{O}, s) \to (\mathsf{tails}\, B, \mathcal{B}, (1))$.

Let A be an $\mathbb{N}$-graded algebra. Then the augmentation ideal $A_{\geq 1}$ is denoted by $\mathfrak{m}$. We define the functor $\underline{\Gamma}_{\mathfrak{m}} : \mathsf{GrMod}\, A \to \mathsf{GrMod}\, A$ by

$$\underline{\Gamma}_{\mathfrak{m}}(-) = \lim_{n \to \infty} \underline{\operatorname{Hom}}_A(A/A_{\geq n}, -).$$

The derived functor of $\underline{\Gamma}_{\mathfrak{m}}$ is denoted by $\mathrm{R}\underline{\Gamma}_{\mathfrak{m}}(-)$, and its cohomologies are denoted by $\underline{\mathrm{H}}_{\mathfrak{m}}^i(-) = h^i(\mathrm{R}\underline{\Gamma}_{\mathfrak{m}}(-))$. For a graded module $M \in \mathsf{GrMod}\, A$, we define

$$\operatorname{depth} M = \inf\{i \mid \underline{\mathrm{H}}_{\mathfrak{m}}^i(M) \neq 0\} \quad \text{and} \quad \operatorname{ldim} M = \sup\{i \mid \underline{\mathrm{H}}_{\mathfrak{m}}^i(M) \neq 0\}.$$

See [Yekutieli 1992; Van den Bergh 1997] for basic properties of $\mathrm{R}\underline{\Gamma}_{\mathfrak{m}}(-)$.

If A is an $\mathbb{N}$-graded algebra with $A_0 = k$, then A is called connected graded. Note that a right noetherian connected graded algebra is locally finite.

Definition 2.3. A two-sided noetherian connected graded algebra A is called AS-Gorenstein (resp. AS-regular) of dimension d and of Gorenstein parameter ℓ if

- $\operatorname{injdim}_A A = \operatorname{injdim}_{A^{\mathrm{op}}} A = d < \infty$ (resp. $\operatorname{gldim} A = d < \infty$), and
- $\underline{\operatorname{Ext}}^i_A(k, A) \cong \underline{\operatorname{Ext}}^i_{A^{\mathrm{op}}}(k, A) \cong \begin{cases} k(\ell) & \text{if } i = d, \\ 0 & \text{if } i \neq d. \end{cases}$

For $M \in \mathsf{GrMod}\, A$ and a graded algebra automorphism $\sigma \in \operatorname{GrAut} A$, we define the twist $M_\sigma \in \mathsf{GrMod}\, A$ by $M_\sigma = M$ as a graded k-vector space with the new right action $m * a = m\sigma(a)$. Let A be an AS-Gorenstein algebra. Then it is well known that A has a balanced dualizing complex $D\,\mathrm{R}\underline{\Gamma}_{\mathfrak{m}}(A) \cong D\,\mathrm{R}\underline{\Gamma}_{\mathfrak{m}^{\mathrm{op}}}(A) \cong A_\nu(-\ell)[d]$ in $\mathsf{D}(\mathsf{GrMod}\, A^e)$ with some graded algebra automorphism $\nu \in \operatorname{GrAut} A$. This graded algebra automorphism $\nu \in \operatorname{GrAut} A$ is called the generalized Nakayama automorphism. We define $\omega_A := A_\nu(-\ell) \in \mathsf{GrMod}\, A^e$.

Definition 2.4. Let A be an AS-Gorenstein algebra of dimension d and $M \in \mathsf{grmod}\, A$. Then M is called a graded maximal Cohen–Macaulay module if $\operatorname{depth} M = \operatorname{ldim} M = d$.

Let A be an AS-Gorenstein algebra. Then A is a graded maximal Cohen–Macaulay A-module. It is well known that $M \in \mathsf{grmod}\, A$ is graded maximal Cohen–Macaulay if and only if $\underline{\operatorname{Ext}}^i_A(M, A) = 0$ for all $i \neq 0$. See [Mori 2003] for basic properties of maximal Cohen–Macaulay modules.

We write $\mathsf{CM}^{\mathsf{gr}}(A)$ for the full subcategory of $\mathsf{grmod}\, A$ consisting of graded maximal Cohen–Macaulay modules. If $M \in \mathsf{grmod}\, A$, then we define $M^\dagger = \underline{\operatorname{Hom}}_A(M, A) \in \mathsf{grmod}\, A^{\mathrm{op}}$. Similarly, if $N \in \mathsf{grmod}\, A^{\mathrm{op}}$, then we define $N^\dagger = \underline{\operatorname{Hom}}_{A^{\mathrm{op}}}(N, A) \in \mathsf{grmod}\, A$. It is well known that the contravariant functors

$$\mathsf{CM}^{\mathsf{gr}}(A) \underset{(-)^\dagger}{\overset{(-)^\dagger}{\rightleftarrows}} \mathsf{CM}^{\mathsf{gr}}(A^{\mathrm{op}}) \tag{2-1}$$

define a duality. For $M \in \mathsf{CM}^{\mathsf{gr}}(A)$, we put $B = \underline{\operatorname{End}}_A(M)$, $C = \underline{\operatorname{End}}_{A^{\mathrm{op}}}(M^\dagger)$. Then

$$C \cong \underline{\operatorname{End}}_{A^{\mathrm{op}}}(M^\dagger) \cong \underline{\operatorname{End}}_A(M)^{\mathrm{op}} \cong B^{\mathrm{op}} \tag{2-2}$$

as graded algebras by the above duality.

The following theorem, called the maximal Cohen–Macaulay approximation theorem, plays a key role in this paper.

Theorem 2.5. *Let A be an AS-Gorenstein algebra. For any $M \in \mathsf{grmod}\, A$, there exists a short exact sequence*

$$0 \to L \to Z \to M \to 0$$

in $\mathsf{grmod}\, A$ such that $Z \in \mathsf{CM}^{\mathsf{gr}}(A)$ and $\operatorname{injdim}_A L < \infty$. Moreover, $\underline{\operatorname{Ext}}^i_A(X, L) = 0$ holds for any $X \in \mathsf{CM}^{\mathsf{gr}}(A)$ and any $i \geq 1$.

Proof. This follows from [Mori 2003, Proposition 5.3] and [Ueyama 2015, Lemma 3.5]. ☐

Recently, Minamoto and Mori [2011] introduced the two notions of an $\mathbb{N}$-graded (not necessarily connected graded) AS-regular algebra.

Definition 2.6 [Minamoto and Mori 2011, Definition 3.1]. A locally finite $\mathbb{N}$-graded algebra B is called AS-regular over $R = B_0$ of dimension d and of Gorenstein parameter ℓ if

- $\operatorname{gldim} B = \operatorname{gldim} B^{\mathrm{op}} = d < \infty$, and
- $\mathrm{R}\underline{\mathrm{Hom}}_B(B_0, B) \cong \mathrm{R}\underline{\mathrm{Hom}}_{B^{\mathrm{op}}}(B_0, B) \cong (DB_0)(\ell)[-d]$ in $\mathsf{D}(\mathsf{GrMod}\, B_0)$ and in $\mathsf{D}(\mathsf{GrMod}\, {B_0}^{\mathrm{op}})$.

Remark 2.7. For the purpose of this paper, we do not require $\operatorname{gldim} B_0 < \infty$.

Definition 2.8 [Minamoto and Mori 2011, Definition 3.9]. A locally finite $\mathbb{N}$-graded algebra B is called ASF-regular of dimension d and of Gorenstein parameter ℓ if

- $\operatorname{gldim} B = \operatorname{gldim} B^{\mathrm{op}} = d < \infty$, and
- $\mathrm{R}\underline{\Gamma}_{\mathfrak{m}}(B) \cong \mathrm{R}\underline{\Gamma}_{\mathfrak{m}^{\mathrm{op}}}(B) \cong (DB)(\ell)[-d]$ in $\mathsf{D}(\mathsf{GrMod}\, B)$ and in $\mathsf{D}(\mathsf{GrMod}\, B^{\mathrm{op}})$.

By expanding the notion of an AS-regular algebra to $\mathbb{N}$-graded algebras, Minamoto and Mori [2011] gave a nice correspondence between $\mathbb{N}$-graded AS-regular algebras over R of dimension d with $\operatorname{gldim} R < \infty$ and quasi-Fano algebras of global dimension $d-1$. This result provides a strong connection between noncommutative projective geometry and representation theory of finite-dimensional algebras. See [Herschend et al. 2014; Minamoto and Mori 2011; Mori 2015] for details.

At the end of this section, we give a comparison theorem for the two notions of AS-regular algebras.

Lemma 2.9. *Let B be a two-sided noetherian ASF-regular algebra, and C a two-sided noetherian locally finite $\mathbb{N}$-graded algebra. Then for any $M \in \mathsf{GrMod}\, C^{\mathrm{op}} \otimes B$,*

$$D\,\mathrm{R}\underline{\Gamma}_{\mathfrak{m}}(M) \cong \mathrm{R}\underline{\mathrm{Hom}}_B(M, D\,\mathrm{R}\underline{\Gamma}_{\mathfrak{m}}(B))$$

in $\mathsf{D}(\mathsf{GrMod}\, B^{\mathrm{op}} \otimes C)$.

Proof. Van den Bergh [1997] gave a theory on local duality for connected graded algebras. One can check that the results in [Van den Bergh 1997, Sections 3–6] hold with no essential change for a noetherian locally finite $\mathbb{N}$-graded algebra (see also [Reyes et al. 2014, Remark 3.6; 2017, Lemma 3.2 (1)]). This follows from the locally finite $\mathbb{N}$-graded version of [Van den Bergh 1997, Theorem 5.1]. ☐

Theorem 2.10. *If B is a two-sided noetherian ASF-regular algebra of dimension d and of Gorenstein parameter ℓ, then B is an AS-regular algebra over B_0 of dimension d and of Gorenstein parameter ℓ.*

Proof. There exists an algebra automorphism $\mu \in \operatorname{GrAut} B$ such that $D\,\mathrm{R}\underline{\Gamma}_{\mathfrak{m}}(B) \cong B_\mu(-\ell)[d]$ in $\mathsf{D}(\mathsf{GrMod}\, B^e)$, so we have

$$\begin{aligned} \mathrm{R}\underline{\mathrm{Hom}}_B(B_0, B) &\cong \mathrm{R}\underline{\mathrm{Hom}}_B(B_0, B_\mu(-\ell)[d])_{\mu^{-1}}(\ell)[-d] \\ &\cong \mathrm{R}\underline{\mathrm{Hom}}_B(B_0, D\,\mathrm{R}\underline{\Gamma}_{\mathfrak{m}}(B))_{\mu^{-1}}(\ell)[-d] \\ &\cong D\,\mathrm{R}\underline{\Gamma}_{\mathfrak{m}}(B_0)_{\mu^{-1}}(\ell)[-d] \\ &\cong D(B_0)_{\mu^{-1}}(\ell)[-d] \end{aligned}$$

in $\mathsf{D}(\mathsf{GrMod}\, B^{\mathrm{op}} \otimes B_0)$, and so $\mathrm{R}\underline{\mathrm{Hom}}_B(B_0, B) \cong D(B_0)(\ell)[-d]$ in $\mathsf{D}(\mathsf{GrMod}\, B_0)$ and in $\mathsf{D}(\mathsf{GrMod}\, {B_0}^{\mathrm{op}})$. Hence the result follows. □

Corollary 2.11. *Let B be a two-sided noetherian locally finite $\mathbb{N}$-graded algebra with* $\operatorname{gldim} B_0 < \infty$. *Then B is ASF-regular if and only if it is AS-regular over B_0.*

Proof. This is a combination of Theorem 2.10 and [Minamoto and Mori 2011, Theorem 3.12]. □

3. Main result

In this section, we prove the main result (Theorem 3.10) and give an example of its use. First we introduce a condition which we require for the main result.

Definition 3.1. Let A be an AS-Gorenstein algebra with the generalized Nakayama automorphism $\nu \in \operatorname{GrAut} A$. Then $X \in \mathsf{GrMod}\, A$ is called ν-stable if $X_\nu \cong X$ as graded right A-modules.

Since $A_\nu \cong A$ in $\mathsf{GrMod}\, A$, A is always ν-stable. Clearly, if A is symmetric, that is, the generalized Nakayama automorphism of A is the identity, then every $M \in \mathsf{GrMod}\, A$ is ν-stable.

Example 3.2. Let S be an AS-regular algebra and G a finite subgroup of $\operatorname{GrAut} S$ such that $\operatorname{char} k$ does not divide $|G|$. If S^G is AS-Gorenstein, then $S \in \mathsf{GrMod}\, S^G$ is ν-stable by [Ueyama 2013, Lemma 5.8].

Lemma 3.3. *Let A be an AS-Gorenstein algebra and let $X \in \mathsf{CM}^{\mathrm{gr}}(A)$. If X is ν-stable, then $X \cong X_{\nu^{-1}}$ in* $\mathsf{GrMod}\, A$ *and $X^\dagger$ is ν-stable in* $\mathsf{GrMod}\, A^{\mathrm{op}}$.

Proof. It is easy to check that $X \cong X_{\nu\nu^{-1}} \cong X_{\nu^{-1}}$ in $\mathsf{GrMod}\, A$, and

$${}_\nu(X^\dagger) \cong \underline{\mathrm{Hom}}_A(X, {}_\nu A) \cong \underline{\mathrm{Hom}}_A(X, A_{\nu^{-1}}) \cong \underline{\mathrm{Hom}}_A(X_\nu, A) \cong \underline{\mathrm{Hom}}_A(X, A) \cong X^\dagger$$

in $\mathsf{GrMod}\, A^{\mathrm{op}}$. □

The notion of an n-cluster tilting module plays an important role in representation theory of orders, especially higher-dimensional analogues of Auslander–Reiten theory. It can be regarded as a natural generalization of the classical notion of Cohen–Macaulay representation-finiteness.

Definition 3.4. Let A be an AS-Gorenstein algebra. For a positive integer $n \in \mathbb{N}^+$, a graded maximal Cohen–Macaulay module $X \in \mathsf{CM}^{\mathrm{gr}}(A)$ is called n-cluster tilting if

$$\begin{aligned}\mathsf{add}\{X(i) \mid i \in \mathbb{Z}\} &= \{M \in \mathsf{CM}^{\mathrm{gr}}(A) \mid \underline{\mathrm{Ext}}^i_A(X, M) = 0 \text{ for } 0 < i < n\} \\ &= \{M \in \mathsf{CM}^{\mathrm{gr}}(A) \mid \underline{\mathrm{Ext}}^i_A(M, X) = 0 \text{ for } 0 < i < n\},\end{aligned}$$

where $\mathsf{add}\{X(i) \mid i \in \mathbb{Z}\}$ is the full subcategory of $\mathsf{grmod}\, A$ consisting of direct summands of finite direct sums of shifts of X.

Let A be a noetherian locally finite $\mathbb{N}$-graded algebra. Then it is known that $\mathsf{grmod}\, A$ has the Krull–Schmidt property, i.e., each finitely generated graded module is a direct sum of a uniquely determined set of indecomposable graded modules. Recall that $M \in \mathsf{grmod}\, A$ is called basic if each indecomposable direct summand occurs exactly once (up to isomorphism and degree shift of grading) in a direct sum decomposition. The following proposition says that if A has a $(d-1)$-cluster tilting module and $\mathrm{gldim}(\mathsf{tails}\, A) < \infty$, then it has a ν-stable $(d-1)$-cluster tilting module.

Proposition 3.5. *Let A be an AS-Gorenstein algebra of dimension $d \geq 2$, Gorenstein parameter ℓ, and $\mathrm{gldim}(\mathsf{tails}\, A) < \infty$. If $X \in \mathsf{CM}^{\mathrm{gr}}(A)$ is a basic $(d-1)$-cluster tilting module, then X is ν-stable.*

Proof. By [Ueyama 2013, Corollary 4.5], we see that the stable category $\underline{\mathsf{CM}}^{\mathrm{gr}}(A)$ has the Serre functor $- \otimes_A A_\nu(-\ell)[d-1]$. Since

$$\underline{\mathrm{Ext}}^i_A(X, X) \cong \bigoplus_{s \in \mathbb{Z}} \mathrm{Hom}_{\underline{\mathsf{CM}}^{\mathrm{gr}}(A)}(X, X(s)[i]) = 0$$

for any $0 < i < d-1$, we have $\underline{\mathrm{Ext}}^i_A(X, X_\nu) = 0$ for any $0 < i < d-1$ by using the Serre functor of $\underline{\mathsf{CM}}^{\mathrm{gr}}(A)$, so $X_\nu \in \mathsf{add}\{X(i) \mid i \in \mathbb{Z}\}$. Since X is basic, X_ν is also basic, so it follows that X_ν is a direct summand of X. Similarly, we can show that $X_{\nu^{-1}}$ is a direct summand of X, so X is a direct summand of X_ν. Hence the result follows. $\square$

Next we prepare some lemmas which we need to prove the main result.

Lemma 3.6. *Let A be a graded algebra and X a graded right A-module containing A as a direct summand. Then X is a finitely generated graded left projective module over $\underline{\mathrm{End}}_A(X)$. Moreover, for any $M \in \mathsf{GrMod}\, A$, there is a natural isomorphism*

$$M \cong \underline{\mathrm{Hom}}_A(X, M) \otimes_{\underline{\mathrm{End}}_A(X)} X$$

in $\mathsf{GrMod}\, A$.

Proof. Put $B := \underline{\mathrm{End}}_A(X)$. Since $X \cong A \oplus Y$ for some $Y \in \mathsf{GrMod}\, A$, we have

$$B = \underline{\mathrm{Hom}}_A(X, X) \cong \underline{\mathrm{Hom}}_A(A, X) \oplus \underline{\mathrm{Hom}}_A(Y, X) \cong X \oplus \underline{\mathrm{Hom}}_A(Y, X)$$

in $\mathsf{GrMod}\, B^{\mathrm{op}}$, so X is finitely generated graded left projective over B. Moreover, one can verify that $\underline{\mathrm{End}}_{B^{\mathrm{op}}}(X) \cong A$ as graded algebras. Thus, for any $M \in \mathsf{GrMod}\, A$, we have

$$\underline{\mathrm{Hom}}_A(X, M) \otimes_B X \cong \underline{\mathrm{Hom}}_A(\underline{\mathrm{End}}_{B^{\mathrm{op}}}(X), M) \cong \underline{\mathrm{Hom}}_A(A, M) \cong M$$

as graded right A-modules. □

Lemma 3.7. *Let A be an AS-Gorenstein algebra, and $X \in \mathsf{CM}^{\mathrm{gr}}(A)$ such that X contains A as a direct summand. If $\underline{\mathrm{End}}_A(X)$ is right noetherian and $M \in \mathsf{grmod}\, A$, then $\underline{\mathrm{Hom}}_A(X, M)$ is a finitely generated graded right $\underline{\mathrm{End}}_A(X)$-module.*

Proof. Put $B := \underline{\mathrm{End}}_A(X)$. By Theorem 2.5, we have an exact sequence

$$0 \to \underline{\mathrm{Hom}}_A(X, L) \to \underline{\mathrm{Hom}}_A(X, Z) \to \underline{\mathrm{Hom}}_A(X, M) \to 0$$

in $\mathsf{GrMod}\, B$, where $Z \in \mathsf{CM}^{\mathrm{gr}}(A)$ and $\mathrm{injdim}_A L < \infty$, so it is enough to show that $\underline{\mathrm{Hom}}_A(X, Z)$ is finitely generated. Since A is AS-Gorenstein and $Z \in \mathsf{CM}^{\mathrm{gr}}(A)$, we have a graded monomorphism $Z \to F$ in $\mathsf{GrMod}\, A$, where F is a finitely generated free A-module. Now A is a direct summand of X, so there exist graded monomorphisms $Z \to \widehat{X}$ in $\mathsf{GrMod}\, A$ and $\underline{\mathrm{Hom}}_A(X, Z) \to \underline{\mathrm{Hom}}_A(X, \widehat{X})$ in $\mathsf{GrMod}\, B$, where $\widehat{X}$ is a finite direct sum of shifts of X. Since $\underline{\mathrm{Hom}}_A(X, \widehat{X})$ is finitely generated and B is right noetherian, $\underline{\mathrm{Hom}}_A(X, Z)$ is also finitely generated. □

Lemma 3.8. *Let B be a two-sided noetherian locally finite $\mathbb{N}$-graded algebra. For any $M \in \mathsf{GrMod}\, B^e$,*

$$\mathrm{R}\underline{\Gamma}_{\mathfrak{m}}(\mathrm{R}\underline{\Gamma}_{\mathfrak{m}^{\mathrm{op}}}(M)) \cong \mathrm{R}\underline{\Gamma}_{\mathfrak{m}^{\mathrm{op}}}(\mathrm{R}\underline{\Gamma}_{\mathfrak{m}}(M))$$

in $\mathsf{D}(\mathsf{GrMod}\, B^e)$.

Proof. One can show the locally finite $\mathbb{N}$-graded version of [Van den Bergh 1997, Lemma 4.5], so the assertion holds. □

Lemma 3.9. *Let B be a two-sided noetherian ASF-regular algebra with an idempotent e. If $M \in \mathsf{grmod}\, B$ is finite-dimensional over k such that $Me = 0$, then $\underline{\mathrm{Ext}}^i_B(M, eB) = 0$ for any i.*

Proof. Since B is ASF-regular, we have

$$\begin{aligned} \mathrm{R}\underline{\mathrm{Hom}}_B(M, eB) &\cong e\,\mathrm{R}\underline{\mathrm{Hom}}_B(M, D\,\mathrm{R}\underline{\Gamma}_{\mathfrak{m}}(B)(\ell)[-d]) \\ &\cong e\,\mathrm{R}\underline{\mathrm{Hom}}_B(M, D\,\mathrm{R}\underline{\Gamma}_{\mathfrak{m}}(B))(\ell)[-d]. \end{aligned}$$

Moreover, by Lemma 2.9,

$$e\,\mathrm{R}\underline{\mathrm{Hom}}_B(M, D\,\mathrm{R}\underline{\Gamma}_{\mathfrak{m}}(B))(\ell)[-d] \cong eD\,\mathrm{R}\underline{\Gamma}_{\mathfrak{m}}(M)(\ell)[-d].$$

Since $M \in \mathsf{grmod}\, B$ is finite-dimensional over k, $\mathrm{R}\underline{\Gamma}_{\mathfrak{m}}(M) \cong M$, so we have $\underline{\mathrm{Ext}}^i_B(M, eB) \cong eD\,\underline{\mathrm{H}}^{d-i}_{\mathfrak{m}}(M)(\ell) = 0$ for all $i \neq d$, and $\underline{\mathrm{Ext}}^d_B(M, eB) \cong eD(M)(\ell) \cong D(Me)(\ell) = 0$ because $Me = 0$. □

Now we are ready to prove the main result of this paper.

Theorem 3.10. *Let A be a two-sided noetherian connected graded algebra satisfying χ on both sides, and let $X \in \mathsf{grmod}\, A$. We consider the following conditions*:

(A) *A and X satisfy*

- (A1) *A is an AS-Gorenstein algebra of dimension $d \geq 2$,*
- (A2) *X is a $(d-1)$-cluster tilting module,*
- (A3) $\underline{\mathrm{End}}_A(X)_{<0} = 0$, *and*
- (A4) $\underline{\mathrm{Ext}}^1_A(X, M)$ *and* $\underline{\mathrm{Ext}}^1_A(M, X)$ *are finite-dimensional over k for any* $M \in \mathsf{CM}^{\mathsf{gr}}(A)$.

(B) $B := \underline{\mathrm{End}}_A(X)$ *satisfies*

- (B1) *B is a two-sided noetherian ASF-regular algebra of dimension $d \geq 2$,*
- (B2) $- \otimes_B X$ *induces an equivalence functor* $\mathsf{tails}\, B \xrightarrow{\sim} \mathsf{tails}\, A$, *and*
- (B3) $X^\dagger \otimes_B -$ *induces an equivalence functor* $\mathsf{tails}\, B^{\mathrm{op}} \xrightarrow{\sim} \mathsf{tails}\, A^{\mathrm{op}}$.

Then:

(1) *If* (A) *is fulfilled and X is either ν-stable or basic*, *then* (B) *holds.*

(2) *If* (B) *is fulfilled and X contains A as a direct summand*, *then* (A) *holds.*

Proof of (1) *in Theorem 3.10.* Suppose that A and X satisfy (A), and X is either ν-stable or basic. Since A is indecomposable, (A2) implies that X is a graded maximal Cohen–Macaulay A-module containing a shift of A as a direct summand. By properties of degree shifts, we may assume that X contains A as a direct summand without loss of generality.

The proof is divided into several steps:

(a) $(\mathcal{X}, (1))$ is ample for $\mathsf{tails}\, A$,

(b) B is right noetherian, and (B2) holds,

(c) $\mathrm{gldim}\, B = d$,

(d) $\underline{\mathrm{H}}^i_{\mathfrak{m}}(B) = 0$ for any $i \neq d$,

(e) $(\mathcal{X}^\dagger, (1))$ is ample for $\mathsf{tails}\, A^{\mathrm{op}}$,

(f) B is left noetherian, and (B3) holds,

(g) $\mathrm{gldim}\, B^{\mathrm{op}} = d$,

(h) $\underline{\mathrm{H}}^i_{\mathfrak{m}^{\mathrm{op}}}(B) = 0$ for any $i \neq d$, and

(i) $\underline{\mathrm{H}}^d_{\mathfrak{m}}(B) \cong \underline{\mathrm{H}}^d_{\mathfrak{m}^{\mathrm{op}}}(B) \cong (DB)(\ell)$ in $\mathsf{GrMod}\, B$ and in $\mathsf{GrMod}\, B^{\mathrm{op}}$.

Proof of (a). We show that $(\mathcal{X}, (1))$ is ample for $\mathsf{tails}\, A$. Since A is a direct summand of X, it is easy to check that the condition (Am1) is satisfied. Let $f : \mathcal{M} \to \mathcal{N}$ be an epimorphism in $\mathsf{tails}\, A$. It gives a short exact sequence $0 \to \mathcal{K} \to \mathcal{M} \to \mathcal{N} \to 0$ in $\mathsf{tails}\, A$. We take a finitely generated graded module K such that $\pi K = \mathcal{K}$. By Theorem 2.5, there exists an exact sequence

$$0 \to L \to Z \to K \to 0$$

in $\mathsf{grmod}\, A$ such that $Z \in \mathsf{CM}^{\mathsf{gr}}(A)$ and $\operatorname{injdim}_A L < \infty$. Furthermore, Theorem 2.5 also implies that $\underline{\operatorname{Ext}}^1_A(X, K) \cong \underline{\operatorname{Ext}}^1_A(X, Z)$, so $\underline{\operatorname{Ext}}^1_A(X, K)$ is finite-dimensional over k by (A4). By [Artin and Zhang 1994, Corollary 7.3 (2)], it follows that $\underline{\operatorname{Ext}}^1_{\mathcal{A}}(\mathcal{X}, \mathcal{K})$ is right bounded; thus we have $\operatorname{Ext}^1_{\mathcal{A}}(\mathcal{X}(-n), \mathcal{K}) = \underline{\operatorname{Ext}}^1_{\mathcal{A}}(\mathcal{X}, \mathcal{K})_n = 0$ for all $n \gg 0$. Since

$$\operatorname{Hom}_{\mathcal{A}}(\mathcal{X}(-n), \mathcal{M}) \to \operatorname{Hom}_{\mathcal{A}}(\mathcal{X}(-n), \mathcal{N}) \to \operatorname{Ext}^1_{\mathcal{A}}(\mathcal{X}(-n), \mathcal{K})$$

is exact, $\operatorname{Hom}_{\mathcal{A}}(\mathcal{X}(-n), \mathcal{M}) \to \operatorname{Hom}_{\mathcal{A}}(\mathcal{X}(-n), \mathcal{N})$ is surjective for all $n \gg 0$, so the condition (Am2) is also satisfied; hence $(\mathcal{X}, (1))$ is ample for $\mathsf{tails}\, A$.

Proof of (b). We show that B is right noetherian and (B2) is satisfied. The basic idea of the proof comes from [Mori and Ueyama 2016a, Section 2]. Since $\operatorname{depth}_A X = d \geq 2$, we have

$$B = \underline{\operatorname{End}}_A(X) = B(\mathsf{grmod}\, A, X, (1)) \cong B(\mathsf{tails}\, A, \mathcal{X}, (1))$$

by [Mori 2013, Lemma 3.3]. By using (A3) and Theorem 2.2, it follows that $B = B_{\geq 0} \cong B(\mathsf{tails}\, A, \mathcal{X}, (1))_{\geq 0}$ is right noetherian locally finite. Moreover, the functor

$$F := \pi \circ \underline{\operatorname{Hom}}_{\mathcal{A}}(\mathcal{X}, -) : \mathsf{tails}\, A \to \mathsf{tails}\, B$$

is an equivalence. For $M \in \mathsf{grmod}\, A$, there exists $n \in \mathbb{Z}$ such that

$$\underline{\operatorname{Hom}}_{\mathcal{A}}(\mathcal{X}, \mathcal{M})_{\geq n} \cong \underline{\operatorname{Hom}}_A(X, M)_{\geq n}$$

in $\mathsf{grmod}\, B$ by [Artin and Zhang 1994, Corollary 7.3 (2)], so the functor F is induced by the functor $\underline{\operatorname{Hom}}_A(X, -) : \mathsf{grmod}\, A \to \mathsf{grmod}\, B$. By using Lemma 3.6, we see that the functor $\mathsf{tails}\, B \to \mathsf{tails}\, A$ induced by $- \otimes_B X : \mathsf{grmod}\, B \to \mathsf{grmod}\, A$ is an equivalence functor quasi-inverse to F.

Proof of (c). Here we show that $\operatorname{gldim} B = d$. First let us explain that we can construct a graded right $\mathsf{add}\{X(i) \mid i \in \mathbb{Z}\}$-approximation of $M \in \mathsf{grmod}\, A$. Since $\underline{\operatorname{Hom}}_A(X, M)$ is a finitely generated graded right B-module by Lemma 3.7, we can take $f_i \in \operatorname{Hom}_{\mathsf{GrMod}\, A}(X(s_i), M)$ such that $f_1, \dots, f_n$ generate $\underline{\operatorname{Hom}}_A(X, M)$. Thus for any $f \in \operatorname{Hom}_{\mathsf{GrMod}\, A}(X(t), M)$, there exist graded homomorphisms $g_i \in$

$\mathrm{Hom}_{\mathsf{GrMod}\,A}(X(t), X(s_i))$ such that

$$\begin{tikzcd} X(s_1)\oplus\cdots\oplus X(s_n) \arrow[rr, "(f_1\ \cdots\ f_n)"] & & M \\ & X(t) \arrow[ul, "\begin{pmatrix} g_1 \\ \vdots \\ g_n \end{pmatrix}"] \arrow[ur, "f"'] & \end{tikzcd}$$

commutes. We can check that $\phi := (f_1\ \ldots\ f_n) : \bigoplus_{i=1}^n X(s_i) \to M$ is surjective, and

$$\underline{\mathrm{Hom}}_A\Big(X, \bigoplus_{i=1}^n X(s_i)\Big) \xrightarrow{\underline{\mathrm{Hom}}(X,\phi)} \underline{\mathrm{Hom}}_A(X, M)$$

is also surjective.

By the above arguments, the proof of gldim $B \le d$ is along the same lines as that of [Dao and Huneke 2013, Theorem 3.6] (see also [Ueyama 2013, Theorem 5.10]). Let $N \in \mathsf{grmod}\, B$ and take a projective presentation $P_1 \to P_0 \to N \to 0$. Since we can write $P_i \cong \underline{\mathrm{Hom}}_A(X, X_i)$, where $X_i \in \mathsf{add}\{X(i) \mid i \in \mathbb{Z}\}$ for each i, we have an exact sequence

$$0 \to \underline{\mathrm{Hom}}_A(X, M_1) \to \underline{\mathrm{Hom}}_A(X, X_1) \to \underline{\mathrm{Hom}}_A(X, X_0) \to N \to 0$$

in $\mathsf{grmod}\, B$ such that $0 \to M_1 \to X_1 \to X_0$ is exact in $\mathsf{grmod}\, A$. By using a graded right $\mathsf{add}\{X(i) \mid i \in \mathbb{Z}\}$-approximation, we have a module $X_2 \in \mathsf{add}\{X(i) \mid i \in \mathbb{Z}\}$ and a surjection $X_2 \to M_1$ such that $\underline{\mathrm{Hom}}_A(X, X_2) \to \underline{\mathrm{Hom}}_A(X, M_1)$ is also surjective. Let M_2 be the kernel of $X_2 \to M_1$. Then it is easy to see that $\underline{\mathrm{Ext}}^1_A(X, M_2) = 0$. Continuing in this way inductively, we can make exact sequences

$$0 \to M_{d-1} \to X_{d-1} \to \cdots \to X_2 \to X_1 \xrightarrow{\xi} X_0 \to \mathrm{Coker}\,\xi \to 0$$

in $\mathsf{grmod}\, A$, and

$$0 \to \underline{\mathrm{Hom}}_A(X, M_{d-1}) \to P_{d-1} \to \cdots \to P_2 \to P_1 \to P_0 \to N \to 0$$

in $\mathsf{grmod}\, B$, where $P_i = \underline{\mathrm{Hom}}_A(X, X_i)$. Furthermore we see $\underline{\mathrm{Ext}}^j_A(X, M_i) = 0$ for any $2 \le i \le d-1$ and any $0 < j < i$. If $M_{d-1} = 0$, then clearly $\mathrm{projdim}_B\, N \le d-1$. We now consider the case $M_{d-1} \ne 0$. Since $X_i \in \mathsf{CM}^{\mathsf{gr}}(A)$, it follows from the depth lemma (see [Bruns and Herzog 1998, Proposition 1.2.9]) that M_{d-1} is graded maximal Cohen–Macaulay. Moreover, since $\underline{\mathrm{Ext}}^j_A(X, M_{d-1}) = 0$ for $0 < j < d-1$, it follows that $M_{d-1} \in \mathsf{add}\{X(i) \mid i \in \mathbb{Z}\}$ by (A2). Thus we obtain $\mathrm{projdim}_B\, N \le d$.

To see that gldim $B = d$, consider a graded projective resolution of $\underline{\mathrm{Hom}}_A(X, k)$ in $\mathsf{grmod}\, B$, namely,

$$\cdots \to P_2 \to P_1 \to P_0 \to \underline{\mathrm{Hom}}_A(X, k) \to 0.$$

Applying $-\otimes_B X$ to the above exact sequence, we have an exact sequence

$$\cdots \to P_2 \otimes_B X \to P_1 \otimes_B X \to P_0 \otimes_B X \to \underline{\mathrm{Hom}}_A(X,k) \otimes_B X \to 0$$

in $\mathsf{grmod}\, A$. By Lemma 3.6, $\underline{\mathrm{Hom}}_A(X,k)\otimes_B X \cong k$. Since each $P_i \otimes_B X$ is a direct summand of finite direct sums of shifts of X, it is graded maximal Cohen–Macaulay, so $\mathrm{projdim}_B\, \underline{\mathrm{Hom}}_A(X,k) \le d-1$ gives a contradiction to the fact that $\mathrm{depth}_A\, k = 0$. Thus $\mathrm{gldim}\, B = d$.

Proof of (d). We next show that $\underline{\mathrm{H}}^i_{\mathfrak{m}}(B) = 0$ for $i \neq d$. By the arguments in the proof of (a) and (b), we see that

$$\begin{tikzcd}
(\mathsf{grmod}\, A, X, (1)) \arrow[r, "\pi"] & (\mathsf{tails}\, A, \mathcal{X}, (1)) \arrow[d, "\cong"', "F"] \\
(\mathsf{grmod}\, B, B, (1)) \arrow[r, "\pi"] \arrow[u, "-\otimes_B X"] & (\mathsf{tails}\, B, \mathcal{B}, (1))
\end{tikzcd}$$

commutes, and the graded algebra homomorphism

$$\underline{\mathrm{End}}_B(B) = B(\mathsf{grmod}\, B, B, (1)) \to B(\mathsf{tails}\, B, \mathcal{B}, (1)) = \underline{\mathrm{End}}_{\mathcal{B}}(\mathcal{B})$$

induced by the natural functor π is an isomorphism. This says that the natural map φ appearing in the exact sequence

$$0 \to \underline{\mathrm{H}}^0_{\mathfrak{m}}(B) \to \underline{\mathrm{End}}_B(B) \xrightarrow{\varphi} \underline{\mathrm{End}}_{\mathcal{B}}(\mathcal{B}) \to \underline{\mathrm{H}}^1_{\mathfrak{m}}(B) \to 0$$

in [Artin and Zhang 1994, Proposition 7.2 (2)] is an isomorphism. Thus $\underline{\mathrm{H}}^0_{\mathfrak{m}}(B) = \underline{\mathrm{H}}^1_{\mathfrak{m}}(B) = 0$.

Since we already have the equivalence functor in (B2), it follows that

$$\underline{\mathrm{Ext}}^i_{\mathcal{B}}(\mathcal{B},\mathcal{B}) \cong \underline{\mathrm{Ext}}^i_{\mathcal{A}}(\mathcal{X},\mathcal{X}) \tag{3-1}$$

for any i. Using $\mathrm{depth}_A\, X = d$ and (A2), we have $\underline{\mathrm{Ext}}^i_{\mathcal{A}}(\mathcal{X},\mathcal{X}) \cong \underline{\mathrm{Ext}}^i_A(X,X) = 0$ for any $1 \le i \le d-2$, so $\underline{\mathrm{Ext}}^i_{\mathcal{B}}(\mathcal{B},\mathcal{B}) = 0$ for any $1 \le i \le d-2$. Thus $\underline{\mathrm{H}}^i_{\mathfrak{m}}(B) = 0$ for $2 \le i \le d-1$ by [Artin and Zhang 1994, Theorem 7.2 (2)]. Furthermore B has global dimension d by (c), so $\underline{\mathrm{H}}^i_{\mathfrak{m}}(B) = 0$ for $i \ge d+1$.

Proof of (e). By the duality (2-1), we see that (A) is equivalent to the following:

(A^{op}) A^{op} and $X^\dagger$ satisfy

- ($\mathrm{A1}^{\mathrm{op}}$) A^{op} is an AS-Gorenstein algebra of dimension $d \ge 2$,
- ($\mathrm{A2}^{\mathrm{op}}$) $X^\dagger$ is a $(d-1)$-cluster tilting module,
- ($\mathrm{A3}^{\mathrm{op}}$) $\underline{\mathrm{End}}_{A^{\mathrm{op}}}(X^\dagger)_{<0} = 0$, and
- ($\mathrm{A4}^{\mathrm{op}}$) $\underline{\mathrm{Ext}}^1_{A^{\mathrm{op}}}(N, X^\dagger)$ and $\underline{\mathrm{Ext}}^1_{A^{\mathrm{op}}}(X^\dagger, N)$ are finite-dimensional over k for any $N \in \mathsf{CM}^{\mathsf{gr}}(A^{\mathrm{op}})$.

Hence the same argument as that in the proof of (a) implies that $(X^\dagger, (1))$ is ample for $\mathsf{tails}\, A^{\mathrm{op}}$.

Proof of (f). Since $(X^\dagger, (1))$ is ample for $\mathsf{tails}\, A^{\mathrm{op}}$, it follows from the same argument as that in the proof of (b) that $C := \underline{\mathrm{End}}_{A^{\mathrm{op}}}(X^\dagger)$ is right noetherian and $- \otimes_C X^\dagger$ induces an equivalence functor $\mathsf{tails}\, C \xrightarrow{\sim} \mathsf{tails}\, A^{\mathrm{op}}$. However, $B^{\mathrm{op}} \cong C$ as graded algebras by (2-2), so we obtain that B is left noetherian and $X^\dagger \otimes_B -$ induces an equivalence functor $\mathsf{tails}\, B^{\mathrm{op}} \xrightarrow{\sim} \mathsf{tails}\, A^{\mathrm{op}}$.

Proof of (g) *and* (h). By the arguments in the proofs of (e) and (f), the proof of (g) (resp. (h)) is obtained in the same way as that of (c) (resp. (d)).

Proof of (i). By (b) and (c), $\mathsf{tails}\, A \cong \mathsf{tails}\, B$ and $\operatorname{gldim} B < \infty$, so it follows that $\mathsf{tails}\, A$ has finite global dimension. Thus the derived category $\mathsf{D}^{\mathsf{b}}(\mathsf{tails}\, A)$ has the Serre functor $- \otimes_{\mathcal{A}} \omega_{\mathcal{A}}[d-1] \cong - \otimes_{\mathcal{A}} \mathcal{A}_\nu(-\ell)[d-1]$ by [de Naeghel and Van den Bergh 2004, Theorem A.4]. Moreover, if X is basic, then it is ν-stable by Proposition 3.5. Using the equivalence in (B2) and the fact that X is ν-stable, for any $\mathcal{M} = \pi M \in \mathsf{tails}\, B$, we have natural isomorphisms

$$\begin{aligned}
\underline{\mathrm{Hom}}_{\mathcal{B}}(\mathcal{M}, \mathcal{B}(-\ell)) &\cong \underline{\mathrm{Hom}}_{\mathcal{A}}(\pi(M \otimes_B X), \pi X(-\ell)) \\
&\cong \underline{\mathrm{Hom}}_{\mathcal{A}}(\pi(M \otimes_B X_{\nu^{-1}}), \pi X_{\nu^{-1}}(-\ell)) \\
&\cong \underline{\mathrm{Hom}}_{\mathcal{A}}(\pi(M \otimes_B X_{\nu^{-1}}), \pi X(-\ell)) \qquad \text{(by Lemma 3.3)} \\
&\cong D\, \underline{\mathrm{Ext}}^{d-1}_{\mathcal{A}}(\pi X(-\ell), \pi(M \otimes_B X_{\nu^{-1}} \otimes_A A_\nu(-\ell))) \\
&\cong D\, \underline{\mathrm{Ext}}^{d-1}_{\mathcal{A}}(\pi X(-\ell), \pi(M \otimes_B X)(-\ell)) \\
&\cong D\, \underline{\mathrm{Ext}}^{d-1}_{\mathcal{B}}(\mathcal{B}, \mathcal{M})
\end{aligned}$$

as graded k-vector spaces. It follows that $\mathcal{B}(-\ell)$ is a dualizing sheaf of $\mathsf{tails}\, B$ in the sense of [Yekutieli and Zhang 1997]. By [Artin and Zhang 1994, Proposition 7.10 (3)], it is easy to check that $\operatorname{cd}(\mathsf{tails}\, B) = d-1$. Moreover, by Theorem 2.2, we see that B is a right noetherian locally finite graded algebra satisfying χ_1, so it follows that

$$\mathrm{Hom}_{\mathcal{B}}(\mathcal{B}, \mathcal{B}(-\ell)) \cong D\, \underline{\mathrm{Ext}}^{d-1}_{\mathcal{B}}(\mathcal{B}, \mathcal{B})$$

in $\mathsf{GrMod}\, B$ by the proof of [Yekutieli and Zhang 1997, Theorem 2.3 (2)]. Hence we obtain

$$B(-\ell) \cong \mathrm{Hom}_{\mathcal{B}}(\mathcal{B}, \mathcal{B})(-\ell) \cong D\, \underline{\mathrm{Ext}}^{d-1}_{\mathcal{B}}(\mathcal{B}, \mathcal{B}) \cong D\, \underline{\mathrm{H}}^d_{\mathfrak{m}}(B) \tag{3-2}$$

in $\mathsf{GrMod}\, B$ by (d). Dually, (f), (g), (h), and Lemma 3.3 imply

$$B(-\ell) \cong D\, \underline{\mathrm{H}}^d_{\mathfrak{m}^{\mathrm{op}}}(B) \tag{3-3}$$

in $\mathsf{GrMod}\, B^{\mathrm{op}}$. In addition, we see that $\underline{\mathrm{H}}^d_{\mathfrak{m}}(B)$ and $\underline{\mathrm{H}}^d_{\mathfrak{m}^{\mathrm{op}}}(B)$ are right bounded graded B^e-modules, so it follows from Lemma 3.8 that

$$\begin{aligned}\underline{\mathrm{H}}^d_{\mathfrak{m}^{\mathrm{op}}}(B) &\cong \mathrm{R}\underline{\Gamma}_{\mathfrak{m}}(\underline{\mathrm{H}}^d_{\mathfrak{m}^{\mathrm{op}}}(B)) \cong \mathrm{R}\underline{\Gamma}_{\mathfrak{m}}(\underline{\mathrm{H}}^d_{\mathfrak{m}^{\mathrm{op}}}(B)[-d])[d] \\ &\cong \mathrm{R}\underline{\Gamma}_{\mathfrak{m}}(\mathrm{R}\underline{\Gamma}_{\mathfrak{m}^{\mathrm{op}}}(B))[d] \cong \mathrm{R}\underline{\Gamma}_{\mathfrak{m}^{\mathrm{op}}}(\mathrm{R}\underline{\Gamma}_{\mathfrak{m}}(B))[d] \\ &\cong \mathrm{R}\underline{\Gamma}_{\mathfrak{m}^{\mathrm{op}}}(\underline{\mathrm{H}}^d_{\mathfrak{m}}(B)[-d])[d] \cong \mathrm{R}\underline{\Gamma}_{\mathfrak{m}^{\mathrm{op}}}(\underline{\mathrm{H}}^d_{\mathfrak{m}}(B)) \cong \underline{\mathrm{H}}^d_{\mathfrak{m}}(B)\end{aligned}$$

in $\mathsf{D}(\mathsf{GrMod}\, B^e)$. Therefore our assertion follows from (3-2) and (3-3).

Hence the proof of Theorem 3.10 (1) is now complete. □

Proof of (2) *in Theorem 3.10.* Suppose that X satisfies (B), and contains A as a direct summand. Clearly (A3) is satisfied by (B1).

First we show that (A1) holds. Since A is a direct summand of X, there exists an idempotent $e \in B$ such that $eBe \cong \underline{\mathrm{End}}_A(A) \cong A$ as graded algebras and $Be \cong \underline{\mathrm{Hom}}_A(A, X) \cong X$ as graded B-A bimodules. Then (B2) says that the functor $\mathsf{tails}\, B \to \mathsf{tails}\, eBe$ induced by $- \otimes_B Be$ is an equivalence, so it follows that $B/(e)$ is finite-dimensional over k by [Mori and Ueyama 2016b, Lemma 3.17]. Let ρ be the composition map

$$Be \otimes^{\mathrm{L}}_{eBe} eB \xrightarrow[\text{nat.}]{} Be \otimes_{eBe} eB \xrightarrow[\text{mult.}]{} B.$$

We have the triangle

$$Be \otimes^{\mathrm{L}}_{eBe} eB \xrightarrow{\rho} B \to C \to Be \otimes^{\mathrm{L}}_{eBe} eB[-1]$$

in $\mathsf{D}(\mathsf{grmod}\, B)$. Applying $- \otimes^{\mathrm{L}}_B Be$ implies $C \otimes^{\mathrm{L}}_B Be = 0$. It follows that $h^i(C)e = 0$ for all i, and thus $\underline{\mathrm{Ext}}^j_B(h^i(C), eB) = 0$ for all i, j by Lemma 3.9. By using a hypercohomology spectral sequence [Weibel 1994, 5.7.9], we obtain for all p that $h^p(\mathrm{R}\underline{\mathrm{Hom}}_B(C, eB)) = 0$, so $\mathrm{R}\underline{\mathrm{Hom}}_B(C, eB) = 0$. Applying $\mathrm{R}\underline{\mathrm{Hom}}_B(-, eB)$ to the above triangle induces

$$\mathrm{R}\underline{\mathrm{Hom}}_B(Be \otimes^{\mathrm{L}}_{eBe} eB, eB) \cong \mathrm{R}\underline{\mathrm{Hom}}_B(B, eB) \cong eB.$$

On the other hand, we have

$$\begin{aligned}\mathrm{R}\underline{\mathrm{Hom}}_B(Be \otimes^{\mathrm{L}}_{eBe} eB, eB) &\cong \mathrm{R}\underline{\mathrm{Hom}}_{eBe}(Be, \mathrm{R}\underline{\mathrm{Hom}}_B(eB, eB)) \\ &\cong \mathrm{R}\underline{\mathrm{Hom}}_{eBe}(Be, eBe),\end{aligned}$$

so it follows that

$$\underline{\mathrm{Ext}}^i_A(X, A) \cong \underline{\mathrm{Ext}}^i_{eBe}(Be, eBe) = 0 \tag{3-4}$$

for all $i > 0$.

Let $M \in \mathsf{grmod}\, A$. Since $\mathrm{gldim}\, B = d$, we can take a projective resolution

$$0 \to P_m \to \cdots \to P_2 \to P_1 \to P_0 \to M \otimes_{eBe} eB \to 0$$

in $\mathsf{grmod}\, B$ with $m \le d$. Applying $- \otimes_B Be$ to the above exact sequence, we have an exact sequence

$$0 \to P_m \otimes_B Be \to \cdots \to P_2 \otimes_B Be \to P_1 \otimes_B Be \to P_0 \otimes_B Be \to M \to 0.$$

Since each $P_j \otimes_B Be \in \mathsf{add}\{Be(i) \mid i \in \mathbb{Z}\} \cong \mathsf{add}\{X(i) \mid i \in \mathbb{Z}\}$, by (3-4), we see that $\underline{\mathrm{Ext}}^{m+1}_A(M, A) = 0$. Thus $\mathrm{injdim}_A A \le d$. By (B1), (B2), and (B3), $\mathrm{gldim}(\mathsf{tails}\, A) < \infty$ and $\mathrm{gldim}(\mathsf{tails}\, A^{\mathrm{op}}) < \infty$. Moreover, A satisfies χ on both sides, so it has a dualizing complex by [Van den Bergh 1997, Theorem 6.3]. It follows from [Dong and Wu 2009, Theorem 3.6] that A is AS-Gorenstein. If $\mathrm{injdim}_A A \le d-1$, then $\mathrm{gldim}(\mathsf{tails}\, B) = \mathrm{gldim}(\mathsf{tails}\, A) \le d-2$ by the Serre duality [de Naeghel and Van den Bergh 2004, Theorem A.4]. This is a contradiction to the fact that $\underline{\mathrm{Ext}}^{d-1}_{\mathcal{B}}(\mathcal{B}, \mathcal{B}) \cong \underline{\mathrm{H}}^d_{\mathfrak{m}}(B) \neq 0$. Hence A is AS-Gorenstein of dimension d.

To show that (A2) holds, it is enough to show that

(a) $X \in \mathsf{CM}^{\mathrm{gr}}(A)$ and $\underline{\mathrm{Ext}}^i_A(X, X) = 0$ for any $0 < i < d-1$,

(b) $M \in \mathsf{CM}^{\mathrm{gr}}(A)$ satisfying $\underline{\mathrm{Ext}}^i_A(X, M) = 0$ for any $0 < i < d-1$ belongs to $\mathsf{add}\{X(i) \mid i \in \mathbb{Z}\}$, and

(c) $M \in \mathsf{CM}^{\mathrm{gr}}(A)$ satisfying $\underline{\mathrm{Ext}}^i_A(M, X) = 0$ for any $0 < i < d-1$ belongs to $\mathsf{add}\{X(i) \mid i \in \mathbb{Z}\}$.

By (3-4), we see that $X \in \mathsf{CM}^{\mathrm{gr}}(A)$. By (B2), the functor $\mathsf{tails}\, B \to \mathsf{tails}\, A$ induced by $- \otimes_A X$ is an equivalence, so $\underline{\mathrm{Ext}}^i_{\mathcal{A}}(\mathcal{X}, \mathcal{X}) \cong \underline{\mathrm{Ext}}^i_{\mathcal{B}}(\mathcal{B}, \mathcal{B})$ for any i. Using $\mathrm{depth}_A X = d$ and [Artin and Zhang 1994, Theorem 7.2 (2)], it follows that

$$\underline{\mathrm{Ext}}^i_A(X, X) \cong \underline{\mathrm{Ext}}^i_{\mathcal{A}}(\mathcal{X}, \mathcal{X}) \cong \underline{\mathrm{Ext}}^i_{\mathcal{B}}(\mathcal{B}, \mathcal{B}) \cong \underline{\mathrm{H}}^{i+1}_{\mathfrak{m}}(B)$$

for any $0 < i < d-1$. Hence (a) holds by (B1).

We now give the proof of (b). Let $M \in \mathsf{CM}^{\mathrm{gr}}(A)$ be such that $\underline{\mathrm{Ext}}^i_A(X, M) = 0$ for any $0 < i < d-1$. Since we know that A is AS-Gorenstein, taking a free resolution of $M^\dagger$ in $\mathsf{grmod}\, A^{\mathrm{op}}$ and applying $(-)^\dagger$, we have an exact sequence

$$0 \to M \to F_0 \to \cdots \to F_{d-3} \to Y \to 0$$

in $\mathsf{grmod}\, A$, where each F_i is a graded free A-module and $Y \in \mathsf{CM}^{\mathrm{gr}}(A)$. Then we can make an exact sequence

$$0 \to \underline{\mathrm{Hom}}_A(X, M) \to \underline{\mathrm{Hom}}_A(X, F_0) \to \cdots \to \underline{\mathrm{Hom}}_A(X, F_{d-3}) \to \underline{\mathrm{Hom}}_A(X, Y) \to 0$$

in $\mathsf{GrMod}\, B$ because $\underline{\mathrm{Ext}}^i_A(X, M) = 0$ for $0 < i < d-1$. Moreover, taking a free presentation of $Y^\dagger \in \mathsf{CM}^{\mathrm{gr}}(A^{\mathrm{op}})$ and applying $\mathrm{Hom}_A(X, (-)^\dagger)$, we have an exact sequence

$$0 \to \mathrm{Hom}_A(X, Y) \to \mathrm{Hom}_A(X, F_{d-2}) \to \mathrm{Hom}_A(X, F_{d-1})$$

in $\mathsf{GrMod}\, B$, where F_{d-2} and F_{d-1} are graded free A-modules. By the assumption that A is a direct summand of X, each $\underline{\mathrm{Hom}}_A(X, F_i)$ is a graded right projective B-module. Since any graded right B-module has projective dimension at most d by (B1), it holds that $\operatorname{projdim}_B \underline{\mathrm{Hom}}_A(X, M) = 0$. This means that $\mathrm{Hom}_A(X, M)$ is finitely generated graded right projective over B by Lemma 3.7, so $M \in \mathsf{add}\{X(i) \mid i \in \mathbb{Z}\}$ by Lemma 3.6. Hence (b) is proved.

We next give the proof of (c). Let $M \in \mathsf{CM}^{\mathsf{gr}}(A)$ be such that $\underline{\mathrm{Ext}}^i_A(M, X) = 0$ for any $0 < i < d-1$. Then $\underline{\mathrm{Ext}}^i_{A^{\mathrm{op}}}(X^\dagger, M^\dagger) = 0$ for any $0 < i < d-1$. Since $B^{\mathrm{op}} \cong \underline{\mathrm{End}}_{A^{\mathrm{op}}}(X^\dagger)$ is also ASF-regular, the same method as that in the proof of (b) yields $M^\dagger \in \mathsf{add}\{X^\dagger(i) \mid i \in \mathbb{Z}\}$. Thus we see $M \in \mathsf{add}\{X(i) \mid i \in \mathbb{Z}\}$ and we obtain (c).

In addition, since we have $\operatorname{gldim} B < \infty$ and $\mathsf{tails}\, B \cong \mathsf{tails}\, A$ by (B1) and (B2), it follows that $\operatorname{gldim}(\mathsf{tails}\, A) < \infty$, so (A4) is satisfied by [Ueyama 2013, Lemma 5.7].

Hence the proof of Theorem 3.10(2) is finished. □

Remark 3.11. By observing the proof of Theorem 3.10, we notice that the condition (A4) in Theorem 3.10 can be replaced by the condition

$$\text{(A4}'\text{)} \qquad \operatorname{gldim}(\mathsf{tails}\, A) < \infty.$$

In this sense, it can be said that (A4) corresponds to the graded isolated singularity property.

Hence we obtain the following result.

Corollary 3.12. *Let A be an AS-Gorenstein algebra of dimension $d \geq 2$ and of Gorenstein parameter ℓ. Assume that* $\operatorname{gldim}(\mathsf{tails}\, A) < \infty$ *and A has a $(d-1)$-cluster tilting module $X \in \mathsf{CM}^{\mathsf{gr}}(A)$ satisfying $\underline{\mathrm{End}}_A(X)_{<0} = 0$. Then a basic $(d-1)$-cluster tilting module Y can be extracted from X, and in addition $B = \underline{\mathrm{End}}_A(Y)$ is a two-sided noetherian AS-regular algebra over B_0 of dimension d and of Gorenstein parameter ℓ such that* $\mathsf{tails}\, B \cong \mathsf{tails}\, A$.

Proof. By the proof of Theorem 3.10 (1), we see that the Gorenstein parameter of B is given by the Gorenstein parameter of A, so the statement follows from Theorem 3.10, Remark 3.11 and Theorem 2.10. □

It is clear that (A2), (A3), (A4) and the ν-stable assumption are conditions for a "right" A-module X. On the other hand, we see that (B1), (B2) and (B3) are "two-sided" conditions for $B = \underline{\mathrm{End}}_A(X)$. Thus Theorem 3.10 (1) asserts that one-sided conditions for X imply two-sided conditions for $B = \underline{\mathrm{End}}_A(X)$.

We now consider the case that A is a noncommutative quotient singularity.

Example 3.13. Let S be an AS-regular algebra of dimension $d \geq 2$. Let G be a finite subgroup of $\mathrm{GrAut}\, S$ such that $\operatorname{hdet} g = 1$ for each $g \in G$ (in the sense of Jørgensen and Zhang [2000]) and $\operatorname{char} k$ does not divide $|G|$. Assume that $S * G/(e)$ is finite-dimensional over k, where $e = \frac{1}{|G|} \sum_{g \in G} 1 * g \in S * G$. Then

- S^G is AS-Gorenstein of dimension $d \geq 2$ by [Jørgensen and Zhang 2000, Theorem 3.3],
- $S \in \mathsf{CM}^{\mathsf{gr}}(S^G)$ is a $(d-1)$-cluster tilting module by [Mori and Ueyama 2016a, Theorems 3.10 and 3.15],
- $\underline{\mathrm{End}}_{S^G}(S)_{<0} = (S * G)_{<0} = 0$ by [Mori and Ueyama 2016a, Theorem 3.10],
- $\mathrm{gldim}(\mathsf{tails}\, S^G) < \infty$ by [Mori and Ueyama 2016a, Theorem 3.10, Lemma 2.12],
- S is a ν-stable S^G-module by Example 3.2,

so the assumption of Theorem 3.10 (1) is satisfied. If fact, it follows from [Mori and Ueyama 2016a, Corollary 3.6, Theorem 3.10] that $B = \underline{\mathrm{End}}_{S^G}(S)$ is a two-sided noetherian AS-regular algebra over $B_0 \cong kG$ of dimension d such that $\mathsf{tails}\, \underline{\mathrm{End}}_{S^G}(S) \cong \mathsf{tails}\, S^G$. Hence Theorem 3.10 is regarded as a detailed version of this phenomenon.

For the rest of this paper, we construct a noncommutative quadric hypersurface which gives a concrete example of Theorem 3.10. See [Smith and Van den Bergh 2013, Section 5] and [Ueyama 2015, Section 4] for detailed information.

Example 3.14. In this example, we assume that k is algebraically closed of characteristic 0. Let

$$S = k\langle x, y, z\rangle/(xy + yx - z^2, xz + zx, yz + zy), \quad \deg x = \deg y = \deg z = 1.$$

Then S is a Koszul AS-regular algebra of dimension $d_S = 3$ and Gorenstein parameter $\ell_S = 3$ with a central regular element $x^2 + y^2 \in S_2$. Let

$$A = S/(x^2 + y^2).$$

Then A is a Koszul AS-Gorenstein algebra of dimension $d_A = 2$ and Gorenstein parameter $\ell_A = 1$. Take $w := x^2 \in A^!_2$ so that $S^! \cong A^!/(w)$. We define a finite-dimensional algebra $C(A)$ by $C(A) := A^![w^{-1}]_0$ (see [Smith and Van den Bergh 2013, Lemma 5.1]). It is easy to check that $C(A) \cong k[t]/(t^4 - 1) \cong k^4$ as algebras. By [Ueyama 2015, Proposition 4.1], A is of finite Cohen–Macaulay representation type, so it has a 1-cluster tilting module. It follows from [Ueyama 2015, Theorem 3.4] that $\mathrm{gldim}(\mathsf{tails}\, A) < \infty$. By [Ueyama 2015, Proposition 4.4] and [Mori 2006, Theorem 3.8], indecomposable nonprojective maximal Cohen–Macaulay A-modules (up to isomorphism and degree shift of grading) are parametrized by points of $\mathrm{Proj}\, A^! = \mathcal{V}(xy + z^2, x^2 - y^2) \subset \mathbb{P}^2$. Using this, one can check that the graded A-module

$$X := A \oplus X_1 \oplus X_2 \oplus X_3 \oplus X_4$$

is a representation generator of $\mathsf{CM}^{\mathsf{gr}}(A)$, that is, a 1-cluster tilting module, where

$$X_1 := A/(x - y + z)A, \quad X_2 := A/(x - y - z)A,$$
$$X_3 := A/(x + y + iz)A, \quad X_4 := A/(x + y - iz)A$$

and i is a primitive fourth root of unity. Clearly X is basic, so it is ν-stable. Since generators of X are concentrated in degree 0, every graded A-module homomorphism $X \to X(-s)$ has to be zero for any positive integer $s \in \mathbb{N}^+$, so we have $\underline{\mathrm{End}}(X)_{<0} = 0$. Hence we obtain that $B = \underline{\mathrm{End}}_A(X)$ is a two-sided noetherian AS-regular algebra over B_0 of dimension 2 and of Gorenstein parameter 1 satisfying $\mathsf{tails}\, B \cong \mathsf{tails}\, A$ by Corollary 3.12. We can calculate that the Hilbert series $H_B(t)$ is $9(1+t)/(1-t)^2$. Furthermore B_0 is isomorphic to the path algebra kQ, where

$$Q = \begin{tikzcd}[row sep=small, column sep=small] \bullet \arrow[dr] & & \bullet \arrow[dl] \\ & \bullet & \\ \bullet \arrow[ur] & & \bullet \arrow[ul] \end{tikzcd},$$

so $\operatorname{gldim} B_0 < \infty$. Hence we obtain

$$\mathsf{D}^{\mathsf{b}}(\mathsf{tails}\, A) \cong \mathsf{D}^{\mathsf{b}}(\mathsf{tails}\, B) \cong \mathsf{D}^{\mathsf{b}}(\mathsf{mod}\, B_0) \cong \mathsf{D}^{\mathsf{b}}(\mathsf{mod}\, kQ)$$

by [Minamoto and Mori 2011, Theorem 4.14]. As a remark, for any 2-dimensional commutative Cohen–Macaulay algebra C of finite Cohen–Macaulay representation type generated in degree 1, it follows that $\mathsf{tails}\, C \cong \mathsf{tails}\, k[x, y] \cong \mathsf{coh}\, \mathbb{P}^1$ by [Eisenbud and Herzog 1988, theorem on page 347], so $\mathsf{D}^{\mathsf{b}}(\mathsf{tails}\, C) \cong \mathsf{D}^{\mathsf{b}}(\mathsf{coh}\, \mathbb{P}^1) \cong \mathsf{D}^{\mathsf{b}}(\mathsf{mod}\, kQ')$, where $Q' = \bullet \rightrightarrows \bullet$. Thus we see that $\mathsf{D}^{\mathsf{b}}(\mathsf{tails}\, A) \cong \mathsf{D}^{\mathsf{b}}(\mathsf{tails}\, B) \not\cong \mathsf{D}^{\mathsf{b}}(\mathsf{tails}\, C)$.

Acknowledgment

The author thanks the referee for a careful reading of the manuscript and helpful suggestions.

References

[Artin and Zhang 1994] M. Artin and J. J. Zhang, "Noncommutative projective schemes", *Adv. Math.* **109**:2 (1994), 228–287. MR Zbl

[Bruns and Herzog 1998] W. Bruns and J. Herzog, *Cohen–Macaulay rings*, 2nd ed., Cambridge Studies in Advanced Mathematics **39**, Cambridge Univ. Press, 1998. Zbl

[Dao and Huneke 2013] H. Dao and C. Huneke, "Vanishing of Ext, cluster tilting modules and finite global dimension of endomorphism rings", *Amer. J. Math.* **135**:2 (2013), 561–578. MR Zbl

[Dong and Wu 2009] Z.-C. Dong and Q.-S. Wu, "Non-commutative Castelnuovo–Mumford regularity and AS-regular algebras", *J. Algebra* **322**:1 (2009), 122–136. MR Zbl

[Eisenbud and Herzog 1988] D. Eisenbud and J. Herzog, "The classification of homogeneous Cohen–Macaulay rings of finite representation type", *Math. Ann.* **280**:2 (1988), 347–352. MR Zbl

[Herschend et al. 2014] M. Herschend, O. Iyama, and S. Oppermann, "n-representation infinite algebras", *Adv. Math.* **252** (2014), 292–342. MR Zbl

[Iyama 2008] O. Iyama, "Auslander–Reiten theory revisited", pp. 349–397 in *Trends in representation theory of algebras and related topics*, edited by A. Skowroński, Eur. Math. Soc., Zürich, 2008. MR Zbl

[Jørgensen and Zhang 2000] P. Jørgensen and J. J. Zhang, "Gourmet's guide to Gorensteinness", *Adv. Math.* **151**:2 (2000), 313–345. MR Zbl

[Minamoto and Mori 2011] H. Minamoto and I. Mori, "The structure of AS-Gorenstein algebras", *Adv. Math.* **226**:5 (2011), 4061–4095. MR Zbl

[Mori 2003] I. Mori, "Homological properties of balanced Cohen–Macaulay algebras", *Trans. Amer. Math. Soc.* **355**:3 (2003), 1025–1042. MR Zbl

[Mori 2006] I. Mori, "Co-point modules over Koszul algebras", *J. London Math. Soc.* (2) **74**:3 (2006), 639–656. MR Zbl

[Mori 2013] I. Mori, "McKay-type correspondence for AS-regular algebras", *J. Lond. Math. Soc.* (2) **88**:1 (2013), 97–117. MR Zbl

[Mori 2015] I. Mori, "Regular modules over 2-dimensional quantum Beilinson algebras of type *S*", *Math. Z.* **279**:3-4 (2015), 1143–1174. MR Zbl

[Mori and Ueyama 2016a] I. Mori and K. Ueyama, "Ample group action on AS-regular algebras and noncommutative graded isolated singularities", *Trans. Amer. Math. Soc.* **368**:10 (2016), 7359–7383. MR Zbl

[Mori and Ueyama 2016b] I. Mori and K. Ueyama, "Stable categories of graded maximal Cohen–Macaulay modules over noncommutative quotient singularities", *Adv. Math.* **297** (2016), 54–92. MR Zbl

[de Naeghel and Van den Bergh 2004] K. de Naeghel and M. Van den Bergh, "Ideal classes of three-dimensional Sklyanin algebras", *J. Algebra* **276**:2 (2004), 515–551. MR Zbl

[Reyes et al. 2014] M. Reyes, D. Rogalski, and J. J. Zhang, "Skew Calabi–Yau algebras and homological identities", *Adv. Math.* **264** (2014), 308–354. MR Zbl

[Reyes et al. 2017] M. Reyes, D. Rogalski, and J. J. Zhang, "Skew Calabi–Yau triangulated categories and Frobenius Ext-algebras", *Trans. Amer. Math. Soc.* **369**:1 (2017), 309–340. MR Zbl

[Smith and Van den Bergh 2013] S. P. Smith and M. Van den Bergh, "Noncommutative quadric surfaces", *J. Noncommut. Geom.* **7**:3 (2013), 817–856. MR Zbl

[Ueyama 2013] K. Ueyama, "Graded maximal Cohen–Macaulay modules over noncommutative graded Gorenstein isolated singularities", *J. Algebra* **383** (2013), 85–103. MR Zbl

[Ueyama 2015] K. Ueyama, "Noncommutative graded algebras of finite Cohen–Macaulay representation type", *Proc. Amer. Math. Soc.* **143**:9 (2015), 3703–3715. MR Zbl

[Van den Bergh 1997] M. Van den Bergh, "Existence theorems for dualizing complexes over noncommutative graded and filtered rings", *J. Algebra* **195**:2 (1997), 662–679. MR Zbl

[Weibel 1994] C. A. Weibel, *An introduction to homological algebra*, Cambridge Studies in Advanced Mathematics **38**, Cambridge Univ. Press, 1994. MR Zbl

[Yekutieli 1992] A. Yekutieli, "Dualizing complexes over noncommutative graded algebras", *J. Algebra* **153**:1 (1992), 41–84. MR Zbl

[Yekutieli and Zhang 1997] A. Yekutieli and J. J. Zhang, "Serre duality for noncommutative projective schemes", *Proc. Amer. Math. Soc.* **125**:3 (1997), 697–707. MR Zbl

Received April 13, 2016. Revised February 6, 2017.

KENTA UEYAMA
DEPARTMENT OF MATHEMATICS
FACULTY OF EDUCATION
HIROSAKI UNIVERSITY
1 BUNKYOCHO, HIROSAKI
AOMORI 036-8560
JAPAN
k-ueyama@hirosaki-u.ac.jp

PACIFIC JOURNAL OF MATHEMATICS
Vol. 289, No. 2, 2017
dx.doi.org/10.2140/pjm.2017.289.469

CONCENTRATION FOR A BIHARMONIC SCHRÖDINGER EQUATION

DONG WANG

We consider the fourth-order problem

$$\begin{cases} \epsilon^4\Delta^2 u + V(x)u = P(x)f(|u|)u, & x \in \mathbb{R}^N, \\ u(x) \to 0 & \text{as } |x| \to \infty, \end{cases}$$

where V and P are spatial distributions of external potentials. We study the concentration phenomena of the solutions as $\epsilon \to 0$ using variational methods.

1. Introduction

This work is devoted to the analysis of solutions that solve the following nonlinear stationary biharmonic Schrödinger equation:

$$\begin{cases} \epsilon^4\Delta^2 u + V(x)u = P(x)f(|u|)u, & x \in \mathbb{R}^N, \\ u(x) \to 0 & \text{as } |x| \to \infty, \end{cases} \tag{1-1}$$

where ϵ denotes Planck's constant, and V, P are spatial distributions of external potentials. To simplify the idea of this work, we are going to describe certain concentration phenomena of the solutions of (1-1) as $\epsilon \to 0$, for physical purposes.

Problem (1-1) is the biharmonic version of the usual Schrödinger equation which has been extensively studied in literature [Floer and Weinstein 1986; Coti Zelati and Rabinowitz 1992; Rabinowitz 1992; Wang 1993; Willem 1996; del Pino and Felmer 1996; Gui 1996; Ambrosetti et al. 1997; Bartsch et al. 2001; Sirakov 2002; Byeon and Wang 2003; Ding and Tanaka 2003; Ni and Wei 2006; Ambrosetti and Malchiodi 2006; Byeon and Jeanjean 2007; Ding and Szulkin 2007; Ding and Wei 2007; Ding and Liu 2013] and references therein. In general, if we omit the exponent 2 of the first term in (1-1), we have an equation

$$(-i\epsilon\nabla + A(x))^2 w + V(x)w = f(x, w), \qquad w \in H^1(\mathbb{R}^N, \mathbb{C}), \tag{1-2}$$

MSC2010: 35J10, 35Q40, 47J30.
Keywords: nonlinear biharmonic Schrödinger equations, standing waves, critical point theory.

which arises when one seeks the standing wave solutions of the Schrödinger equation

$$i\hbar \frac{\partial \varphi}{\partial t} = (-i\hbar\nabla + A(x))^2\varphi + W(x)\varphi - n(x, |\varphi|)\varphi. \tag{1-3}$$

Considerable effort has gone into the study of the nonlinear Schrödinger equations without the magnetic field (i.e., $A(x) = 0$) for studying the existence, multiplicity and qualitative properties of standing wave solutions. When the magnetic vector $A(x) \not\equiv 0$, the first work seems to be involved in [Esteban and Lions 1989] in which the existence of solutions of (1-2) via a constrained minimization argument with $\epsilon = 1$ is studied. Later, under certain assumptions, the existence and multiplicity of solutions of (1-2) ($\epsilon = 1$) were obtained in [Arioli and Szulkin 2003; Pankov 2003; Wang 2008; Liang and Zhang 2011]. The existence and concentration phenomena of semiclassical solutions of (1-2) were studied in [Kurata 2000], where $f(x, w) = g(|w|^2)w$ is subcritical and $\inf_{x\in\mathbb{R}^N} V(x) > 0$ such that the Palais–Smale condition holds for any energy level and for any $\epsilon > 0$. For the case $A(x) = 0$, Floer and Weinstein [1986], proved in the one dimensional case and for $f(w) = w^3$ that a single spike solution concentrates around any given nondegenerate critical point of the linear potential $V(x)$. Oh [1988; 1990] extended this result in higher dimensions and for $f(u) = |u|^{p-1}u$ $(1 < p < N+2/N-2)$. Subsequently, variational methods were found suitable for such issues and the existence of spike layer solutions in the semiclassical limit were established under various conditions of $V(x)$. Particularly, initiated by Rabinowitz [1992], the existence of positive solutions of the Schrödinger equation for small $\epsilon > 0$ is proved whenever

$$\liminf_{|x|\to\infty} V(x) > \inf_{x\in\mathbb{R}^N} V(x).$$

These solutions concentrate around the global minimum points of V when $\epsilon \to 0$, as was shown by Wang [1993]. It should be pointed out that M. del Pino and P. Felmer [1996] first succeeded in proving a localized version of the concentration behavior of semiclassical solutions.

By using a combination of stability analysis and numerical simulations, the role of small fourth-order dispersion has been considered in a series of papers by Karpman and Shagalov ([2000] and the references therein), who studied the equation

$$i\psi_t(t, x) + \Delta\psi + |\psi|^{2\sigma}\psi + \epsilon\Delta^2\psi = 0,$$

in the case when $\epsilon < 0$. Later, in [Ben-Artzi et al. 2000], Ben-Artzi, Koch, and Saut obtained sharp dispersive estimates for the biharmonic Schrödinger operator in

$$i\partial_t u + \Delta^2 u + \epsilon\Delta u + f(|u|^2)u = 0,$$

namely for the linear group associated to $i\partial_t + \Delta^2 \pm \Delta$. Parallel to this, some specific nonlinear fourth-order Schrödinger equations have received deep consideration. Fibich, Ilan, and Papanicolaou, in [Fibich et al. 2002], analyzed self-focusing and singularity formation in the nonlinear Schrödinger equation (NLS) with high-order dispersion $i\psi_t \pm \Delta^q \psi + |\psi|^{2\sigma}\psi = 0$, in the isotropic mixed-dispersion NLS $i\psi_t + \Delta\psi + \epsilon\Delta^2\psi + |\psi|^{2\sigma}\psi = 0$, and in nonisotropic mixed-dispersion NLS equations which model propagation in fiber arrays. Almost at the same time, Guo and Wang [2002] studied the existence and scattering theory for the nonlinear Schrödinger equations $iu_t + (-\Delta)^m u + f(u) = 0$, with $u(0, x) = \phi(x)$, where $u(t, x)$ defined on $\mathbb{R} \times \mathbb{R}^n$ is a complex valued function, $m \geq 1$ is an integer and f is a scalar nonlinear function. Not much later, Hao, Hsiao and Wang, in [Hao et al. 2006; 2007], discussed the Cauchy problem in a high regularity setting. Subsequently, Segata [2006] proved scattering in the case the space dimension is one and considered the three-dimensional motion of an isolated vortex filament by using the method of Fourier restriction norm.

Motivated by the previously mentioned works, we are mainly interested in (1-1) with the biharmonic operator

$$\Delta^2 u = \sum_{i=1}^{N} \frac{\partial^4}{\partial x_i^4} u + \sum_{i \neq j}^{N} \frac{\partial^4}{\partial x_i^2 x_j^2} u.$$

The biharmonic Schrödinger equation (1-1) appears when one considers the stationary solutions $w(t, x) = e^{i\lambda t} u(x)$ of the t-dependent equation of the form

$$i\epsilon\partial_t w - \epsilon^4 \Delta^2 w - M(x)w + P(x) f(|w|)w = 0, \tag{1-4}$$

where $\lambda \in \mathbb{R}$. Such stationary solutions, also called standing waves, are finite energy waveguide solutions of (1-1) after rearranging terms in (1-4). It is worth pointing out that although there are many works dealing with problems related to (1-2), so many problems appear when dealing with the fourth-order problem. The main reason for this difficulty is the lack of a general maximum principle to the biharmonic operator. This leads to a series of technical problems in trying to adapt some second-order classical arguments.

Precisely, we formulate the fundamental assumption on the potential functions V, P as

(VP) V, $P \in L^\infty(\mathbb{R}^N)$ are uniformly continuous such that $\inf_{x\in\mathbb{R}^N} V(x) > 0$ and $\inf_{x\in\mathbb{R}^N} P(x) > 0$.

To obtain the concentration results, let us introduce the following restrictions on the nonlinear function f:

(f_1) $f(0)=0$, $f \in C^1(0,\infty)$, $f'(s)>0$ for $s>0$, and there is a $p \in (2, 2N/(N-4))$ such that $\lim_{s\to\infty} f(s)/s^{p-2} < \infty$,

(f_2) denoting $F(s) = \int_0^s f(t)t\,dt$, there is $\theta > 2$ such that $0 < F(s) \le (1/\theta) f(s)s^2$ for $s > 0$.

Now we are ready to describe our concentration results. First let us set

$$\tau := \inf_{x\in\mathbb{R}^N} V(x), \qquad \tau_\infty := \lim_{|x|\to\infty} \inf V(x), \qquad \tau_p := \inf_{x\in\mathcal{P}} V(x),$$
$$\gamma := \sup_{x\in\mathbb{R}^N} P(x), \qquad \gamma_\infty := \lim_{|x|\to\infty} \sup P(x), \qquad \gamma_v := \sup_{x\in\mathcal{V}} P(x),$$
$$\mathcal{V} := \{x \in \mathbb{R}^N : V(x) = \tau\}, \qquad \mathcal{P} := \{x \in \mathbb{R}^N : P(x) = \gamma\},$$

and suppose that

(V) V leads the behavior, i.e., $\tau < \tau_\infty$ and there exists $R > 0$ such that $\gamma_v \ge P(x)$ for any $|x| \ge R$,

(P) P leads the behavior, i.e., $\gamma > \gamma_\infty$ and there exists $R > 0$ such that $\tau_p \le V(x)$ for any $|x| \ge R$.

Let us define for (V)

$$\mathscr{A}_v := \{x \in \mathcal{V} : P(x) = \gamma_v\} \cup \{x \notin \mathcal{V} : P(x) > \gamma_v\},$$

and for (P)

$$\mathscr{A}_p := \{x \in \mathcal{P} : V(x) = \tau_p\} \cup \{x \notin \mathcal{P} : V(x) < \tau_p\}.$$

Remark that, generally, $\mathcal{V} \cap \mathcal{P} = \varnothing$. And notice that, for example, $\mathcal{V} \cap \mathcal{P} \ne \varnothing$ implies $\tau = \tau_p$ and $\gamma = \gamma_v$, from which we deduce that

$$\mathscr{A}_v = \mathcal{V} \cap \mathcal{P}.$$

Under our assumptions (V) and (P), $\mathscr{A}_v$ and $\mathscr{A}_p$ are nonempty bounded sets in $\mathbb{R}^N$, and $\mathscr{A}_v = \mathscr{A}_p = \mathcal{V} \cap \mathcal{P}$ if and only if $\mathcal{V} \cap \mathcal{P} \ne \varnothing$; see also [Ding and Liu 2013].

To give a better description of our results, let us set

$$\mathscr{C} = \begin{cases} \mathscr{A}_v & \text{if } (V) \text{ holds,} \\ \mathscr{A}_p & \text{if } (P) \text{ holds.} \end{cases}$$

We have the following results:

Theorem 1.1. *Let (VP), (f_1) and (f_2) hold. Assume additionally that either (V) or (P) holds. Then* (1-1) *has* (*at least*) *one ground state solution for all small* ϵ.

And for the concentration of the solutions of (1-1) as $\epsilon \to 0$, we have:

Theorem 1.2. *Let (VP), (f_1) and (f_2) hold. Assume additionally that either (V) or (P) holds. Let u_ϵ be the solution of* (1-1), *given by Theorem 1.1. For any $\epsilon_n > 0$ with $\lim_{n\to\infty} \epsilon_n = 0$, up to a subsequence, there exists $x_{\epsilon_n} \in \mathbb{R}^N$ which is a maximum point of $|u_{\epsilon_n}|$, such that*

$$\mathrm{dist}(\epsilon_n x_{\epsilon_n}, \mathscr{C}) \to 0 \quad \textit{as} \quad n \to \infty.$$

Let $w_{\epsilon_n}(x) := u_{\epsilon_n}(\epsilon_n(x + x_{\epsilon_n}))$. Then $w_{\epsilon_n} \to w_0$ in $H^2(\mathbb{R}^N)$, where w_0 is a ground state solution of

$$\Delta^2 w + V(x_0)w = P(x_0) f(|w|)w, \qquad x_0 \in \mathscr{C}.$$

Our arguments are variational with a mixture of the mountain pass technique and Nehari Manifolds. The paper is organized as follows. In the next section, we introduce some notations and the variational framework for such problem. We prove the existence and concentration results for (1-1) in the remaining two sections.

2. Variational framework

Hereafter we use the following notation:

- $E := H^2(\mathbb{R}^N)$ is the usual Sobolev space endowed with the standard scalar product and norm

$$\langle u, v\rangle = \int_{\mathbb{R}^N} (\Delta u \Delta v + uv)\, dx, \qquad \|u\|^2 = \int_{\mathbb{R}^N} (|\Delta u|^2 + u^2)\, dx.$$

- $L^q(\Omega)$, $1 \le q \le +\infty$, denotes a Lebesgue space. The norm in $L^q(\Omega)$ is denoted by $|u|_{q,\Omega}$ when Ω is a proper subset of $\mathbb{R}^N$, by $|\cdot|_p$ when $\Omega = \mathbb{R}^N$.
- For any $\rho > 0$ and for any $z \in \mathbb{R}^N$, $B_\rho(z)$ denotes the ball of radius ρ centered at z, $|B_\rho(z)|$ denotes its Lebesgue measure and $\partial B_\rho(z)$ denotes its boundary.
- For ease of notation, let us set $2^* = 2N/(N-4)$. Without loss of generality, we assume that $0 \in \mathscr{C}$.

By assumption (VP), the following energy functional I of (1-1) defined in E is well defined,

$$I(u) = \frac{1}{2}\epsilon^4 \int_{\mathbb{R}^N} |\Delta u|^2\, dx + \frac{1}{2}\int_{\mathbb{R}^N} V(x)|u|^2\, dx - \int_{\mathbb{R}^N} P(x)F(|u|)\, dx.$$

Moreover, the solutions of (1-1) are the critical points of I.

Equivalent problem. Making the change of variable $x \to \epsilon x$, (1-1) becomes

$$\begin{cases} \Delta^2 z + V_\epsilon(x)z = P_\epsilon(x) f(|z|)z, & x \in \mathbb{R}^N, \\ z \in E, \end{cases} \tag{2-1}$$

where $V_\epsilon(x) = V(\epsilon x)$, $P_\epsilon(x) = P(\epsilon x)$, and $z(x) = u(\epsilon x)$.

In the sequel, we will in fact focus on finding the critical points of the energy functional associated to (2-1) which is defined by

$$\phi_\epsilon(z)=\frac{1}{2}\int_{\mathbb{R}^N}|\Delta z|^2\,dx+\frac{1}{2}\int_{\mathbb{R}^N}V_\epsilon(x)|z|^2\,dx-\int_{\mathbb{R}^N}P_\epsilon(x)F(|z|)\,dx.$$

Remark. Denote $\|z\|_\epsilon^2=\int_{\mathbb{R}^N}[|\Delta z|^2+V_\epsilon(x)|z|^2]\,dx$. Recall that $\tau=\inf_{x\in\mathbb{R}^N}V(x)$. Here norm $\|\cdot\|_\epsilon$ is equivalent to $\|\cdot\|$. Indeed, by assumption (VP),

$$\|z\|_\epsilon^2\geq\int_{\mathbb{R}^N}[|\Delta z|^2+\tau|z|^2]\,dx\geq\delta\int_{\mathbb{R}^N}(|\Delta z|^2+|z|^2)\,dx=\delta\|z\|^2,$$

where

$$\delta=\min\{1,\tau\}>0$$

and also we have

$$\begin{aligned}\|z\|_\epsilon^2&\leq\int_{\mathbb{R}^N}(|\Delta z|^2+|V|_\infty\cdot|z|^2)\,dx\\&\leq(1+|V|_\infty)\int_{\mathbb{R}^N}(|\Delta z|^2+|z|^2)\,dx=(1+|V|_\infty)\|z\|^2.\end{aligned}$$

For notational convenience, let us write $\phi_\epsilon(z)=\frac{1}{2}\|z\|_\epsilon^2-\int_{\mathbb{R}^N}P_\epsilon(x)F(|z|)\,dx$. We observe that ϕ_ϵ satisfies the so-called mountain pass structure. For details, recall that by (f_1), (f_2), we have

$$\hat{F}(s):=\frac{1}{2}f(s)s^2-F(s)\geq\frac{\theta-2}{2\theta}f(s)s^2\quad\text{for any s>0.}$$

Moreover there exists δ_1 small enough and $c_1>0$ such that

$$f(s)\leq\delta_1+c_1s^{p-2}.$$

Hence

$$F(s)\leq\frac{\delta_1}{2}s^2+c_1's^p,$$

which implies

$$\phi_\epsilon(z)=\frac{1}{2}\|z\|_\epsilon^2-\int_{\mathbb{R}^N}P_\epsilon(x)F(|z|)\,dx\geq\frac{1}{4}\|z\|_\epsilon^2-c_1''\|z\|_\epsilon^p,$$

where we have used the Sobolev embedding theorems, $E\hookrightarrow L^p(\mathbb{R}^N)$, $|z|_p\leq C\cdot\|z\|_\epsilon$ for some positive constant C. Notice that when $p>2$, then by direct computation, we have $\phi_\epsilon(z)\geq\alpha>0$, where $z\in\partial B_\rho(0)=\{z\in E:\|z\|=\rho\}$ for some $\rho>0$.

By (f_2) we have $F(s)\geq cs^\theta-s^2$ for some $c>0$. Thus, for any $z\in E\setminus\{0\}$ and for any positive real number t, we have

$$\phi_\epsilon(tz)\leq c_2t^2\|z\|_\epsilon^2-c_3t^\theta\int_{\mathbb{R}^N}|z|^\theta\,dx.$$

Since $p > 2$, we know $\phi_\epsilon(tz) \to -\infty$ as $t \to \infty$. Also, $\phi_\epsilon(0) = 0$. Thus there exists a sufficiently large positive real number t_z, such that $\|t_z z\| > \rho$ and $\max\{\phi_\epsilon(0), \phi_\epsilon(t_z z)\} \le 0 < \alpha$.

Based on the above discussion, by the classical mountain pass theorem, there exists a sequence $\{z_{n,\epsilon} : n = 1, 2, \dots\} \subseteq E \setminus \{0\}$ such that

$$\phi_\epsilon(z_{n,\epsilon}) \to c_\epsilon \quad \text{as } n \to \infty, \quad \text{and} \quad \phi'_\epsilon(z_{n,\epsilon}) \to 0 \quad \text{as } n \to \infty, \tag{2-2}$$

where

$$c_\epsilon = \inf_{\gamma\in\Gamma} \max_{t\in[0,1]} \phi_\epsilon(\gamma(t))$$

and

$$\Gamma = \{\gamma \in C([0, 1], E) : \gamma(0) = 0, \phi_\epsilon(\gamma(1)) < \alpha\}.$$

Moreover, by (2-2), we have

$$o_n(1)\|z_{n,\epsilon}\| = \phi'_\epsilon(z_{n,\epsilon})z_{n,\epsilon} = \|z_{n,\epsilon}\|^2_\epsilon - \int_{\mathbb{R}^N} P_\epsilon(x) f(|z_{n,\epsilon}|)z^2_{n,\epsilon}\,dx,$$

$$c_\epsilon + o_n(1)\|z_{n,\epsilon}\| = \phi_\epsilon(z_{n,\epsilon}) - \frac{1}{2}\phi'_\epsilon(z_{n,\epsilon})z_{n,\epsilon} = \int_{\mathbb{R}^N} P_\epsilon(x)\hat{F}(|z_{n,\epsilon}|)\,dx.$$

It follows that

$$\|z_{n,\epsilon}\|^2_\epsilon \le M + o_n(1)\|z_{n,\epsilon}\|$$

for some $M > 0$. Thus $\{z_{n,\epsilon} : n = 1, 2, \dots\}$ is a bounded sequence and $\|z_{n,\epsilon}\| \ge D$ for some positive constant D, $n = 1, 2, \dots$.

The limiting equation. Now we consider the special form of the equation (2-1)

$$\begin{cases} \Delta^2 z + \mu z = \lambda f(|z|)z, & x \in \mathbb{R}^N, \\ z \in E, \end{cases} \tag{2-3}$$

where $\mu > 0$ and $\lambda > 0$ are both constants.

Then the energy functional of (2-3) is

$$\phi_{\mu,\lambda}(z) = \frac{1}{2}\int_{\mathbb{R}^N} |\Delta z|^2\,dx + \frac{\mu}{2}\int_{\mathbb{R}^N} |z|^2\,dx - \lambda\int_{\mathbb{R}^N} F(|z|)\,dx.$$

Denote $\|z\|^2_\mu = \int_{\mathbb{R}^N}(|\Delta z|^2 + \mu|z|^2)\,dx$. Then similarly, norm $\|\cdot\|_\mu$ is also equivalent to $\|\cdot\|$.

Rewrite $\phi_{\mu,\lambda}(z) = \frac{1}{2}\|z\|^2_\mu - \lambda\int_{\mathbb{R}^N} F(|z|)\,dx$. Similarly, $\phi_{\mu,\lambda}$ also satisfies the conditions of the mountain pass theorem, i.e.,

- there exists $\rho, \alpha > 0$ such that $\phi_{\mu,\lambda}|_{\partial B_\rho(0)} \ge \alpha$,
- there exists $e \in E$, $\|e\| > \rho$ such that $\max\{\phi_{\mu,\lambda}(0), \phi_{\mu,\lambda}(e)\} < \alpha$.

By the mountain pass lemma, there exists a sequence $\{z_{n,\mu,\lambda} : n=1,2,\dots\} \subseteq E\backslash\{0\}$ such that

$$\phi_{\mu,\lambda}(z_{n,\mu,\lambda}) \to c_{\mu,\lambda} \quad \text{as } n \to \infty, \quad \text{and} \quad \phi'_{\mu,\lambda}(z_{n,\mu,\lambda}) \to 0 \quad \text{as } n \to \infty, \tag{2-4}$$

where

$$c_{\mu,\lambda} = \inf_{\gamma\in\Gamma} \max_{t\in[0,1]} \phi_{\mu,\lambda}(\gamma(t))$$

and

$$\Gamma = \{\gamma \in C([0, 1], E) : \gamma(0) = 0, \gamma(1) = e\}.$$

Similarly, $\{z_{n,\mu,\lambda} : n = 1, 2, \dots\}$ is also a bounded sequence, and $\|z_{n,\mu,\lambda}\| \geq D_{\mu,\lambda}$ for some positive constant $D_{\mu,\lambda}$, $n = 1, 2, \dots$.

A standard concentration compactness argument shows that there exists $R > 0$, $\alpha > 0$ and $\{x_n : n = 1, 2, \dots\} \subseteq \mathbb{R}^N$, such that

$$\int_{B_R(x_n)} |z_{n,\mu,\lambda}|^2 \, dx \geq \alpha.$$

Let be $\tilde{z}_{n,\mu,\lambda}(x) = z_{n,\mu,\lambda}(x + x_n)$. Then $\{\tilde{z}_{n,\mu,\lambda} : n = 1, 2, \dots\}$ is also a bounded sequence and $\|\tilde{z}_{n,\mu,\lambda}\| \geq D_{\mu,\lambda}$, $n = 1, 2, \dots$. Thus, up to a subsequence, we have $\tilde{z}_{n,\mu,\lambda} \rightharpoonup z_{\mu,\lambda} \in E$. Particularly, $z_{\mu,\lambda} \neq 0$. Indeed, by Sobolev embedding theorems, $\tilde{z}_{n,\mu,\lambda} \rightharpoonup z_{\mu,\lambda}$ in E implies $\tilde{z}_{n,\mu,\lambda} \to z_{\mu,\lambda}$ in $L^2_{loc}(\mathbb{R}^N)$. Together with

$$\int_{B_R(0)} |\tilde{z}_{n,\mu,\lambda}|^2 \, dx = \int_{B_R(x_n)} |z_{n,\mu,\lambda}|^2 \, dx \geq \alpha,$$

we have

$$\int_{B_R(0)} |z_{\mu,\lambda}|^2 \, dx \geq \alpha,$$

i.e., $z_{\mu,\lambda} \neq 0$.

By using the weak sequential continuity of $\phi'_{\mu,\lambda} : E \to E^*$, we know that $\phi'_{\mu,\lambda}(z_{\mu,\lambda})z_{\mu,\lambda} = 0$, i.e., $z_{\mu,\lambda}$ is a critical point of the functional $\phi_{\mu,\lambda}$ which is a weak solution of (2-3).

3. Proof of Theorem 1.1

In this section, we prove Theorem 1.1. Initially, we would like to collect some useful results concerning the limiting equation. A first observation is that if $\gamma_{\mu,\lambda}$ denotes the ground state energy of $\phi_{\mu,\lambda}$, we have

$$c_{\mu,\lambda} = \gamma_{\mu,\lambda} = \phi_{\mu,\lambda}(z_{\mu,\lambda}).$$

Indeed, since $\phi'_{\mu,\lambda}(z_{\mu,\lambda})z_{\mu,\lambda} = 0$, we have

$$\begin{aligned}
\phi_{\mu,\lambda}(z_{\mu,\lambda}) &= \phi_{\mu,\lambda}(z_{\mu,\lambda}) - \tfrac{1}{2}\phi'_{\mu,\lambda}(z_{\mu,\lambda})z_{\mu,\lambda} \\
&= \lambda \int_{\mathbb{R}^N} \hat{F}(|z_{\mu,\lambda}|)\,dx \\
&\le \liminf_{n\to\infty} \lambda \int_{\mathbb{R}^N} \hat{F}(|\tilde{z}_{n,\mu,\lambda}|)\,dx \\
&= \liminf_{n\to\infty} \big[\phi_{\mu,\lambda}(\tilde{z}_{n,\mu,\lambda}) - \tfrac{1}{2}\phi'_{\mu,\lambda}(\tilde{z}_{n,\mu,\lambda})\tilde{z}_{n,\mu,\lambda}\big] \\
&= c_{\mu,\lambda},
\end{aligned}$$

the above inequality follows from the Fatou's lemma. Now we consider the Nehari manifold

$$\mathcal{N}_{\mu,\lambda} := \{z \in E \setminus \{0\} : \phi'_{\mu,\lambda}(z)z = 0\}.$$

By a direct observation, for any $z \in E \setminus \{0\}$, there exists $t_z > 0$, such that $t_z z \in \mathcal{N}_{\mu,\lambda}$. Notice that

$$\max_{t>0} \phi_{\mu,\lambda}(tz) = \phi_{\mu,\lambda}(t_z z).$$

Indeed, denote $f(t) = \phi_{\mu,\lambda}(tz)$. Then it is easy to check that $f(t) \to -\infty$ as $t \to \infty$, $f(0) = 0$ and there is small t such that $f(t) > 0$. Since $f'(t_z) = \phi'_{\mu,\lambda}(t_z z)z = 0$ and $f''(t) < 0$, we know that t_z is the unique critical point of $f(t)$. Therefore $\max_{t>0} \phi_{\mu,\lambda}(tz) = \phi_{\mu,\lambda}(t_z z)$. Let

$$\bar{c}_{\mu,\lambda} = \inf_{z\in E\setminus\{0\}} \max_{t>0} \phi_{\mu,\lambda}(tz).$$

Then a standard argument shows that

$$\phi_{\mu,\lambda}(z_{\mu,\lambda}) \le c_{\mu,\lambda} \le \bar{c}_{\mu,\lambda} = \inf_{z\in\mathcal{N}_{\mu,\lambda}} \phi_{\mu,\lambda}(z) \le \phi_{\mu,\lambda}(z_{\mu,\lambda}).$$

Thus $\phi_{\mu,\lambda}(z_{\mu,\lambda}) = c_{\mu,\lambda}$.

The following lemma is a direct application of the above observation:

Lemma 3.1. *If* $\mu_2 \ge \mu_1$ *and* $\lambda_2 \le \lambda_1$, *then* $c_{\mu_2,\lambda_2} \ge c_{\mu_1,\lambda_1}$. *Particularly, if*

$$\max\{\mu_2 - \mu_1, \lambda_1 - \lambda_2\} > 0,$$

then $c_{\mu_2,\lambda_2} > c_{\mu_1,\lambda_1}$.

Proof. Let $z_{\mu_2,\lambda_2} \in E \setminus \{0\}$ be the critical point such that

$$\phi_{\mu_2,\lambda_2}(z_{\mu_2,\lambda_2}) = c_{\mu_2,\lambda_2}.$$

A standard argument implies that

$$c_{\mu_2,\lambda_2} = \max_{t>0} \phi_{\mu_2,\lambda_2}(tz_{\mu_2,\lambda_2}).$$

Observe that if $\mu_2 \geq \mu_1$ and $\lambda_2 \leq \lambda_1$, then

$$\phi_{\mu_2,\lambda_2}(z) \geq \phi_{\mu_1,\lambda_1}(z), \quad \text{for all } z \in E \setminus \{0\}.$$

Thus

$$c_{\mu_2,\lambda_2} = \max_{t>0} \phi_{\mu_2,\lambda_2}(tz_{\mu_2,\lambda_2}) \geq \max_{t>0} \phi_{\mu_1,\lambda_1}(tz_{\mu_2,\lambda_2}) \geq \inf_{z\in E\setminus\{0\}} \max_{t>0} \phi_{\mu_1,\lambda_1}(tz) = c_{\mu_1,\lambda_1}.$$

Particularly, if $\max\{\mu_2 - \mu_1, \lambda_1 - \lambda_2\} > 0$, then $c_{\mu_2,\lambda_2} > c_{\mu_1,\lambda_1}$ follows from the following equality:

$$\phi_{\mu_2,\lambda_2}(z) = \phi_{\mu_1,\lambda_1}(z) + \frac{\mu_2 - \mu_1}{2}|z|_2^2 + (\lambda_1 - \lambda_2)\int_{\mathbb{R}^N} F(|z|)\,dx. \qquad \square$$

By assumption (VP), we have

$$V_\epsilon(x) = V(\epsilon x) \to V(0) \quad \text{and} \quad P_\epsilon(x) = P(\epsilon x) \to P(0) \quad \text{in } L^2_{loc}(\mathbb{R}^N) \tag{3-1}$$

as $\epsilon \to 0$. Let be $\mu_0 = V(0)$ and $\lambda_0 = P(0)$. The following lemma is the key to the concentration behavior:

Lemma 3.2. $$\limsup_{\epsilon\to 0} c_\epsilon \leq c_{\mu_0,\lambda_0}.$$

Proof. It is equivalent to prove that

$$c_\epsilon \leq c_{\mu_0,\lambda_0} + o_\epsilon(1).$$

Let z_0 be the critical point such that $\phi_{\mu_0,\lambda_0}(z_0) = c_{\mu_0,\lambda_0}$. A standard argument implies that there exists $t_0 > 0$, such that $\phi_{\mu_0,\lambda_0}(t_0z_0) < -1$. For any $t \in [0, t_0]$, let

$$f_0(t) = \phi_{\mu_0,\lambda_0}(tz_0) \quad \text{and} \quad f_\epsilon(t) = \phi_\epsilon(tz_0).$$

Claim: For each $z \in E$ fixed,

$$\phi_\epsilon(z) \to \phi_{\mu_0,\lambda_0}(z). \tag{3-2}$$

First observe that, in view of the Sobolev embedding theorems, $z \in E$ implies that $z \in L^p(\mathbb{R}^N)$, $p \in [2, 2^*]$. Thus, for any $\eta > 0$, we have for large ρ

$$|z|^2_{2,\mathbb{R}^N\setminus B_\rho(0)} < \eta \quad \text{and} \quad |z|^p_{p,\mathbb{R}^N\setminus B_\rho(0)} < \eta. \tag{3-3}$$

Furthermore, considering (3-1), we can assert that for any $x \in B_\rho(0)$, the relations

$$|V_\epsilon(x) - \mu_0| < \eta \quad \text{and} \quad |P_\epsilon(x) - \lambda_0| < \eta \tag{3-4}$$

hold for small ϵ.

Hence, by using (3-3) and (3-4), we deduce, for small ϵ,

$$\begin{aligned}
&|\phi_\epsilon(z)-\phi_{\mu_0,\lambda_0}(z)| \\
&\qquad = \left|\frac{1}{2}\int_{\mathbb{R}^N}[V_\epsilon(x)-\mu_0]|z|^2\,dx-\int_{\mathbb{R}^N}[P_\epsilon(x)-\lambda_0]F(|z|)\,dx\right| \\
&\qquad \le \frac{1}{2}\int_{\mathbb{R}^N}|V_\epsilon(x)-\mu_0|\cdot|z|^2\,dx+\int_{\mathbb{R}^N}|P_\epsilon(x)-\lambda_0|F(|z|)\,dx \\
&\qquad = \frac{1}{2}\int_{B_\rho(0)}|V_\epsilon(x)-\mu_0|\cdot|z|^2\,dx+\frac{1}{2}\int_{\mathbb{R}^N\setminus B_\rho(0)}|V_\epsilon(x)-\mu_0|\cdot|z|^2\,dx \\
&\qquad\qquad +\int_{B_\rho(0)}|P_\epsilon(x)-\lambda_0|F(|z|)\,dx+\int_{\mathbb{R}^N\setminus B_\rho(0)}|P_\epsilon(x)-\lambda_0|F(|z|)\,dx \\
&\qquad \le c\cdot\eta,
\end{aligned}$$

proving (3-2).

Now together with $\phi_{\mu_0,\lambda_0}(t_0z_0)<-1$, we have for small ϵ, $\phi_\epsilon(t_0z_0)<-\frac{1}{2}$. This implies that f_ϵ admits a maximum in $(0,t_0)$. Observe that $\{f_\epsilon\}_{\epsilon>0}$ and $\{f'_\epsilon\}_{\epsilon>0}$ are both uniformly bounded.

By a simple application of Arzela–Ascoli theorem, we have that $\{f_\epsilon\}_{\epsilon>0}\subseteq C([0,t_0])$ is compact. Our claim implies that f_ϵ converges pointwise to f_0. The compactness of $\{f_\epsilon\}_{\epsilon>0}$ tells us $|f_\epsilon-f_0|_\infty\to 0$ as $\epsilon\to 0$. Together with $t_0>1$ and $\phi_\epsilon(t_0z_0)<-\frac{1}{2}$, we deduce that

$$\begin{aligned}
c_\epsilon=\inf_{z\in E\setminus\{0\}}\max_{t>0}\phi_\epsilon(tz) &\le \max_{t\in[0,t_0]}\phi_\epsilon(tz_0) \\
&= \max_{t\in[0,t_0]} f_\epsilon(t) \\
&\le \max_{t\in[0,t_0]} f_0(t)+o_\epsilon(1) \\
&= \max_{t\in[0,t_0]}\phi_{\mu_0,\lambda_0}(tz_0)+o_\epsilon(1)=c_{\mu_0,\lambda_0}+o_\epsilon(1). \qquad \square
\end{aligned}$$

The following lemma is our main result, which shows that (2-1) has (at least) one ground state solution for all small ϵ.

Lemma 3.3. *Under the assumptions of Theorem 1.1, if $\epsilon>0$ is small enough, then c_ϵ is attained.*

Proof. Ekeland's variational principle implies that there exists

$$\{z_n : n=1,2,\dots\}\subseteq\mathcal{N}_\epsilon,$$

such that

$$\phi_\epsilon(z_n)\to c_\epsilon \quad \text{as } n\to\infty, \quad \text{and} \quad \phi'_\epsilon(z_n)\to 0 \quad \text{as } n\to\infty. \tag{3-5}$$

We also know that $\{z_n : n = 1, 2, \ldots\}$ is bounded. Thus $z_n \rightharpoonup z_\epsilon$ for some $z_\epsilon \in E$.

Notice that if $z_\epsilon \neq 0$, we are done. Indeed, since $\phi'_\epsilon(z_n) \to 0$ as $n \to \infty$, for any $v \in E$, we have

$$\langle z_n, v\rangle_\epsilon - \int_{\mathbb{R}^N} P_\epsilon(x) f(|z_n|) z_n v\, dx = \phi'_\epsilon(z_n)v \to 0$$

as $n \to \infty$. By using of $z_n \rightharpoonup z_\epsilon$ and the Sobolev embedding theorem, we have that

$$\begin{aligned}\phi'_\epsilon(z_\epsilon)v &= \langle z_\epsilon, v\rangle_\epsilon - \int_{\mathbb{R}^N} P_\epsilon(x) f(|z_\epsilon|) z_\epsilon v\, dx \\ &= \lim_{n\to\infty}\left[\langle z_n, v\rangle_\epsilon - \int_{\mathbb{R}^N} P_\epsilon(x) f(|z_n|) z_n v\, dx\right] = 0\end{aligned}$$

holds for any $v \in E$, which implies that $z_\epsilon \in \mathcal{N}_\epsilon$. Furthermore, in view of Fatou's lemma, (3-5) implies

$$\begin{aligned}\phi_\epsilon(z_\epsilon) \geq \inf_{z\in\mathcal{N}_\epsilon} \phi_\epsilon(z) = c_\epsilon &= \phi_\epsilon(z_n) + o_n(1) \\ &= \phi_\epsilon(z_n) - \tfrac{1}{2}\phi'_\epsilon(z_n)z_n + o_n(1) \\ &= \int_{\mathbb{R}^N} P_\epsilon(x)\hat{F}(|z_n|)\, dx + o_n(1) \\ &\geq \int_{\mathbb{R}^N} P_\epsilon(x)\hat{F}(|z_\epsilon|)\, dx = \phi_\epsilon(z_\epsilon).\end{aligned}$$

Thus $c_\epsilon = \phi_\epsilon(z_\epsilon)$.

Now assume for contradiction that $z_\epsilon = 0$ for all small ϵ. Here we assume that V leads the behavior. Similarly, we can deal with the case where P leads the behavior.

Take $\kappa \in (\tau, \tau_\infty)$ and let be $\eta := \gamma_v$. Denote

$$\begin{aligned} V^\kappa(x) &= \max\{\kappa, V(x)\}, \quad V^\kappa_\epsilon(x) = V^\kappa(\epsilon x). \\ P^\eta(x) &= \min\{\eta, P(x)\}, \quad P^\eta_\epsilon(x) = P^\eta(\epsilon x).\end{aligned}$$

Let

$$\phi^{\kappa,\eta}_\epsilon(z) = \frac{1}{2}\int_{\mathbb{R}^N} [|\Delta z|^2 + V^\kappa_\epsilon(x)|z|^2]\, dx - \int_{\mathbb{R}^N} P^\eta_\epsilon(x) F(|z|)\, dx,$$

and $c^{\kappa,\eta}_\epsilon = \inf_{z\in\mathcal{N}^{\kappa,\eta}_\epsilon} \phi^{\kappa,\eta}_\epsilon(z)$, where $\mathcal{N}^{\kappa,\eta}_\epsilon$ is the Nehari manifold corresponding to $\phi^{\kappa,\eta}_\epsilon$. Then

$$\phi_\epsilon(z) = \phi^{\kappa,\eta}_\epsilon(z) + \frac{1}{2}\int_{\mathbb{R}^N} [V_\epsilon(x) - V^\kappa_\epsilon(x)]|z|^2\, dx - \int_{\mathbb{R}^N} [P_\epsilon(x) - P^\eta_\epsilon(x)] F(|z|)\, dx.$$

Denote

$$\begin{aligned}\mathcal{O} &= \{x \in \mathbb{R}^N : V(x) \leq \kappa\}, \qquad \mathcal{O}_\epsilon = \{x \in \mathbb{R}^N : \epsilon x \in \mathcal{O}\}. \\ \mathcal{O}' &= \{x \in \mathbb{R}^N : P(x) > \eta\}, \qquad \mathcal{O}'_\epsilon = \{x \in \mathbb{R}^N : \epsilon x \in \mathcal{O}'\}.\end{aligned}$$

Notice that $\mathcal{O}_\epsilon$ and $\mathcal{O}'_\epsilon$ are both bounded for given ϵ, and for any z_n, there exists a positive real number t_n, such that $t_n z_n \in \mathcal{N}_\epsilon^{\kappa,\eta}$. Since $\{z_n : n = 1, 2, \ldots\}$ is bounded and the distance between 0 and $\mathcal{N}_\epsilon^{\kappa,\eta}$ is strictly positive, we know that the sequence of positive numbers $\{t_n : n = 1, 2, \ldots\}$ is also bounded. Without loss of generality, assume that $t_n \leq D$ for some constant D, $n = 1, 2, \ldots$.

By the Sobolev embedding theorem, we have a compact embedding $E \hookrightarrow L^q_{loc}(\mathbb{R}^N)$, where $q \in [2, 2^*)$, thus $z_n \rightharpoonup z_\epsilon = 0$ implies that $z_n \to 0$ in $L^q_{loc}(\mathbb{R}^N)$. Since V and P are both L^∞-functions, we have by Hölder's inequalities

$$\left|\frac{1}{2}\int_{\mathcal{O}_\epsilon}[V_\epsilon(x) - V_\epsilon^\kappa(x)]|t_n z_n|^2\,dx\right| \leq D^2|V|_\infty \int_{\mathcal{O}_\epsilon}|z_n|^2\,dx \to 0,$$

as $n \to \infty$. Thus

$$\frac{1}{2}\int_{\mathcal{O}_\epsilon}[V_\epsilon(x) - V_\epsilon^\kappa(x)]|t_n z_n|^2\,dx = o_n(1).$$

Similarly,

$$\left|\int_{\mathcal{O}'_\epsilon}[P_\epsilon(x) - P_\epsilon^\eta(x)]F(|t_n z_n|)\,dx\right| \to 0 \quad \text{as } n \to \infty.$$

Now taking all the above into one package, together with $\phi_\epsilon(t_n z_n) \leq \phi_\epsilon(z_n)$ and (3-5), we have

$$\begin{aligned} c_\epsilon^{\kappa,\eta} &= \inf_{z\in\mathcal{N}_\epsilon^{\kappa,\eta}} \phi_\epsilon^{\kappa,\eta}(z) \leq \phi_\epsilon^{\kappa,\eta}(t_n z_n) \\ &= \phi_\epsilon(t_n z_n) - \frac{1}{2}\int_{\mathbb{R}^N}[V_\epsilon(x) - V_\epsilon^\kappa(x)]|t_n z_n|^2\,dx \\ &\qquad + \int_{\mathbb{R}^N}[P_\epsilon(x) - P_\epsilon^\eta(x)]F(|t_n z_n|)\,dx \\ &\leq \phi_\epsilon(z_n) - \frac{1}{2}\int_{\mathcal{O}_\epsilon}[V_\epsilon(x) - V_\epsilon^\kappa(x)]|t_n z_n|^2\,dx \\ &\qquad + \int_{\mathcal{O}'_\epsilon}[P_\epsilon(x) - P_\epsilon^\eta(x)]F(|t_n z_n|)\,dx \\ &= c_\epsilon + o_n(1). \end{aligned}$$

Thus $c_\epsilon^{\kappa,\eta} \leq c_\epsilon$.

Observe that

$$\phi_\epsilon^{\kappa,\eta}(z) \geq \frac{1}{2}\int_{\mathbb{R}^N}(|\Delta z|^2 + \kappa|z|^2)\,dx - \eta\int_{\mathbb{R}^N}F(|z|)\,dx = \phi_{\kappa,\eta}(z).$$

From this, we deduce that

$$c_{\kappa,\eta} = \inf_{z\in E\setminus\{0\}}\max_{t>0}\phi_{\kappa,\eta}(tz) \leq \inf_{z\in E\setminus\{0\}}\max_{t>0}\phi_\epsilon^{\kappa,\eta}(tz) = c_\epsilon^{\kappa,\eta}.$$

This holds for any small ϵ, which implies that

$$c_{\kappa,\eta} \le \liminf_{\epsilon\to 0} c_\epsilon^{\kappa,\eta} \le \limsup_{\epsilon\to 0} c_\epsilon^{\kappa,\eta} \le c_{\kappa,\eta},$$

the last inequality followed by our key lemma (Lemma 3.2) since

$$V^\kappa(0) = \max\{\kappa, V(0)\} = \max\{\kappa, \tau\} = \kappa$$

and

$$P^\eta(0) = \min\{\eta, P(0)\} = \min\{\eta, \gamma_v\} = \eta.$$

Thus $c_\epsilon^{\kappa,\eta} \to c_{\kappa,\eta}$ as $\epsilon \to 0$. We have already seen that $c_\epsilon^{\kappa,\eta} \le c_\epsilon$. Letting $\epsilon \to 0$, we have

$$c_{\kappa,\eta} \le \liminf_{\epsilon\to 0} c_\epsilon \le \limsup_{\epsilon\to 0} c_\epsilon \le c_{\tau,\eta},$$

the last inequality is also followed by Lemma 3.2. The above inequality contradicts our Lemma 3.1 since $\kappa > \tau$. Thus $z_\epsilon \ne 0$, which is a ground state solution. □

Proof of Theorem 1.1. This theorem is another description of Lemma 3.3; that is, (2-1) has (at least) one ground state solution for small ϵ. □

4. Proof of Theorem 1.2

Now, using the same notation as in the previous section, we are now ready to show the concentration of the ground state solution given in Theorem 1.1.

Here we recall the description of Theorem 1.2.

Proposition 4.1. *Let* (VP), (f_1) *and* (f_2) *hold. Assume additionally that either* (V) *or* (P) *holds. Let* z_ϵ *be the solution of* (2-1), *given by Lemma 3.3. For any* $\epsilon_{\epsilon_n} > 0$ *with* $\lim_{n\to\infty} \epsilon_n = 0$, *up to a subsequence*, *there exists* $x_{\epsilon_n} \in \mathbb{R}^N$ *which is a maximum point of* $|z_{\epsilon_n}|$, *such that*

$$\operatorname{dist}(\epsilon_n x_{\epsilon_n}, \mathscr{C}) \to 0 \quad \textit{as} \quad n \to \infty.$$

Let $w_{\epsilon_n}(x) := z_{\epsilon_n}(x + x_{\epsilon_n})$. *Then* $w_{\epsilon_n} \to w_0$ *in* E, *where* w_0 *is a ground state solution of*

$$\Delta^2 z + V(x_0) z = P(x_0) f(|z|) z, \qquad x_0 \in \mathscr{C}.$$

Proof. Without loss of generality, we assume (V) holds. Clearly, $\{z_\epsilon\} \subseteq \mathcal{N}_\epsilon$ is bounded.

Claim 1: $\{z_\epsilon\}$ is nonvanishing.

Suppose for contradiction $\{z_\epsilon\}$ is vanishing. This means $z_\epsilon \to 0$ as $\epsilon \to 0$ in $L^p(\mathbb{R}^N)$. Then

$$\|z_\epsilon\|_\epsilon^2 = \int_{\mathbb{R}^N} P_\epsilon(x) f(|z_\epsilon|) z_\epsilon^2 \, dx \to 0, \quad \text{as } \epsilon \to 0,$$

which contradicts the fact that 0 is bounded away from $\mathcal{N}_\epsilon$. Hence, the sequence $\{z_\epsilon\}$ is nonvanishing. Thus we know there exist $R > 0$, $\alpha > 0$ and $\{x_\epsilon\} \subseteq \mathbb{R}^N$ such that

$$\int_{B_R(x_\epsilon)} |z_\epsilon|^2 \, dx \geq \alpha.$$

Claim 2: Let $\{x_\epsilon\}$ be the nonvanishing points found in Claim 1. Then $\{\epsilon x_\epsilon\}$ is bounded.

Suppose for contradiction that $|\epsilon x_\epsilon| \to \infty$. Up to a subsequence, we have

$$V_\epsilon(x_\epsilon) \to V_\infty \quad \text{and} \quad P_\epsilon(x_\epsilon) \to P_\infty,$$

for some positive numbers V_∞ and P_∞. Let

$$w_\epsilon(x) := z_\epsilon(x + x_\epsilon), \quad \tilde{V}_\epsilon(x) := V_\epsilon(x + x_\epsilon) \quad \text{and} \quad \tilde{P}_\epsilon(x) := P_\epsilon(x + x_\epsilon).$$

Notice that the boundedness of z_ϵ implies that of $\{w_\epsilon\}$, thus, up to a subsequence, we have $w_\epsilon \rightharpoonup w_0 \in E$. Particularly, $w_0 \neq 0$. Indeed, by Sobolev embedding theorems, $w_\epsilon \rightharpoonup w_0$ in E implies $w_\epsilon \to w_0$ in $L^2_{loc}(\mathbb{R}^N)$. Together with

$$\int_{B_R(0)} |w_\epsilon|^2 \, dx = \int_{B_R(x_\epsilon)} |z_\epsilon|^2 \, dx \geq \alpha,$$

we have $\int_{B_R(0)} |w_0|^2 \, dx \geq \alpha$, i.e., $w_0 \neq 0$.

Observe that w_ϵ is a ground state solution to the following equation:

$$\Delta^2 w + \tilde{V}_\epsilon(x) w = \tilde{P}_\epsilon(x) f(|w|) w.$$

Now for any $\varphi \in C_c^\infty(\mathbb{R}^N)$, by using the dominated convergence theorem, we have

$$\begin{aligned} 0 &= \lim_{\epsilon \to 0} \int_{\mathbb{R}^N} [\Delta^2 w_\epsilon + \tilde{V}_\epsilon(x) w_\epsilon - \tilde{P}_\epsilon(x) f(|w_\epsilon|) w_\epsilon] \varphi \, dx \\ &= \int_{\mathbb{R}^N} (\Delta^2 w_0 + V_\infty w_0 - P_\infty f(|w_0|) w_0) \varphi \, dx \\ &= \phi'_{V_\infty, P_\infty}(w_0) \varphi. \end{aligned}$$

It follows from this that

$$\begin{aligned} c_\epsilon = \phi_\epsilon(z_\epsilon) &= \frac{1}{2} \int_{\mathbb{R}^N} [|\Delta z_\epsilon|^2 + V_\epsilon(x) |z_\epsilon|^2] \, dx - \int_{\mathbb{R}^N} P_\epsilon(x) F(|z_\epsilon|) \, dx \\ &= \frac{1}{2} \int_{\mathbb{R}^N} [|\Delta w_\epsilon|^2 + \tilde{V}_\epsilon(x) |w_\epsilon|^2] \, dx - \int_{\mathbb{R}^N} \tilde{P}_\epsilon(x) F(|w_\epsilon|) \, dx \\ &= \int_{\mathbb{R}^N} \tilde{P}_\epsilon(x) \hat{F}(|w_\epsilon|) \, dx \\ &\geq \int_{\mathbb{R}^N} P_\infty \hat{F}(|w_0|) \, dx = \phi_{V_\infty, P_\infty}(w_0) \geq c_{V_\infty, P_\infty}. \end{aligned}$$

Together with $c_\epsilon \le c_{\tau,\eta} + o_\epsilon(1)$ and $\max\{V_\infty - \tau, \eta - P_\infty\} > 0$, we have a contradiction with Lemma 3.1. Thus $\{\epsilon x_\epsilon\}$ is bounded.

Claim 3: $\operatorname{dist}(\epsilon x_\epsilon, \mathscr{A}_v) \to 0$.

Suppose for contradiction, up to a subsequence, $\epsilon x_\epsilon \to x_0 \notin \mathscr{A}_v$ as $\epsilon \to 0$ since $\{\epsilon x_\epsilon\}$ is bounded.

Denote

$$V_\epsilon(x_\epsilon) \to V(x_0) \quad \text{and} \quad P_\epsilon(x_\epsilon) \to P(x_0),$$

as $\epsilon \to 0$. Then similarly to the proof of Claim 2, we have $c_\epsilon \ge c_{V(x_0),P(x_0)}$. Recall that

$$\mathscr{A}_v := \{x \in \mathcal{V} : P(x) = \gamma_v\} \cup \{x \notin \mathcal{V} : P(x) > \gamma_v\}.$$

It is easy to check that $x_0 \notin \mathscr{A}_v$ implies that $\max\{V(x_0) - \tau, \eta - P(x_0)\} > 0$. Together with $c_\epsilon \le c_{\tau,\eta} + o_\epsilon(1)$, we have a contradiction with Lemma 3.1. Thus $\operatorname{dist}(\epsilon x_\epsilon, \mathscr{A}_v) \to 0$. At this point, as was argued in Claim 2, the transformed solution $w_\epsilon(x) := z_\epsilon(x + x_\epsilon)$ will converge weakly (up to a subsequence) to w_0 which is a ground state solution of

$$\Delta^2 z + V(x_0) z = P(x_0) f(|z|) z, \qquad x_0 \in \mathscr{A}_v.$$

Claim 4: For any $\epsilon_n > 0$ with $\lim_{n\to\infty} \epsilon_n = 0$, up to a subsequence, we have $w_{\epsilon_n} \to w_0$ as $n \to \infty$ in $E \setminus \{0\}$.

Let $y_n := w_{\epsilon_n} - w_0$, then $y_n \rightharpoonup 0$ as $n \to \infty$. Recall that we have

$$\begin{gathered}\Delta^2 w_{\epsilon_n} + \tilde{V}_{\epsilon_n}(x) w_{\epsilon_n} = \tilde{P}_{\epsilon_n}(x) f(|w_{\epsilon_n}|) w_{\epsilon_n},\\ \Delta^2 w_0 + V(x_0) w_0 = P(x_0) f(|w_0|) w_0\end{gathered}$$

It follows that

$$\begin{aligned}\int_{\mathbb{R}^N} [\Delta w_{\epsilon_n} \Delta y_n + \tilde{V}_{\epsilon_n}(x) w_{\epsilon_n} y_n]\,dx &= \int_{\mathbb{R}^N} \tilde{P}_{\epsilon_n}(x) f(|w_{\epsilon_n}|) w_{\epsilon_n} y_n\,dx,\\ \int_{\mathbb{R}^N} [\Delta w_0 \Delta y_n + V(x_0) w_0 y_n]\,dx &= \int_{\mathbb{R}^N} P(x_0) f(|w_0|) w_0 y_n\,dx\end{aligned}$$

Notice that

$$\begin{aligned}\langle V(x_0) w_0, y_n\rangle_2 &= \langle w_0, V(x_0) y_n\rangle_2\\ &= \langle w_0, \tilde{V}_{\epsilon_n}(x) y_n + [V(x_0) - \tilde{V}_{\epsilon_n}(x)] y_n\rangle_2\\ &= \langle w_0, \tilde{V}_{\epsilon_n}(x) y_n\rangle_2 + \langle w_0, [V(x_0) - \tilde{V}_{\epsilon_n}(x)] y_n\rangle_2\\ &= \langle \tilde{V}_{\epsilon_n}(x) w_0, y_n\rangle_2 + o_{\epsilon_n}(1).\end{aligned}$$

Now we have

$$\int_{\mathbb{R}^N} [\Delta w_{\epsilon_n} \Delta y_n + \tilde{V}_{\epsilon_n}(x) w_{\epsilon_n} y_n]\, dx = \int_{\mathbb{R}^N} \tilde{P}_{\epsilon_n}(x) f(|w_{\epsilon_n}|) w_{\epsilon_n} y_n\, dx,$$

$$\int_{\mathbb{R}^N} [\Delta w_0 \Delta y_n + \tilde{V}_{\epsilon_n}(x) w_0 y_n]\, dx = \int_{\mathbb{R}^N} P(x_0) f(|w_0|) w_0 y_n\, dx + o_{\epsilon_n}(1)$$

Then, by (f_1), (f_2), it is easy to check

$$\int_{\mathbb{R}^N} [\Delta y_n \Delta y_n + \tilde{V}_{\epsilon_n}(x) y_n^2]\, dx = o_{\epsilon_n}(1),$$

which implies $\|y_n\|_{\epsilon_n} = o_{\epsilon_n}(1)$, ending the proof. □

Final remark

Our approach can be described in a more abstract way to deal with some general variational problems. Indeed, if we rewrite (2-1) as

$$\begin{cases} (\Delta^2 + \alpha) z + (V_\epsilon(x) - \alpha) z = P_\epsilon(x) f(|z|) z, & x \in \mathbb{R}^N, \\ z(x) \to 0 & \text{as } |x| \to \infty, \end{cases}$$

where $0 < \alpha < \inf_{x \in \mathbb{R}^N} V(x)$, then we are led to the abstract equation of the form

$$Lz + M_\epsilon(x) z = P_\epsilon(x) \nabla G(z), \tag{4-1}$$

in which L is a positive defined differential operator on $L^2(\mathbb{R}^N)$ and $M(x)$, $P(x)$ satisfy the condition (VP). Let $E := \mathcal{D}(L^{1/2})$ be equipped with the scalar product

$$\langle u, v \rangle = \langle L^{1/2} u, L^{1/2} v \rangle_2$$

and the induced norm $\|u\| = \langle u, u \rangle^{1/2}$. Then the associated energy functional of (4-1) is of the form

$$\phi_\epsilon(z) = \frac{1}{2} \|z\|^2 + \frac{1}{2} \int_{\mathbb{R}^N} M_\epsilon(x) |z|^2\, dx - \int_{\mathbb{R}^N} P_\epsilon(x) G(z)\, dx,$$

and our arguments are in general feasible to such problems under some suitable assumptions on the nonlinear function G.

In fact, (4-1) is related to several equations appearing in quantum physics, including the Schrödinger equations and the fractional Schrödinger equations, etc. Therefore, our approach covers the semiclassical behavior of different equations under a general class of subcritical nonlinearities.

References

[Ambrosetti and Malchiodi 2006] A. Ambrosetti and A. Malchiodi, *Perturbation methods and semilinear elliptic problems on* $\mathbb{R}^n$, Progress in Mathematics **240**, Birkhäuser, Basel, 2006. MR Zbl

[Ambrosetti et al. 1997] A. Ambrosetti, M. Badiale, and S. Cingolani, "Semiclassical states of nonlinear Schrödinger equations", *Arch. Rational Mech. Anal.* **140**:3 (1997), 285–300. MR Zbl

[Arioli and Szulkin 2003] G. Arioli and A. Szulkin, "A semilinear Schrödinger equation in the presence of a magnetic field", *Arch. Ration. Mech. Anal.* **170**:4 (2003), 277–295. MR Zbl

[Bartsch et al. 2001] T. Bartsch, A. Pankov, and Z.-Q. Wang, "Nonlinear Schrödinger equations with steep potential well", *Commun. Contemp. Math.* **3**:4 (2001), 549–569. MR Zbl

[Ben-Artzi et al. 2000] M. Ben-Artzi, H. Koch, and J.-C. Saut, "Dispersion estimates for fourth order Schrödinger equations", *C. R. Acad. Sci. Paris Sér. I Math.* **330**:2 (2000), 87–92. MR Zbl

[Byeon and Jeanjean 2007] J. Byeon and L. Jeanjean, "Standing waves for nonlinear Schrödinger equations with a general nonlinearity", *Arch. Ration. Mech. Anal.* **185**:2 (2007), 185–200. MR Zbl

[Byeon and Wang 2003] J. Byeon and Z.-Q. Wang, "Standing waves with a critical frequency for nonlinear Schrödinger equations, II", *Calc. Var. Partial Differential Equations* **18**:2 (2003), 207–219. MR Zbl

[Coti Zelati and Rabinowitz 1992] V. Coti Zelati and P. H. Rabinowitz, "Homoclinic type solutions for a semilinear elliptic PDE on $\mathbb{R}^n$", *Comm. Pure Appl. Math.* **45**:10 (1992), 1217–1269. MR Zbl

[Ding and Liu 2013] Y. Ding and X. Liu, "Semiclassical solutions of Schrödinger equations with magnetic fields and critical nonlinearities", *Manuscripta Math.* **140**:1-2 (2013), 51–82. MR Zbl

[Ding and Szulkin 2007] Y. Ding and A. Szulkin, "Bound states for semilinear Schrödinger equations with sign-changing potential", *Calc. Var. Partial Differential Equations* **29**:3 (2007), 397–419. MR Zbl

[Ding and Tanaka 2003] Y. Ding and K. Tanaka, "Multiplicity of positive solutions of a nonlinear Schrödinger equation", *Manuscripta Math.* **112**:1 (2003), 109–135. MR Zbl

[Ding and Wei 2007] Y. Ding and J. Wei, "Semiclassical states for nonlinear Schrödinger equations with sign-changing potentials", *J. Funct. Anal.* **251**:2 (2007), 546–572. MR Zbl

[Esteban and Lions 1989] M. J. Esteban and P.-L. Lions, "Stationary solutions of nonlinear Schrödinger equations with an external magnetic field", pp. 401–449 in *Partial differential equations and the calculus of variations*, vol. I, edited by F. Colombini et al., Progr. Nonlinear Differential Equations Appl. **1**, Birkhäuser, Boston, 1989. MR Zbl

[Fibich et al. 2002] G. Fibich, B. Ilan, and G. Papanicolaou, "Self-focusing with fourth-order dispersion", *SIAM J. Appl. Math.* **62**:4 (2002), 1437–1462. MR Zbl

[Floer and Weinstein 1986] A. Floer and A. Weinstein, "Nonspreading wave packets for the cubic Schrödinger equation with a bounded potential", *J. Funct. Anal.* **69**:3 (1986), 397–408. MR Zbl

[Gui 1996] C. Gui, "Existence of multi-bump solutions for nonlinear Schrödinger equations via variational method", *Comm. Partial Differential Equations* **21**:5-6 (1996), 787–820. MR Zbl

[Guo and Wang 2002] B. Guo and B. Wang, "The global Cauchy problem and scattering of solutions for nonlinear Schrödinger equations in H^s", *Differential Integral Equations* **15**:9 (2002), 1073–1083. MR Zbl

[Hao et al. 2006] C. Hao, L. Hsiao, and B. Wang, "Wellposedness for the fourth order nonlinear Schrödinger equations", *J. Math. Anal. Appl.* **320**:1 (2006), 246–265. MR Zbl

[Hao et al. 2007] C. Hao, L. Hsiao, and B. Wang, "Well-posedness of Cauchy problem for the fourth order nonlinear Schrödinger equations in multi-dimensional spaces", *J. Math. Anal. Appl.* **328**:1 (2007), 58–83. MR Zbl

[Karpman and Shagalov 2000] V. I. Karpman and A. G. Shagalov, "Stability of solitons described by nonlinear Schrödinger-type equations with higher-order dispersion", *Phys. D* **144**:1-2 (2000), 194–210. MR Zbl

[Kurata 2000] K. Kurata, "Existence and semi-classical limit of the least energy solution to a nonlinear Schrödinger equation with electromagnetic fields", *Nonlinear Anal. Ser. A* **41**:5-6 (2000), 763–778. MR Zbl

[Liang and Zhang 2011] S. Liang and J. Zhang, "Solutions of perturbed Schrödinger equations with electromagnetic fields and critical nonlinearity", *Proc. Edinb. Math. Soc.* (2) **54**:1 (2011), 131–147. MR Zbl

[Ni and Wei 2006] W.-M. Ni and J. Wei, "On positive solutions concentrating on spheres for the Gierer–Meinhardt system", *J. Differential Equations* **221**:1 (2006), 158–189. MR Zbl

[Oh 1988] Y.-G. Oh, "Existence of semiclassical bound states of nonlinear Schrödinger equations with potentials of the class $(V)_a$", *Comm. Partial Differential Equations* **13**:12 (1988), 1499–1519. MR Zbl

[Oh 1990] Y.-G. Oh, "On positive multi-lump bound states of nonlinear Schrödinger equations under multiple well potential", *Comm. Math. Phys.* **131**:2 (1990), 223–253. MR Zbl

[Pankov 2003] A. A. Pankov, "On nontrivial solutions of the nonlinear Schrödinger equation with a magnetic field", *Funktsional. Anal. i Prilozhen.* **37**:1 (2003), 88–91. In Russian; translated in *Funct. Anal. Appl.* **37**:1 (2003), 75–77. MR Zbl

[del Pino and Felmer 1996] M. del Pino and P. L. Felmer, "Local mountain passes for semilinear elliptic problems in unbounded domains", *Calc. Var. Partial Differential Equations* **4**:2 (1996), 121–137. MR Zbl

[Rabinowitz 1992] P. H. Rabinowitz, "On a class of nonlinear Schrödinger equations", *Z. Angew. Math. Phys.* **43**:2 (1992), 270–291. MR Zbl

[Segata 2006] J.-i. Segata, "Modified wave operators for the fourth-order non-linear Schrödinger-type equation with cubic non-linearity", *Math. Methods Appl. Sci.* **29**:15 (2006), 1785–1800. MR

[Sirakov 2002] B. Sirakov, "Standing wave solutions of the nonlinear Schrödinger equation in $\mathbb{R}^N$", *Ann. Mat. Pura Appl.* (4) **181**:1 (2002), 73–83. MR Zbl

[Wang 1993] X. Wang, "On concentration of positive bound states of nonlinear Schrödinger equations", *Comm. Math. Phys.* **153**:2 (1993), 229–244. MR Zbl

[Wang 2008] F. Wang, "On an electromagnetic Schrödinger equation with critical growth", *Nonlinear Anal.* **69**:11 (2008), 4088–4098. MR Zbl

[Willem 1996] M. Willem, *Minimax theorems*, Progress in Nonlinear Differential Equations and their Applications **24**, Birkhäuser, Boston, 1996. MR Zbl

Received July 31, 2015. Revised December 20, 2016.

DONG WANG
SCHOOL OF MATHEMATICS AND PHYSICS
CHANGZHOU UNIVERSITY
213164 JIANGSU
CHINA
wdflight@126.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 289, No. 2, 2017
dx.doi.org/10.2140/pjm.2017.289.489

GLOBAL EXISTENCE OF SMOOTH SOLUTIONS TO EXPONENTIAL WAVE MAPS IN FLRW SPACETIMES

CHANG-HUA WEI AND NING-AN LAI

We consider the Cauchy problem of exponential wave maps in Friedmann–Lemaître–Robertson–Walker (FLRW) spacetimes. Using a weighted energy estimate, we show that the smooth solution of the Cauchy problem for exponential wave maps in FLRW spacetimes exists globally for small initial data.

1. Introduction

Wave maps are the Lorentzian counterparts of harmonic maps. Due to their significance in geometry and physics, wave maps have attracted much attention, and many achievements have been made in recent decades [Choquet-Bruhat 2000; Gu 1980; Klainerman and Machedon 1993; 1995; Shatah and Struwe 1998; 2002; Tao 2001a; 2001b; 2008]. In this paper, we investigate exponential wave maps which were introduced by Eells and Lemaire [1992]. The relationship between wave maps and exponential wave maps has been studied by Chiang and Yang [2007] from the geometric point of view. In this work we study the PDE aspect of the exponential wave maps on a curved Lorentzian manifold.

In this paper, we consider exponential wave maps on a special class of FLRW spacetimes, whose metric takes the form $g = -dt^2 + a^2(t)\,d\sigma^2$ for suitable $a(t)$. The FLRW metric is an exact solution of Einstein's field equations; it describes a homogeneous and isotropic universe. For more details on this metric, we refer to the book [Hawking and Ellis 1973]. In particular, in local coordinates $(t, x_1, \dots, x_m)$, with $m \geq 1$, we consider the following metric

$$g = -dt^2 + t^l \sum_{i=1}^{m} dx_i^2, \tag{1-1}$$

where $l > 2$ and $t > 0$. Here we remark that for the case $a(t) = e^t$, the metric g turns to be the de Sitter metric and much progress has been made on the study of wave type equations on de Sitter spacetimes; see [Kong and Wei 2013; Yagdjian 2012; 2009; Yagdjian and Galstian 2008; 2009].

MSC2010: 35L05.
Keywords: exponential wave map, FLRW, weighted energy estimate.

Before our main result, we give a brief introduction to exponential wave maps.

Basic equation and the main result.

Definition 1.1. An exponential wave map $u : M^{1+m} \to N$ from a $(1+m)$-dimensional Lorentzian manifold $(M^{1+m}, g_{\mu\nu})$ to an n-dimensional manifold $(N, h_{\alpha\beta})$ is a critical point of the exponential energy functional

$$E(u) := \int_M \exp(|du|^2)\, d\,\mathrm{Vol}_M = \int_M \exp(h_{\alpha\beta}\partial_\mu u^\alpha\, \partial_\nu u^\beta g^{\mu\nu})\, d\,\mathrm{Vol}_M, \tag{1-2}$$

where $d\,\mathrm{Vol}_M$ is the volume element of M.

We consider the exponential wave map taking M^{1+m} as the background manifold and $\mathbb{R}$ as the target manifold N in Definition 1.1. Therefore, the exponential energy functional takes the explicit form

$$E(u) = \int_M \exp\biggl(-(\partial_t u)^2 + t^{-l}\sum_{i=1}^m (\partial_i u)^2\biggr) t^{lm/2} dt\, dx_1 \cdots dx_m, \tag{1-3}$$

where $\partial_t u = \frac{\partial u}{\partial t}$ and $\partial_i u = \frac{\partial u}{\partial x_i}$; the corresponding Euler–Lagrange equation reads

$$\begin{aligned}\frac{\partial}{\partial t}\biggl[-2\partial_t u \exp\biggl(-(\partial_t u)^2 + t^{-l}\sum_{i=1}^m (\partial_i u)^2\biggr) t^{lm/2}\biggr] \\ + \sum_{i=1}^m \frac{\partial}{\partial x_i}\biggl[2t^{-l}\partial_i u \exp\biggl(-(\partial_t u)^2 + t^{-l}\sum_{i=1}^m (\partial_i u)^2\biggr) t^{lm/2}\biggr] = 0,\end{aligned} \tag{1-4}$$

i.e., with $\Delta := \sum_{i=1}^m \partial_i^2$,

$$\begin{aligned}-\partial_t^2 u + t^{-l}\Delta u - \frac{lm}{2t}\partial_t u = \partial_t u\, \partial_t\biggl[-(\partial_t u)^2 + t^{-l}\sum_{i=1}^m (\partial_i u)^2\biggr] \\ - \sum_{i=1}^m t^{-l}\partial_i u\, \partial_i\biggl[-(\partial_t u)^2 + t^{-l}\sum_{i=1}^m (\partial_i u)^2\biggr].\end{aligned} \tag{1-5}$$

Remark 1.2. When $l = 0$, (1-5) is exactly the exponential wave map on Minkowski spacetime. Huh [2013] obtained a global existence result for sufficiently small initial data.

Define the wave operator as

$$\Box_g = \frac{1}{\sqrt{|\det g|}}\partial_\mu\bigl(\sqrt{|\det g|}\, g^{\mu\nu}\partial_\nu\bigr) = -\partial_t^2 - \frac{lm}{2t}\partial_t + t^{-l}\sum_{i=1}^m \partial_i^2, \tag{1-6}$$

where the Greek index μ ranges from 0 to m and $\partial_0 := \partial_t$. In (1-6) and throughout the paper, we use the Einstein summation convention, which says that one always takes the summation over the repeated upper and lower indices.

Let

$$Q(\varphi, \psi) = -(\partial_t\varphi)(\partial_t\psi) + t^{-l}\sum_{i=1}^{m}(\partial_i\varphi)(\partial_i\psi). \tag{1-7}$$

Then (1-5) can be rewritten as

$$\Box_g u = -Q(u, Q(u, u)). \tag{1-8}$$

From now on, we consider (1-8) with the following initial data

$$t = 1: \quad u(1, x) = \epsilon f(x), \quad u_t(1, x) = \epsilon q(x), \tag{1-9}$$

where ϵ is a small parameter, f and q are smooth functions with compact support.

The main result of this paper is as follows:

Theorem 1.3. *There exists a positive constant* ϵ_0, *such that the Cauchy problem* (1-8)-(1-9) *has a unique smooth solution on* $[1, +\infty) \times \mathbb{R}^m$ *for any* $\epsilon \in [0, \epsilon_0]$.

The structure of the paper. The paper is organized as follows. In Section 2, we attain pointwise decay estimates of the smooth solution to the linear wave equation by weighted energy estimates. The main theorem is proved in Section 3 by the continuity method and a careful analysis of the nonlinearities. Several further discussions are presented in Section 4.

2. Decay estimates for the linear wave equation

In this section, we investigate the pointwise decay estimates of smooth solution to the following linear wave equation on FLRW spacetimes

$$\Box_g \varphi = 0, \tag{2-1}$$

where the metric g is given by (1-1). It will play an important role in the study of nonlinear (1-8).

For multi-index $I = (I_1, \ldots, I_m)$ with $|I| = \sum_{j=1}^{m}|I_j|$, denote

$$D = \{\partial_1, \ldots, \partial_m\} \quad \text{and} \quad D^I = \partial_1^{I_1} \cdots \partial_m^{I_m}.$$

Furthermore, $\tilde{I} \le I$ means $\tilde{I}_i \le I_i$ for $i = 1, \ldots, m$.

For any function $\varphi(x) = \varphi(x_1, \ldots, x_m)$, we define

$$\|\varphi(x)\|_{L^2} = \left(\int_{\mathbb{R}^m} |\varphi(x)|^2 dx\right)^{1/2}, \quad \|\varphi(x)\|_{L^\infty} := \operatorname{ess}\sup_{x\in\mathbb{R}^m} |\varphi(x)|,$$

and

$$\|\varphi(x)\|_{H^s} = \left(\sum_{i=0}^{s}\big(\|D^i\varphi(x)\|_{L^2}\big)^2\right)^{1/2},$$

where s is an integer and $dx = dx^1 \cdots dx^m$.

From here on, we also use $A \lesssim B$ to denote $A \le CB$ for some positive constant C.

Corresponding to the (2-1) and the specific metric g given by (1-1), we define the energy momentum tensor as

$$T_{\mu\nu}(\varphi) = (\partial_\mu \varphi)(\partial_\nu \varphi) - \frac{1}{2} g_{\mu\nu} |\nabla \varphi|^2, \tag{2-2}$$

where

$$|\nabla\varphi|^2 = g^{\kappa\lambda} \partial_\kappa \varphi \, \partial_\lambda \varphi = -(\partial_t \varphi)^2 + \sum_{i=1}^{m} t^{-l} (\partial_i \varphi)^2.$$

Given a vector field $V = V^\mu \partial_\mu$, define the compatible currents

$$J_\mu^V(\varphi) = T_{\mu\nu}(\varphi) \, V^\nu \tag{2-3}$$

and

$$K^V(\varphi) = \Pi_{\mu\nu}^V T^{\mu\nu}(\varphi), \tag{2-4}$$

where $\Pi_{\mu\nu}^V$ denotes the deformation tensor defined by

$$\Pi_{\mu\nu}^V = \frac{1}{2} (\nabla_\mu V_\nu + \nabla_\nu V_\mu), \tag{2-5}$$

and

$$\nabla_\mu V_\nu = g(\nabla_\mu V, \partial_\nu).$$

In a constant t-slice, the induced volume form is

$$d \,\mathrm{Vol}_t = t^{lm/2} dx_1 \cdots dx_m. \tag{2-6}$$

With above notations, we have for $i, j = 1, \ldots, m$ and $i \neq j$

$$T_{00}(\varphi) = \frac{1}{2} \biggl((\partial_t \varphi)^2 + \sum_{i=1}^{m} t^{-l} (\partial_i \varphi)^2 \biggr), \tag{2-7}$$

$$T_{0i} = \partial_t \varphi \, \partial_i \varphi, \quad T_{ij} = \partial_i \varphi \, \partial_j \varphi, \tag{2-8}$$

and

$$T_{ii} = \frac{1}{2} \biggl(t^l (\partial_t \varphi)^2 + (\partial_i \varphi)^2 - \sum_{j \neq i} (\partial_j \varphi)^2 \biggr). \tag{2-9}$$

We recall the following energy identity without proof; see [Alinhac 2010].

Lemma 2.1. *For a solution φ of the equation $\Box_g \varphi = f$, we have*

$$\nabla^\mu T_{\mu\nu} = (\Box_g \varphi)(\partial_\nu \varphi) \quad \textit{and} \quad \nabla^\mu J_\mu^V(\varphi) = K^V(\varphi) + \Box_g \varphi \cdot V(\varphi). \tag{2-10}$$

The energy density $e(V, \upsilon)$ of the function φ at time t with respect to a past pointed timelike vector field V is the nonnegative number given by

$$e(V, \upsilon) = J^V_\alpha \upsilon^\alpha = T_{\alpha\beta}(\varphi) V^\beta \upsilon^\alpha, \tag{2-11}$$

where υ^α is the α-th component of the past oriented unit normal vector $\upsilon = -\partial_t$.

Given a past pointed vector field V, by Lemma 2.1 and the divergence theorem, we easily get the following lemma:

Lemma 2.2. *Integrating* (2-10) *over the spacetime domain* $D = \{[1, t] \times \mathbb{R}^m\}$,

$$\int_{\Sigma_t} J^V_\alpha \upsilon^\alpha \, d\,\mathrm{Vol}_t - \int_{\Sigma_1} J^V_\alpha \upsilon^\alpha \, d\,\mathrm{Vol}_1 = \int_1^t \int_{\Sigma_\tau} (K^V(\varphi) + \Box_g \varphi V(\varphi)) \, d\,\mathrm{Vol}_\tau \, d\tau, \tag{2-12}$$

where Σ_t denotes the spacelike hypersurface with $t =$ constant.

From now on, we take $V = -\partial_t$ so that

$$\Pi^V_{\mu\nu} = -\frac{1}{2}(g_{\nu\kappa}\Gamma^\kappa_{\mu t} + g_{\mu\kappa}\Gamma^\kappa_{\nu t}). \tag{2-13}$$

Then for $i = 1, \dots, m$,

$$\Pi^V_{ii} = -\frac{l}{2} t^{l-1}, \tag{2-14}$$

and for $i \neq j$,

$$\Pi^V_{ij} = 0, \quad \Pi^V_{0i} = 0, \quad \text{and} \quad \Pi^V_{00} = 0, \tag{2-15}$$

where we have assumed that $x_0 = t$. By (2-4), we have

$$K^{-\partial_t}(\varphi) = \frac{l}{4}\left[(m-2)\sum_{i=1}^m \frac{1}{t^{l+1}}(\partial_i \varphi)^2 - \frac{m}{t}(\partial_t \varphi)^2\right]. \tag{2-16}$$

By (2-11), we have

$$e(V, \upsilon) = \frac{1}{2}\left((\partial_t \varphi)^2 + \sum_{i=1}^m t^{-l}(\partial_i \varphi)^2\right). \tag{2-17}$$

For a constant t-slice, we define

$$\begin{cases} E^{I,I_0}_0(t) = \dfrac{1}{2}\displaystyle\int_{\Sigma_t} (\partial_t \partial_t^{I_0} D^I \varphi)^2 \, d\,\mathrm{Vol}_t, \\ E^{I,I_0}_1(t) = \dfrac{1}{2}\displaystyle\int_{\Sigma_t} \left(\sum_{i=1}^m t^{-l}(\partial_i \partial_t^{I_0} D^I \varphi)^2\right) d\,\mathrm{Vol}_t \end{cases} \tag{2-18}$$

and

$$E^{I,I_0}(t) = E^{I,I_0}_0(t) + E^{I,I_0}_1(t), \tag{2-19}$$

where I_0 and $|I|$ are nonnegative integers, $|I|$ is the magnitude of any multi-index I.

Then, by above calculations and Lemma 2.2, the following zeroth-order energy identity holds:

Lemma 2.3. *The energy identity* (2-12) *can be rewritten as*

$$E^{0,0}(t)-E^{0,0}(1)=\int_1^t\Big[-\frac{l}{\tau}\Big(\frac{m}{2}\Big)E_0^{0,0}(\tau)+\frac{l}{\tau}\Big(\frac{m-2}{2}\Big)E_1^{0,0}(\tau)\Big]\,d\tau. \tag{2-20}$$

Corollary 2.4. *We have*

$$\frac{d}{dt}E^{0,0}(t)=-\frac{lm}{2t}E_0^{0,0}(t)+\frac{l(m-2)}{2t}E_1^{0,0}(t). \tag{2-21}$$

According to the explicit expression of FLRW metric (1-1), it is easy to see that the operator D is a Killing vector field, which means

$$\Pi^D_{\mu\nu}=0,$$

where D denotes the spacial derivatives. Thus, the structure of the equation (2-1) will not change if we take D^J as a commutator, i.e.,

$$\Box_g(D^J\varphi)=0. \tag{2-22}$$

By (2-22) and Corollary 2.4, for $I_0=0$, we have the following corollary.

Corollary 2.5. *For arbitrary nonnegative multi-index J, we have*

$$\frac{d}{dt}E^{J,0}(t)=-\frac{lm}{2t}E_0^{J,0}(t)+\frac{(m-2)l}{2t}E_1^{J,0}(t). \tag{2-23}$$

Define

$$\begin{cases} f^{I,I_0}(t)=E^{I,I_0}(t)t^{(2-m)l/2},\\ f_0^{I,I_0}(t)=E_0^{I,I_0}(t)t^{(2-m)l/2},\\ f_1^{I,I_0}(t)=E_1^{I,I_0}(t)t^{(2-m)l/2}.\end{cases} \tag{2-24}$$

Then we obtain the lemma:

Lemma 2.6. *The function $f^{I,0}(t)$ is uniformly bounded, provided that $E^{I,0}(1)$ is bounded for arbitrary multi-index I.*

Proof. By (2-24), since $m\ge 1$,

$$\begin{aligned}\frac{d}{dt}f^{I,0}(t)&=\Big(\frac{d}{dt}E^{I,0}(t)\Big)t^{(2-m)l/2}+\frac{(2-m)l}{2}E^{I,0}(t)t^{(2-m)l/2}t^{-1}\\ &=t^{(2-m)l/2}\Big[-\frac{lm}{2t}E_0^{I,0}(t)+\frac{(m-2)l}{2t}E_1^{I,0}(t)+\frac{(2-m)l}{2t}E^{I,0}(t)\Big]\\ &=-\frac{l(m-1)}{t}t^{(2-m)l/2}E_0^{I,0}(t)\le 0.\end{aligned} \tag{2-25}$$

Thus, $f^{I,0}(t)$ is monotonically decreasing and we have

$$f^{I,0}(t)\le f^{I,0}(1)=E^{I,0}(1).$$

□

Next, we consider the equation with higher-order derivatives on t.

Lemma 2.7. *For $I_0 > 0$, differentiating (2-22) by $\partial_t^{I_0}$, we have*

$$\begin{aligned}\text{(2-26)}\quad &\Box_g(\partial_t^{I_0} D^J \varphi)\\ &= \sum_{I_0'=1}^{I_0} C_{I_0',l,m}\, t^{-(I_0-I_0'+2)} \partial_t^{I_0'} D^J \varphi + \sum_{i=1}^{m} \sum_{I_0''=0}^{I_0-1} C_{I_0'',l,m} t^{-l-(I_0-I_0'')} \partial_i^2 \partial_t^{I_0''} D^J \varphi,\end{aligned}$$

where $C_{I_0',l,m}$ and $C_{I_0'',l,m}$ are constants depending on I_0', I_0'', l, and m.

Proof. With the notation $D^J \varphi = v$, it suffices to prove

$$\text{(2-27)}\quad \Box_g(\partial_t^{I_0} v) = \sum_{I_0'=1}^{I_0} C_{I_0',l,m}\, t^{-(I_0-I_0'+2)} \partial_t^{I_0'} v + \sum_{i=1}^{m} \sum_{I_0''=0}^{I_0-1} C_{I_0'',l,m}\, t^{-l-(I_0-I_0'')} \partial_i^2 \partial_t^{I_0''} v.$$

By (2-22), we know that

$$\text{(2-28)}\qquad \Box_g v = 0.$$

Since ∂_t is not a Killing vector field, it does not commute with the operator $\Box_g$, and a direct calculation gives

$$\text{(2-29)}\qquad [\Box_g, \partial_t] = -\frac{lm}{2} t^{-2} \partial_t + l t^{-l-1} \sum_{i=1}^{m} \partial_i^2.$$

Now, we can prove the lemma by induction on I_0.

When $I_0 = 1$,

$$\text{(2-30)}\qquad \Box_g(\partial_t v) = [\Box_g, \partial_t] v + \partial_t(\Box_g v) = -\frac{lm}{2} t^{-2} \partial_t v + l t^{-l-1} \sum_{i=1}^{m} \partial_i^2 v.$$

Thus, the lemma holds for $I_0 = 1$.

Suppose the lemma holds for $I_0 - 1$. Then, for I_0,

$$\begin{aligned}\text{(2-31)}\quad \Box_g(\partial_t^{I_0} v) &= [\Box_g, \partial_t](\partial_t^{I_0-1} v) + \partial_t(\Box_g \partial_t^{I_0-1} v)\\ &= \Big(-\frac{lm}{2} t^{-2} \partial_t + l t^{-l-1} \sum_{i=1}^{m} \partial_i^2\Big) \partial_t^{I_0-1} v\\ &\quad + \partial_t\Big(\sum_{I_0'=1}^{I_0-1} C_{I_0',l,m}\, t^{-(I_0-1-I_0'+2)} \partial_t^{I_0'} v\\ &\qquad\qquad + \sum_{i=1}^{m} \sum_{I_0''=0}^{I_0-2} C_{I_0'',l,m}\, t^{-l-(I_0-1-I_0'')} \partial_i^2 \partial_t^{I_0''} v\Big),\end{aligned}$$

which we rewrite as

$$\Box_g(\partial_t^{I_0}v) = -\frac{lm}{2}t^{-2}\partial_t^{I_0}v + lt^{-l-1}\sum_{i=1}^{m}\partial_i^2\partial_t^{I_0-1}v + \sum_{I_0'=1}^{I_0}C_{I_0',l,m}t^{-(I_0-I_0'+2)}\partial_t^{I_0'}v + \sum_{i=1}^{m}\sum_{I_0''=0}^{I_0-1}C_{I_0'',l,m}t^{-l-(I_0-I_0'')}\partial_i^2\partial_t^{I_0''}v.$$

By grouping together the last two terms on the right-hand-side, we see that (2-27) holds for arbitrary I_0. ☐

We define the total energy of M-th order as

$$F^M(t) = \sum_{|I|+I_0\le M} f^{I,I_0}(t). \tag{2-32}$$

Lemma 2.8. *We have*

$$F^M(t) \le C_M F^M(1), \tag{2-33}$$

where C_M is a positive constant depending only on M, and $F^M(1)$ is determined by the initial data.

Proof. For arbitrary $|I| + I_0 \le M$ and $I_0 \ge 1$, by Lemmas 2.2, 2.6, 2.7, and Corollary 2.5,

$$\begin{aligned}\frac{d}{dt}E^{I,I_0}(t) = -\frac{lm}{2t}E_0^{I,I_0}(t) + \frac{l(m-2)}{2t}E_1^{I,I_0}(t) \\ -\int_{\Sigma_t}\Bigl(\sum_{I_0'=1}^{I_0}C_{I_0',l,m}t^{-(I_0-I_0'+2)}\partial_t^{I_0'}D^J\varphi \\ + \sum_{i=1}^{m}\sum_{I_0''=0}^{I_0-1}C_{I_0'',l,m}t^{-l-(I_0-I_0'')}\partial_i^2\partial_t^{I_0''}\varphi\Bigr)\partial_t^{I_0+1}D^I\varphi\, d\,\mathrm{Vol}_t.\end{aligned} \tag{2-34}$$

Then, by Lemma 2.6

$$\begin{aligned}\frac{d}{dt}f^{I,I_0}(t) &= \Bigl(\frac{d}{dt}E^{I,I_0}(t)\Bigr)t^{(2-m)l/2} + \frac{(2-m)l}{2}E^{I,I_0}(t)\,t^{(2-m)l/2}t^{-1} \\ &\le \Bigl| t^{(2-m)l/2}\int_{\Sigma_t}\Bigl(\sum_{I_0'=1}^{I_0}C_{I_0',l,m}t^{-(I_0-I_0'+2)}\partial_t^{I_0'}D^J\varphi \\ &\qquad + \sum_{i=1}^{m}\sum_{I_0''=0}^{I_0-1}C_{I_0'',l,m}t^{-l-(I_0-I_0'')}\partial_i^2\partial_t^{I_0''}\varphi\Bigr)\partial_t^{I_0+1}D^I\varphi\, d\,\mathrm{Vol}_t\Bigr|.\end{aligned} \tag{2-35}$$

By Hölder's inequality, for arbitrary $|I| + I_0 \le M$,

$$\frac{d}{dt}f^{I,I_0}(t) \le C_M\bigl(t^{-(I_0-I_0'+2)}F^M(t) + t^{-l/2-(I_0-I_0'')}F^M(t)\bigr), \tag{2-36}$$

where $1 \le I_0' \le I_0$, $0 \le I_0'' \le I_0 - 1$, which is given by Lemma 2.7.

Summing over all $|I| + I_0 \le M$ and using (2-36),

$$\frac{d}{dt} F^M(t) \le C_M \big(t^{-(I_0 - I_0' + 2)} F^M(t) + t^{-l/2 - (I_0 - I_0'')} F^M(t) \big). \tag{2-37}$$

By virtue of (2-37), one gets

$$F^M(t) \le F^M(1) \exp\bigg(\int_1^t C_M \big(t^{-(I_0 - I_0' + 2)} + t^{-l/2 - (I_0 - I_0'')} \big)\, dt \bigg) \le C_M F^M(1), \tag{2-38}$$

where for the second inequality we use the facts that $I_0 - I_0' + 2 \ge 2$, $I_0 - I_0'' \ge 1$, and $l > 0$. ☐

Remark 2.9. The constant C_M may vary from line to line in the above proof.

We define

$$e_0^{I,I_0}(t) = \frac{1}{2} \| \partial_t (\partial_t^{I_0} D^I \varphi) \|_{L^2}^2 = \frac{1}{2} \int_{\mathbb{R}^m} (\partial_t \partial_t^{I_0} D^I \varphi)^2 \, dx, \tag{2-39}$$

$$e_1^{I,I_0}(t) = \frac{1}{2} \bigg\| \sum_{i=1}^m \partial_i (\partial_t^{I_0} D^I \varphi) \bigg\|_{L^2}^2 = \frac{1}{2} \int_{\mathbb{R}^m} \sum_{i=1}^m (\partial_i \partial_t^{I_0} D^I \varphi)^2 \, dx \tag{2-40}$$

and

$$e^{I,I_0}(t) = e_0^{I,I_0}(t) + e_1^{I,I_0}(t). \tag{2-41}$$

By (2-18), (2-39), and (2-40),

$$E_0^{I,I_0}(t) = 1/2 \int_{\mathbb{R}^m} (\partial_t \partial_t^{I_0} D^I \varphi)^2 t^{lm/2} \, dx = t^{lm/2} e_0^{I,I_0}(t) \tag{2-42}$$

and

$$E_1^{I,I_0}(t) = \frac{1}{2} \int_{\mathbb{R}^m} \sum_{i=1}^m (\partial_i \partial_t^{I_0} D^I \varphi)^2 t^{-l} t^{lm/2} \, dx = t^{(m-2)l/2} e_1^{I,I_0}(t). \tag{2-43}$$

We obtain the following lemma directly from Lemma 2.8 and (2-42)–(2-43):

Lemma 2.10. *The following decay estimates hold*:

$$e_0^{I,I_0}(t) \le t^{-l} f^{I,I_0}(1) \quad \textit{and} \quad e_1^{I,I_0}(t) \le f^{I,I_0}(1). \tag{2-44}$$

By Lemma 2.10 and the Sobolev embedding, one easily obtain the following L^∞ estimates:

Lemma 2.11. *For any* $|I| \ge |J| + \lceil \frac{m}{2} + 1 \rceil$ *and* $i = 1, \dots, m$,

$$\| \partial_t (\partial_t^{I_0} D^J \varphi)(t) \|_{L^\infty} \le C_I t^{-l/2} \bigg(\sum_{K=0}^{I} f^{K,I_0}(t) \bigg)^{1/2} \quad \textit{for } I_0 \ge 0 \tag{2-45}$$

and

$$\|\partial_i (D^J \varphi)(t)\|_{L^\infty} \le C_I \biggl(\sum_{K=0}^{I} f^{K,0}(t) \biggr)^{1/2}, \tag{2-46}$$

provided that $f^{K,I_0}(t)$ $(K = 0, \dots, I)$ *is bounded.*

3. The proof of Theorem 1.3

In this section, we will prove the global existence of smooth solutions to exponential wave maps (1-8) for small initial data (1-9). Since (1-8) can be reduced to a symmetric hyperbolic system, via standard results [Majda 1984; Alinhac 2010], there exists a unique local solution in the function space H^s, for $s > 1 + \frac{m}{2}$. Thus, for the global existence result, it suffices to prove a priori estimate; see Lemma 3.7.

To obtain the global existence of smooth solution to Eq. (1-8), one needs to derive the equation satisfied by the higher-order derivatives of the solution first.

Lemma 3.1. *Differentiating* (1-8) *by the spacial derivatives* D^I,

$$\Box_g (D^I u) = \sum_{I_1+I_2+I_3=I} -Q(D^{I_1} u, Q(D^{I_2} u, D^{I_3} u)). \tag{3-1}$$

Proof. Since D is a Killing vector field, it suffices to prove

$$D^I Q(u, v) = \sum_{I_1+I_2=I} C_{I_1,I_2} Q(D^{I_1} u, D^{I_2} v). \tag{3-2}$$

By Leibniz's rule,

$$\begin{aligned} D^I Q(u, v) &= D^I \biggl(-\partial_t u \, \partial_t v + t^{-l} \sum_{i=1}^{m} (\partial_i u \partial_i v) \biggr) \\ &= \sum_{I_1+I_2=I} C_{I_1,I_2} \biggl(\partial_t D^{I_1} u \, \partial_t D^{I_2} v + t^{-l} \sum_{i=1}^{m} (\partial_i D^{I_1} u)(\partial_i D^{I_2} v) \biggr) \\ &= \sum_{I_1+I_2=I} C_{I_1,I_2} Q(D^{I_1} u, D^{I_2} v). \end{aligned} \tag{3-3}$$

Let

$$v = Q(u, u). \tag{3-4}$$

By (3-2), we have

$$\begin{aligned} D^I Q(u, Q(u, u)) &= \sum_{I_1+I_2=I} C_{I_1,I_2} Q(D^{I_1} u, D^{I_2} Q(u, u)) \\ &= \sum_{I_1+I_{21}+I_{22}=I_1+I_2} C_{I_1,I_{21},I_{22}} Q(D^{I_1} u, Q(D^{I_{21}} u, D^{I_{22}} u)). \end{aligned} \tag{3-5}$$

Thus,

$$\Box_g(D^I u) = [\Box_g, D^I]u + D^I(\Box_g u) = -D^I Q(u, Q(u, u)). \tag{3-6}$$

From (3-5) and (3-6), we complete the proof of the lemma. $\square$

For higher-order derivatives with respect to t, we have the following lemmas:

Lemma 3.2. *For any smooth functions $v(t, x)$ and $w(t, x)$,*

$$\begin{aligned} &\partial_t^{I_0} Q(v, w) \\ &\quad = \sum_{I_{01}+I_{02}=I_0} Q(\partial_t^{I_{01}} v, \partial_t^{I_{02}} w) + \sum_{i=1}^m \sum_{I_{01}+I_{02}<I_0} t^{-l-(I_0-I_{01}-I_{02})} \partial_i \partial_t^{I_{01}} v\, \partial_i \partial_t^{I_{02}} w. \end{aligned} \tag{3-7}$$

Proof. We prove it by induction. For $I_0 = 1$, we have

$$\begin{aligned} &\partial_t Q(v, w) \\ &\quad = \partial_t \Big[-\partial_t v\, \partial_t w + \sum_{i=1}^m t^{-l} \partial_i v\, \partial_i w \Big] \\ &\quad = -\partial_t^2 v\, \partial_t w - \partial_t v\, \partial_t^2 w + \sum_{i=1}^m t^{-l} \partial_{ti}^2 v\, \partial_i w + \sum_{i=1}^m t^{-l} \partial_i v\, \partial_{ti}^2 w - \sum_{i=1}^m l t^{-l-1} \partial_i v\, \partial_i w \\ &\quad = Q(\partial_t v, w) + Q(v, \partial_t w) - \sum_{i=1}^m t^{-l-(1-0)} \partial_i v\, \partial_i w. \end{aligned}$$

Therefore, the lemma holds for $I_0 = 1$.

Assume that the lemma holds for $I_0 - 1$ for arbitrary $I_0 > 1$; then for I_0,

$$\begin{aligned} &\partial_t^{I_0} Q(v, w) \\ &\quad = \partial_t(\partial_t^{I_0-1} Q(v, w)) \\ &\quad = \partial_t \Big(\sum_{I_{01}+I_{02}=I_0-1} Q(\partial_t^{I_{01}} v, \partial_t^{I_{02}} w) + \sum_{i=1}^m \sum_{I_{01}+I_{02}<I_0-1} t^{-l-(I_0-I_{01}-I_{02})} \partial_t^{I_{01}} \partial_i v\, \partial_t^{I_{02}} \partial_i w \Big) \\ &\quad = \sum_{I_{01}+I_{02}=I_0-1} \big(Q(\partial_t^{I_{01}+1} v, \partial_t^{I_{02}} w) + Q(\partial_t^{I_{01}} v, \partial_t^{I_{02}+1} w) \big) \\ &\qquad + \sum_{i=1}^m \sum_{I_{01}+I_{02}<I_0-1} t^{-1-(I_0+1-I_{01}-I_{02})} \partial_t^{I_{01}} \partial_i v\, \partial_t^{I_{02}} \partial_i w \\ &\qquad + \sum_{i=1}^m \sum_{I_{01}+I_{02}<I_0-1} t^{-1-(I_0+1-I_{01}-I_{02})} (\partial_t^{I_{01}+1} \partial_i v\, \partial_t^{I_{02}} \partial_i w + \partial_t^{I_{01}} \partial_i v\, \partial_t^{I_{02}+1} \partial_i w) \\ &\quad = \sum_{I_{01}+I_{02}=I_0} Q(\partial_t^{I_{01}} v, \partial_t^{I_{02}} w) + \sum_{i=1}^m \sum_{I_{01}+I_{02}<I_0} t^{-l-(I_0-I_{01}-I_{02})} \partial_t^{I_{01}} \partial_i v\, \partial_t^{I_{02}} \partial_i w. \end{aligned}$$

Thus, the lemma holds for arbitrary I_0. □

We define

$$S(I_0) = \{(I_0, I_{01}, I_{02}, I_{021}, I_{022}) \mid I_{01} + I_{02} < I_0,\ I_{021} + I_{022} < I_{02}\}.$$

Lemma 3.3. *For any smooth functions $u(t, x)$, $v(t, x)$, and $w(t, x)$, and any nonnegative integer I_0,*

$$\begin{aligned}
\text{(3-8)}\quad & \partial_t^{I_0} Q(u, Q(v, w)) \\
&= \sum_{I_{01}+I_{02}+I_{03}=I_0} Q(\partial_t^{I_{01}} u, Q(\partial_t^{I_{02}} v, \partial_t^{I_{03}} w)) \\
&\quad + \sum_{i=1}^{m} \sum_{S(I_0)} t^{-l-(I_{02}-I_{021}-I_{022})} Q(\partial_t^{I_{01}} u, \partial_t^{I_{021}} \partial_i v\, \partial_t^{I_{022}} \partial_i w) \\
&\quad + \sum_{i=1}^{m} \sum_{S(I_0)} t^{-l-1-(I_{02}-I_{021}-I_{022})} \partial_t^{I_{01}} \partial_t u\, \partial_t^{I_{021}} \partial_i v\, \partial_t^{I_{022}} \partial_i w \\
&\quad + \sum_{i=1}^{m} \sum_{S(I_0)} t^{-l-(I_0-I_{01}-I_{02})} \partial_t^{I_{01}} \partial_i u\, \partial_i Q(\partial_t^{I_{021}} v, \partial_t^{I_{022}} w) \\
&\quad + \sum_{i=1}^{m} \sum_{S(I_0)} t^{2l+(I_0-I_{01}-I_{02})+(I_{02}-I_{021}-I_{022})} \partial_t^{I_{01}} \partial_i u\, \partial_i (\partial_t^{I_{021}} \partial_i v\, \partial_t^{I_{022}} \partial_i w).
\end{aligned}$$

Proof. By Lemma 3.2, let $p = Q(v, w)$, so

$$\begin{aligned}
\text{(3-9)}\quad & \partial_t^{I_0} Q(u, Q(v, w)) \\
&= \partial_t^{I_0} Q(u, p) \\
&= \sum_{I_{01}+I_{02}=I_0} Q(\partial_t^{I_{01}} u, \partial_t^{I_{02}} p) + \sum_{I_{01}+I_{02}<I_0} t^{-l-(I_0-I_{01}-I_{02})} \partial_t^{I_{01}} \partial_i u\, \partial_t^{I_{02}} \partial_i p,
\end{aligned}$$

since

$$\text{(3-10)}\quad \partial_t^{I_{02}} p = \sum_{I_{021}+I_{022}=I_{02}} Q(\partial_t^{I_{021}} v, \partial_t^{I_{022}} w) + \sum_{I_{021}+I_{022}<I_{02}} t^{-l-(I_{02}-I_{021}-I_{022})} \partial_t^{I_{021}} \partial_i v\, \partial_t^{I_{022}} \partial_i w.$$

Inserting (3-10) into (3-9), we obtain the lemma. □

Combining Lemmas 2.1, 3.1, 3.2 and 3.3, we obtain an important lemma:

Lemma 3.4. *Differentiating* (1-8) *by $\partial_t^{I_0} D^I$ for any nonnegative integer I_0 and multi-index I, we have*

$$
\begin{aligned}
(3\text{-}11)\quad & \square_g(\partial_t^{I_0} D^I u) \\
&= \sum_{I_0'=1}^{I_0} t^{-(I_0-I_0'+2)} \partial_t^{I_0'} D^I u + \sum_{i=1}^{m} \sum_{I_0''=0}^{I_0-1} t^{-l-(I_0-I_0'')} \partial_i^2 \partial_t^{I_0''} D^I u \\
&\quad + \sum_{I_{01}+I_{02}+I_{03}=I_0} Q(\partial_t^{I_{01}} D^{I_1} u,\, Q(\partial_t^{I_{02}} D^{I_2} u,\, \partial_t^{I_{03}} D^{I_3} u)) \\
&\quad + \sum_{S(I_0)} t^{-l-(I_{02}-I_{021}-I_{022})} Q(\partial_t^{I_{01}} D^{I_1} u,\, \partial_t^{I_{021}} \partial_i (D^{I_2} u)\, \partial_t^{I_{022}} \partial_i (D^{I_3} u)) \\
&\quad + \sum_{S(I_0)} t^{-l-1-(I_{02}-I_{021}-I_{022})} \partial_t^{I_{01}} (D^{I_1} \partial_t u)\, \partial_t^{I_{021}} (D^{I_2} \partial_i u)\, \partial_t^{I_{022}} (D^{I_3} \partial_i u) \\
&\quad + \sum_{S(I_0)} t^{-l-(I_0-I_{01}-I_{02})} \partial_t^{I_{01}} (D^{I_1} \partial_i u)\, \partial_i Q(\partial_t^{I_{021}} D^{I_2} u,\, \partial_t^{I_{022}} (D^{I_3} u)) \\
&\quad + \sum_{S(I_0)} t^{-E} \partial_t^{I_{01}} \partial_i (D^{I_1} u)\, \partial_i (\partial_t^{I_{021}} (D^{I_2} \partial_i u) \partial_t^{I_{022}} (D^{I_3} \partial_i u)) \\
&:= \sum_{I_0'=1}^{I_0} t^{-(I_0-I_0'+2)} \partial_t^{I_0'} D^I u + \sum_{i=1}^{m} \sum_{I_0''=0}^{I_0-1} t^{-l-(I_0-I_0'')} \partial_i^2 \partial_t^{I_0''} D^I u \\
&\quad + Q(u,\, Q(u,\, \partial_t^{I_0} D^I u)) + R,
\end{aligned}
$$

where $E = 2l + (I_0 - I_{01} - I_{02}) + (I_{02} - I_{021} - I_{022})$, $I_1 + I_2 + I_3 = I$, *and* R *denotes the remaining terms.*

Remark 3.5. For convenience, in Lemmas 3.1–3.4, we have omitted the numerical constants in front of each term at the right hand of (3-1), (3-7), (3-8), and (3-11). It does not affect the estimates in the sequel since they are all universal constants.

Remark 3.6. In the following estimates, we distinguish the $Q(u,\, Q(u,\, \partial_t^{I_0} D^I u))$ term from others, since it contains the highest-order derivatives of $\partial_t^{I_0} D^I u$.

Based on Lemma 3.4, we have the following energy estimate. It plays a key role in the proof of Theorem 1.3.

Lemma 3.7. *The following generalized energy inequality holds in the maximal development of the smooth solution of Cauchy problem* (1-8), (1-9):

$$
\begin{aligned}
(3\text{-}12)\quad & F^M(t) \\
&\lesssim F^M(1) + \int_1^t \big(\max\{\tau^{-(I_0-I_0'+2)},\, \tau^{-l/2-(I_0-I_0'')}\} + \tau^{-l/2} F^M(\tau)\big) F^M(\tau)\, d\tau.
\end{aligned}
$$

provided $F^M(t) \ll 1$, *where* $F^M(t)$ *is defined by* (2-32) *with* $M \geq m+2$.

Proof. We will prove the lemma in three steps.

***Energy estimates based on* (3-11).** Taking $V = -\partial_t$ in Lemma 2.2, one gets

(3-13)
$$E^{I,J}(t) - E^{I,J}(1) = \int_1^t \int_{\Sigma_\tau} \big(K^V(\partial_\tau^J D^I u) + \Box_g(\partial_\tau^J D^I u) V(\partial_\tau^J D^I u)\big)\, d\,\mathrm{Vol}_\tau\ d\tau,$$

which is equivalent to

$$\frac{d}{dt} E^{I,J}(t) = \int_{\Sigma_t} \big(K^V(\partial_t^J D^I u(t)) + \Box_g(\partial_t^J D^I u) V(\partial_t^J D^I u)\big)\, d\,\mathrm{Vol}_t. \tag{3-14}$$

By (3-14) and (2-24),

$$\begin{aligned}
\frac{d}{dt} f^{I,J}(t) \\
&= t^{(2-m)l/2}\Big(\frac{d}{dt}E^{I,J}(t)\Big) + \frac{(2-m)l}{2} t^{(2-m)l/2} t^{-1} E^{I,J}(t) \\
&= t^{(2-m)l/2}\int_{\Sigma_t} \big(K^V(\partial_t^J D^I u(t)) + \Box_g(\partial_t^J D^I u) V(\partial_t^J D^I u)\big)\, d\,\mathrm{Vol}_t \\
&\qquad + \frac{(2-m)l}{2} t^{-1} t^{(2-m)l/2} E^{I,J}(t) \\
&\le t^{(2-m)l/2}\int_{\Sigma_t} \Big(\sum_{J'=1}^{J} t^{-(J-J'+2)} \partial_t^{J'} D^I u + \sum_{i=1}^{m}\sum_{J''=0}^{J-1} t^{-l-(J-J'')} \partial_i^2 \partial_t^{J''} D^I u\Big) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \times V(\partial_t^J D^I u)\, d\,\mathrm{Vol}_t \\
&\qquad + t^{(2-m)l/2}\int_{\Sigma_t} \big(Q(u, Q(u, \partial_t^J D^I u)) + R\big) V(\partial_t^J D^I u)\, d\,\mathrm{Vol}_t.
\end{aligned} \tag{3-15}$$

We first estimate the term that contains the second-order derivatives of $\partial_t^J D^I u$, i.e.,

$$\int_{\mathbb{R}^m} Q(u, Q(u, \partial_t^J D^I u)) V(\partial_t^J D^I u) t^{(2-m)l/2} t^{ml/2}\, dx. \tag{3-16}$$

***Energy estimates based on* (3-16).** Let

$$v = \partial_t^J D^I u. \tag{3-17}$$

Expanding the term $Q(u, Q(u, v))$, one obtains

$$\begin{aligned}
&Q(u, Q(u, v)) \\
&= -\partial_t u\, \partial_t\Big(-\partial_t u\, \partial_t v + t^{-l}\sum_{i=1}^m \partial_i u\, \partial_i v\Big) + t^{-l}\sum_{j=1}^m \partial_j u\, \partial_j\Big(-\partial_t u\, \partial_t v + t^{-l}\sum_{i=1}^m \partial_i u\, \partial_i v\Big)
\end{aligned}$$

which yields

$$
\begin{aligned}
&Q(u,\,Q(u,v))\\
&\quad= (\partial_t u)^2\partial_t^2 v - 2t^{-l}\sum_{i=1}^{m}\partial_t u\,\partial_i u\,\partial_{ti}^2 v + \sum_{i,j=1}^{m} t^{-2l}\partial_i u\,\partial_j u\,\partial_{ij}^2 v + \partial_t u\,\partial_t^2 u\,\partial_t v\\
&\qquad\qquad - t^{-l}\sum_{i=1}^{m}\partial_t u\,\partial_{ti}^2 u\,\partial_i v + l t^{-l-1}\sum_{i=1}^{m}\partial_t u\,\partial_i u\,\partial_i v\\
&\qquad\qquad - t^{-l}\sum_{j=1}^{m}\partial_j u\,\partial_{jt}^2 u\,\partial_t v + t^{-2l}\sum_{i,j=1}^{m}\partial_j u\,\partial_{ij}^2 u\,\partial_i v\\
&\quad= A+B+C+L,
\end{aligned}
$$

where

$$
A=(\partial_t u)^2\partial_t^2 v,\qquad B=-2t^{-l}\sum_{i=1}^{m}\partial_t u\,\partial_i u\,\partial_{ti}^2 v,\qquad C=\sum_{i,j=1}^{m} t^{-2l}\partial_i u\,\partial_j u\,\partial_{ij}^2 v, \tag{3-18}
$$

and

$$
\begin{aligned}
L = \partial_t u\,\partial_t^2 u\,\partial_t v - t^{-l}\sum_{i=1}^{m}\partial_t u\,\partial_{ti}^2 u\,\partial_i v + l t^{-l-1}\sum_{i=1}^{m}\partial_t u\,\partial_i u\,\partial_i v&\\
- t^{-l}\sum_{j=1}^{m}\partial_j u\,\partial_{jt}^2 u\,\partial_t v + t^{-2l}\sum_{i,j=1}^{m}\partial_j u\,\partial_{ij}^2 u\,\partial_i v.&
\end{aligned} \tag{3-19}
$$

In view of (3-16) and (3-18), and integrating by parts, we have

$$
\begin{aligned}
&\int_{\mathbb{R}^m} A\,\partial_t v\, t^{(2-m)l/2} t^{ml/2} dx\\
&\qquad= \int_{\mathbb{R}^m} (\partial_t u)^2\,\partial_t^2 v\,\partial_t v\, t^l\,dx\\
&\qquad= \frac{1}{2}\frac{d}{dt}\int_{\mathbb{R}^m} (\partial_t u)^2(\partial_t v)^2 t^l\,dx\\
&\qquad\quad - \int_{\mathbb{R}^m} \partial_t u\,\partial_t^2 u\,(\partial_t v)^2 t^l\,dx - \frac{l}{2}\int_{\mathbb{R}^m}(\partial_t u)^2(\partial_t v)^2 t^{-1} t^l\,dx\\
&\qquad= \frac{1}{2}\frac{d}{dt}\int_{\mathbb{R}^m} (\partial_t u)^2(\partial_t v)^2 t^l\,dx + A_1 + A_2.
\end{aligned} \tag{3-20}
$$

and

$$\begin{aligned}
(3\text{-}21)\quad & \int_{\mathbb{R}^m} B\,\partial_t v\, t^l\, dx \\
&= \int_{\mathbb{R}^m}\biggl(2t^{-l}\sum_{i=1}^m \partial_t u\,\partial_i u\,\partial_{it}^2 v\biggr)\partial_t v\, t^l\, dx \\
&= \sum_{i=1}^m \partial_i\biggl(\int_{\mathbb{R}^m} \partial_t u\,\partial_i u(\partial_i v)^2\, dx\biggr) \\
&\qquad -\sum_{i=1}^m \int_{\mathbb{R}^m} t^{-l}\partial_{ti}^2 u\,\partial_i u(\partial_t v)^2 t^l\, dx - \sum_{i=1}^m \int_{\mathbb{R}^m} t^{-l}\partial_t u\,\partial_i^2 u(\partial_t v)^2 t^l\, dx \\
&:= \sum_{i=1}^m \partial_i\biggl(\int_{\mathbb{R}^m} \partial_t u\,\partial_i u(\partial_i v)^2\, dx\biggr) + B_1 + B_2.
\end{aligned}$$

Also,

$$\begin{aligned}
(3\text{-}22)\quad & \int_{\mathbb{R}^m} C\,\partial_t v\, t^l\, dx \\
&= -\sum_{i,j=1}^m \int_{\mathbb{R}^m} t^{-2l}\partial_i u\,\partial_j u\,\partial_{ij}^2 v\,\partial_t v\, t^l\, dx \\
&= \sum_{i,j=1}^m \frac{1}{2}\frac{d}{dt}\int_{\mathbb{R}^m} t^{-2l}\partial_i u\,\partial_j u\,\partial_i v\,\partial_j v\, t^l\, dx \\
&\quad + \sum_{i,j=1}^m \int_{\mathbb{R}^m} t^{-2l}\partial_{ij}^2 u\,\partial_j u\,\partial_i v\,\partial_t v\, t^l\, dx + \int_{\mathbb{R}^m} t^{-2l}\partial_i u\,\partial_j^2 u\,\partial_i v\,\partial_t v\, t^l\, dx \\
&\quad - \sum_{i,j=1}^m \int_{\mathbb{R}^m} t^{-2l}\partial_{ti}^2 u\,\partial_j u\,\partial_i v\,\partial_j v\, t^l\, dx - \sum_{i,j=1}^m \int_{\mathbb{R}^m} \frac{l}{2} t^{-2l-1}\partial_i u\,\partial_j u\,\partial_i v\,\partial_j v\, t^l\, dx \\
&= \sum_{i,j=1}^m \frac{1}{2}\frac{d}{dt}\int_{\mathbb{R}^m} t^{-2l}\partial_i u\,\partial_j u\,\partial_i v\,\partial_j v\, t^l\, dx + C_1 + C_2 + C_3 + C_4.
\end{aligned}$$

Next, we estimate these terms via $F^M(t)$. By Lemma 2.11 and the assumption $M \ge m+2$, we obtain easily

$$|A_i| \lesssim t^{-l}[F^M(t)]^2, \qquad i = 1, 2, \tag{3-23}$$

$$|B_i| \lesssim t^{-l}[F^M(t)]^2, \qquad i = 1, 2, \tag{3-24}$$

$$|C_i| \lesssim t^{-3l/2}[F^M(t)]^2, \qquad i = 1, 2, 3, \tag{3-25}$$

and

$$|C_4| \lesssim t^{-l-1}[F^M(t)]^2. \tag{3-26}$$

For the remaining terms L, we have

$$\left| \int_{\mathbb{R}^m} L \partial_t v \, t^{(2-m)l/2} t^{lm/2} \, dx \right| \lesssim (t^{-l} + t^{-2l} + t^{-l-1}) [F^M(t)]^2. \tag{3-27}$$

At last, we estimate the lower-order terms

$$\int_{\mathbb{R}^m} R \partial_t v \, t^{(2-m)l/2} t^{ml/2} \, dx, \tag{3-28}$$

where R is defined by (3-11).

***Energy estimates based on* (3-28).** We have

$$\int_{\mathbb{R}^m} R \partial_t v t^l dx = \int_{\mathbb{R}^m} \left(\sum_{i=1}^{5} R_i \right) \partial_t v t^l dx, \tag{3-29}$$

where R_i $(i = 1, \dots, 5)$ is defined by (3-11).

For R_1, we have

$$\begin{aligned} &Q(\partial_t^{I_{01}} D^{I_1} u, Q(\partial_t^{I_{02}} D^{I_2} u, \partial_t^{I_{03}} D^{I_3} u)) \\ &\quad = -\partial_t (\partial_t^{I_{01}} D^{I_1} u) \, \partial_t \Big[-\partial_t (\partial_t^{I_{02}} D^{I_2} u) \, \partial_t (\partial_t^{I_{03}} D^{I_3} u) \\ &\qquad\qquad + t^{-l} \sum_{i=1}^{m} \partial_i (\partial_t^{I_{02}} D^{I_2} u) \, \partial_i (\partial_t^{I_{03}} D^{I_3} u) \Big] \\ &\qquad + t^{-l} \sum_{j=1}^{m} \partial_j (\partial_t^{I_{01}} D^{I_1} u) \, \partial_j \Big[-\partial_t (\partial_t^{I_{02}} D^{I_2} u) \, \partial_t (\partial_t^{I_{03}} D^{I_3} u) \\ &\qquad\qquad + t^{-l} \sum_{i=1}^{m} \partial_i (\partial_t^{I_{02}} D^{I_2} u) \, \partial_i (\partial_t^{I_{03}} D^{I_3} u) \Big], \end{aligned} \tag{3-30}$$

so

$$\left| \int_{\mathbb{R}^m} R_1 \partial_t v \, t^{(2-m)l/2} t^{ml/2} \, dx \right| \lesssim (t^{-l} + t^{-l/2}) [F^M(t)]^2. \tag{3-31}$$

For R_2, by direct calculations,

$$\begin{aligned} &Q(\partial_t^{I_{01}} D^{I_1} u, \partial_t^{I_{021}} \partial_i (D^{I_2} u) \, \partial_t^{I_{022}} \partial_i (D^{I_3} u)) \\ &\quad = -\partial_t (\partial_t^{I_{01}} D^{I_1} u) \, \partial_t (\partial_t^{I_{021}} \partial_i (D^{I_2} u) \, \partial_t^{I_{022}} \partial_i (D^{I_3} u)) \\ &\qquad + t^{-l} \sum_{i=1}^{m} \partial_i (\partial_t^{I_{01}} D^{I_1} u) \, \partial_i (\partial_t^{I_{021}} \partial_i (D^{I_2} u) \, \partial_t^{I_{022}} \partial_i (D^{I_3} u)), \end{aligned} \tag{3-32}$$

so

$$\left|\int_{\mathbb{R}^m} R_2 \partial_t v\, t^{-l-(I_{02}-I_{021}-I_{022})}\, t^{(2-m)l/2}\, t^{ml/2}\, dx\right| \lesssim (t^{-3l/2-(I_{02}-I_{021}-I_{022})} + t^{-l/2-(I_{02}-I_{021}-I_{022})})\,[F^M(t)]^2. \tag{3-33}$$

For R_3, we have

$$\left|\int_{\mathbb{R}^m} R_3 \partial_t v\, t^{-2l-(I_0-I_{01}-I_{02})-(I_{02}-I_{021}-I_{022})}\, t^{(2-m)l/2}\, t^{ml/2}\, dx\right| \lesssim t^{-l/2-(I_0-I_{01}-I_{02})-(I_{02}-I_{021}-I_{022})}[F^M(t)]^2. \tag{3-34}$$

For R_4,

$$\begin{aligned} Q(\partial_t^{I_{02}} D^{I_2} u, \partial_t^{I_{03}} D^{I_3} u) &= -\partial_t(\partial_t^{I_{02}} D^{I_2} u)\,\partial_t(\partial_t^{I_{03}} D^{I_3} u) + t^{-l}\sum_{i=1}^m \partial_i(\partial_t^{I_{02}} D^{I_2} u)\partial_i(\partial_t^{I_{03}} D^{I_3} u), \end{aligned} \tag{3-35}$$

so

$$\left|\int_{\mathbb{R}^m} R_4 \partial_t v\, t^{-l-(I_0-I_{01}-I_{02})}\, t^{(2-m)l/2}\, t^{ml/2}\, dx\right| \lesssim t^{-l/2-(I_0-I_{01}-I_{02})}\,[F^M(t)]^2. \tag{3-36}$$

For R_5, we obtain

$$\left|\int_{\mathbb{R}^m} R_5 \partial_t v\, t^{-2l-(I_0-I_{01}-I_{02})-(I_{02}-I_{021}-I_{022})}\, t^{(2-m)l/2}\, t^{ml/2}\, dx\right| \lesssim t^{-l/2-(I_0-I_{01}-I_{02})-(I_{02}-I_{021}-I_{022})}\,[F^M(t)]^2. \tag{3-37}$$

Combining these estimates above, (2-37) and (3-15)–(3-37), and summing up $|I|+J \le M$, we have

$$\begin{aligned} \frac{d}{dt}\tilde{F}^M(t) \lesssim{}& \max\{t^{-I_0-I_0'+2}, t^{-l/2-(I_0-I_0'')}\}F^M(t) \\ &+ \max\{t^{-l}, t^{-l-1}, t^{-l/2}, t^{-2l}, t^{-3l/2-(I_{02}-I_{021}-I_{022})}, \\ &\quad t^{-l/2-(I_0-I_{01}-I_{02})-(I_{02}-I_{021}-I_{022})}\}\,[F^M(t)]^2. \end{aligned} \tag{3-38}$$

Here, we define $\tilde{F}^M(t)$ to be

$$\tilde{F}^M(t) = \sum_{|I|+J\le M} f^{I,J}(t) + \frac{1}{2}\int_{\mathbb{R}^m}\big((\partial_t u)^2(\partial_t v)^2 t^l + \partial_i u\,\partial_j u\,\partial_i v\partial_j v\, t^{-2l}\big)\, dx.$$

It is easy to see that, if $F(t) \ll 1$, then there exists a positive constant C such that

$$C^{-1}F^M(t) \le \tilde{F}^M(t) \le C F^M(t).$$

Integrating (3-38) over $[1, t)$,

$$F^M(t) \lesssim F^M(1) + \int_1^t \bigl(\max\{\tau^{-(I_0-I_0'+2)}, \tau^{-l/2-(I_0-I_0'')}\} + \tau^{-l/2} F^M(\tau)\bigr) F^M(\tau)\, d\tau. \tag{3-39}$$ □

Based on Lemma 3.7, we can prove Theorem 1.3 by the bootstrap method.

Proof of Theorem 1.3. Set

$$E = \{t \in [1, T) : F^M(s) \le A\epsilon^2 \quad \text{for } 1 \le s \le t\}. \tag{3-40}$$

For the proof of the main result, it suffices to show that for any $t \in E$, the assumption $F^M(t) \le A\epsilon^2 \ll 1$ will imply $F^M(t) \le (A/2)\epsilon^2$, provided that A is sufficiently large and ϵ is sufficiently small.

By (3-12) and Gronwall's lemma, there exists a positive constant C such that

$$F^M(t) \le C F^M(1) \exp \int_1^t \bigl(\max\{\tau^{-(I_0-I_0'+2)}, \tau^{-l/2-(I_0-I_0'')}\} + \tau^{-l/2} A\epsilon^2\bigr)\, d\tau. \tag{3-41}$$

Choosing

$$\epsilon_0 = \sqrt{\frac{\ln(2)}{100A}} \quad \text{and} \quad A \ge 16 C D^M(1),$$

where $D^M(1)\epsilon^2 = F^M(1)$, then for any $\epsilon \in [0, \epsilon_0]$ and $l > 2$, we have

$$\begin{aligned} F^M(t) &\le C F^M(1) \exp \int_1^\infty \bigl(\max\{\tau^{-(I_0-I_0'+2)}, \tau^{-l/2-(I_0-I_0'')}\} + \tau^{-l/2} A\epsilon\bigr)\, d\tau \\ &\le \frac{A}{2}\epsilon^2. \end{aligned} \tag{3-42}$$

Then we can argument by contradiction and get the global existence result. □

4. Some discussions

In this paper, we have proved the global existence of smooth solutions to exponential wave maps on some special FLRW spacetimes, which are important and interesting in geometry and physics. Along with the development of Lorentzian geometry, the study of classical field theory and evolution equations is generalized to curved spacetimes. We believe that this field will attract better attention in the near future. Equations on a curved background, especially the solutions to the Einstein field equation, take a much more important role in physics. Confined by our knowledge, we only consider the easy case here, but there still exist a lot of problems that are worth focusing on.

For large initial data or large initial energy, whether the smooth solution of exponential wave maps exists globally or not is an interesting problem in the field of PDEs. Since wave maps are a class of equations with good structure, the study of the large data problem is under consideration now; one can refer to [Wang and Yu 2013; Yang 2015] on Minkowski spacetime.

Acknowledgements

Wei was supported by the Scientific Research Foundation of Zhejiang Sci-Tech University (Grant No. 16062021-Y). Lai was supported by NSFC (Grant No. 11501273) and high level talent project of Lishui City (Grant No. 2016RC25).

References

[Alinhac 2010] S. Alinhac, *Geometric analysis of hyperbolic differential equations: an introduction*, London Mathematical Society Lecture Note Series **374**, Cambridge University Press, 2010. MR Zbl

[Chiang and Yang 2007] Y.-J. Chiang and Y.-H. Yang, "Exponential wave maps", *J. Geom. Phys.* **57**:12 (2007), 2521–2532. MR Zbl

[Choquet-Bruhat 2000] Y. Choquet-Bruhat, "Global wave maps on Robertson–Walker spacetimes", *Nonlinear Dynam.* **22**:1 (2000), 39–47. MR Zbl

[Eells and Lemaire 1992] J. Eells and L. Lemaire, "Some properties of exponentially harmonic maps", pp. 129–136 in *Partial differential equations* (Warsaw, 1990), vol. 1, edited by B. Bojarski et al., Banach Center Publ. **27-1**, Polish Acad. Sci., Warsaw, 1992. MR Zbl

[Gu 1980] C. H. Gu, "On the Cauchy problem for harmonic maps defined on two-dimensional Minkowski space", *Comm. Pure Appl. Math.* **33**:6 (1980), 727–737. MR Zbl

[Hawking and Ellis 1973] S. W. Hawking and G. F. R. Ellis, *The large scale structure of space-time*, Cambridge Monographs on Mathematical Physics **1**, Cambridge Univ. Press, London, 1973. MR Zbl

[Huh 2013] H. Huh, "Global solutions of the exponential wave equation with small initial data", *Bull. Korean Math. Soc.* **50**:3 (2013), 811–821. MR Zbl

[Klainerman and Machedon 1993] S. Klainerman and M. Machedon, "Space-time estimates for null forms and the local existence theorem", *Comm. Pure Appl. Math.* **46**:9 (1993), 1221–1268. MR Zbl

[Klainerman and Machedon 1995] S. Klainerman and M. Machedon, "Smoothing estimates for null forms and applications", *Duke Math. J.* **81**:1 (1995), 99–133. MR Zbl

[Kong and Wei 2013] D.-X. Kong and C.-H. Wei, "Lifespan of smooth solutions for timelike extremal surface equation in de Sitter spacetime", preprint, 2013. To appear in *J. Math. Phys.* arXiv

[Majda 1984] A. Majda, *Compressible fluid flow and systems of conservation laws in several space variables*, Applied Mathematical Sciences **53**, Springer, New York, 1984. MR Zbl

[Shatah and Struwe 1998] J. Shatah and M. Struwe, *Geometric wave equations*, Courant Lecture Notes in Mathematics **2**, Courant Institute of Mathematical Sciences, New York, 1998. MR Zbl

[Shatah and Struwe 2002] J. Shatah and M. Struwe, "The Cauchy problem for wave maps", *Int. Math. Res. Not.* **2002**:11 (2002), 555–571. MR Zbl

[Tao 2001a] T. Tao, "Global regularity of wave maps, I: Small critical Sobolev norm in high dimension", *Internat. Math. Res. Notices* **2001**:6 (2001), 299–328. MR Zbl

[Tao 2001b] T. Tao, "Global regularity of wave maps, II: Small energy in two dimensions", *Comm. Math. Phys.* **224**:2 (2001), 443–544. MR Zbl

[Tao 2008] T. Tao, "Global regularity of wave maps, III: Large energy from $\mathbb{R}^{1+2}$ to hyperbolic spaces", preprint, 2008. arXiv

[Wang and Yu 2013] J. Wang and P. Yu, "Long time solutions for wave maps with large data", *J. Hyperbolic Differ. Equ.* **10**:2 (2013), 371–414. MR Zbl

[Yagdjian 2009] K. Yagdjian, "The semilinear Klein–Gordon equation in de Sitter spacetime", *Discrete Contin. Dyn. Syst. Ser. S* **2**:3 (2009), 679–696. MR Zbl

[Yagdjian 2012] K. Yagdjian, "Global existence of the scalar field in de Sitter spacetime", *J. Math. Anal. Appl.* **396**:1 (2012), 323–344. MR Zbl

[Yagdjian and Galstian 2008] K. Yagdjian and A. Galstian, "Fundamental solutions of the wave equation in Robertson–Walker spaces", *J. Math. Anal. Appl.* **346**:2 (2008), 501–520. MR Zbl

[Yagdjian and Galstian 2009] K. Yagdjian and A. Galstian, "Fundamental solutions for the Klein–Gordon equation in de Sitter spacetime", *Comm. Math. Phys.* **285**:1 (2009), 293–344. MR Zbl

[Yang 2015] S. Yang, "Global solutions of nonlinear wave equations with large data", *Selecta Math.* (*N.S.*) **21**:4 (2015), 1405–1427. MR Zbl

Received February 23, 2015. Revised January 27, 2016.

CHANG-HUA WEI
DEPARTMENT OF MATHEMATICS
ZHEJIANG SCI-TECH UNIVERSITY
HANGZHOU, 310018
CHINA
changhuawei1986@gmail.com

NING-AN LAI
DEPARTMENT OF MATHEMATICS, SCIENCES COLLEGE
LISHUI UNIVERSITY
LISHUI, 323000
CHINA
hyayue@gmail.com

CONTENTS

Volume 289, no. 1 and no. 2

Guidelines for Authors

Authors may submit articles at msp.org/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 289 No. 2 August 2017

0030-8730(201708)289:2;1-O