Pacific Journal of Mathematics

Volume 209 No. 1 March 2003

PACIFIC JOURNAL OF MATHEMATICS

http://www.pjmath.org

Founded in 1951 by

E. F. Beckenbach (1906–1982) F. Wolf (1904–1989)

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

Regular subscription rate for 2006: $425.00 a year (10 issues). Special rate: $212.50 a year to individual members of supporting institutions. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163, U.S.A. Prior back issues are obtainable from Periodicals Service Company, 11 Main Street, Germantown, NY 12526-5635. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, PASCAL CNRS Index, Referativnyi Zhurnal, Current Mathematical Publications and the Science Citation Index.

The Pacific Journal of Mathematics (ISSN 0030-8730) at the University of California, c/o Department of Mathematics, 969 Evans Hall, Berkeley, CA 94720-3840 is published monthly except July and August. 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.

PUBLISHED BY PACIFIC JOURNAL OF MATHEMATICS
at the University of California, Berkeley 94720-3840
A NON-PROFIT CORPORATION
Typeset in LaTeX

PACIFIC JOURNAL OF MATHEMATICS
Vol. 209, No. 1, 2003

BOOLEAN ALGEBRAS OF PROJECTIONS & ALGEBRAS OF SPECTRAL OPERATORS

H.R. Dowson, M.B. Ghaemi, and P.G. Spain

We show that, given a weak compactness condition which is always satisfied when the underlying space does not contain an isomorphic copy of c_0, all the operators in the weakly closed algebra generated by the real and imaginary parts of a family of commuting scalar-type spectral operators on a Banach space will again be scalar-type spectral operators, provided that (and this is a necessary condition with even only two operators) the Boolean algebra of projections generated by their resolutions of the identity is uniformly bounded.

1. Introduction.

The problem we address, raised by Dunford [**8**] in 1954, is to find conditions under which the sum and product of a pair of commuting scalar-type spectral operators on a Banach space is also a scalar-type spectral operator.

Two difficulties arise when working on an arbitrary Banach space, as opposed to a Hilbert space: the unit ball of the algebra of bounded linear operators need not be weakly compact; and the Boolean algebra generated by two uniformly bounded Boolean algebras of projections need not be bounded [**15**].

In view of this we must restrict ourselves to the case where the Boolean algebra generated by the resolutions of the identities is uniformly bounded.

Previous treatments of this problem [to show that the sum of two commuting scalar-type spectral operators is a scalar-type spectral operator] have focussed on identifying the resolution of the identity of the sum [**11**, **16**, **20**]. These methods have worked essentially only when X contains no copy of c_0. However, this is precisely the case when one can exploit Grothendieck's theorem on the automatic weak compactness of linear mappings from a C*-algebra into X, and prove somewhat more: that all operators in the weakly closed involutory algebra generated by them are scalar-type spectral operators. An advantage of this approach is that one does not have to identify the resolutions of the identity of the sums, or products, or limits, directly.

2. C*-algebras on Banach spaces.

The properties of scalar-type spectral operators and the involutory algebras they generate seem best explained in the context of numerical range, of hermitian operators, and of C*-algebras. For the sake of completeness, and the convenience of the reader, we present a résumé of the key results.

Consider a complex Banach space X; write $L(X)$ for the Banach algebra of bounded linear operators on X, endowed with the operator norm.

We write A_1 for the unit ball of a subset A of a normed space.

We write $\langle x,\, x' \rangle$ for the value of the functional x' in X' at x in X. Let $\boldsymbol{\omega}$ be the linear span of the functionals $\omega_{x,x'} : L(X) \to \mathbb{C} : T \mapsto \langle Tx,\, x' \rangle$. Let Π be the set

$$\bigl\{(x, x') \in X \times X' : \bigl\langle x,\, x' \bigr\rangle = \|x\| = \bigl\|x'\bigr\| = 1\bigr\}$$

and let $\boldsymbol{\omega}_\Pi$ be the set of functionals

$$\{\omega_{x,x'} : (x, x') \in \Pi\}.$$

The *strong operator topology* and *weak operator topology* on $L(X)$ are of paramount importance: important here too are the BWO *topology* and BSO *topology,* the strongest topologies coinciding with the weak and strong topologies on bounded subsets of $L(X)$ — see [**9**, VI, 9].

The *ultraweak operator topology* on $L(X)$ is the topology generated by the seminorms $T \mapsto |\sum_n \langle Tx_n,\, x'_n \rangle|$ where $\{x_n\}$ and $\{x'_n\}$ range over pairs of sequences in X and X' subject to $\sum_n \|x_n\| \, \|x'_n\| < \infty$. The *ultrastrong operator topology* on $L(H)$ is the topology generated by the seminorms $T \mapsto \left\{\sum_n \|Tx_n\|^2\right\}^{\frac{1}{2}}$ where $\{x_n\}$ ranges over sequences for which $\sum_n \|x_n\|^2 < \infty$.

The BWO topology coincides with the ultraweak topology, the BSO topology with the ultrastrong topology, on $L(H)$, when H is a Hilbert space.

The (*spatial*) *numerical range* $V(T)$ of an operator T is defined to be

$$V(T) \triangleq \bigl\{ \bigl\langle Tx,\, x' \bigr\rangle : (x, x') \in \Pi \bigr\}.$$

An operator R on X is *hermitian* if its numerical range is real i.e., if $V(R) \subset \mathbb{R}$; equivalently, if

$$\bigl\{ \|\exp(irR)\| : r \in \mathbb{R} \bigr\}$$

is bounded. The set of hermitian operators is closed in the norm, strong and weak operator topologies.

The following result is crucial:

Theorem 2.1 (Vidav-Palmer Theorem)**.** *Suppose that $\mathcal{A}$ is a unital subalgebra of $L(X)$ [the unit being the identity operator on X]. Let $\mathcal{H}$ be the set of hermitian elements of $\mathcal{A}$. Then $\mathcal{A} = \mathcal{H} + i\mathcal{H}$ if and only if $\mathcal{A}$ is a pre-*C*-algebra under the operator norm and the natural involution*

$$* : \mathcal{A} \to \mathcal{A} : R + iJ \mapsto R - iJ \qquad (R,\, J \in \mathcal{H}).$$

It then follows that $\mathcal{B} \triangleq \overline{\mathcal{A}}$ is a C*-algebra on X, containing the identity I_X on X. (See [**3**, §38] for a discussion of these topics.)

When $\mathcal{B}$ is a C*-algebra on X the family $\boldsymbol{\omega}_\Pi$ is a *separating* family of *states* on $\mathcal{B}$.

We shall use the following terminology: a *von Neumann algebra* is a weakly closed C*-algebra of operators on a Hilbert space, while a W*-*algebra* is a C*-algebra which has a realisation as a von Neumann algebra [equivalently, is a dual space of a Banach space].

Unital *-isomorphisms of C*-algebras are isometric.

Theorem 2.2 (BWO Closure Theorem)**.** *Suppose that* $\mathcal{B}$ *is a* C*-*algebra on* X *and that its unit ball* $\mathcal{B}_1$ *is relatively weakly compact. Then the* BWO *closure of* $\mathcal{B}$*,*

$$\mathcal{B}^{\frown} \triangleq \bigcup_{n=1}^{\infty} n\overline{\mathcal{B}_1}^{w},$$

is a W*-*algebra; and* $(\mathcal{B}^{\frown})_1 = \overline{\mathcal{B}_1}^{w}$*. Moreover, any faithful representation of* $\mathcal{B}^{\frown}$ *as a concrete von Neumann algebra is* BWO *bicontinuous.*

The proof [**24**] rests on the fact that, by the identity of comparable compact Hausdorff topologies, the weak topology on $\overline{\mathcal{B}_1}^{w}$ is the weak topology induced by the states $\boldsymbol{\omega}_\Pi$.

It remains open, in general, to decide whether $\mathcal{B}^{\frown} = \overline{\mathcal{B}}^{w}$.

2.1. Commutative C*-algebras on X.

The remaining results in this section apply to any commutative unital C*-subalgebra $\mathcal{B}$ of $L(X)$, and in particular to any algebra generated by a Boolean algebra of (hermitian) projections: see §3.

The operators in a commutative C*-subalgebra of $L(X)$ are called *normal* (sometimes *strongly normal*). *Abstractly,* they enjoy all the properties of normal operators on Hilbert spaces.

Let Λ be the maximal ideal space of $\mathcal{B}$ and Θ the *inverse Gelfand map*

$$\Theta : C(\Lambda) \to \mathcal{B}$$

which is a unital isometric *-isomorphism: Θ is also called the *functional calculus* for $\mathcal{B}$.

On restricting Θ to the C*-subalgebra generated by I, T (for any $T \in \mathcal{B}$) we obtain a functional calculus for a (strongly) normal T: a unital isometric *-isomorphism

$$\Theta_T : C(\mathrm{sp}(T)) \to \mathcal{B}$$

such that

$$\begin{aligned} \Theta_T(z \mapsto 1) &= I \\ \Theta_T(z \mapsto z) &= T \\ \Theta_T(z \mapsto \overline{z}) &= T^* \\ \|\Theta_T(f)\| &= \|f\|_{\mathrm{sp}(T)} \qquad \big(f \in C(\mathrm{sp}(T))\big). \end{aligned}$$

The following two lemmas demonstrate how to some extent normal operators on a Banach space mimic normal operators on a Hilbert space:

Lemma 2.3. *Let $\mathcal{B}$ be a commutative* C*-*algebra on X and let $\mathcal{H}$ be the set of hermitian elements of $\mathcal{B}$. Suppose that ${}^{H}_{K} \in \mathcal{H}$ and $0 \le H \le K$. Then*

$$\|Hx\| \le \|Kx\| \qquad \big(x \in X\big).$$

For any $\varepsilon > 0$ the operator $L = H/(K + \varepsilon I)$ is defined in $\mathcal{H}$, and, by the functional calculus, $0 \le L \le 1$; so $\|L\| \le 1$. It follows that $\|Hx\| = \|L(K + \varepsilon I)x\| \le \|(K + \varepsilon I)x\|$. Now let $\varepsilon \to 0$.

The next result, originally due to Palmer [**18**, Lemma 2.7], helps us extend the C* structure from $\mathcal{B}$ to $\mathcal{C} \triangleq \overline{\mathcal{B}}^w$. The following short proof is taken from [**4**]:

Lemma 2.4. *For all $B \in \mathcal{B}$ and $x \in X$ we have*

$$\|Bx\| = \|B^*x\|.$$

Proof. For $\varepsilon > 0$ the functional calculus gives

$$\left\|B - B^2(B^*B + \varepsilon I)^{-1}B^*\right\| = \left\|\varepsilon B(B^*B + \varepsilon I)^{-1}\right\| \le \sqrt{\varepsilon}/2,$$

and

$$\left\|B^2(B^*B + \varepsilon I)^{-1}\right\| \le 1.$$

Thus, for any $x \in X$,

$$\|Bx\| = \lim_{\varepsilon \to 0} \left\|B^2(B^*B + \varepsilon I)^{-1}B^*x\right\| \le \|B^*x\|,$$

and then $\|B^*x\| \le \|B^{**}x\| = \|Bx\|$. □

The weak closure of a *commutative* C*-algebra on X is also a C*-algebra on X.

Theorem 2.5. *Let $\mathcal{B}$ be a commutative* C*-*algebra on X and let $\mathcal{H}$ be the set of hermitian elements of $\mathcal{B}$. Let $\overline{\mathcal{H}}^w$ be the weak operator topology closure of $\mathcal{H}$, and $\overline{\mathcal{B}}^w$ the weak operator topology closure of $\mathcal{B}$. Then*

$$\overline{\mathcal{B}}^w = \overline{\mathcal{H}}^w + i\overline{\mathcal{H}}^w$$

and so $\overline{\mathcal{B}}^w$ is a C*-*algebra. Moreover, $\big(\overline{\mathcal{B}}^w\big)_1 = \overline{\mathcal{B}_1}^w$. Hence $\mathcal{B}^\sim = \overline{\mathcal{B}}^w$.*

Proof. First note that the weak and strong closures coincide for $\mathcal{H}$ and $\mathcal{B}$ (they are both convex sets). Now Lemma 2.4 shows that $\overline{\mathcal{B}}^s = \overline{\mathcal{H}}^s + i\overline{\mathcal{H}}^s$, so $\overline{\mathcal{B}}^w$ is a C*-algebra.

Consider $H \in (\overline{\mathcal{H}}^w)_1$. Then $K = \big(I - [I - H^2]^{\frac{1}{2}}\big)/H \ \in \overline{\mathcal{H}}^w$, and $H = 2K/(I + K^2)$. Take a net K_α in $\mathcal{H}$ converging strongly to K: put $H_\alpha = 2K_\alpha/(I + K_\alpha^2)$. Then

$$H_\alpha - H = 2(I + K_\alpha^2)^{-1}\,(K_\alpha - K)\,(I + K^2)^{-1} + \frac{1}{2}H_\alpha(K - K_\alpha)H$$

so $H \in \overline{\mathcal{H}_1}^w$. By the Russo-Dye Theorem [**3**, §38] we have $(\overline{\mathcal{B}}^w)_1 \subseteq \overline{\mathcal{B}_1}^w$. □

Corollary 2.6. *If, further, the unit ball of $\mathcal{B}$ is relatively weakly compact, then $\overline{\mathcal{B}}^w$ is a* W**-algebra and any faithful representation of $\overline{\mathcal{B}}^w$ as a concrete von Neumann algebra on a Hilbert space is* BWO *bicontinuous* (*that is, weakly bicontinuous on bounded sets*).

Proof. Use Theorem 2.2. □

Remark 2.7. We show later (§4) that any such faithful representation is also BSO bicontinuous (that is, strongly bicontinuous on bounded sets). The proof (maybe the result) depends on being able to represent $\overline{\mathcal{B}}^w$ by a spectral measure: and the presence of c_0 as a subspace of X seems to be the natural obstruction to this: see §6 below.

3. Boolean algebras of projections & the algebras they generate.

Let X be a complex Banach space, and $\mathcal{E}$ a *bounded Boolean algebra of projections on* X:

$$\begin{array}{rcll} I \in \mathcal{E} & \subseteq & L(X) & \\ E \in \mathcal{E} & \Longrightarrow & E^2 = E & \\ E \in \mathcal{E} & \Longrightarrow & I - E \in \mathcal{E} & \\ E,\, F \in \mathcal{E} & \Longrightarrow & EF = FE \in \mathcal{E} & \\ \|E\| & \leq & K_{\mathcal{E}} & \qquad (E \in \mathcal{E}) \end{array}$$

for some constant $K_{\mathcal{E}}$. Write $\operatorname{aco}\mathcal{E}$ for the *absolutely convex hull* of $\mathcal{E}$ in $L(X)$.

It can be shown (see [**6**, 5.4]) that then

$$\mathcal{S} = \left\{ \sum_{\text{finite}} \lambda_j E_j : |\lambda_j| \leq 1, E_j \in \mathcal{E}, E_j E_k = 0 \ (j \neq k) \right\}$$

is a bounded multiplicative semigroup of operators on X. If we define

$$\|x\|_{\mathcal{E}} = \sup\left\{ \|Sx\| : S \in \mathcal{S} \right\} \qquad (x \in X)$$

we obtain a norm $\|\cdot\|_{\mathcal{E}}$ on X, equivalent to the original norm on X, with respect to which each element of $\mathcal{E}$ is hermitian. Thus, without loss of generality,

we shall assume that all elements of $\mathcal{E}$ are hermitian.

Remark 3.1. By Sinclair's Theorem $\|E\| = 1$ for any nonzero hermitian projection.

Theorem 3.2. *Let $\mathcal{E}$ be a Boolean algebra of hermitian projections on a complex Banach space X. Then $\mathcal{A}$, the linear span of $\mathcal{E}$, is the *-algebra generated by $\mathcal{E}$: $\mathcal{A}$ is a commutative unital algebra, and $\mathcal{A} = \mathcal{H} + i\mathcal{H}$, where $\mathcal{H}$ is the set of hermitian elements of $\mathcal{A}$. So $\mathcal{B}$, $\triangleq \overline{\mathcal{A}}$, is a commutative* C*-*algebra on X.*

Proof. Immediate from the Vidav-Palmer Theorem (Theorem 2.1). ☐

Lemma 3.3. *Let $S \in \mathcal{A}$ and suppose that $-I \leq S \leq I$. Then*

$$S \in 2\,\mathrm{aco}\,\mathcal{E}.$$

Proof. Suppose first that $0 \leq S \leq I$. Write S in *$\mathcal{E}$-step-form* as $S = \sum_{j=1}^{M} \lambda_j E_j$, where the E_j are pairwise disjoint. Then $0 \leq \lambda_j \leq 1$. Arrange the λ_j in descending order: then $\|S\| = \lambda_1$. Define $\lambda_{M+1} = 0$ and use Abel summation —

$$S = \sum_{j=1}^{M} \lambda_j E_j = \sum_{j=1}^{M} (\lambda_j - \lambda_{j+1}) \left(\sum_{h=1}^{j} E_h \right) \in \mathrm{aco}\,\mathcal{E}.$$

If $-I \leq S \leq I$, split S into its positive and negative parts. ☐

Theorem 3.4. *Let $\mathcal{E}$ be a Boolean algebra of hermitian projections on a complex Banach space X, and let $\mathcal{B}$ be the* C*-*algebra it generates. Let $\mathcal{B}_1$ be the closed unit ball of $\mathcal{B}$. Then*

$$\mathcal{B}_1 \subseteq 4\,\overline{\mathrm{aco}}\,\mathcal{E}.$$

Proof. Consider an element $B \in \mathcal{B}$ such that $\|B\| < 1$. Given $\varepsilon > 0$ we can find $S = R + iJ$ in $\mathcal{A}$ such that $\|B - R - iJ\| \leq \min\{\varepsilon, 1 - \|B\|\}$. Now $\begin{smallmatrix}\|R\| \\ \|J\|\end{smallmatrix} \leq 1$, so that, by Lemma 3.3, $\begin{smallmatrix}R \\ J\end{smallmatrix} \in 2\,\mathrm{aco}\,\mathcal{E}$. ☐

Corollary 3.5. *The following are equivalent:*

1) *$\mathcal{B}_1$ is relatively weakly compact.*
2) aco $\mathcal{E}$ *is relatively weakly compact.*
3) *$\mathcal{E}$ is relatively weakly compact.*

Proof. Use the Krein-Šmulian Theorem. ☐

We can now state the main theorem of this section.

Theorem 3.6. *Let $\mathcal{E}$ be a relatively weakly compact Boolean algebra of hermitian projections on a complex Banach space X, and let $\mathcal{B}$ be the* C*-*algebra generated by $\mathcal{E}$. Then $\overline{\mathcal{B}}^w$ is a* W*-*algebra and any faithful representation of $\overline{\mathcal{B}}^w$ as a concrete von Neumann algebra on a Hilbert space is* BWO *bicontinuous (that is, weakly bicontinuous on bounded sets).*

Proof. This follows from Corollary 3.5 and Theorem 2.2. □

4. σ-complete Boolean algebras of projections & spectral measures.

The fundamental results on Boolean algebras of projections on a Banach space were developed by Bade and are to be found in [**10**, XVII]. Much interesting material on this topic is also to be found in [**21**].

Following [**10**] we say that an abstract Boolean algebra $\mathcal{E}$ is (σ-)*complete* if each (countable) subset of $\mathcal{E}$ has a supremum and infimum in $\mathcal{E}$.

$\mathcal{E}$, a Boolean algebra of projections on X, is (σ-)*complete on X* if each (countable) subset $\mathcal{F}$ of $\mathcal{E}$ has a supremum and infimum in $\mathcal{E}$ such that

$$\left(\bigvee \mathcal{F}\right) X = \overline{\text{lin}}\{F\,X \,:\, F \in \mathcal{F}\}, \qquad \left(\bigwedge \mathcal{F}\right) X = \bigcap_{F\in\mathcal{F}} F\,X.$$

It has been shown that $\mathcal{E}$ is (σ-)complete on X if and only if every bounded monotone (sequence) net in $\mathcal{E}$ converges strongly to a limit [**10**, XVII.3.4]. In this case $\mathcal{E}$ must be bounded [**10**, XVII.3.3].

On Hilbert space. On a Hilbert space $\mathcal{H}$ the following two *facts* are classical. We sketch their (elementary) proofs for the convenience of the reader.

Fact 4.1. Any monotone net of hermitian projections on $\mathcal{H}$ has a supremum, to which it converges strongly.

Proof. Let $\left(E_\alpha\right)_{\alpha\in A}$ be such a net. The generalized Cauchy-Schwarz inequality $\left\langle P^2\xi,\, \xi\right\rangle \le \langle P\xi,\, \xi\rangle \left\langle P^3\xi,\, \xi\right\rangle$, which holds for any positive operator P on $\mathcal{H}$ and any element $\xi \in \mathcal{H}$, shows that the net $\left(E_\alpha\right)_{\alpha\in A}$ is strongly Cauchy. Also, its limit must be the supremum. □

Fact 4.2. Suppose that $\left(E_\alpha\right)_{\alpha\in A}$ is a net of hermitian projections that converges *weakly* to a projection E. Then it converges *strongly.*

Proof. This is immediate from the calculation

$$\begin{aligned}\|(E - E_\alpha)\,\xi\|^2 &= \left\langle (E - E_\alpha)^2\,\xi,\, \xi\right\rangle \\ &= \left\langle E^2\xi,\, \xi\right\rangle - \langle EE_\alpha\,\xi,\, \xi\rangle - \langle E_\alpha E\,\xi,\, \xi\rangle + \left\langle E_\alpha^2\xi,\, \xi\right\rangle \\ &\to \left\langle (E - E^2)\,\xi,\, \xi\right\rangle = 0.\end{aligned}$$

□

It follows that *on a Hilbert space* every Boolean algebra $\mathcal{E}$ of hermitian projections can be extended to a *complete* one; that $\overline{\mathcal{E}}^s$ is the smallest such complete extension; and that $\overline{\mathcal{E}}^s = \overline{\mathcal{E}}^w \bigcap \{\text{projections on } \mathcal{H}\}$.

On a Banach space the situation is more delicate. It has been shown that if $\mathcal{E}$ is σ-complete on X then $\overline{\mathcal{E}}^s$ is complete on X [**10**, XVII.3.23], and that the family of projections in $\overline{\mathcal{E}}^w$ coincides with $\overline{\mathcal{E}}^s$. See Corollary 4.10 below for a proof [independent of Bade's original methods].

We shall require the following result, proposed as an exercise in [**9**]:

Lemma 4.3. *If $\mathcal{S} \subset L(X)$ then $\mathcal{S}$ is relatively compact in the weak operator topology if and only if the sets $\mathcal{S}\,x$ are relatively weakly compact for all $x \in X$.*

Proof. See [**9**, VI.9.2]. □

4.1. Spectral measures. Let Σ be a σ-algebra of subsets of a set Ω and Γ a total subset of X'. A *spectral measure of class* (Σ, Γ) is a Boolean algebra homomorphism $\sigma \mapsto E(\sigma)$ from Σ into $L(X)$ such that $\langle E(\sigma)x,\, x'\rangle$ is countably additive for each $x \in X$ and $x' \in \Gamma$: by the Banach-Orlicz-Pettis theorem any spectral measure of class X' is strongly countably additive.

A σ-complete Boolean algebra of projections $\mathcal{E}$ on X can be identified with the range of a spectral measure of class X' on the Borel sets of the Stone space of $\mathcal{E}$ ([**5**, Chapter I]): then each vector measure $\mathcal{E}\,x$ is strongly countably additive.

Lemma 4.4. *If μ is a strongly countably additive vector measure with values in X then* $\operatorname{aco}\{\mu(\sigma) : \sigma \in \Sigma\}$ *is relatively weakly compact.*

Proof. Essentially this is a result of Bartle, Dunford and Schwartz [**1**, 2.3]: see also [**5**, I.2.7 & I.5.3]. □

Corollary 4.5. *If $\mathcal{E}$ is σ-complete then the set $\overline{\operatorname{aco}}^{\,w}(\mathcal{E}x)$ is weakly compact for each $x \in X$.*

Theorem 4.6. *Let $\mathcal{E}$ be a σ-complete Boolean algebra of hermitian projections. Then $\mathcal{C}$, $\triangleq \overline{\mathcal{B}}^w$, the commutative* C**-algebra generated by $\mathcal{E}$ in the weak operator topology, is a* W**-algebra, and $\mathcal{C}_1 = \overline{\mathcal{B}_1}^w \subseteq 4\,\overline{\operatorname{aco}}^{\,w}\,\mathcal{E}$. Furthermore, any faithful representation of $\mathcal{C}$ as a von Neumann algebra on a Hilbert space is weakly bicontinuous on bounded sets.*

Proof. $\overline{\operatorname{aco}}^{\,w}(\mathcal{E}x)$ is weakly compact for each $x \in X$ (Corollary 4.5) so $\operatorname{aco}(\mathcal{E})$ is relatively weakly compact, by Lemma 4.3. Apply Theorem 3.6. □

Theorem 4.7. *Let $\mathcal{B}$ be a commutative* C**-algebra on X such that $\mathcal{B}_1$ is relatively weakly compact. Let $\mathcal{C} = \overline{\mathcal{B}}^w$. Then there is a representing spectral measure $E(\cdot)$ defined on the Borel sets of the Gelfand space Λ of $\mathcal{C}$ such that*

$$\Theta(f) = \int_\Lambda f(\lambda) E(d\lambda) \qquad (f \in C(\Lambda)).$$

Proof. Let $\pi : \mathcal{C} \to L(H)$ be a BWO continuous representation of $\mathcal{C}$ as a concrete W*-algebra. Let $\widetilde{E}(\cdot)$ be a representing spectral measure for $\pi(\mathcal{C})$:

$$\pi \circ \Theta(f) = \int_\Lambda f(\lambda)\widetilde{E}(d\lambda) \qquad (f \in C(\Lambda)).$$

Now define $E(\cdot) = \pi^{-1}\widetilde{E}(\cdot)$: this yields a spectral measure on X $\big[E(\cdot)$ is weakly countably additive, and so, by the Banach-Orlicz-Pettis theorem, strongly countably additive$\big]$: and then

$$\Theta(f) = \int_\Lambda f(\lambda)E(d\lambda) \quad (f \in C(\Lambda)).$$

□

It is immediate that for a bounded net $\big(T_\alpha\big)_{\alpha\in A}$ of operators *on a Hilbert space* we have

$$\big(T_\alpha\big)_{\alpha\in A} \to_{\text{strongly}} 0 \iff \big(T_\alpha^* T_\alpha\big)_{\alpha\in A} \to_{\text{weakly}} 0.$$

A similar result holds for normal operators on a Banach space provided that they belong to a *common* W*-algebra.

Theorem 4.8. *Let $\mathcal{C}$ be a commutative* W*-*algebra on X. Suppose that $\big(S_\alpha\big)_{\alpha\in A}$ is a bounded net in $\mathcal{C}$. Then*

$$\big(S_\alpha\big)_{\alpha\in A} \to_{\text{strongly}} 0 \iff \big(S_\alpha^* S_\alpha\big)_{\alpha\in A} \to_{\text{weakly}} 0.$$

Proof. Clearly $S_\alpha \to_{\text{strongly}} 0$ implies that $S_\alpha^* S_\alpha \to_{\text{strongly}} 0$, whence $S_\alpha^* S_\alpha \to_{\text{weakly}} 0$.

Let $E(\cdot)$ be the representing spectral measure for $\mathcal{C}$ guaranteed by Theorem 4.7.

Suppose that $S_\alpha^* S_\alpha \to_{\text{weakly}} 0$. Let $f_\alpha = \Theta^{-1} S_\alpha$. Then

$$\lim_\alpha \left\langle S_\alpha^* S_\alpha x,\, x'\right\rangle = \lim_\alpha \int_\Lambda |f_\alpha|^2 \left\langle E(d\lambda)x,\, x'\right\rangle \qquad (x \in X,\ x' \in X').$$

Therefore $\lim_\alpha f_\alpha = 0$ in var $\langle E(\cdot)x,\, x'\rangle$ measure and $\lim_\alpha \int_\Lambda f_\alpha \langle E(d\lambda)x,\, x'\rangle = 0$. For fixed $x \in X$ the set $\{\langle E(\cdot)x,\, x'\rangle : \|x'\| \le 1\}$ is a relatively weakly compact set of measures [**9**, IV.10.2]: hence $\lim_\alpha \int_\Lambda f_\alpha \langle E(d\lambda)x,\, x'\rangle = 0$ uniformly for $\|x'\| \le 1$ [**14**, Théorème 2]. Therefore $\lim_\alpha \int_\Lambda f_\alpha E(d\lambda)x = 0$; that is, $S_\alpha \to_{\text{strongly}} 0$. □

Corollary 4.9. *Let $\mathcal{C}$ be a commutative* W*-*algebra on X. Then any faithful concrete representation of $\mathcal{C}$ as a von Neumann algebra is weakly and strongly bicontinuous on bounded sets.*

Corollary 4.10. *Let $\mathcal{E}$ be a σ-complete Boolean algebra of hermitian projections, and let $\big(E_\alpha\big)_{\alpha\in A}$ be a monotone net of hermitian projections in* *the commutative* W*-*algebra* $\mathcal{C}$ *generated on* X *by* $\mathcal{E}$*. Then* $\left(E_\alpha\right)_{\alpha\in A}$ *converges strongly to a projection in* $\mathcal{C}$*. So* $\overline{\mathcal{E}}^s$ *is complete on* X*. What is more,* $\overline{\mathcal{E}}^s = \overline{\mathcal{E}}^w \bigcap \{$*projections in* $\mathcal{C}\}$*.*

Proof. This follows immediately from the known results on Hilbert spaces and from the strong bicontinuity of faithful representations guaranteed by the theorem. □

The next corollary complements [**23**, Theorem 5] and [**12**, Theorems 1, 2].

Corollary 4.11. *Let* $\mathcal{E}$ *be a bounded Boolean algebra of projections on a Banach space* X *and suppose that* $\mathcal{E}$ *is relatively weakly compact. Then* $\mathcal{E}$ *has a* (σ-)*complete extension contained in* $\overline{\mathcal{E}}^s$*.*

Remark 4.12. This happens automatically when $X \not\supset c_0$ (see §6).

Corollary 4.13 ([**10**, XVII.3.7])**.** *Let* $\mathcal{E}$ *be a complete bounded Boolean algebra of projections on a Banach space* X*. Then* $\mathcal{E}$ *is strongly closed.*

Remark 4.14. The results of [**7**] overlap with ours.

5. Spectral operators.

An operator $T \in L(X)$ is *prespectral of class* Γ if there is a spectral measure $E(\cdot)$ of class (Σ_p, Γ) (here Σ_p is the family of Borel subsets of the complex plane) such that for all $\sigma \in \Sigma_p$:

$$T\,E(\sigma) = E(\sigma)\,T, \tag{1}$$

$$\operatorname{sp}(T\big|E(\sigma)X) \subseteq \overline{\sigma}. \tag{2}$$

The spectral measure $E(\cdot)$ is called a *resolution of the identity of class* Γ for T. If, further, $T = \int_{\operatorname{sp}(T)} \lambda E(d\lambda)$, then T is a *scalar-type operator of class* Γ.

Remark 5.1. Given a scalar-type spectral operator $T = \int_{\operatorname{sp}(T)} \lambda E(d\lambda)$ we can define its *real part* $\Re\, T = \int_{\operatorname{sp}(T)} \Re\,\lambda\; E(d\lambda)$, and its *imaginary part* $\Im\, T = \int_{\operatorname{sp}(T)} \Im\,\lambda\; E(d\lambda)$. By the (closed) *-algebra generated by T we mean the (closed) algebra generated by $\Re\, T$ and $\Im\, T$.

An operator $T \in L(X)$ is a *spectral operator* if it is prespectral of class X': that is, if there is a spectral measure $E(\cdot)$ of class X' satisfying Conditions (1) and (2) above, and if also

$$E(\cdot) \text{ is strongly countably additive on } \Sigma_p.$$

An important property of spectral operators is that if T is spectral and S commutes with T, then S commutes with the resolution of the identity of T [**6**, Theorem 6.6].

Scalar-type spectral operators have been characterised as follows:

Theorem 5.2 ([**17**] & [**22**, Theorem])**.** *The operator $T \in L(X)$ is a scalar-type spectral operator if and only if it satisfies the following two conditions:*

(1) *T has a functional calculus,* and
(2) *for every $x \in X$ the map $\Theta_x : C(\mathrm{sp}(T)) \to X : f \mapsto \Theta(f)x$ is weakly compact.*

Note that by Lemma 4.3 Property (2) is equivalent to:

(2′) *The functional calculus $\Theta : C(\mathrm{sp}(T)) \to L(X)$ is weakly compact* in the sense that $\Theta\left(\left\{f \in C(\mathrm{sp}(T)) : \|f\|_{\mathrm{sp}(T)} \leq 1\right\}\right)$ is relatively compact in the weak operator topology of $L(X)$.

6. In the absence of c_0.

The following theorem goes back to Grothendieck, Bartle-Dunford-Schwartz, and others. See [**5**, VI, Notes] for an interesting discussion of its genesis and development.

Theorem 6.1. *If $\mathcal{B}$ is a* C*-*algebra, if $\Theta : \mathcal{B} \to X$ is a bounded operator, and X does not contain a subspace isomorphic to c_0, then Θ is a weakly compact mapping.*

Remarks on the proof. A stronger version of this theorem, where $\mathcal{B}$ may be any complete Jordan algebra of operators, not necessarily commutative, can be found in [**25**, Theorem 2]. That proof relies on James's characterisation of weakly compact sets and the Bessaga-Pełczyński result that X contains no copy of c_0 if and only if all series $\sum_n x_n$ in X with $\sum_n |\langle x_n, x'\rangle|$ convergent for all $x' \in X'$ are unconditionally norm convergent.

Corollary 6.2. *Let T be a normal operator on a Banach space X that does not contain a subspace isomorphic to c_0. Then T is a scalar-type spectral operator.*

Proof. T has a functional calculus (see §2) which, by the theorem, is weakly compact. Apply Theorem 6.1. □

We can now present a theorem which is stronger than any other known to us in this area.

Theorem 6.3. *Let $\mathcal{E}$ be a bounded Boolean algebra of hermitian projections on a Banach space X and suppose that X does not contain a subspace isomorphic to c_0. Then the weakly closed algebra $\overline{\mathcal{B}}^w$ generated by $\mathcal{E}$ is a* W*-*algebra and any faithful representation of $\overline{\mathcal{B}}^w$ as a concrete von Neumann algebra on a Hilbert space is* BWO *and* BSO *bicontinuous. Moreover, every operator in $\overline{\mathcal{B}}^w$ is a scalar-type spectral operator.*

Proof. Theorem 6.1 shows that $\mathcal{E}$ is relatively weakly compact. The result follows from Theorem 3.6, Corollary 4.9, and Corollary 6.2. □

Corollary 6.4. *Let $\mathcal{T}$ be a commuting family of scalar-type spectral operators on a Banach space X that does not contain a subspace isomorphic to c_0. Suppose that the Boolean algebra generated by the resolutions of the identity of T for each $T \in \mathcal{T}$ is uniformly bounded. Then every operator in the weakly closed *-algebra generated by $\mathcal{T}$ is a scalar-type spectral operator.*

It has recently been shown [**13**, Theorem 2.5] that on a Banach lattice the Boolean algebra generated by two commuting bounded Boolean algebras of projections is itself bounded. Hence:

Corollary 6.5. *Let X be a complex Banach lattice not containing a copy of c_0, and let $\mathcal{T}$ be a finite commuting family of scalar-type spectral operators on X. Then every operator in the weakly closed *-algebra generated by $\mathcal{T}$ is a scalar-type spectral operator.*

c_0 as the natural obstruction. If X contains c_0 then there is a strongly closed bounded Boolean algebra $\mathcal{F}$ of projections on X that is not complete [**12**, Theorem 2]. Then the weakly closed algebra generated by $\mathcal{F}$ cannot have relatively weakly compact unit ball, and there can be no BWO bicontinuous faithful representation of this algebra on a Hilbert space.

7. Boolean algebras with countable basis.

As remarked above, c_0 seems to be the natural essential obstruction to extending the results of the previous section. It is of course conceivable that closer analysis will lead to a proof that the sum and product of a pair of commuting scalar-type spectral operators must be scalar-type spectral operators so long as the Boolean algebra generated by their resolutions of the identity is bounded.

We shall say that a Boolean algebra $\mathcal{E}$ has a *countable basis* if it contains a countable orthogonal subfamily $\mathcal{F} = \left(F_m\right)_{m\in\mathbb{N}}$ such that every $E \in \mathcal{E}$ can be written as the strong *sum* of a subset of this family. Note that then $I = \sum_{m=1}^{\infty} F_m$, the sum being strongly convergent.

Lemma 7.1. *Let $\mathcal{C}$ be a commutative* C**-algebra on X and $\left(F_m\right)_{m\in\mathbb{N}}$ a countable family of positive elements of $\mathcal{C}$ such that $\sum_{m=1}^{\infty} F_m$ converges in the strong topology. Let C_m be any sequence in $\mathcal{C}$ for which $0 \le C_m \le I$ $(\forall m)$. Then*

$$\sum_{m=1}^{\infty} C_m F_m$$

converges strongly.

Proof. Note that $0 \le C_m F_m \le F_m$ $(\forall m)$. Then, for $M < N$,

$$0 \le \sum_{m=M+1}^{N} C_m F_m \le \sum_{m=M+1}^{N} F_m;$$

so, by Lemma 2.3, the sequence $\big(C_m F_m\big)_{m=\in\mathbb{N}}$ is a strongly Cauchy sequence, and hence strongly convergent. □

The following theorem generalises [**13**, Theorem 3.6]:

Theorem 7.2. *Suppose that $\mathcal{E}^{(1)}$ and $\mathcal{E}^{(2)}$ are two commuting σ-complete Boolean algebras of projections on X and that the Boolean algebra $\mathcal{E}$ generated by $\mathcal{E}^{(1)}$ and $\mathcal{E}^{(2)}$ is bounded. Assume, further, that $\mathcal{E}^{(2)}$ has a countable basis $\mathcal{F} = \big(F_m\big)_{m\in\mathbb{N}}$. Then $\mathcal{E}$ has a σ-complete extension, and hence a complete extension.*

Proof. As remarked in §3 we may, and shall, assume that all the elements of $\mathcal{E}^{(1)}$ and $\mathcal{E}^{(2)}$ are hermitian. Let $\mathcal{C}$ be the weakly closed C*-algebra generated by $\mathcal{E}$.

For each sequence of projections $\big(E_m^{(1)}\big)_{m\in\mathbb{N}}$ taken from $\mathcal{E}^{(1)}$ we can, by Lemma 7.1, define $E = \sum_{m=1}^{\infty} E_m^{(1)} F_m \in \mathcal{C}$. Each such E is a hermitian projection in $\mathcal{C}$ so has norm ≤ 1.

Consider

$$\mathcal{G} \triangleq \left\{ \sum_{m=1}^{\infty} E_m^{(1)} F_m : E_m^{(1)} \in \mathcal{E}^{(1)} \right\}.$$

It is clear that $F_m \in \mathcal{G}$ $(\forall m)$, so $\mathcal{E}^{(2)} \subseteq \mathcal{G}$. Note also that for any $E^{(1)} \in \mathcal{E}^{(1)}$ we have $E^{(1)} = \sum_m E^{(1)} F_m$, so $E^{(1)} \in \mathcal{G}$. Thus $\mathcal{E}^{(1)} \vee \mathcal{E}^{(2)} \subseteq \mathcal{G}$.

It is clear that $\mathcal{G}$ is closed under products. Further, for any

$$E = \sum_{m=1}^{\infty} E_m^{(1)} F_m \in \mathcal{G}$$

we have

$$I - E = \sum_{m=1}^{\infty} \big[I - E_m^{(1)}\big] F_m \in \mathcal{G},$$

and so $\mathcal{G}$ is a Boolean algebra of hermitian projections on X.

Note that for any such $E \in \mathcal{G}$ we have $EF_m = E_m^{(1)} F_m$ $(\forall m)$: thus any element of $\mathcal{G}$, which can be written, though not in a unique manner, as an (orthogonal) sum

$$E = \sum_{m=1}^{\infty} E_m^{(1)} F_m,$$

satisfies

$$E = \sum_{m=1}^{\infty} E_m^{(1)} F_m = \sum_{m=1}^{\infty} EF_m.$$

Now consider a sequence $\left(E_h\right)_{h\in\mathbb{N}}$ of pairwise orthogonal projections in $\mathcal{G}$:

$$E_h = \sum_{m=1}^{\infty} E_{h,m}^{(1)} F_m = \sum_{m=1}^{\infty} E_h F_m.$$

For each k and m define

$$G_{k,m} \triangleq \bigvee_{h=1}^{k} E_{h,m}^{(1)} \in \mathcal{E}^{(1)}$$

and then define

$$G_m \triangleq \bigvee_{k=1}^{\infty} G_{k,m} = \bigvee_{h=1}^{\infty} E_{h,m}^{(1)} \in \mathcal{E}^{(1)}.$$

Note that for each k and m

$$G_{k,m}F_m = \bigvee_{h=1}^{k} E_{h,m}^{(1)} F_m = \sum_{h=1}^{k} E_{h,m}^{(1)} F_m = \left(\sum_{h=1}^{k} E_h\right) F_m.$$

Suppose that $x \in X$ and $\varepsilon > 0$. Then there exists an M such that

$$\left\| x - \sum_{m=1}^{M} F_m\, x \right\| < \varepsilon$$

and so we can find N such that for $1 \leq m \leq M$ and $k \geq N$

$$\|(G_m - G_{k,m})x\| < \varepsilon/M.$$

Suppose that $j < k$: then $0 \leq \sum_{h=j+1}^{k} E_h \leq I$, and so, by Lemma 2.3,

$$\begin{aligned}
\left\| \left(\sum_{h=j+1}^{k} E_h \right) x \right\| &\leq \left\| \left(\sum_{h=j+1}^{k} E_h \right) \left(x - \sum_{m=1}^{M} F_m\, x \right) \right\| \\
&\quad + \sum_{m=1}^{M} \left\| \left(\sum_{h=j+1}^{k} E_h \right) F_m x \right\| \\
&\leq \left\| x - \sum_{m=1}^{M} F_m\, x \right\| + \sum_{m=1}^{M} \|(G_{k,m} - G_{j,m})\, F_m\, x\| \\
&\leq \left\| x - \sum_{m=1}^{M} F_m\, x \right\| + \sum_{m=1}^{M} \|(G_{k,m} - G_{j,m})\, x\| \\
&\leq \varepsilon + \varepsilon = 2\varepsilon.
\end{aligned}$$

This shows that $\mathcal{G}$ is σ-complete. Then $\overline{\mathcal{E}}^s$ is complete (Corollary 4.10). □

From this we obtain the following results.

Theorem 7.3. *Let $\mathcal{E}^{(1)}$ and $\mathcal{E}^{(2)}$ be two commuting σ-complete Boolean algebras of hermitian projections on X. Suppose that the Boolean algebra $\mathcal{E}$ generated by $\mathcal{E}^{(1)}$ and $\mathcal{E}^{(2)}$ is bounded, and that $\mathcal{E}^{(2)}$ has a countable basis. Then the weakly closed *-algebra $\mathcal{C}$ generated by $\mathcal{E}$ is a* W**-algebra.*

Corollary 7.4 (Extension of [**13**, 3.6])**.** *Let X be a Banach space and T_1, T_2 be commuting scalar-type spectral operators on X with resolutions of the identity $\mathcal{E}^{(1)}$, $\mathcal{E}^{(2)}$ such that $\mathcal{E}^{(1)} \vee \mathcal{E}^{(1)}$ is bounded. Suppose further that one of these operators has countable spectrum. Then all operators in the weakly closed *-algebra generated by T_1 and T_2 are scalar-type spectral operators.*

References

[1] R.G. Bartle, N. Dunford and J.T. Schwartz, *Weak compactness and vector measures*, Canad. J. Math., **7** (1955), 289-305, MR 16,1123c, Zbl 0068.09301.

[2] C. Bessaga and A. Pełczyński, *On bases and unconditional convergence of series in Banach spaces*, Studia Math., **17** (1958), 151-164, MR 22 #5872, Zbl 0084.09805.

[3] F.F. Bonsall and J. Duncan, *Complete Normed Algebras*, Springer-Verlag, 1973, MR 54 #11013, Zbl 0271.46039.

[4] M.J. Crabb and P.G. Spain, *Commutators and normal operators*, Glasgow Math. J., **18** (1977), 197-198, MR 56 #1115, Zbl 0351.47025.

[5] J. Diestel and J.J. Uhl Jr., *Vector Measures*, Amer. Math. Surveys, **15**, Amer. Math. Soc., 1977, MR 56 #12216, Zbl 0369.46039.

[6] H.R. Dowson, *Spectral Theory of Linear Operators*, Academic Press, 1978, MR 80c:47022, Zbl 0384.47001.

[7] H.R. Dowson and T.A. Gillespie, *A representation theorem for a complete Boolean algebra of projections*, Proc. Roy Soc. Edin., **83A** (1979), 225-237, MR 80m:47039, Zbl 0417.47014.

[8] N. Dunford, *Spectral operators*, Pacific J. Math., **4** (1954), 321-354, MR 16,142d, Zbl 0056.34601.

[9] N. Dunford and J.T. Schwartz, *Linear Operators.* I. *General Theory*, Interscience, New York, 1958, MR 22 #8302, Zbl 0084.10402.

[10] N. Dunford and J.T. Schwartz, *Linear Operators, Part* III: *Spectral Operators*, Interscience (Wiley), New York, 1971, MR 54 #1009.

[11] S.R. Foguel, *Sums and products of commuting spectral operators*, Ark. Mat., **3** (1957), 449-461, MR 21 #2914, Zbl 0081.12301.

[12] T.A. Gillespie, *Strongly closed bounded Boolean algebras of projections*, Glasgow Math. J., **22** (1981), 73-75, MR 82a:46018, Zbl 0455.47024.

[13] ———, *Boundedness criteria for Boolean algebras of projections*, J. Functional Anal., **148** (1997), 70-85, MR 98h:47048, Zbl 0909.46017.

[14] A. Grothendieck, *Sur les applications linéaires faiblement compactes d'espaces du type* $C(K)$, Canad. J. Math., **5** (1953), 129-173, MR 15,438b, Zbl 0050.10902.

[15] S. Kakutani, *An example concerning uniform boundedness of spectral measures*, Pacific J. Math., **4** (1954), 363-372, MR 16,143b, Zbl 0056.34702.

[16] S. Kantorovitz, *On the characterization of spectral operators*, Trans. Amer. Math. Soc., **110** (1964), 519-537, MR 28 #3329, Zbl 0139.08702.

[17] I. Kluvanek, *Characterization of scalar-type spectral operators*, Arch. Math. (Brno), **2** (1966), 153-156, MR 35 #2163, Zbl 0206.13704.

[18] T.W. Palmer, *Unbounded normal operators on Banach spaces*, Trans. Amer. Math. Soc., **133** (1968), 385-414, MR 37 #6768, Zbl 0169.16901.

[19] G.K. Pedersen, C^**-Algebras and their Automorphism Groups*, Academic Press, 1979, MR 81e:46037, Zbl 0416.46043.

[20] N.W. Pedersen, *The resolutions of the identity for sums and products of commuting spectral operators*, Math. Scand., **11** (1962), 123-130, MR 28 #479.

[21] W. Ricker, *Operator Algebras generated by Commuting Projections: A Vector Measure Approach*, Lecture Notes in Mathematics, **1711**, 1999, MR 2001b:47055, Zbl 0936.47020.

[22] P.G. Spain, *On scalar-type spectral operators*, Proc. Camb. Phil. Soc., **69** (1971), 409-410, MR 44 #7338, Zbl 0211.44702.

[23] ______, *On commutative* V^**-algebras* II, Glasgow Math. J., **13** (1972), 129-134, MR 47 #7461, Zbl 0245.46099.

[24] ______, *The* W^**-closure of a* V^**-algebra*, J. London Math. Soc. (2), **7** (1973), 385-386, MR 49 #9649, Zbl 0272.46047.

[25] ______, *A generalisation of a theorem of Grothendieck*, Quarterly J. Math., **27** (1976), 475-479, MR 56 #1007, Zbl 0341.46007.

Received October 1, 2001.

Mathematics Department
University of Glasgow
Glasgow G12 8QW
Scotland
E-mail address: hrd@maths.gla.ac.uk

Mathematics Department
Birjand University
Birjand
Iran
E-mail address: mohammadbg@yahoo.com

Mathematics Department
University of Glasgow
Glasgow G12 8QW
Scotland
E-mail address: pgs@maths.gla.ac.uk

PACIFIC JOURNAL OF MATHEMATICS
Vol. 209, No. 1, 2003

ON RINGS WHICH ARE SUMS OF TWO PI-SUBRINGS: A COMBINATORIAL APPROACH

B. Felzenszwalb, A. Giambruno, and G. Leal

We study the following open question: If a ring R is the sum of two subrings A and B both satisfying a polynomial identity, does R itself satisfy a polynomial identity? We give a positive answer to this question in case R satisfies a special "mixed" identity or $(AB)^k \subseteq A$ for some $k \geq 1$ or A or B is a Lie ideal. Our approach is based on a comparative analysis of the sequences of codimensions of the three rings and their asymptotics. As a reward we obtain a bound on the degree of a polynomial identity satisfied by R as a function of the degree of an identity satisfied by A and B.

1. Introduction.

Let R be a ring and suppose that A and B are two subrings of R such that $R = A + B$ is their sum. Here we consider the following open question:

If both A and B satisfy a polynomial identity (*i.e.,* PI-*rings*), *does R itself satisfy a polynomial identity*?

The answer to this question is known to be positive in several cases. The first result that can be read in this setting is due to Kegel [**K**]. He showed that if A and B are both nilpotent rings (so, they satisfy an identity of the type $x_1x_2 \dots x_n \equiv 0$) then the same conclusion holds in R. Bahturin and Giambruno in [**BG**] proved that if A and B are commutative rings then R satisfies the identity $[x_1, x_2][x_3, x_4] \equiv 0$, where $[x, y] = xy - yx$ is the Lie bracket. This result was later generalized by Beidar and Mikhalev in [**BM**]. They proved that if both, A and B satisfy an identity of the form $[x_1, x_2] \dots [x_{2n-1}, x_{2n}] \equiv 0$ for some $n \geq 2$, then R is a PI-ring. By extending Kegel's result, Kepczyk and Puczylowski in [**KP1**] showed that if A and B are nil of bounded exponent (so, they satisfy an identity of the form $x^n \equiv 0$) then so is R. This result was later pushed further in [**KP2**] by proving that if one of the two subrings is nil of bounded exponent and the other is PI, then R is PI.

To our knowledge these are the only results proved so far which hold without any further assumption on the structure of the ring R (or A or B).

On the other hand, in [**Ro1**] Rowen proved that if A and B are both right (left) ideals of R, then an identity on both A and B forces R to be a PI-ring.

This result was later extended in [**KP3**] where the authors proved that the same conclusion holds if one requires that only A or B is a one-sided ideal. In case the ring R is semiprime, the also showed that if A is nil PI and B is PI then R is PI.

We remark that, except for some results about semiprime rings, only in [**K**] and [**BG**] an explicit identity of R was exhibited. In all the other cases proved so far, the authors have shown the existence of an identity for R without providing any information on its explicit form or on its degree (as a function of the degrees of an identity of A and B). The reason for such failure is essentially the following: Most of these results use a reduction technique to the prime case where structure theory can be applied through the Martindale ring of quotients. Such reduction makes essentially use of the so called “Amitsur’s trick” (see [**Ro2**]) which allows to pass from the semiprime case to the general case but gives no information on the degree of the identities so far found.

In this paper we answer this question in some special cases. Let $R = A + B$, where A and B are subrings satisfying a polynomial identity.

We prove that if for some $k \geq 1$, either $(AB)^k \subseteq A$ or $(BA)^k \subseteq A$, then R is PI. As a corollary we obtain the case when A is a one-sided ideal or the case when $AB = BA$ and $B^k \subseteq A$. Moreover, R is still PI if either A or B is a Lie ideal of R. We shall remark that this last result can also be derived from a theorem on special Lie algebras.

One can also consider, in a natural fashion, “mixed” identities or semi-identities for R, i.e., polynomials in two distinct sets of variables

$$f(y_1, \dots, y_n, z_1, \dots, z_m)$$

that vanish when we evaluate the y_i’s into elements of A and the z_i’s into elements of B. We prove that R is a PI-ring provided R satisfies a k-special semi-identity i.e., an identity of the type $f(y_1, \dots, y_k, z_1, \dots, z_k)$, for some $k \geq 1$ where only one monomial of the type $y_{\sigma(1)}z_{\tau(1)} \cdots y_{\sigma(k)}z_{\tau(k)}$ appears with nonzero coefficient, for all $\sigma, \tau \in S_k$.

In all these results, we obtain an explicit function giving the degree of an identity for R in terms of the degree of an identity of A and B. Through this function, an explicit identity for R can be constructed as it has been shown by Regev in [**R2**]. More precisely, suppose that A and B satisfy an identity of degree d and one of the above hypotheses holds. Then we prove that R satisfies an identity of degree d' where d' is the least integer greater than a^a where a has the following value: $a = 8e(d-1)^4$, if A or B is a Lie ideal; $a = 8e(kd(d-1)-1)^2(d-1)^2$, if $(AB)^k \subseteq A$; $a = 8e(d-1)^4$, if A is a one-sided ideal; $a = 8ek(d-1)^2$, if R satisfies a k-special mixed identity of degree k (here e is the basis of the natural logarithms).

Our technique is based on a combinatorial approach to the problem using the sequence of codimensions of a ring. This sequence was introduced and exploited by Regev in [**R1**]. He proved, through this method, that the tensor product of two PI-rings is a PI-ring. In this paper we follow that approach and we attach to each of the rings R, A and B its codimension sequence and through the study of the relations among these sequences and their asymptotic behaviour we are able to prove our results. Unfortunately our approach does not solve the problem in its generality. We feel that one needs a better understanding through a deeper and throughout analysis of the sequences of codimensions and their relations.

2. Preliminaries.

Throughout we shall assume that all rings are algebras over a fixed field F. We make this assumption in order to simplify the notation. On the other hand it is easily verified that our results are still valid for general rings if one assumes that all polynomials have integer coefficients and that A and B satisfy an identity which is a monic polynomial. To this end we recall that by a theorem of Amitsur (see [**Ro2**]), if a ring satisfies an identity which is proper for all its homomorphic images, then it satisfies an identity of the type $(St_k)^l$, for some k, l, where St_k is the standard polynomial of degree k.

Let $X = \{x_1, x_2, \dots\}$ be a countable set and let $F\langle X\rangle$ be the free algebra on X over F. Recall that a polynomial $f(x_1, \dots, x_n)$ is an identity for the algebra R if $f(r_1, \dots, r_n) = 0$ for all $r_1, \dots, r_n \in R$ (in this case we write $f \equiv 0$ on R). In case R satisfies a nontrivial identity f, i.e., $f \neq 0$, we say that R is a PI-algebra.

In general one defines

$$\mathrm{Id}(R) = \{f \in F\langle X\rangle \mid f \equiv 0 \text{ on } R\},$$

the set of polynomial identities of R. $\mathrm{Id}(R)$ has a structure of T-ideal of $F\langle X\rangle$ i.e., an ideal invariant under all endomorphisms of $F\langle X\rangle$; it is obvious that R is a PI-algebra if and only if $\mathrm{Id}(R) \neq 0$.

Recall that a polynomial $f(x_1, \dots, x_n)$ is multilinear if each variable x_i, $i = 1, \dots, n$, appears in every monomial of f with degree one. Multilinear polynomials are important; in fact it is well-known that if R satisfies an identity of degree d then it satisfies a multilinear identity of degree $\leq d$.

For every $n \geq 1$ we define

$$V_n = \mathrm{Span}_F\{\, x_{\sigma(1)} \dots x_{\sigma(n)} \mid \sigma \in S_n\}$$

where S_n is the symmetric group of degree n. V_n is the space of multilinear polynomials in $x_1, \dots, x_n$. Since $\dim_F V_n = n!$, from the above observation we easily get the following:

Remark 2.1. The algebra R satisfies a polynomial identity if and only if there exists $n \geq 1$ such that

$$\dim_F \frac{V_n}{V_n \cap \mathrm{Id}(R)} < n!$$

The above remark, though trivial, will be essential not only in the proof of the existence of an identity for R but also for the explicit computation of the corresponding degree.

3. The basic reduction.

From now on we shall assume that R is an F-algebra such that $R = A + B$ for suitable subalgebras A and B. We shall also assume that A and B are PI-algebras. For the sake of simplicity let us denote by d the degree of an identity satisfied by A and B.

Our first aim is to relate the valuations of polynomials in A, B to those in R. To this end, we introduce two new countable sets $Y = \{y_1, y_2, \dots\}$ and $Z = \{z_1, z_2, \dots\}$. Then, we let $F\langle Y \cup Z\rangle$ be the free algebra on the set $Y \cup Z$ over F. We relate $F\langle Y \cup Z\rangle$ to $F\langle X\rangle$ by assuming that $x_i = y_i + z_i$, $i = 1, 2, \dots$.

We can now define the notion of s-identity (or semi-identity) of R. A polynomial $f(y_1, \dots, y_n, z_1, \dots, z_m) \in F\langle Y \cup Z\rangle$ is an s-identity of R if $f(a_1, \dots, a_n, b_1, \dots, b_m) = 0$ for all $a_1, \dots, a_n \in A$, $b_1, \dots, b_m \in B$. Accordingly one defines

$$\mathrm{Id}^s(R) = \{f \in F\langle Y \cup Z\rangle \mid f \text{ is an } s\text{-identity of } R\},$$

the ideal of s-identities of R. It is clear that $\mathrm{Id}^s(R)$ is an ideal invariant under all endomorphisms of $F\langle Y \cup Z\rangle$ that leave $F\langle Y\rangle$ and $F\langle Z\rangle$ invariant. Also $\mathrm{Id}(R)$, $\mathrm{Id}(A)$, $\mathrm{Id}(B) \subseteq \mathrm{Id}^s(R)$.

Now we need the notion of multilinear polynomial in $F\langle Y \cup Z\rangle$. To this end, we give the same degree one to the variables y_i and z_i for all $i = 1, 2, \dots$. Then

$$W_n = \mathrm{Span}_F\{\, w_{\sigma(1)} \dots w_{\sigma(n)} \mid \sigma \in S_n,\ w_i = y_i \text{ or } z_i, \text{ for all } i = 1, \dots, n\}$$

is the space of multilinear polynomials in $y_1, z_1, \dots, y_n, z_n$. It is clear that $\dim_F W_n = 2^n n!$ and

$$V_n \subseteq W_n.$$

Since

$$\frac{V_n}{V_n \cap \mathrm{Id}(R)} = \frac{V_n}{V_n \cap (W_n \cap \mathrm{Id}^s(R))} \cong \frac{V_n + (W_n \cap \mathrm{Id}^s(R))}{W_n \cap \mathrm{Id}^s(R)} \subseteq \frac{W_n}{W_n \cap \mathrm{Id}^s(R)},$$

we have the following:

Lemma 3.1. $\dim_F \dfrac{V_n}{V_n \cap \mathrm{Id}(R)} \le \dim_F \dfrac{W_n}{W_n \cap \mathrm{Id}^s(R)}$.

At the light of Remark 2.1, we can now make the following reduction. Recall that R satisfies an identity of degree n if and only if $\dim_F \frac{V_n}{V_n \cap \mathrm{Id}(R)} < n!$

Remark 3.2. If there exists $n \ge 1$ such that $\dim_F \dfrac{W_n}{W_n \cap \mathrm{Id}^s(R)} < n!$, then R is a PI-algebra and satisfies an identity of degree n.

The spaces W_n are still too large for our computations. Hence we next make one further reduction.

Let $t \geq 0$ and fix integers $1 \leq r_1 \leq \cdots \leq r_t \leq n$. Then define

$$V_{r_1,\ldots,r_t} = \mathrm{Span}_F\{w_{\sigma(1)} \ldots w_{\sigma(n)} \mid \sigma \in S_n,\ w_i = y_i,\ \text{for } i \in \{r_1, \ldots, r_t\},\ w_j = z_j,\ \text{for } j \notin \{r_1, \ldots, r_t\}\}.$$

Clearly

$$W_n = \bigoplus_{1 \leq r_1 \leq \cdots \leq r_t \leq n} V_{r_1,\ldots,r_t}.$$

Also, it is easy to see that $W_n \cap \mathrm{Id}^s(R) = \bigoplus_{1\leq r_1 \leq \cdots \leq r_t \leq n}(V_{r_1,\ldots,r_t} \cap \mathrm{Id}^s(R))$. It follows that

$$\dim_F \frac{W_n}{W_n \cap \mathrm{Id}^s(R)} = \sum_{1 \leq r_1 \leq \cdots \leq r_t \leq n} \dim_F \frac{V_{r_1,\ldots,r_t}}{V_{r_1,\ldots,r_t} \cap \mathrm{Id}^s(R)}.$$

Write for simplicity $V_{1,\ldots,t} = V_{t,n-t}$ and notice that for all $1 \leq r_1 \leq \cdots \leq r_t \leq n$, $V_{r_1,\ldots,r_t} \cong V_{t,n-t}$ and $V_{r_1,\ldots,r_t} \cap \mathrm{Id}^s(R) \cong V_{t,n-t} \cap \mathrm{Id}^s(R)$. Since for every $t = 0, \ldots, n$, there exist $\binom{n}{t}$ subspaces $V_{r_1,\ldots,r_t}$ isomorphic to $V_{t,n-t}$, we get:

Lemma 3.3. $\dim_F \dfrac{W_n}{W_n \cap \mathrm{Id}^s(R)} = \displaystyle\sum_{t=0}^{n} \binom{n}{t} \dim_F \frac{V_{t,n-t}}{V_{t,n-t} \cap \mathrm{Id}^s(R)}.$

We can now prove the final reduction.

Remark 3.4. In order to prove that R is a PI-algebra, it is enough to prove that there exists $n \geq 1$ such that for all $t = 0, 1, \ldots, n$,

$$\dim_F \frac{V_{t,n-t}}{V_{t,n-t} \cap \mathrm{Id}^s(R)} < \frac{n!}{2^n}.$$

In this case R satisfies an identity of degree n.

Proof. Suppose $\dim_F \frac{V_{t,n-t}}{V_{t,n-t} \cap \mathrm{Id}^s(R)} < \frac{n!}{2^n}$. Then, by the previous lemma,

$$\dim_F \frac{W_n}{W_n \cap \mathrm{Id}^s(R)} < \sum_{t=0}^{n} \binom{n}{t} \frac{n!}{2^n} = 2^n \frac{n!}{2^n} = n!$$

and we are done by Remark 3.2. □

4. Ordering monomials.

Let $1 \leq d \leq n$. Recall (see [**R2**]) that a permutation $\sigma \in S_n$ is called d-bad if there exist $1 \leq k_1 < \cdots < k_d \leq n$ such that $\sigma(k_1) > \cdots > \sigma(k_d)$. We say that σ is d-good if it is not d-bad. The d-good permutations are quite spare in S_n; in fact, as a consequence of a theorem of Dilworth one can prove the following:

Lemma 4.1 ([**R2**, Theorem 1.8])**.** *In S_n the number of d-good permutations is $\leq \frac{(d-1)^{2n}}{(d-1)!}$.*

The d-good permutations were used in PI-theory for finding a bound to $\dim_F \frac{V_n}{V_n \cap \mathrm{Id}(C)}$ for a PI-algebra C.

We say that a monomial $x_{\sigma(1)} \ldots x_{\sigma(n)}$ is d-good if the corresponding permutation σ is d-good. The result is the following:

Theorem 4.2 ([**R2**, Theorem 1.3])**.** *Let C be an algebra satisfying an identity of degree d. Then every monomial in V_n can be written* (mod. $\mathrm{Id}(C)$) *as a linear combination of d-good monomials. Hence* $\dim_F \frac{V_n}{V_n \cap \mathrm{Id}(C)} \leq \frac{(d-1)^{2n}}{(d-1)!}$.

Next step is to generalize the above theorem by adapting it to our situation $R = A + B$. Recall that we are assuming throughout that A and B satisfy an identity of degree d. In order to simplify the notation, we make the following:

Definition 4.3. Let $t \geq 0$ and write $w \in V_{t,n-t}$ in the following form:

$$w = w_1 y_{\sigma(1)} \cdots y_{\sigma(i_1)} w_2 y_{\sigma(i_1+1)} \cdots y_{\sigma(i_2)} w_3 \cdots w_r y_{\sigma(i_{r-1}+1)} \cdots y_{\sigma(i_r)} w_{r+1}$$

where $w_1, \ldots, w_{r+1}$ are (eventually trivial) monomials in the variables z_i. If the permutation σ is d-good (d-bad) we say that w is d-y-good (d-y-bad resp.).

Recall that an additive subgroup U of a ring R is a Lie ideal of R if for all $u \in U, r \in R$, we have that $[u, r] \in U$.

Lemma 4.4. *Let A be a Lie ideal of R. Then, for all $t = 0, 1, \ldots, n$, $V_{t,n-t}$ is spanned* (mod. $\mathrm{Id}^s(R)$) *by all d-y-good monomials.*

Proof. Suppose that the conclusion of the lemma is false. We first order the monomials of $V_{t,n-t}$ according to the left lexicographic order of the variables y_i. Then, among all monomials which do not satisfy the conclusion of the lemma, we pick a smallest one (in the given order). Let such monomial be

$$w = w_1 y_{\sigma(1)} \cdots y_{\sigma(i_1)} w_2 \cdots w_r y_{\sigma(i_{r-1}+1)} \cdots y_{\sigma(i_r)} w_{r+1}.$$

In the monomial w we first make the formal substitution

$$y_{\sigma(i_s)} w_{s+1} = \bar{y}_{\sigma(i_s)} + w_{s+1} y_{\sigma(i_s)}$$

for $s = 1, \ldots, r$, where $\bar{y}_{\sigma(i_s)} = [y_{\sigma(i_s)}, w_{s+1}]$. Since A ia a Lie ideal of R, the elements $\bar{y}_{\sigma(i_s)}$ evaluate to elements of A. It follows that we can write w as a linear combination of monomials in the variables $y_i, \bar{y}_i, z_i$ where either some monomial w_i has been absorbed in a $\bar{y}_i$ or it has been moved to the left past some y_i.

A repeated application of this process allows us to write w as a linear combination of monomials of the type

$$w' = w'_1 y'_{\sigma(1)} \cdots y'_{\sigma(i_1)} y'_{\sigma(i_1+1)} \cdots y'_{\sigma(i_2)} \cdots y'_{\sigma(i_r)}$$

where $y'_j = [y_j, w_{u_1}, \ldots, w_{u_s}]$ for some $u_1, \ldots, u_s$, $s \geq 0$ and w'_1 is a monomial in the variables z_i. Note that in our terminology the Lie commutators are left normed i.e., $[x_1, \ldots, x_n] = [[x_1, x_2], \ldots, x_n]$. Also, in each monomial w', the permutation of the indices of the variables y'_i is still σ.

Since the conclusion of the lemma does not hold for the monomial w, then in particular σ is a d-bad permutation. It follows that there exist $1 \leq j_1 < \cdots < j_d \leq t$ such that $\sigma(j_1) > \cdots > \sigma(j_d)$. Write

$$w' = ay''_1 y''_2 \ldots y''_d$$

where $a = w'_1 y'_{\sigma(1)} \cdots y'_{\sigma(j_1-1)}$ and

$$y''_1 = y'_{\sigma(j_1)} \cdots y'_{\sigma(j_2-1)}, \ \ldots, \ y''_d = y'_{\sigma(j_d)} \cdots y'_{\sigma(i_r)}.$$

Let $f(y_1, \ldots, y_d) = \sum_{\tau \in S_d} \alpha_\tau y_{\tau(1)} \cdots y_{\tau(d)}$ be a multilinear identity of degree d satisfied by A. We may clearly assume that $\alpha_1 = 1$. Since A is a subring and a Lie ideal of R, then the polynomials $y''_1, \ldots, y''_d$ evaluate to elements of A. It follows that $f(y''_1, \ldots, y''_d)$ is an s-identity of R. Hence $f(y''_1, \ldots, y''_d) \in \mathrm{Id}^s(R) \cap V_{t,n-t}$.

Write

$$w' = ay''_1 \ldots y''_d \equiv -\sum a\alpha_\tau y''_{\tau(1)} \cdots y''_{\tau(d)} \pmod{\mathrm{Id}^s(R)}. \tag{1}$$

By the definition of $y''_1, \ldots, y''_d$, since σ is d-bad, it follows that each monomial $ay''_{\tau(1)} \cdots y''_{\tau(d)}$, to the right-hand side of (1), is smaller that $ay''_1 y''_2 \ldots y''_d$ (in the left lexicographic order of the y''_i' s).

If we now recall the definition of the y_i's and we open up all the brackets, we obtain that w' and, so, the original monomial w, can be written (mod. $\mathrm{Id}^s(R)$) as a linear combination of monomials (in the variables y_i and z_i) which are smaller than w in the left lexicographic order of the y_i's. By the minimality of w, it follows that the lemma holds for such monomials. Hence each of them can be written as a linear combination (mod. $\mathrm{Id}^s(R)$) of monomials which are d-y-good. But then the same conclusion holds for w and this is a contradiction. □

Lemma 4.5. *Suppose that for some $k \geq 1$, $(AB)^k \subseteq A$. Then, for all $t = 0, 1, \ldots, n$, $V_{t,n-t}$ is spanned* (mod. $\mathrm{Id}^s(R)$) *by all $kd(d-1)$-y-good monomials.*

Proof. Suppose that the conclusion of the lemma is false and take, as before, a monomial w which does not satisfy the conclusion of the lemma and is smallest in the order of the y_i's. Let

$$w = w_1 y_{\sigma(1)} \cdots y_{\sigma(i_1)} w_2 \cdots w_r y_{\sigma(i_{r-1}+1)} \cdots y_{\sigma(i_r)} w_{r+1}.$$

By the choice of w, in particular, σ is $kd(d-1)$-bad and let $1 \leq j_1 < \cdots < j_{kd(d-1)} \leq t$ be such that $\sigma(j_1) > \cdots > \sigma(j_{kd(d-1)})$. Suppose first that we can write

$$w = uy_{\sigma(j_m)}a_1y_{\sigma(j_{m+1})}a_2\dots a_dy_{\sigma(j_{m+d-1})}v$$

for some $m \in \{1, 2, \dots, kd(d-1) - d + 1\}$, where $a_1, \dots, a_d$ are monomials in the only variables y_i and u and v are suitable monomials.

In this case set $y_{\sigma(j_m)}a_1 = y'_1$, $y_{\sigma(j_{m+1})}a_2 = y'_2, \dots, y_{\sigma(j_{m+d-1})}v = y'_d$. Then

$$w = uy'_1y'_2\dots y'_d.$$

If $f(y_1, \dots, y_d)$ is the multilinear identity of degree d satisfied by A, then $f(y'_1, \dots, y'_d)$ is still an identity of A and, by applying it as in the previous lemma, we can write

$$w \equiv -\sum_{\tau\in S_d} \alpha_\tau uy'_{\tau(1)}\dots y'_{\tau(d)} \pmod{\mathrm{Id}^s(R)} \tag{2}$$

for some $\alpha_\tau \in F$. But each monomial on the right-hand side of (2) is smaller than w in the order of the variables y_i; hence we get, by the minimality of w, that each of them can be written (mod. $\mathrm{Id}^s(R)$) as a linear combination of $kd(d-1)$-y-good monomials. But then the same conclusion holds for w, a contradiction.

Hence we may assume that for any d indices $j_m < j_{m+1} < \dots < j_{m+d-1}$ (of the sequence giving the $kd(d-1)$-y-badness of w), in the monomial w, at least one variable z_i appears between the variables $y_{\sigma(j_m)}$ and $y_{\sigma(j_{m+d-1})}$.

Let us write w in the form

$$w = a_0y_{\sigma(j_{p_1})}a_1y_{\sigma(j_{p_2})}a_2\dots y_{\sigma(j_{p_d})}a_d$$

where $p_1 = 1$, $a_0, a_1, \dots, a_d$ are monomials in the variables y_i and z_i and $y_{\sigma(j_{p_m})}a_m$ evaluates to either $(AB)^kA$ or $(AB)^k$, for all $m = 1, \dots, d$. In order for this decomposition to hold, we have to show that $p_d \le kd(d-1)$. In fact, by the assumption made above, in w, at least one variable z_i appears between two variables $y_{\sigma(j_m)}$ and $y_{\sigma(j_{m+d-1})}$. Hence, for $i = 1, \dots, d-1$,

$$p_{i+1} - p_i \le k(d-1),$$

but then, $p_d = 1 + (p_2 - p_1) + \dots + (p_d - p_{d-1}) \le kd(d-1)$ as claimed. Write now

$$w = a_0y'_1y'_2\dots y'_d$$

where $y'_1 = y_{\sigma(j_{p_1})}a_1, \dots, y'_d = y_{\sigma(j_{p_d})}a_d$ and let $f(y_1, \dots, y_d)$ be the identity of A. Recall that $y'_i = y_{\sigma(j_{p_i})}a_i$ evaluates to $(AB)^kA$ or $(AB)^k$ and both these sets lie in A by hypothesis. Hence $f(y'_1, \dots, y'_d)$ is an s-identity of R and $f(y'_1, \dots, y'_d) \in \mathrm{Id}^s(R) \cap V_{t,n-t}$.

Through the identity $f(y'_1, \dots, y'_d)$ we can now rearrange (mod. $\mathrm{Id}^s(R)$) the variables y'_i in w and, as above, this leads to a contradiction due to the minimality of w. □

Corollary 4.6. *If A is a one-sided ideal of R then, for all $t = 0, 1, \dots, n$, $V_{t,n-t}$ is spanned* (mod. $\mathrm{Id}^s(R)$) *by all d-y-good monomials.*

Proof. From the previous lemma we get that $V_{t,n-t}$ is spanned by the $d(d-1)$-y-good monomials. We next show that this result can be improved as claimed in the conclusion of the corollary. Suppose A is a one-sided ideal of R.

As in the previous lemma we take w smallest in the order of the y_i's for which the conclusion does not hold. We take $1 \le j_1 < \cdots < j_d \le t$ such that $\sigma(j_1) > \cdots > \sigma(j_d)$.

Then we write
$$w = ay_1'y_2' \dots y_d'$$
where $a = w_1 y_{\sigma(1)} \dots y_{\sigma(j_1-1)} w_1'$ and
$$y_1' = y_{\sigma(j_1)} \dots y_{\sigma(j_2-1)} w_2', \dots, y_d' = w_1 y_{\sigma(j_d)} \dots w_{r+1}$$
with $w_1', w_2', \dots$ eventually trivial monomials in the variables z_i. Since A is a right ideal of R, the monomials $y_1', \dots, y_d'$ evaluate to elements of A. Hence they can be rearranged (mod. $\mathrm{Id}^s(R)$) using the identity of A. This completes the proof as in the previous lemmas. □

5. Mixed identities.

In this section we examine the case when R further satisfies a semi-identity $f \in \mathrm{Id}^s(R)$ of a special type. It is clear, by the standard multilinearization process, that if R satisfies a nontrivial semi-identity of degree m, then it also satisfies a multilinear one of degree $\le m$. If $\{i_1, \dots, i_k\}$ is a subset of $\{1, \dots, n\}$ we denote by $S_k(i_1, \dots, i_k)$ the subgroup of S_n of all permutations fixing $\{1, \dots, n\} \backslash \{i_1, \dots, i_k\}$. Also, in order to simplify the notation, we write the variables $z_{t+1}, \dots, z_n$ of $V_{t,n-t}$ as $z_1, \dots, z_{n-t}$, respectively. We now make the formal definition.

Definition 5.1. Let $f(y_1, \dots, y_k, z_1, \dots, z_k) \in F\langle Y \cup Z\rangle$ be a multilinear polynomial. We say that f is k-special if
$$f(y_1, \dots, y_k, z_1, \dots, z_k) = y_1 z_1 \dots y_k z_k + \sum_{\substack{\sigma \in S_{2k} \\ \sigma \notin T}} \alpha_\sigma w_{\sigma(1)} \dots w_{\sigma(2k)}$$
for some $\alpha_\sigma \in F$, where, for $i = 1, \dots k$, $w_{2i} = z_i$, $w_{2i-1} = y_i$ and $T = S_k(1, 3, \dots, 2k-1) \times S_k(2, 4, \dots, 2k) \subseteq S_{2k}$.

In few words, f is k-special if the only monomial of the type
$$y_{\sigma(1)} z_{\tau(1)} \cdots y_{\sigma(k)} z_{\tau(k)}$$
appearing in f with nonzero coefficient is $y_1 z_1 \dots y_k z_k$.

We next prove that in the presence of a k-special semi-identity for R we can bound exponentially the dimension of $V_{t,n-t}$ (mod. $\mathrm{Id}^s(R)$). Recall that if $n \geq 1$ is an integer such that $n = r_1 + \cdots + r_p$, $r_i > 0$, then the multinomial coefficient is defined as $\binom{n}{r_1, \ldots, r_p} = \frac{n!}{r_1! \ldots r_p!}$.

Lemma 5.2. *Suppose that R satisfies a k-special semi-identity. Then for all $t = 0, 1, \ldots, n$,*

$$\dim_F \frac{V_{t,n-t}}{V_{t,n-t} \cap \mathrm{Id}^s(R)} \leq \frac{(4k(d-1)^2)^n}{(d-1)!}.$$

Proof. We decompose the space $V_{t,n-t}$ as follows: Write $t = q_1 + \cdots + q_p$, $n - t = r_1 + \cdots + r_p$ where $q_1 \geq 0, r_p \geq 0$, for $p > 1$ we have that $q_2, \ldots, q_p, r_1, \ldots, r_{p-1}$ are positive integers and at most one between q_1 and r_p can eventually be zero. Then define $(\mathbf{q}, \mathbf{r}) = (q_1, \ldots, q_p,\ r_1, \ldots, r_p)$ and

$$U^{(p)}_{(\mathbf{q},\mathbf{r})} = \mathrm{Span}_{\mathrm{F}}\{y_{\sigma(1)} \cdots y_{\sigma(q_1)} z_{\tau(1)} \cdots z_{\tau(r_1)} \cdots y_{\sigma(q_1+\cdots+q_{p-1}+1)} \cdots y_{\sigma(t)} z_{\tau(r_1+\cdots+r_{p-1}+1)} \cdots z_{\tau(n-t)} \mid \sigma \in S_t,\ \tau \in S_{n-t}\}.$$

We write

$$\bigoplus_{(\mathbf{q},\mathbf{r})} U^{(p)}_{(\mathbf{q},\mathbf{r})} = U^{(p)}.$$

Clearly $V_{t,n-t} = \oplus_{p \geq 1} U^{(p)}$.

We claim that for every $(\mathbf{q}, \mathbf{r})$ and for every p,

$$U^{(p)}_{(\mathbf{q},\mathbf{r})} \subseteq \bigoplus_{s \leq 2k} U^{(s)}. \tag{3}$$

In fact, suppose the above inclusion is false and pick a subspace $U^{(p)}_{(\mathbf{q},\mathbf{r})}$ with p minimal such that $U^{(p)}_{(\mathbf{q},\mathbf{r})}$ is not contained in the right-hand side of (3). Since $p > 2k$, every monomial in $U^{(p)}_{(\mathbf{q},\mathbf{r})}$ is of the form

$$v = v_1 u_1 \ldots v_p u_p$$

where $v_1, \ldots, v_p$ are monomials in the variables y_i, $u_1, \ldots, u_p$ are monomials in the variables z_i and either v_1 or u_p is eventually trivial. Suppose for short that $v_1 \neq 1$. Since, by hypothesis, R satisfies a k-special semi-identity and A and B are subrings, we get that

$$v_1 u_1 \ldots v_k u_k \equiv \sum_{\substack{\sigma \in S_{2k} \\ \sigma \notin T}} \alpha_\sigma w_{\sigma(1)} \ldots w_{\sigma(2k)} \pmod{\mathrm{Id}^s(R)},$$

where $w_{2i} = v_i$ and $w_{2i-1} = u_i$, for $i = 1, \ldots, k$. But then, this says that (mod. $\mathrm{Id}^s(R)$) v is a linear combination of monomials belonging to some $U^{(p_1)}_{(\mathbf{s_1},\mathbf{r_1})}$ with $p_1 < p$. By the minimality of p, $U^{(p_1)}_{(\mathbf{s_1},\mathbf{r_1})} \subseteq \bigoplus_{q \leq 2k} U^{(p)}$ and, since $U^{(p)}_{(\mathbf{s},\mathbf{r})} \subseteq U^{(p_1)}_{(\mathbf{s_1},\mathbf{r_1})}$, we get a contradiction.

We now compute $\dim_F U^{(p)}_{(\mathbf{q},\mathbf{r})}$ for every $p \le 2k$. Since both A and B satisfy an identity of degree d, by Theorem 4.2, any monomial in the only variables y_i (in the only variables z_i) can be written (mod. $\mathrm{Id}^s(R)$) as a linear combination of d-good monomials. Moreover, by Lemma 4.1, the number of d-good monomials in m variables is bounded by $\frac{(d-1)^{2m}}{(d-1)!}$. Since in every monomial of $U^{(p)}_{(\mathbf{q},\mathbf{r})}$ occur q_1 consecutive variables y_i, …, q_p consecutive variables y_i (r_1 consecutive variables z_i, …, r_p consecutive variables z_i), we get that $U^{(p)}_{(\mathbf{q},\mathbf{r})}$ is spanned (mod. $\mathrm{Id}^s(R)$) by at most

$$\begin{aligned}&\frac{(d-1)^{2q_1}}{(d-1)!}\cdots\frac{(d-1)^{2q_p}}{(d-1)!}\frac{t!}{q_1!\dots q_p!}\frac{(d-1)^{2r_1}}{(d-1)!}\cdots\frac{(d-1)^{2r_p}}{(d-1)!}\frac{(n-t)!}{r_1!\dots r_p!}\\ &\le \frac{(d-1)^{2n}}{((d-1)!)^{2p-1}}\binom{t}{q_1,\dots,q_p}\binom{n-t}{r_1,\dots,r_p}\end{aligned}$$

monomials.

Recall that if $m, m_1, \dots, m_p$ are positive integers such that $m = m_1+\cdots+m_p$, by the multinomial theorem (see [**Bi**]), we have that $\binom{m}{m_1,\dots,m_p} \le p^m$. Hence, since $p \le 2k$, we get that $U^{(p)}_{(\mathbf{q},\mathbf{r})}$ is spanned (mod. $\mathrm{Id}^s(R)$) by at most

$$\frac{(d-1)^{2n}}{(d-1)!}(2k)^t(2k)^{n-t} = \frac{(d-1)^{2n}(2k)^n}{(d-1)!}$$

monomials. Since there are at most 2^n spaces $U^{(p)}_{(\mathbf{q},\mathbf{r})}$ inside $V_{t,n-t}$, the conclusion follows. ☐

6. The main results.

In the next lemma we prove that the results obtained in Lemma 4.4, Lemma 4.5 and Corollary 4.6 allow us to get a suitable upper bound to the dimension of $V_{t,n-t}$ (modulo $V_{t,n-t} \cap \mathrm{Id}^s(R)$).

Lemma 6.1. *Let* $0 \le t \le n$ *and suppose that* $V_{t,n-t}$ *is spanned* (mod. $\mathrm{Id}^s(R)$) *by the* c-y-*good monomials, for some* $c \ge 1$. *Then,*

$$\dim_F \frac{V_{t,n-t}}{V_{t,n-t}\cap \mathrm{Id}^s(R)} \le 2^n(c-1)^{2t}(d-1)^{2(n-t)}(t+1)^{n-t}.$$

Proof. For $t = q_1+\cdots+q_p$ and $n-t = r_1+\cdots+r_p$, let $U^{(p)}_{(\mathbf{s},\mathbf{r})}$ be the subspace of $V_{t,n-t}$ defined in the proof of the Lemma 5.2.

We shall compute the dimension of the spaces $U^{(p)}_{(\mathbf{q},\mathbf{r})}$ and, so, the dimension of $V_{t,n-t}$ (mod. $\mathrm{Id}^s(R)$), for $t = 0, 1, \dots$. By hypothesis $U^{(p)}_{(\mathbf{q},\mathbf{r})}$ is generated (mod. $\mathrm{Id}^s(R)$) by all c-y-good monomials. By Lemma 4.1 the number of such monomials is bounded by $\frac{(c-1)^{2t}}{(c-1)!}$. On the other hand, by recalling that B satisfies an identity of degree d, by Theorem 4.2, every monomial in the only variables z_i can be written (mod. $\mathrm{Id}^s(R)$) as a linear combination of d-good monomials. Since in every monomial of $U^{(p)}_{(\mathbf{q},\mathbf{r})}$ occur r_1 consecutive z_i, …, r_p consecutive z_i, it follows that $U^{(p)}_{(\mathbf{s},\mathbf{r})}$ is spanned (mod. $\mathrm{Id}^s(R)$) by at most

$$\begin{aligned}&\frac{(c-1)^{2t}}{(c-1)!}\frac{(n-t)!}{r_1!\dots r_p!}\frac{(d-1)^{2r_1}}{(d-1)!}\cdots\frac{(d-1)^{2r_p}}{(d-1)!}\\ &\le \frac{(c-1)^{2t}(d-1)^{2(n-t)}}{((c-1)!(d-1)!)^{p-1}}\binom{n-t}{r_1,\dots,r_p}\\ &\le (c-1)^{2t}(d-1)^{2(n-t)}p^{n-t}\end{aligned}$$

monomials (note that the last inequality holds since by the multinomial theorem $\binom{n-t}{r_1,\dots,r_p}\le p^{n-t}$).

Recalling that $p\le t+1$ and that there exist at most 2^n subspaces $U^{(p)}_{(\mathbf{q},\mathbf{r})}$, we get the desired conclusion. □

Theorem 6.2. *Let $R = A + B$ be a ring which is the sum of two subrings A and B and suppose that A and B both satisfy an identity of degree d. We have:*

1) *If $(AB)^k \subseteq A$, for some $k \ge 1$, then R is PI and satisfies an identity whose degree is the least integer greater than a^a where $a = 8e(kd(d-1)-1)^2(d-1)^2$; in case A is a one-sided ideal then we can take $a = 8e(d-1)^4$;*
2) *if A is a Lie ideal, then R is PI and satisfies an identity whose degree is the least integer greater than a^a where $a = 8e(d-1)^4$;*
3) *if R satisfies a k-special semi-identity, for some $k \ge 1$, then R is PI and satisfies an identity whose degree is the least integer greater than $8ek(d-1)^2$.*

Proof. By Remark 3.4 it is enough to prove that there exists n such that

$$\dim_F \frac{V_{t,n-t}}{V_{t,n-t}\cap \mathrm{Id}^s(R)} < \frac{n!}{2^n},$$

for all $t = 0,\dots,n$. The smallest n for which this inequality holds will also give us the desired degree of an identity for R.

In order to get a bound for this smallest n, we are going to use the well-known inequality (see, for instance, [**FR**, p. 105]) that holds for any $x \ge 1$:

$$\left(\frac{x}{e}\right)^x < \frac{\Gamma(x+1)}{\sqrt{2\pi x}} < \Gamma(x+1)$$

where $\Gamma(x+1)$ is the gamma function (recall that $\Gamma(n+1) = n!$ for every natural number n) and e is the basis of the natural logarithms.

Suppose that $V_{t,n-t}$ is spanned (mod. $\mathrm{Id}^s(R)$) by the c-y-good monomials, for some $c \geq 1$. Then, according to Lemma 6.1, it is enough to find a natural number n such that

$$2^n(c-1)^{2t}(d-1)^{2(n-t)}(t+1)^{n-t} < \frac{n!}{2^n}$$

or, in view of the above remark,

$$(4e)^n(c-1)^{2t}(d-1)^{2(n-t)}(t+1)^{n-t} \leq n^n.$$

If we define $a = 8e(c-1)^2(d-1)^2$ then, since $t+1 \leq 2t$ and $t \leq n$, all we need is n such that

$$a^n t^{n-t} \leq n^n.$$

Here we have two possibilities: If $t \leq n/a$ clearly the inequality holds for every natural number $n \geq a$. If $n \geq t > n/a$ it is easy to check that the above inequality still holds for every natural number $n \geq a^a$.

By Lemma 4.4, in case A is a Lie ideal of R, $V_{t,n-t}$ is spanned (mod. $\mathrm{Id}^s(R)$) by the d-y-good monomials. Hence by what we have just proved, by taking $c = d$, it follows that in this case R is PI and it satisfies an identity of degree the least integer greater than a^a where $a = 8e(d-1)^4$. Similarly, in case $(AB)^k \subseteq A$, by invoking Lemma 4.5 and $c = kd(d-1)$ above, we get an identity for R of degree the least integer greater than a^a where $a = 8e(kd(d-1)-1)^2(d-1)^2$.

Suppose now that R satisfies a k-special semi-identity. Then, by Lemma 5.2, we know that

$$\dim_F \frac{V_{t,n-t}}{V_{t,n-t} \cap \mathrm{Id}^s(R)} \leq \frac{(4k(d-1)^2)^n}{(d-1)!}.$$

Therefore it is enough to take $n \geq 8ek(d-1)^2 > \frac{8ek(d-1)^2}{\sqrt[n]{(d-1)!}}$. □

One last remark is in order. As the referee has pointed out, the result: If A and B are PI and one of them is a Lie ideal then R is PI, can also be proved directly by an application of the theory of special Lie algebras (see [**B**, Section 6.3]). In order to see this, regard R as a Lie algebra under the bracket operation $[\ ,\ ]$ and A and B as Lie subalgebras. Since a Lie subalgebra of an associative PI-algebra is Lie PI ([**B**, Section 6.3]), both A and B satisfy a nontrivial identity as Lie algebras. If, say, A is a Lie ideal, then R/A is isomorphic to B which is Lie PI. Then the Lie algebra R, being an extension of a PI-ideal A by a PI-quotient algebra B is PI. This clearly implies that R is PI as an associative algebra.

References

[B] Y. Bahturin, *Identical Relations in Lie Algebras*, VNU Science Press, Utrecht, 1987, MR 88f:17032, Zbl 0691.17001.

[BG] Y. Bahturin and A. Giambruno, *Identities of sums of commutative subalgebras,* Rend. Cir. Mat. Palermo (2), **43** (1994), 250-258, MR 96g:16032, Zbl 0829.16014.

[BM] K.I. Beidar and A.V. Mikhalev, *Generalised polynomial identities and rings which are sums of two subrings,* Algebra i Logica, **34** (1995), 3-11 (Russian); English translation: Algebra and Logica, **34** (1995), 1-5, MR 96i:16038, Zbl 0846.16016.

[Bi] N.L. Biggs, *Discrete Mathematics*, Clarendon Press, Oxford, 1989, MR 91h:00002, Zbl 0682.05001.

[FR] O.J. Farrell and B. Ross, *Solved Problems: Gamma and Beta Functions, Legendre Polinomials, Bessel Functions*, Dover Publications, New York, 1971, Zbl 0285.33003.

[K] O.H. Kegel, *Zur nilpotenz gewisser assoziativer ringe,* Math. Ann., **149** (1962/63), 258-260, MR 28 #3049, Zbl 0106.25402.

[KP1] M. Kepczyk and E.R. Puczylowski, *On radicals of rings which are sums of two subrings,* Arch. Math., **66** (1996), 8-12, MR 96k:16033, Zbl 0860.16016.

[KP2] ______, *Rings which are sums of two subrings,* J. Pure Applied Algebra, **133** (1998), 151-162, MR 2000a:16050, Zbl 0929.16017.

[KP3] ______, *Rings which are sums of two subrings satisfying a polynomial identity,* Comm. Algebra, **29** (2001), 2059-2065, MR 2002g:16038, Zbl 0993.16017.

[R1] A. Regev, *Existence of identities in $A \otimes B$,* Israel J. Math., **11** (1972), 131-152, MR 47 #3442, Zbl 0249.16007.

[R2] ______, *The representations of S_n and explicit identities of* PI-*algebras,* J. Algebra, **51** (1978), 25-40, MR 57 #9745, Zbl 0374.16009.

[Ro1] L.H. Rowen, *Generalized polynomial identities* II, J. Algebra, **38** (1976), 380-392, MR 57 #3189a, Zbl 0324.16016.

[Ro2] ______, *Polynomial Identities in Ring Theory*, Academic Press, New York, 1980, MR 82a:16021, Zbl 0461.16001.

Received October 15, 2001 and revised April 16, 2002. The second author was partially supported by Murst of Italy and CNPq of Brazil; the third author was partially supported by CNPq of Brazil.

Instituto de Matemática
Universidade Federal do Rio de Janeiro
Rio de Janeiro
Brazil
E-mail address: bfel@terra.com.br

Dipartimento di Matematica ed Applicazioni
Università di Palermo
Via Archirafi 34
90123 Palermo
Italy

E-mail address: a.giambruno@unipa.it

Instituto de Matemática
Universidade Federal do Rio de Janeiro
Rio de Janeiro
Brazil
E-mail address: gleal@acd.ufrj.br

PACIFIC JOURNAL OF MATHEMATICS
Vol. 209, No. 1, 2003

A METHOD OF WASHINGTON APPLIED TO THE DERIVATION OF A TWO-VARIABLE p-ADIC L-FUNCTION

Glenn J. Fox

We derive the existence of a specific two-variable p-adic L-function by means of a method provided by Washington. This two-variable function is a generalization of the one-variable p-adic L-function of Kubota and Leopoldt, yielding the one-variable function when the second variable vanishes.

1. Introduction.

In [**5**] Kubota and Leopoldt prove the existence of meromorphic functions, $L_p(s;\chi)$, defined over the p-adic number field, that serve as p-adic equivalents of the Dirichlet L-series. These p-adic L-functions interpolate the values

$$L_p(1-n;\chi) = -\frac{1}{n}\left(1-\chi_n(p)p^{n-1}\right)B_{n,\chi_n},$$

whenever n is a positive integer. Here, $B_{n,\chi}$ denotes the n^{th} generalized Bernoulli number associated with the primitive Dirichlet character χ, and $\chi_n = \chi\omega^{-n}$, with ω the Teichmüller character. Since the time of that publication, a number of individuals have derived the existence of these functions by various means. In particular, Washington [**8**] derives the functions by elementary means and expresses them in an explicit form:

Theorem 1. *Let F be a positive integral multiple of q and f_χ, and let*

$$L_p(s;\chi) = \frac{1}{s-1}\frac{1}{F}\sum_{\substack{a=1\\(a,p)=1}}^{F}\chi(a)\langle a\rangle^{1-s}\sum_{m=0}^{\infty}\binom{1-s}{m}\left(\frac{F}{a}\right)^m B_m.$$

Then $L_p(s;\chi)$ is analytic for $s\in\mathfrak{D}$ when $\chi\neq 1$, and meromorphic for $s\in\mathfrak{D}$, with a simple pole at $s=1$ having residue $1-1/p$, when $\chi=1$. Furthermore, for each $n\in\mathbf{Z}$, $n\geq 1$,

$$L_p(1-n;\chi) = -\frac{1}{n}\left(1-\chi_n(p)p^{n-1}\right)B_{n,\chi_n}.$$

Thus, $L_p(s;\chi)$ vanishes identically if $\chi(-1)=-1$.

Recently, a particular two-variable extension, $L_p(s,t;\chi)$, of the p-adic L-functions was produced—one in which interpolating values of the two-variable functions yield expressions in terms of the generalized Bernoulli polynomials [**3**]. For positive integers n, these functions satisfy

$$L_p(1-n,t;\chi) = -\frac{1}{n}\left(B_{n,\chi_n}(qt) - \chi_n(p)p^{n-1}B_{n,\chi_n}\left(p^{-1}qt\right)\right),$$

with the restriction that $t \in \mathbf{C}_p$, $|t|_p \leq 1$. It has been shown that these interpolating values share certain congruence properties with the corresponding interpolating values of the one-variable functions [**2**]. By applying the method that Washington used to derive Theorem 1, we obtain $L_p(s,t;\chi)$ by elementary means and express the functions in an explicit form.

Theorem 2. *Let F be a positive integral multiple of q and f_χ, and let*

$$L_p(s,t;\chi) = \frac{1}{s-1}\frac{\chi(-1)}{F}\sum_{\substack{a=1\\(a,p)=1}}^{F}\chi(a)\langle a-qt\rangle^{1-s}\sum_{m=0}^{\infty}\binom{1-s}{m}\left(\frac{F}{a-qt}\right)^m B_m.$$

Then $L_p(s,t;\chi)$ is analytic for $t \in \mathbf{C}_p$, $|t|_p \leq 1$, provided $s \in \mathfrak{D}$, except $s \neq 1$ when $\chi = 1$. Also, if $t \in \mathbf{C}_p$, $|t|_p \leq 1$, this function is analytic for $s \in \mathfrak{D}$ when $\chi \neq 1$, and meromorphic for $s \in \mathfrak{D}$, with a simple pole at $s = 1$ having residue $1 - 1/p$, when $\chi = 1$. Furthermore, for each $n \in \mathbf{Z}$, $n \geq 1$,

$$L_p(1-n,t;\chi) = -\frac{1}{n}\left(B_{n,\chi_n}(qt) - \chi_n(p)p^{n-1}B_{n,\chi_n}\left(p^{-1}qt\right)\right).$$

Thus, $L_p(s,0;\chi) = L_p(s;\chi)$ for each $s \in \mathfrak{D}$, with $s \neq 1$ if $\chi = 1$.

By an analysis of the formula for $L_p(s;\chi)$ given in Theorem 1, one can obtain Diamond's formula for the value of $L_p'(0;\chi)$ (see [**6**, p. 393]):

Theorem 3. *Let χ be a primitive Dirichlet character, and let F be a positive integral multiple of q and f_χ. Then*

$$L_p'(0;\chi) = \sum_{\substack{a=1\\(a,p)=1}}^{F}\chi_1(a)G_p\left(\frac{a}{F}\right) - L_p(0;\chi)\log_p(F).$$

Here, the function G_p is the Diamond function, and $\log_p$ is the p-adic logarithm function of Iwasawa.

Young [**9**] derives a similar formula for $(\partial/\partial s)L_p(0,t;\chi)$ by means of a p-adic integral representation of $L_p(s,t;\chi)$. However, his work is restricted to those characters χ such that the conductor of χ_1 is not a power of p. The explicit formula given in Theorem 2 enables one to derive a formula for $(\partial/\partial s)L_p(0,t;\chi)$, similar to that obtained by Young, but for all primitive Dirichlet characters χ.

Theorem 4. *Let χ be a primitive Dirichlet character, and let F be a positive integral multiple of q and f_χ. Then for any $t \in \mathbf{C}_p$, $|t|_p \leq 1$,*

$$\frac{\partial}{\partial s} L_p(0, t; \chi) = \chi(-1) \sum_{\substack{a=1 \\ (a,p)=1}}^{F} \chi_1(a) G_p\left(\frac{a - qt}{F}\right) - L_p(0; \chi) \log_p(F).$$

Note that, if $\chi(-1) = -1$, then the function $L_p(s; \chi)$ vanishes identically. However, $L_p(s, t; \chi)$ is not identically 0 for any character χ. Thus, $L_p(s, t; \chi)$ provides us with a p-adic L-function that does not vanish identically for those χ such that $\chi(-1) = -1$. This may prove to be of use in the study of structures associated with such characters.

2. Preliminaries.

Let χ be a Dirichlet character, defined modulo its conductor f_χ. Then $\chi(a)^{\phi(f_\chi)} = 1$ for any $a \in \mathbf{Z}$ with $(a, f_\chi) = 1$, and $\chi(a) = 0$ otherwise. For two such characters χ and ψ, having conductors f_χ and f_ψ, respectively, let $\chi\psi$ denote the primitive character associated to the product of the characters. The conductor $f_{\chi\psi}$ then divides $\mathrm{lcm}(f_\chi, f_\psi)$.

The generalized Bernoulli polynomials associated with χ, $B_{n,\chi}(t)$, are defined by the generating function

$$\sum_{a=1}^{f_\chi} \frac{\chi(a) x e^{(a+t)x}}{e^{f_\chi x} - 1} = \sum_{n=0}^{\infty} B_{n,\chi}(t) \frac{x^n}{n!}. \tag{1}$$

The corresponding generalized Bernoulli numbers can then be defined by $B_{n,\chi} = B_{n,\chi}(0)$. With this definition, the generalized Bernoulli polynomials are expressed more precisely in terms of the expansion

$$B_{n,\chi}(t) = \sum_{m=0}^{n} \binom{n}{m} B_{n-m,\chi} t^m,$$

which is derived from (1).

The classical Bernoulli polynomials, $B_n(t)$, are defined by

$$\frac{x e^{tx}}{e^x - 1} = \sum_{n=0}^{\infty} B_n(t) \frac{x^n}{n!},$$

and the classical Bernoulli numbers by $B_n = B_n(0)$. This yields the values $B_0 = 1$, $B_1 = -1/2$, $B_2 = 1/6$, $B_4 = -1/30, \ldots$, with $B_n = 0$ for odd $n \geq 3$. The Bernoulli numbers are rational numbers, and the von Staudt-Clausen theorem states that for even $n \geq 2$,

$$B_n + \sum_{\substack{p \text{ prime} \\ (p-1)|n}} \frac{1}{p} \in \mathbf{Z}.$$

Thus, the denominator of each B_n must be square-free. We also have the relation

$$B_n(t) = \sum_{m=0}^{n} \binom{n}{m} B_{n-m} t^m. \tag{2}$$

The classical Bernoulli polynomials are related to the generalized Bernoulli polynomials in that $B_{n,1}(t) = (-1)^n B_n(-t)$, where $\chi = 1$ is the unique character having conductor 1 and satisfying $\chi(a) = 1$ for each $a \in \mathbf{Z}$.

Let p be a fixed prime. We will use $\mathbf{Z}_p$ to represent the p-adic integers, and $\mathbf{Q}_p$ the p-adic rationals. Let $\mathbf{C}_p$ denote the completion of the algebraic closure of $\mathbf{Q}_p$ under the p-adic absolute value $|\cdot|_p$, normalized so that $|p|_p = p^{-1}$. Fix an embedding of the algebraic closure of $\mathbf{Q}$ into $\mathbf{C}_p$. Since each value of a Dirichlet character χ is either 0 or a root of unity, we may consider the values of χ as lying in $\mathbf{C}_p$.

Denote $q = 4$ if $p = 2$ and $q = p$ otherwise. Let ω denote the Teichmüller character, having conductor $f_\omega = q$. For an arbitrary character χ we then define the character $\chi_n = \chi\omega^{-n}$, where $n \in \mathbf{Z}$, in the sense of the product of characters.

Let $\langle a \rangle = \omega^{-1}(a)a$ whenever $(a, p) = 1$. Then $\langle a \rangle \equiv 1 \pmod{q\mathbf{Z}_p}$ for these values of a. For our purposes, we extend this by defining $\langle a + qt \rangle = \omega^{-1}(a)(a + qt)$ for all $a \in \mathbf{Z}$, with $(a, p) = 1$, and $t \in \mathbf{C}_p$ such that $|t|_p \leq 1$. Then $\langle a + qt \rangle = \langle a \rangle + q\omega^{-1}(a)t$, so that $\langle a + qt \rangle \equiv 1 \pmod{q\mathbf{Z}_p[t]}$.

The p-adic logarithm function [**4**], $\log_p$, is the unique function mapping $\mathbf{C}_p^\times \to \mathbf{C}_p$ that satisfies $\log_p(1 + x) = \sum_{n=1}^{\infty} (-1)^{n-1} x^n / n$ for $|x|_p < 1$, $\log_p(xy) = \log_p(x) + \log_p(y)$ for all $x, y \in \mathbf{C}_p^\times$, and $\log_p(p) = 0$. Note that these conditions imply that this function vanishes at any rational power of p. The Diamond function [**1**] is defined by

$$G_p(x) = \left(x - \frac{1}{2}\right) \log_p(x) - x + \sum_{j=2}^{\infty} \frac{B_j}{j(j-1)} x^{1-j}.$$

The domain of this function is $|x|_p > 1$, with the p-adic convergence of this sum being for each x in this domain.

Recall that whenever $m \in \mathbf{Z}$, $m \geq 0$, the power of p that divides $m!$ is given by the sum

$$\sum_{j=1}^{\infty} \left[\frac{m}{p^j}\right] \leq \frac{m}{p-1},$$

where $[x]$ is the unique integer n satisfying $n \leq x < n + 1$. The bound on this sum then implies that $|m!|_p \geq |p|_p^{m/(p-1)}$.

For each $n \in \mathbf{Z}$, $n \geq 0$, the quantity $\binom{x}{n}$ is defined in like manner as the binomial coefficients, denoting $\binom{x}{0} = 1$ and

$$\binom{x}{n} = \frac{1}{n!}x(x-1)\dots(x-(n-1))$$

for $n > 0$. Note that each such quantity is a polynomial in x.

Consider the following result from [**8**] (see also Chapter 5 of [**7**]):

Lemma 5. *Let $A_j(X) = \sum_{n=0}^{\infty} a_{n,j}x^n$, $a_{n,j} \in \mathbf{C}_p$, $j = 0, 1, \dots$, be a sequence of power series, each of which converges in a fixed subset D of $\mathbf{C}_p$, such that:*

(1) *$a_{n,j} \to a_{n,0}$ as $j \to \infty$ for each n; and*
(2) *for each $s \in D$ and $\epsilon > 0$, there exists $n_0 = n_0(s, \epsilon)$ such that $|\sum_{n \geq n_0} a_{n,j}s^n|_p < \epsilon$ for all j.*

Then $\lim_{j\to\infty} A_j(s) = A_0(s)$ for all $s \in D$.

This lemma is used by Washington to show that each of the functions $\langle a \rangle^s$ and $\sum_{m=0}^{\infty} \binom{s}{m}(F/a)^m B_m$, where F is a multiple of q and f_χ, is analytic in $\mathfrak{D} = \{s \in \mathbf{C}_p : |s-1|_p < |p|_p^{1/(p-1)}|q|_p^{-1}\}$. This, along with an identity concerning generalized Bernoulli polynomials, enables the proof of the main theorem of [**8**].

By the same means, we derive a similar result for a two-variable p-adic L-function $L_p(s,t;\chi)$. This function is defined for $s \in \mathfrak{D}$, except $s \neq 1$ if $\chi = 1$, and $t \in \mathbf{C}_p$, $|t|_p \leq 1$, and it interpolates the values

$$L_p(1-n,t;\chi) = -\frac{1}{n}\left(B_{n,\chi_n}(qt) - \chi_n(p)p^{n-1}B_{n,\chi_n}\left(p^{-1}qt\right)\right),$$

where $n \in \mathbf{Z}$, $n \geq 1$. It is related to the one-variable function $L_p(s;\chi)$ in that $L_p(s,0;\chi) = L_p(s;\chi)$ for each s in the domain of $L_p(s;\chi)$.

3. The two-variable p-adic L-function.

This now brings us to our main result. We will construct our function $L_p(s,t;\chi)$ for $s \in \mathfrak{D}$, except $s \neq 1$ if $\chi = 1$, and $t \in \mathbf{C}_p$, $|t|_p \leq 1$, and in the process derive an explicit formula for this function. Before we begin this derivation, we need the following result concerning generalized Bernoulli polynomials:

Lemma 6. *Let g be a positive integral multiple of f_χ. Then for each $n \in \mathbf{Z}$, $n \geq 0$,*

$$B_{n,\chi}(t) = (-1)^n g^{n-1} \sum_{a=0}^{g-1} \chi(-a) B_n\left(\frac{a-t}{g}\right).$$

A version of this result appears in Chapter 2 of [**4**], and can be derived by a manipulation of the appropriate generating functions.

Proof of Theorem 2. Let $a \in \mathbf{Z}$, $(a,p)=1$. For $t \in \mathbf{C}_p$, $|t|_p \le 1$, the same argument as that given in the proof of the main theorem of [**8**] can be applied to show that each of the functions $\sum_{m=0}^{\infty}\binom{s}{m}(F/(a-qt))^m B_m$ and $\langle a-qt\rangle^s = \sum_{m=0}^{\infty}\binom{s}{m}(\langle a-qt\rangle - 1)^m$ is analytic for $s \in \mathfrak{D}$. This method can also be used to show that the function $\sum_{m=0}^{\infty}\binom{s}{m}(F/(a-qt))^m B_m$ is analytic for $t \in \mathbf{C}_p$, $|t|_p < |q|_p^{-1}$, whenever $s \in \mathfrak{D}$. It readily follows that $\langle a-qt\rangle^s = \langle a\rangle^s \sum_{m=0}^{\infty}\binom{s}{m}(-a^{-1}qt)^m$ is analytic for $t \in \mathbf{C}_p$, $|t|_p \le 1$, when $s \in \mathfrak{D}$. Thus, since $(s-1)L_p(s,t;\chi)$ is a finite sum of products of these two functions, it must also be analytic for $s \in \mathfrak{D}$ given $t \in \mathbf{C}_p$, $|t|_p \le 1$, and for $t \in \mathbf{C}_p$, $|t|_p \le 1$, whenever $s \in \mathfrak{D}$. Note that

$$\lim_{s\to 1}(s-1)L_p(s,t;\chi) = \frac{\chi(-1)}{F}\sum_{\substack{a=1\\(a,p)=1}}^{F}\chi(a) = \begin{cases} 1-p^{-1}, & \text{if } \chi = 1,\\ 0, & \text{if } \chi \ne 1.\end{cases}$$

Thus, our conclusions on when $L_p(s,t;\chi)$ is analytic or meromorphic follow.

Now let $n \in \mathbf{Z}$, $n \ge 1$, and fix $t \in \mathbf{C}_p$, $|t|_p \le 1$. Since F must be a multiple of f_{χ_n}, Lemma 6 implies that

$$B_{n,\chi_n}(qt) = (-1)^n F^{n-1}\sum_{a=0}^{F-1}\chi_n(-a)B_n\left(\frac{a-qt}{F}\right).$$

If $\chi_n(p) \ne 0$, then $(p, f_{\chi_n}) = 1$, so that F/p is a multiple of f_{χ_n}. Therefore,

$$\begin{aligned}
&\chi_n(p)p^{n-1}B_{n,\chi_n}\left(p^{-1}qt\right)\\
&= (-1)^n\chi_n(p)F^{n-1}\sum_{a=0}^{F/p-1}\chi_n(-a)B_n\left(\frac{a-p^{-1}qt}{Fp^{-1}}\right)\\
&= (-1)^n F^{n-1}\sum_{\substack{a=0\\p\mid a}}^{F-1}\chi_n(-a)B_n\left(\frac{a-qt}{F}\right).
\end{aligned}$$

The difference of these quantities yields

$$B_{n,\chi_n}(qt) - \chi_n(p)p^{n-1}B_{n,\chi_n}\left(p^{-1}qt\right) = \chi(-1)F^{n-1}\sum_{\substack{a=1\\(a,p)=1}}^{F}\chi_n(a)B_n\left(\frac{a-qt}{F}\right).$$

By using (2), we can rewrite the Bernoulli polynomial $B_n(t)$ in this expression as

$$B_n\left(\frac{a-qt}{F}\right) = F^{-n}(a-qt)^n\sum_{m=0}^{n}\binom{n}{m}\left(\frac{F}{a-qt}\right)^m B_m.$$

Since $\chi_n(a) = \chi(a)\omega^{-n}(a)$ and $\omega^{-1}(a)(a-qt) = \langle a-qt\rangle$ for $(a,p)=1$ and $t \in \mathbf{C}_p$, $|t|_p \leq 1$, we obtain

$$\begin{aligned} &B_{n,\chi_n}(qt) - \chi_n(p)p^{n-1}B_{n,\chi_n}\left(p^{-1}qt\right) \\ &= \frac{\chi(-1)}{F} \sum_{\substack{a=1\\(a,p)=1}}^{F} \chi(a)\langle a-qt\rangle^n \sum_{m=0}^{\infty} \binom{n}{m}\left(\frac{F}{a-qt}\right)^m B_m. \end{aligned}$$

Therefore,

$$-\frac{1}{n}\left(B_{n,\chi_n}(qt) - \chi_n(p)p^{n-1}B_{n,\chi_n}\left(p^{-1}qt\right)\right) = L_p(1-n,t;\chi),$$

completing the proof. □

Note that the proof of the main theorem of [**8**] infers the existence of the factor $\chi(-1)$ in the formula for $L_p(s;\chi)$. However, since $\chi(-1) \neq 1$ implies that $L_p(s;\chi)$ is identically 0, this quantity is not needed in the given expression. As $L_p(s,t;\chi)$ is not identically 0 for any character χ, the factor $\chi(-1)$ is needed in the expression corresponding to this function.

In [**7**], Washington modifies the derivation of $L_p(s;\chi)$ by first defining the function

$$H_p(s,a,F) = \frac{1}{s-1}\frac{1}{F}\langle a\rangle^{1-s}\sum_{m=0}^{\infty}\binom{1-s}{m}\left(\frac{F}{a}\right)^m B_m, \tag{3}$$

where $s \in \mathfrak{D}$, $s \neq 1$, $a \in \mathbf{Z}$ with $(a,p)=1$, and F is a multiple of q. The function $L_p(s;\chi)$ can then be expressed as the sum

$$L_p(s;\chi) = \sum_{\substack{a=1\\(a,p)=1}}^{F} \chi(a)H_p(s,a,F),$$

provided F is a multiple of both q and f_χ. The function $H_p(s,a,F)$ is meromorphic for $s \in \mathfrak{D}$ with a simple pole at $s=1$, having residue $1/F$, and it interpolates the values

$$H_p(1-n,a,F) = -\frac{1}{n}\omega^{-n}(a)F^{n-1}B_n\left(\frac{a}{F}\right),$$

where $n \in \mathbf{Z}$, $n \geq 1$.

It is obvious that we can express $L_p(s,t;\chi)$ in a similar manner. Using (3) to define $H_p(s,a-qt,F)$ for all $a \in \mathbf{Z}$, $(a,p)=1$, and $t \in \mathbf{C}_p$, $|t|_p \leq 1$, we obtain

$$L_p(s,t;\chi) = \chi(-1)\sum_{\substack{a=1\\(a,p)=1}}^{F} \chi(a)H_p(s,a-qt,F).$$

From the proof of Theorem 2, it follows that $H_p(s, a-qt, F)$ is analytic for $t \in \mathbf{C}_p$, $|t|_p \leq 1$, when $s \in \mathfrak{D}$, $s \neq 1$, and meromorphic for $s \in \mathfrak{D}$, with a simple pole at $s = 1$, when $t \in \mathbf{C}_p$, $|t|_p \leq 1$.

4. The value of $(\partial/\partial s)L_p(0, t; \chi)$.

Let us now consider the values of the first partial derivatives of the function $L_p(s, t; \chi)$ at $s = 0$.

In [**3**], it is shown that whenever $n \in \mathbf{Z}$, $n \geq 1$,

$$\frac{\partial^n}{\partial t^n} L_p(s, t; \chi) = n! q^n \binom{-s}{n} L_p(s+n, t; \chi_n)$$

for all $s \in \mathfrak{D}$, $s \neq 1$ if $\chi = 1$, and $t \in \mathbf{C}_p$ with $|t|_p \leq 1$. Furthermore, we have

$$\lim_{s \to 1-n} \binom{-s}{n} L_p(s+n, t; \chi) = -\frac{1}{n}\left(1 - \chi(p)p^{-1}\right) B_{0,\chi}.$$

Therefore,

$$\frac{\partial^n}{\partial t^n} L_p(1-n, t; \chi) = -(n-1)! q^n \left(1 - \chi_n(p)p^{-1}\right) B_{0,\chi_n}.$$

Since $B_{0,\chi} = 0$ whenever $\chi \neq 1$, this becomes

$$\frac{\partial^n}{\partial t^n} L_p(1-n, t; \chi) = \begin{cases} -(n-1)! q^n \left(1 - p^{-1}\right), & \text{if } \chi_n = 1, \\ 0, & \text{if } \chi_n \neq 1. \end{cases}$$

Thus, when $n = 1$, we have the value of $(\partial/\partial t)L_p(0, t; \chi)$.

The value of $(\partial/\partial s)L_p(0, t; \chi)$ is given in Theorem 4. The proof of this result follows in much the same manner as the proof of Theorem 3, given in [**6**, pp. 393-394].

Proof of Theorem 4. The value of $(\partial/\partial s)L_p(0, t; \chi)$ is the coefficient of s in the expansion of $L_p(s, t; \chi)$ about $s = 0$. We find this by determining the constant and linear terms in the corresponding expansions of each of three functions of s that make up the expression given in Theorem 2.

Expanding $1/(1-s)$ about $s = 0$ yields

$$\frac{1}{1-s} = 1 + s + \cdots,$$

while expanding $\langle a - qt \rangle^{1-s}$ about $s = 0$ yields

$$\langle a - qt \rangle^{1-s} = \langle a - qt \rangle \left(1 - s \log_p \langle a - qt \rangle + \cdots \right).$$

The expansion of $\binom{1-s}{m}$ about $s = 0$ is given by

$$\binom{1-s}{m} = \frac{(-1)^{m+1}}{m(m-1)} s + \cdots,$$

provided $m \geq 2$. Employing these expansions, along with some algebraic manipulations, we obtain

$$\frac{\partial}{\partial s} L_p(0, t; \chi) = \chi(-1) \sum_{\substack{a=1 \\ (a,p)=1}}^{F} \chi_1(a) \left(\left(\frac{a - qt}{F} - \frac{1}{2} \right) \log_p \langle a - qt \rangle - \frac{a - qt}{F} \right.$$
$$\left. + \sum_{m=2}^{\infty} \frac{1}{m(m-1)} \left(\frac{a - qt}{F} \right)^{1-m} B_m \right).$$

Since $\omega(a)$ is a root of unity for $(a, p) = 1$, we see that $\log_p \langle a - qt \rangle = \log_p(a - qt)$. Therefore,

$$\frac{\partial}{\partial s} L_p(0, t; \chi) = \chi(-1) \sum_{\substack{a=1 \\ (a,p)=1}}^{F} \chi_1(a) \left(F^{-1} \log_p(F) \cdot a + G_p \left(\frac{a - qt}{F} \right) \right).$$

By evaluating the sum

$$F^{-1} \sum_{\substack{a=1 \\ (a,p)=1}}^{F} \chi_1(a) a = (1 - \chi_1(a)) B_{1,\chi_1} = -L_p(0; \chi),$$

we obtain the result. □

By means similar to those used in the proof of Theorem 4, one can derive the following formula for the value of $L_p(1, t; \chi)$, whenever $\chi \neq 1$:

$$L_p(1, t; \chi)$$
$$= \frac{\chi(-1)}{F} \sum_{\substack{a=1 \\ (a,p)=1}}^{F} \chi(a) \left(-\log_p \langle a - qt \rangle + \sum_{m=1}^{\infty} \frac{(-1)^m}{m} \left(\frac{F}{a - qt} \right)^m B_m \right),$$

where F is a positive integral multiple of q and f_χ. This is a generalization of a similar formula for $L_p(1; \chi)$ (see [**6**, p. 85]).

Acknowledgements. Finally, the author would like to express his gratitude to his graduate advisor, Professor Andrew Granville of the University of Georgia, Athens, GA, under whose direction the two-variable p-adic L-function was derived, and to Paul Young of the University of Charleston, Charleston, SC, for suggesting the consideration of the partial derivative of this function at $s = 0$.

References

[1] J. Diamond, *The p-adic log gamma function and p-adic Euler constants*, Trans. Amer. Math. Soc., **233** (1977), 321-337, MR 58 #16610, Zbl 0382.12008.

[2] G.J. Fox, *Kummer congruences for expressions involving generalized Bernoulli polynomials*, J. Théor. Nombres Bordeaux, **14** (2002), 1-18, CMP 1 925 997.

[3] ———, *A p-adic L-function of two variables*, Enseign. Math. (2), **46** (2000), 225-278, MR 2002c:11162.

[4] K. Iwasawa, *Lectures on p-Adic L-Functions*, Ann. of Math. Studies, **74**, Princeton Univ. Press, Princeton, NJ, 1972, MR 50 #12974, Zbl 0236.12001.

[5] T. Kubota and H.-W. Leopoldt, *Eine p-adische Theorie der Zetawerte*, I. *Einführung der p-adischen Dirichletschen L-Funktionen*, J. Reine Angew. Math., **214/215** (1964), 328-339, MR 29 #1199, Zbl 0186.09103.

[6] S. Lang, *Cyclotomic Fields* I *and* II, Combined second edition, Graduate Texts in Math., **121**, Springer, New York, 1990, MR 91c:11001, Zbl 0704.11038.

[7] L.C. Washington, *Introduction to Cyclotomic Fields*, Second edition, Graduate Texts in Math., **83**, Springer, New York, 1997, MR 97h:11130, Zbl 0966.11047.

[8] ———, *A note on p-adic L-functions*, J. Number Theory, **8** (1976), 245-250, MR 53 #10766, Zbl 0329.12017.

[9] P.T. Young, *On the behavior of some two-variable p-adic L-functions*, J. Number Theory, **98** (2003), 67-88.

Received December 7, 2001 and revised May 24, 2002.

Department of Mathematics and Computer Science
Emory University
Atlanta GA 30322
E-mail address: fox@mathcs.emory.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 209, No. 1, 2003

BASS NUMBERS OF SEMIGROUP-GRADED LOCAL COHOMOLOGY

David Helm and Ezra Miller

Given a module M over a ring R that has a grading by a semigroup Q, we present a spectral sequence that computes the local cohomology $H^i_I(M)$ at any graded ideal I in terms of Ext modules. We use this method to obtain finiteness results for the local cohomology of graded modules over semigroup rings. In particular we prove that for a semigroup Q whose saturation Q^{sat} is simplicial, and a finitely generated module M over $k[Q]$ that is graded by Q^{gp}, the Bass numbers of $H^i_I(M)$ are finite for any Q-graded ideal I of $k[Q]$. Conversely, if Q^{sat} is not simplicial, we find a graded ideal I and graded $k[Q]$-module M such that the local cohomology module $H^i_I(M)$ has infinite-dimensional socle. We introduce and exploit the combinatorially defined *essential set* of a semigroup.

1. Introduction.

The local cohomology modules $H^i_I(M)$ for finitely generated modules M over noetherian rings R have been studied for several decades. When I is a maximal ideal of R the local cohomology of M is fairly well-understood (and reasonably well-behaved), but for general ideals I much less is known, and the behavior can be quite bad. For instance, Hartshorne [**Har70**] has shown that the Bass numbers of $H^i_I(M)$ need not be finite for general I and R.

Recently, however, some progress has been made in special cases. When R is a regular local ring containing a field, Lyubeznik [**Lyu93**], and Huneke and Sharp [**HS93**] have shown that $H^i_I(M)$ has finite Bass numbers. In the same spirit (albeit by different techniques), Yanagawa [**Yan01**] has shown that if ω is the canonical module of a simplicial and normal semigroup ring and I is a monomial ideal, then $H^i_I(\omega)$ has finite Bass numbers.

Our approach is a substantial generalization of that found in [**Yan01**]. We consider a noetherian ring R graded by a semigroup Q, and modules over R graded by Q^{gp}. In this setting, we introduce a functor called the *Čech hull* (Section 2), which allows us to recover the full local cohomology of a finitely generated module M from the portion of the local cohomology that lies in those graded degrees that are elements of Q (Section 3). This piece of the local cohomology is often easier to understand than the entire module; in particular it is finitely generated when $R = k[Q]$ for an affine semigroup Q (see Proposition 4.3, or Corollary 5.3 along with Proposition 5.4). Using this fact we prove several finiteness results for the local cohomology of Q^{gp}-graded modules over affine semigroup rings (Section 5). In particular we show (without Cohen-Macaulay hypotheses) that when the saturation Q^{sat} is simplicial, $H^i_I(M)$ has finite Bass numbers.

The converse holds, as well: Our construction of a local cohomology module with infinite-dimensional socle when Q^{sat} is not simplicial (Section 7) contains Hartshorne's counterexample to Grothendieck's conjecture [**Har70**] as the simplest special case. The constructions are polyhedral in nature, exploiting a new combinatorial structure, the *essential set* of a semigroup (Section 6). Properties of the essential set govern the associated primes, and to some extent the module structure, of the local cohomology of the canonical module.

The reader interested in affine semigroup rings rather than general semigroup gradings need not endure anything in Sections 2-4 except Proposition 4.1 (which can be taken as a definition) and the two paragraphs preceding it. Instead, begin with Section 7 and the first half of Section 6 (through Example 6.5)—in either order—noting especially Theorem 7.1, which contains our main results on local cohomology over affine semigroup rings. Then continue with Proposition 4.1 and Section 5. The required results from other parts of the paper can be referred to as necessary.

2. The Čech hull.

Let Q be a cancellative, commutative semigroup, and Q^{gp} its Grothendieck group, i.e., the group obtained from Q by adjoining an inverse for every element. (The reader may safely assume for the purposes of this paper that Q is *affine*—that is, a finitely generated submonoid of $\mathbb{Z}^d$ with $Q^{\mathrm{gp}} = \mathbb{Z}^d$; indeed, we will make this assumption starting in Section 4. However, we hope that the extra generality in this section and the next will be useful for more general semigroup gradings, such as those arising in the Cox homogeneous coordinate rings of toric geometry [**Cox95**].) We say a ring R is *Q-graded* and an R-module M is Q^{gp}-graded if we are given direct sum decompositions

$$R = \bigoplus_{a \in Q} R_a \quad \text{and} \quad M = \bigoplus_{\alpha \in Q^{\mathrm{gp}}} M_\alpha$$

such that $R_a R_b \subseteq R_{a+b}$ and $R_a M_\beta \subseteq M_{a+\beta}$. The category of Q^{gp}-graded modules is henceforth denoted by $\mathcal{M}$.

A morphism $M \to N$ in $\mathcal{M}$ is a degree-preserving R-module homomorphism; i.e., a map f of R-modules such that $f(M_\alpha) \subseteq N_\alpha$. We denote by $\mathrm{Hom}_R(M, N)$ the R_0-module of such morphisms and by $\underline{\mathrm{Hom}}_R(M, N)$ the

Q^{gp}-graded R-module

$$\underline{\mathrm{Hom}}_R(M,N) = \bigoplus_{\alpha\in Q^{\mathrm{gp}}} \mathrm{Hom}_R(M,N(\alpha)) = \bigoplus_{\alpha\in Q^{\mathrm{gp}}} \mathrm{Hom}_R(M(-\alpha),N).$$

Here, the Q^{gp}-graded R-module $N(\alpha)$ is the *shift of N by α*, defined by $N(\alpha)_\beta = N_{\alpha+\beta}$.

For a subset $S \subseteq Q^{\mathrm{gp}}$ closed under the action of Q, we define the *S-graded part* $M_S \subseteq M$ to be

$$M_S = \bigoplus_{\alpha\in S} M_\alpha.$$

We frequently consider the case in which $S = \alpha + Q$ for some $\alpha \in Q^{\mathrm{gp}}$. In particular, taking M_Q yields the part of M graded by elements of Q. Taking S-graded parts is functorial and exact, for any S.

The theory we develop below revolves around the following question: To what extent can we recover a module M from its Q-graded part M_Q? If M is finitely generated, for example, then although we may not be able to get M from M_Q, we can shift by some $a \in Q$ to get $M(-a) = M(-a)_Q$. Therefore, the question is more meaningful for infinitely generated modules, such as the local cohomology modules of a finitely generated module. We will find that these belong to a certain class of modules that can be recovered from Q-graded parts of other modules by a functor $\check{C}$ that we call the *Čech hull*, as defined by the next result.

Theorem 2.1. *The functor $(-)_Q : \mathcal{M} \to \mathcal{M}$ taking Q-graded parts has a right adjoint $\check{C}$; that is, there exists a functor $\check{C}$ and natural isomorphisms*

$$\mathrm{Hom}_R(M_Q, N) = \mathrm{Hom}_R(M, \check{C}N)$$

for any $M, N \in \mathcal{M}$.

Proof. Given N, we explicitly construct $\check{C}N$ by defining

$$(1)\qquad \begin{aligned}(\check{C}N)_\alpha &= \mathrm{Hom}_R(R_{Q-\alpha}, N(\alpha))\\ &= \mathrm{Hom}_R(R_{Q-\alpha}(-\alpha), N).\end{aligned}$$

The multiplication maps

$$R_b \otimes_{R_{\mathbf{0}}} (\check{C}N)_\alpha \to (\check{C}N)_{b+\alpha}$$

are given by taking $r \otimes \phi$ to $(x \mapsto \phi(rx))$. This is well-defined since multiplication by $r \in R_b$ is a degree zero map $R_{Q-\alpha-b}(-\alpha-b) \to R_{Q-\alpha}(-\alpha)$.

Note that if $a \in Q$ then $R_{Q-a} = R$, so

$$(\check{C}N)_a = \mathrm{Hom}_R(R_{Q-a}(-a), N) = \mathrm{Hom}_R(R(-a), N) = N_a,$$

whence $(\check{C}N)_Q = N_Q$. Therefore, given an element of $\mathrm{Hom}_R(M, \check{C}N)$, taking its Q-graded part gives an element of $\mathrm{Hom}_R(M_Q, N_Q)$. This last module is $\mathrm{Hom}_R(M_Q, N)$ (since degree zero maps from M_Q to N must land in N_Q), so we have produced a natural map $\mathrm{Hom}_R(M, \check{C}N) \to \mathrm{Hom}_R(M_Q, N)$.

Conversely, if $f \in \mathrm{Hom}_R(M_Q, N)$ then for each $\alpha \in Q^{\mathrm{gp}}$ we have a map

$$M_\alpha \to (\check{C}N)_\alpha = \mathrm{Hom}_R(R_{Q-\alpha}, N(\alpha)) \quad \text{defined by} \quad x \mapsto (r \mapsto f(rx)).$$

This is well-defined since if $r \in R_{Q-\alpha}$ and $x \in M_\alpha$, then $rx \in M_Q$, so we can evaluate $f(rx)$. We thus obtain a well-defined element of $\mathrm{Hom}_R(M, \check{C}N)$, whose Q-graded part is just f. This gives the natural inverse map for our bijection. □

Remark 2.2. It is clear by looking at the graded pieces of $\check{C}$ that it is left exact but not right exact, and that its derived functors are given in terms of Ext modules; that is $((R^i\check{C})M)_\alpha = \mathrm{Ext}^i_R(R_{Q-\alpha}(-\alpha), M)$. Since $R_{Q-\alpha}(-\alpha)$ is supported on $Q \subset Q^{\mathrm{gp}}$, both the Čech hull and its derived functors depend only on the Q-graded part of M.

Remark 2.3. The Čech hull was defined in [**Mil00**] for polynomial rings. In this case $Q = \mathbb{N}^d$, $R = k[Q]$, and $\mathrm{Hom}_R(R_{Q-\alpha}(-\alpha), M) = M_{\alpha^+}$, where α^+ is obtained from α by zeroing out the negative coordinates. Thus our definition of $\check{C}$ agrees with the one in [**Mil00**]. Note that the Čech hull is exact in this case, since $R_{Q-\alpha}$ is free for all α, so the Ext modules that make up the graded pieces of $R^i\check{C}$ vanish.

Example 2.4. Let $Q \subset \mathbb{Z}^2$ be the semigroup generated by $(0,2)$,$(1,1)$, and $(2,0)$, so that $Q^{\mathrm{gp}} \subset \mathbb{Z}^2$ is a lattice of index 2. Take $R = k[Q]$, graded by Q. If $\alpha = (x,y) \in Q^{\mathrm{gp}}$, then x and y have the same parity. If x and y are even, or if x and y have the same sign, then $R_{Q-\alpha}(-\alpha)$ is free. On the other hand, if x is odd and negative while y is odd and positive, then $R_{Q-\alpha}(-\alpha)$ is generated in degrees $(1,y)$ and $(0,y+1)$. Moreover, one has an exact sequence:

$$0 \to R_{Q-\alpha}(-\alpha-(1,1)) \longrightarrow R(-1,-y) \oplus R(0,-y-1) \longrightarrow R_{Q-\alpha}(-\alpha) \to 0.$$

Splicing homological and graded shifts of this short exact sequence together gives a free resolution $F_{\bullet}$ of $R_{Q-\alpha}(-\alpha)$ such that $F_i = R(-1-i,-y-i) \oplus R(-i,-y-i-1)$. Since Q is symmetric in x and y, a similar result holds with x and y reversed.

We now have free resolutions of $R_{Q-\alpha}(-\alpha)$ for every $\alpha \in Q^{\mathrm{gp}}$, and we can use them to compute the Cech hull and its derived functors. For instance, consider the module $k(-u,-v)$ consisting of a single copy of the residue field k, supported in a degree (u,v) satisfying $v > u > 1$. Then for $\alpha = (x,y)$, we find that $\mathrm{Hom}_R(R_{Q-\alpha}(-\alpha), k(-u,-v))$ is only nontrivial if $\alpha = (u,v)$. Since Equation (1) implies that

$$R^i\check{C}(k(-u,-v))_\alpha \cong \mathrm{Ext}^i_R(R_{Q-\alpha}(-\alpha), k(-u,-v)),$$

it follows that $\check{C}(k(-u,-v)) = k(-u,-v)$. In contrast, $R^i\check{C}(k(-u,-v))$ for $i > 1$ can only be nonzero in degrees α for which x and y are odd and of differing sign, since $R_{Q-\alpha}(-\alpha)$ is free otherwise.

Suppose we have such an α, and let $F_\bullet$ be the corresponding free resolution of α constructed above. Then $\mathrm{Hom}_R(F_i, k(-u,-v))$ is nonzero if (and only if) one of the generators of F_i sits in degree (u,v). Referring to the expression for the degrees of the F_i in terms of α (and remembering that we assumed $v > u > 1$), we find that $\mathrm{Hom}_R(F_i, k(-u,-v))$ is nonzero if and only if $i = u$ and $y = v-u-1$ or if $i = u-1$ and $y = v-u+1$. In other words, $R^i\check{C}(k(-u,-v))$ vanishes except when $i \in \{u-1, u\}$. Moreover, $R^{u-1}\check{C}(k(-u,-v))$ is supported in those degrees α such that x is odd and negative and $y = v-u+1$, while $R^u\check{C}(k(-u,-v))$ is supported in those degrees α such that x is odd and negative and $y = v-u-1$.

To summarize, when $v > u > 1$, we have:

- $\check{C}(k(-u,-v)) = k(-u,-v)$
- $R^{u-1}\check{C}(k(-u,-v)) = k[\mathbf{x}^{-(2,0)}](-1,-v+u-1)$
- $R^u\check{C}(k(-u,-v)) = k[\mathbf{x}^{-(2,0)}](-1,-v+u+1)$

and all other derived functors of $\check{C}$ vanish. Here, $\mathbf{x}^\alpha \in k[Q^{\mathrm{gp}}]$ is the element corresponding to $\alpha \in Q^{\mathrm{gp}}$.

3. Local cohomology.

In this section we study the interaction of the Čech hull with the functor Γ_I, which takes a module M to the submodule annhilated elementwise by some power of the ideal I. If I is graded, then Γ_I takes the category $\mathcal{M}$ of Q^{gp}-graded modules to itself.

Proposition 3.1. *If R is noetherian, then $\check{C}$ and Γ_I commute: $\Gamma_I\check{C} = \check{C}\Gamma_I$.*

Proof.

$$\begin{aligned}(\Gamma_I\check{C}M)_\alpha &= \bigcup_n \{x \in \mathrm{Hom}_R(R_{Q-\alpha}(-\alpha), M) \mid I^n x = 0\} \\ &= \mathrm{Hom}_R(R_{Q-\alpha}(-\alpha), \Gamma_I M) \\ &= (\check{C}\Gamma_I M)_\alpha.\end{aligned}$$

□

Henceforth we assume R is noetherian. In Proposition 3.3 we shall apply the spectral sequence of a composite functor to the functors $\Gamma_I\check{C}$ and $\check{C}\Gamma_I$. In order to do this we use the fact that both Γ_I and $\check{C}$ take injectives to injectives. For Γ_I this is standard; for $\check{C}$ this follows from Lemma 3.2, whose extra precision is vital for Section 5.

Lemma 3.2. *Let J be an indecomposable injective in $\mathcal{M}$. Then $\check{C}J = 0$ if $J_Q = 0$, and $\check{C}J = J$ otherwise. In particular, $\check{C}$ takes injectives to injectives.*

Proof. The first statement is clear, since $\check{C}$ depends only on the Q-graded portion of a module. The last statement follows from the others, using the fact that every injective is a direct sum of indecomposable injectives (this uses the noetherian hypothesis).

For the remaining statement, write $J = \underline{E}\,(R/\mathfrak{p})(\alpha)$ for some prime $\mathfrak{p}$ of R and $\alpha \in Q^{\mathrm{gp}}$. Then $(R/\mathfrak{p})(\alpha)$ is an essential submodule of J, so since J_Q is nonzero, $(R/\mathfrak{p})(\alpha)_Q$ is nonzero. Now $(R/\mathfrak{p})(\alpha)_Q$ is an essential submodule of $(R/\mathfrak{p})(\alpha)$ (since $R/\mathfrak{p}$ is an integral domain), so it is an essential submodule of J. Thus in particular J_Q is an essential submodule of J.

The inclusion $J_Q \to J$ induces a map $\phi : J \to \check{C}J$, by the adjointness property of $\check{C}$. Moreover, ϕ is injective, since it restricts to the identity on the essential submodule J_Q. Thus J is a direct summand of $\check{C}J$, since J is injective. We claim that $\check{C}J$ is an essential extension of J, from which the result follows immediately.

Let $x \in (\check{C}J)_\alpha = \mathrm{Hom}_R(R_{Q-\alpha}(-\alpha), J_Q)$ be a nonzero homogeneous element. Then there is an $r \in R$ such that $x(r)$ is nonzero, and then $rx = x(r)$ is a nonzero element of J_Q. Thus $\check{C}J$ is an essential extension of J_Q and hence of J as well. □

Proposition 3.3. *Let M be a graded R-module. There are spectral sequences $E(M)$ and $F(M)$ described by*

$$E_2^{p,q}(M) = R^p\check{C}H_I^q(M) \Rightarrow R^{p+q}(\Gamma_I\check{C})M$$
$$F_2^{p,q}(M) = H_I^p(R^q\check{C}M) \Rightarrow R^{p+q}(\Gamma_I\check{C})M.$$

Proof. These are spectral sequences for the composite functors $\check{C}\Gamma_I$ and $\Gamma_I\check{C}$ [**Wei94**], using Proposition 3.1 along with Lemma 3.2 and the fact that Γ_I takes injectives to injectives. □

Example 3.4. We return to the case of Remark 2.3; that is, $R = k[\mathbb{N}^d]$. Here $\check{C}$ is exact, and so the spectral sequences $E(M)$ and $F(M)$ both collapse. The proposition simply says that $H_I^i(\check{C}M) = \check{C}H_I^i(M)$. The right derived functors of the Čech hull measure the degree to which this equality fails in other rings.

Ultimately, the goal of this section is Theorem 3.10, which describes how the local cohomology modules of M can be reconstructed from a finite collection of submodules thereof, using the Čech hull and its derived functors. After a choosing a certain $\alpha \in Q^{\mathrm{gp}}$, this is accomplished by a spectral sequence $E(M(-\alpha))$ that only depends on the (finitely generated) Q-graded parts of the local cohomology modules of $M(-\alpha)$, since $R^p\check{C}H_I^q(-)$ depends only on the Q-graded part of $H_I^q(-)$.

For this approach to work, we of course need $R^{p+q}(\Gamma_I\check{C})M$ to be a local cohomology module. To this end, we use the spectral sequence $F(M)$.

Although the filtration that arises from $F(M)$ is generally nontrivial, we avoid this nuisance by replacing M with a suitable Q^{gp}-graded shift, forcing F to collapse in low cohomological degree. We find this suitable shift in Corollary 3.6 using Proposition 3.5, which is interesting in its own right.

Proposition 3.5. *Let $J^{\bullet}$ be a minimal injective resolution of a finitely generated module $M \in \mathcal{M}$. Let $\mathfrak{p}$ be a homogeneous prime of R, let $\mathfrak{m}$ be a homogeneous maximal ideal containing $\mathfrak{p}$, and let $c = \dim(R/\mathfrak{p}) - \dim(R/\mathfrak{m})$. If every indecomposable summand of $\Gamma_{\mathfrak{m}} J^{i+c}$ has nonzero Q-graded part, then every indecomposable summand of J^i isomorphic to a shift of $\underline{E}(R/\mathfrak{p})$ has nonzero Q-graded part.*

Proof. Inverting all homogeneous elements outside $\mathfrak{m}$ fixes all shifts of $\underline{E}(R/\mathfrak{p})$ as well as $\Gamma_{\mathfrak{m}} J^i$, so we assume henceforth that $\mathfrak{m}$ is the unique maximal homogeneous ideal of R.

We begin with the case $c = 1$. Using $(-)_{(\mathfrak{p})}$ to denote the localization by all homogeneous elements outside of $\mathfrak{p}$, it is a standard fact (in [**BH93**, p. 101], for instance) that $(\underline{\mathrm{Ext}}^i_R(R/\mathfrak{p}, M))_{(\mathfrak{p})} = (\underline{\mathrm{Hom}}_R(R/\mathfrak{p}, J^i))_{(\mathfrak{p})} \subset (\Gamma_{\mathfrak{p}} J^i)_{(\mathfrak{p})}$ is an essential extension. Therefore, we need only show that every indecomposable submodule of the free $(R/\mathfrak{p})_{(\mathfrak{p})}$-module $(\underline{\mathrm{Ext}}^i_R(R/\mathfrak{p}, M))_{(\mathfrak{p})}$ has nonzero Q-graded part.

Choose a homogeneous element $x \in \mathfrak{m} \setminus \mathfrak{p}$, and define L by the exact sequence

$$0 \to R/\mathfrak{p} \xrightarrow{x} R/\mathfrak{p} \longrightarrow L \to 0,$$

where $\xrightarrow{x}$ is multiplication by x. The long exact sequence for $\underline{\mathrm{Ext}}^{\bullet}_R(-, M)$ provides a right exact sequence

$$\underline{\mathrm{Ext}}^i_R(R/\mathfrak{p}, M) \xrightarrow{x} \underline{\mathrm{Ext}}^i_R(R/\mathfrak{p}, M) \longrightarrow L(i, x, M) \to 0$$

for the appropriate submodule $L(i, x, M) \subseteq \underline{\mathrm{Ext}}^{i+1}_R(L, M)$. Tensoring with $R/\mathfrak{m}$ yields an isomorphism $R/\mathfrak{m} \otimes \underline{\mathrm{Ext}}^i_R(R/\mathfrak{p}, M) \cong R/\mathfrak{m} \otimes L(i, x, M)$ since multiplication by x becomes the zero map. Nakayama's lemma implies that $\underline{\mathrm{Ext}}^i_R(R/\mathfrak{p}, M)$, and hence $(\underline{\mathrm{Ext}}^i_R(R/\mathfrak{p}, M))_{(\mathfrak{p})}$, is generated by elements in degrees $\gamma \in Q^{\mathrm{gp}}$ such that $L(i, x, M)_\gamma \subseteq \underline{\mathrm{Ext}}^{i+1}_R(L, M)_\gamma \neq 0$.

Now $\underline{\mathrm{Ext}}^{i+1}_R(L, M)$ is the $(i+1)^{\mathrm{st}}$ cohomology of the complex $\underline{\mathrm{Hom}}_R(L, J^{\bullet}) \subset \Gamma_{\mathfrak{m}} J^{\bullet}$, whose socle subcomplex $\underline{\mathrm{Hom}}_R(R/\mathfrak{m}, J^{\bullet}) \subseteq \underline{\mathrm{Hom}}_R(L, J^{\bullet})$ (the inclusion being induced by the surjection $L \twoheadrightarrow R/\mathfrak{m}$) is equal to $\underline{\mathrm{Ext}}^{\bullet}_R(R/\mathfrak{m}, M)$. The hypothesis on $(\Gamma_{\mathfrak{m}} J^i)_Q$ in the Proposition implies that $\underline{\mathrm{Ext}}^{i+1}_R(R/\mathfrak{m}, M)$ must in fact equal its Q-graded part. Given any degree $\gamma \in Q^{\mathrm{gp}}$ for which $\underline{\mathrm{Ext}}^{i+1}_R(L, M)_\gamma \neq 0$, we therefore can find a homogeneous element $r \in R$ such that $\gamma + \deg(r) \in Q$. Note that $\mathfrak{p}$ annihilates L and hence also $\underline{\mathrm{Hom}}_R(L, J^{\bullet})$, so we can always find our element r outside of $\mathfrak{p}$. When we invert r to form the localization $(\underline{\mathrm{Ext}}^i_R(R/\mathfrak{p}, M))_{(\mathfrak{p})}$, any generator y in degree γ can be replaced by the generator ry whose degree is in Q. This concludes the case where $c = 1$.

The general case proceeds by induction on c, replacing R with its homogeneous localization at a prime containing $\mathfrak{p}$ and having dimension $\dim(R/\mathfrak{p}) - 1$. □

Recall that the *Bass number* of a module M at the prime $\mathfrak{p}$ in cohomological degree i is the number of indecomposable summands isomorphic to a shift of $\underline{E}(R/\mathfrak{p})$ appearing at the i^{th} stage in any minimal injective resolution of M. These numbers are always finite if M is finitely generated, but may in general be infinite.

Corollary 3.6. *Suppose R has a unique maximal homogeneous ideal $\mathfrak{m}$. Let $M \in \mathcal{M}$ be a finitely generated R-module, and n be a positive integer. Then there exists $\alpha \in Q^{\mathrm{gp}}$ such that for all $\beta \in \alpha + Q$,*

1. *$\check{C}(M(-\beta)) = M(-\beta)$, and*
2. *$R^j\check{C}(M(-\beta)) = 0$ if $1 \leq j < n$.*

Proof. Since M is finitely generated, the Bass numbers of M at $\mathfrak{m}$ are finite. Thus $\Gamma_{\mathfrak{m}} J^i$ is a finite direct sum of indecomposables for each $i \leq n + \dim(R)$. These can be moved to have nonzero Q-graded part by some shift $(-\alpha)$. Lemma 3.2 and Proposition 3.5 together then imply that $\check{C}$ fixes $J^{\bullet}(-\beta)$ in cohomological degree n and less for all $\beta \in \alpha + Q$. □

Remark 3.7. If R is a ring with only finitely many homogeneous primes (e.g., a semigroup ring), then the conclusion of Corollary 3.6 holds for any M with finite Bass numbers, as then M has an injective resolution with finitely many summands in each cohomological degree.

Example 3.8. Let $Q \subset \mathbb{Z}^d$ be a finitely generated semigroup, and let $R = k[Q]$, graded by Q. Ishida [**Ish88**] constructed a dualizing complex for R, in which each indecomposable injective appears without shift. When R is Cohen-Macaulay, this is an injective resolution of the canonical module ω_R that is fixed by the Čech hull. Hence Corollary 3.6 holds for ω_R with $\alpha = 0$. It follows that for $q > 0$ we have

$$F_2^{p,q}(\omega_R) = H^p_I(R^q\check{C}\omega_R) = 0,$$

so $F(\omega_R)$ converges to $H^{p+q}_I(\omega_R)$. Thus $E(\omega_R)$ likewise converges to a filtration of $H^{p+q}_I(\omega_R)$. We will use this fact in the next section (Proposition 4.9) to compute $H^{p+q}_I(\omega_R)$ for I prime.

Example 3.9. The phenomenon predicted by Corollary 3.6 is clearly illustrated in Example 2.4: As u and v increase, the derived functors of $\check{C}(k(-u, -v))$ in positive cohomological degrees $< u - 1$ vanish.

Theorem 3.10. *Suppose R has a unique homogeneous maximal ideal. Let M be a finitely generated R-module, and n be a positive integer. Then there exists $\alpha \in Q^{\mathrm{gp}}$ such that for all $\beta \in \alpha + Q$, the spectral sequence*

$$E_2^{p,q}(M(-\beta)) = R^p\check{C}H_I^q(M(-\beta)) \Rightarrow H_I^{p+q}(M)(-\beta)$$

converges to a local cohomology module for $p+q<n$.

Proof. Choose α as in Corollary 3.6. Then for all $\beta \in \alpha + Q$,

$$F_2^{p,q}(M(-\beta)) = H_I^p(R^q\check{C}M(-\beta)) = \begin{cases} 0 & \text{if } q>0 \\ H_I^p(M)(-\beta) & \text{if } q=0 \end{cases}.$$

Hence if $p+q<n$, $R^{p+q}(\Gamma_I\check{C})(M(-\beta)) = H_I^{p+q}(M(-\beta))$. Since $E(M(-\beta))$ converges to the former by Proposition 3.3, the result follows. □

Injective resolutions are rarely finite, so no matter which α is chosen in Theorem 3.10, $F(M)$ really can converge to something other than $H_I^{p+q}(M)$ in large cohomological degrees. For example, if we take Q, R, and $k(-u,-v)$ as in Example 2.4, then $H_I^i(k(-u,-v))$ vanishes for $i \geq 1$ and any I. On the other hand, we have $F_2^{p,q}(k(-u,-v)) = 0$ for $p \geq 1$ and $F_2^{0,q}$ nonzero for $q \in \{u-1, u\}$, so the nonvanishing derived functors of $\check{C}$ cause $F(k(-u,-v))$ to fail to converge to local cohomology in these degrees.

However, since $H_I^{p+q}(M)$ vanishes in sufficiently high cohomological degrees, choosing n large in the theorem does show how the collection of Q-graded parts $H_I^j(M)(-\beta)_Q$ for all j determine the entire local cohomology modules. As we shall see in Section 4, the Q-graded portion of a local cohomology module is often much easier to understand than the local cohomology module itself.

4. Semigroup rings.

One of the ways of understanding local cohomology $H_I^\bullet(-)$ in terms of finitely generated modules is by taking limits (over m) of modules $\mathrm{Ext}_R^\bullet(R/I^m, -)$. Unfortunately, these limits are frequently quite badly behaved (see[**EMS00**], for example). Here, we bypass them entirely, in the case where $R = k[Q]$ is an affine semigroup algebra over a field k, by constructing the graded pieces of $H_I^\bullet(M)$ in terms of the derived functors $\underline{\mathrm{Ext}}_R^\bullet(R/I^m, M)$ of $\underline{\mathrm{Hom}}_R(R/I^m, M)$ for a *single fixed* m, using Theorem 3.10. In order for this to work, we need to know what the Q-graded part of local cohomology looks like.

In Sections 4-7, we set $R = k[Q]$, an affine semigroup algebra graded by $Q \subseteq \mathbb{Z}^d$, which is not assumed normal. Such a ring satisfies the hypotheses of Corollary 3.6, so that all of the machinery of the previous sections applies. For $k[Q]$ we also have a simpler expression for the Čech hull. In what follows, Q is viewed as contained in $k[Q]$ via $a \mapsto \mathbf{x}^a$.

Proposition 4.1. *When $R = k[Q]$ and $M \in \mathcal{M}$, we have*

$$(\check{C}M)_\beta \cong \mathrm{Hom}_R(R_{Q+\beta}, M).$$

If $a \in Q$, we have a commutative diagram:

$$\begin{array}{ccc} (\check{C}M)_\beta & \xrightarrow{\cdot \mathbf{x}^a} & (\check{C}M)_{\beta+a} \\ \downarrow & & \downarrow \\ \mathrm{Hom}_R(R_{Q+\beta}, M) & \longrightarrow & \mathrm{Hom}_R(R_{Q+a+\beta}, M), \end{array}$$

where the vertical arrows are isomorphisms and the bottom arrow is induced by the inclusion of $R_{Q+a+\beta}$ in $R_{Q+\beta}$.

Proof. By definition, $(\check{C}M)_\beta = \mathrm{Hom}_R(R_{Q-\beta}(-\beta), M)$. Multiplication by $\mathbf{x}^\beta$ induces an injection $R_{Q-\beta}(-\beta) \to R_{Q+\beta}$; this is an isomorphism since both of these modules are supported in the same degrees.

Moreover, one has the following commutative diagram:

$$\begin{array}{ccc} R_{Q-\beta-a}(-\beta-a) & \xrightarrow{\cdot \mathbf{x}^a} & R_{Q-\beta}(-\beta) \\ \downarrow & & \downarrow \\ R_{Q+\beta+a} & \longrightarrow & R_{Q+\beta} \end{array}$$

from which the rest of the proposition follows immediately. ☐

Lemma 4.2. *Let $J = \underline{E}(R/\mathfrak{p})(-\alpha)$ be an indecomposable injective, and I an ideal of R. There exists $n \in \mathbb{N}$ such that for all $m > n$, $(\Gamma_I J)_Q = \underline{\mathrm{Hom}}_R(R/I^m, J)_Q$.*

Proof. Suppose I is not contained in $\mathfrak{p}$. Then some element of I acts as a unit on $R/\mathfrak{p}$, so $\Gamma_I J = \underline{\mathrm{Hom}}_R(R/I^m, J) = 0$ and the result is trivial. Thus it suffices to show this result for I contained in $\mathfrak{p}$. In this case $\Gamma_I J = J$, so it suffices to show that $\underline{\mathrm{Hom}}_R(R/I^m, J)_Q = J_Q$; i.e., that every element of J_Q is killed by I^m.

Let τ be a linear functional that takes nonnegative values on Q, such that if $b \in Q$, then $\tau(b) > 0 \Leftrightarrow b \in \mathfrak{p}$. Then $\underline{E}(R/\mathfrak{p})$ is supported in those degrees β such that $\tau(\beta) \le 0$. Thus J is supported in those degrees β such that $\tau(\beta) \le \tau(\alpha) =: n$.

Suppose $m > n$. Let $y \in J_Q$ and $x \in I^m$ be nonzero homogeneous elements of degrees b and c, respectively. Then $\tau(b) \ge 0$ because $b \in Q$ and $\tau(c) \ge m$ because $x \in I^m \subseteq \mathfrak{p}^m$. Thus xy lies in degree $b+c$ and $\tau(b+c) > n$, so $xy = 0$, as required. ☐

Now we can apply the Lemma to describe Q-graded parts of local cohomology.

Proposition 4.3. *Let $M \in \mathcal{M}$ be finitely generated, and I be a graded ideal of R. Fix a nonnegative integer i. Then there exists $m_0 \in \mathbb{N}$ such that for any $m \ge m_0$,*

$$H^i_I(M)_Q \cong \underline{\mathrm{Ext}}^i_R(R/I^m, M)_Q.$$

Proof. Let $J^\bullet$ be an injective resolution for M, and choose m_0 sufficiently large that $(\Gamma_I N)_Q = \underline{\mathrm{Hom}}_R(R/I^m, N)_Q$ agree for every $m \geq m_0$ and every indecomposable injective summand N appearing in cohomological degree i or lower in $J^\bullet$. Then the first i right derived functors of $(\Gamma_I -)_Q$ and $\underline{\mathrm{Hom}}_R(R/I^m, -)_Q$ agree on M; since $(-)_Q$ is exact this means $H^i_I(M)_Q \cong \underline{\mathrm{Ext}}^i_R(R/I^m, -)_Q$. □

Example 4.4. If Q is saturated and $M = \omega_R$ then the power of I in Proposition 4.3 can be set equal to 1; i.e.,

$$\underline{\mathrm{Ext}}^p_R(R/I, \omega_R)_Q \cong H^p_I(\omega_R)_Q,$$

since the indecomposable summands of the injective resolution of ω_R are unshifted $\underline{E}\,(R/\mathfrak{p})$'s.

Corollary 4.5. *If $M \in \mathcal{M}$ is finitely generated, $H^i_I(M)_Q$ is finitely generated.*

Corollary 4.6. *Let $M \in \mathcal{M}$ be finitely generated. Then $H^i_I(M)$ has a finitely generated essential submodule if and only if for some $\beta \in Q^{\mathrm{gp}}$ the natural map $H^i_I(M)(-\beta) \to \check{C}(H^i_I(M)(-\beta))$ is an injection.*

Proof. Suppose $H^i_I(M)$ has a finitely generated essential submodule N. Then we can shift N so that all of its generators have degrees in Q; i.e., there exists β such that $N(-\beta) \subset (H^i_I(M)(-\beta))_Q$. Since the map $H^i_I(M)(-\beta) \to \check{C}H^i_I(M)(-\beta)$ is injective on its Q-graded part, it is injective on $N(-\beta)$; since $N(-\beta)$ is essential the map is injective everywhere.

Conversely, $H^i_I(M)(-\beta)_Q$ is an essential submodule of $\check{C}(H^i_I(M)(-\beta))$ and hence of $H^i_I(M)(-\beta)$. By Corollary 4.5 it is finitely generated. □

The upshot of the above is that since the spectral sequence $E(M)$ of Proposition 3.3 depends only on the Q-graded parts of the local cohomology modules whic appear in it, we can just replace these local cohomology modules with the corresponding $\underline{\mathrm{Ext}}$-modules.

Theorem 4.7. *Let M be a finitely generated module over $R = k[Q]$, and n be a positive integer. Then there exists $\alpha \in Q^{\mathrm{gp}}$ such that for all $\beta \in \alpha + Q$, there exists $m \in \mathbb{Z}$ making the spectral sequence*

$$E_2^{p,q}(M(-\beta)) = R^p\check{C}\underline{\mathrm{Ext}}^q_R(R/I^m, M(-\beta)) \Rightarrow H^{p+q}_I(M)(-\beta)$$

from Proposition 3.3 converge to a local cohomology module for $p + q < n$. Taking degree γ parts for any $\gamma \in Q^{\mathrm{gp}}$ yields a spectral sequence of iterated Ext *modules,*

$$E_2^{p,q}(M(-\beta))_\gamma = \mathrm{Ext}^p_R(R_{Q+\gamma}(\beta), \underline{\mathrm{Ext}}^q_R(R/I^m, M)) \Rightarrow H^{p+q}_I(M)_{\gamma-\beta}.$$

Proof. This is immediate from Theorem 3.10 and Proposition 4.3. □

Example 4.8. Returning to the setting of Example 3.4, we find using this theorem that if M is a $\mathbb{Z}^d$-graded module over a polynomial ring in d variables and I is a monomial ideal, then there exist $m \in \mathbb{Z}$ and $\beta \in \mathbb{Z}^d$ such that $H_I^i(M) = \check{C}\underline{\mathrm{Ext}}_R^i(R/I^m, M(-\beta))$. This generalizes a result proved independently by Mustaţă [**Mus00**] and Terai [**Ter99**].

For saturated Q, Theorem 4.7 takes an especially nice form for canonical modules. Recall that a *face* of Q is the set of degrees of elements outside a prime ideal of R.

Proposition 4.9. *Suppose R is normal and of dimension d. Let $\mathfrak{p}$ be a prime of R, corresponding to an n-dimensional face of Q. Then $H_{\mathfrak{p}}^{d-i}(\omega_R) \cong R^{n-i}\check{C}(\omega_{R/\mathfrak{p}})$.*

Proof. $R^q\check{C}\underline{\mathrm{Ext}}_R^p(R/\mathfrak{p}, \omega_R) \Rightarrow H_{\mathfrak{p}}^{p+q}(\omega_R)$ by Example 4.4 and Theorem 4.7. Since $R/\mathfrak{p}$ is a dimension n Cohen-Macaulay quotient of the Cohen-Macaulay ring R of dimension d, the module $\underline{\mathrm{Ext}}_R^p(R/\mathfrak{p}, \omega_R)$ is nonzero only when $p = d-n$, in which case it is $\omega_{R/\mathfrak{p}}$. Thus the spectral sequence degenerates, and $R^q\check{C}(\omega_{R/\mathfrak{p}}) \cong H_{\mathfrak{p}}^{q+d-n}(\omega_R)$. □

Example 4.10. Let Q be the semigroup on four generators $\{x, y, u, v\}$ and one relation $x+u = y+v$, and $R = k[Q]$. In [**Har70**], Hartshorne shows that for the ideal $I = (\mathbf{x}^u, \mathbf{x}^v)$, the local cohomology module $H_I^1(\omega_R)$ has a finitely generated essential submodule while $H_I^2(\omega_R)$ has an infinite dimensional socle, supported in degrees $n(x-v)$ for $n > 0$. This is consistent with Proposition 4.9, which says that infinite-dimensional socles must arise from a nonvanishing higher derived functor of the Čech hull. See Section 7 for a combinatorial explanation of why this bad behavior occurs, in the context of its generalization to arbitrary affine semigroup rings.

5. Finiteness for simplicial semigroups.

Letting $Q \subseteq \mathbb{Z}^d$ be affine as in the previous section, the above machinery allows us to make strong statements about local cohomology over $R = k[Q]$. Indeed, the fact that local cohomology modules "come from" the derived functors of the Čech hull forces certain structure on them. This structure is codified in the notion of a *straight* module, a common generalization of notions due to Miller [**Mil00**] (who defined **a**-*determined modules* over a polynomial ring for $\mathbf{a} \in \mathbb{N}^n$) and Yanagawa [**Yan01**, **Yan00**] (who defined straightness for a restrictive class of modules over semigroup rings). One obtains from Theorem 4.7 that local cohomology modules over semigroup rings are straight when shifted appropriately. Over a simplicial (and not necessarily normal) semigroup ring this forces them to have finite Bass numbers (Theorem 5.8). The key to all this is the following definition:

Definition 5.1. Let $Q \subseteq \mathbb{Z}^d$ be affine and $R = k[Q]$. The category $\mathcal{S}$ of *straight* R-modules is the smallest subcategory of the Q^{gp}-graded modules $\mathcal{M}$ such that (1) all indecomposable injectives J satisfying $J_Q \neq 0$ are in $\mathcal{S}$; (2) finite direct sums of modules in $\mathcal{S}$ are in $\mathcal{S}$; and (3) if ϕ is a homomorphism of straight modules, then $\ker(\phi)$ and $\operatorname{coker}(\phi)$ are straight.

Proposition 5.2. *Let $M \in \mathcal{M}$ be finitely generated. Then $R^i\check{C}M$ is straight.*

Proof. By definition, a finite direct sum of straight modules is straight. Thus if $C^\bullet$ is a complex of finite direct sums of indecomposable injectives, and each indecomposable injective in $C^\bullet$ has nontrivial Q-graded part, then $C^\bullet$ is a complex of straight modules, and the cohomology of $C^\bullet$ is straight. In particular, if $C^\bullet = \check{C}J^\bullet$ is the Čech hull of an injective resolution $J^\bullet$ of M, then $\check{C}J^\bullet$ is a complex of straight modules by Lemma 3.2. Thus its cohomology is straight, as required. □

Corollary 5.3. *Let $M \in \mathcal{M}$ be finitely generated. Then there exists an element $a \in Q$ such that $H^i_I(M)(-a)$ is straight.*

Proof. By Theorem 4.7 and Proposition 5.2, we have $a \in Q$ such that $H^i_I(M)(-a)$ is the limit of a spectral sequence of straight modules. This spectral sequence yields a *finite* filtration of $H^i_I(M)(-a)$ whose associated graded modules are therefore straight. □

Straight modules have a number of useful properties. In particular, the fact that they can be "built out of" indecomposable injectives with nontrivial Q-graded part by taking kernels, cokernels, and finite direct sums forces many of their graded pieces to be isomorphic to each other.

Proposition 5.4. *Let M be a straight module over $R = k[Q]$. Then:*

1. *M_Q is finitely generated.*
2. *$M(-a)$ is straight for all $a \in Q$.*
3. *Multiplication by $\mathbf{x}^a$ is an isomorphism $M_\beta \to M_{a+\beta}$ whenever $\beta \in Q^{\mathrm{gp}}$ and $a \in Q$ satisfy $(\beta + Q) \cap Q = (a + \beta + Q) \cap Q$.*

Proof. If the above three properties hold for M and N, then they also hold for $M \oplus N$, as well as $\ker(\phi)$ and $\operatorname{coker}(\phi)$ for any $\phi : M \to N$. Thus it suffices to check that if J is an indecomposable injective, and J_Q is nonzero, then J has the above properties. J_Q is clearly finitely generated, and if J_Q is nonzero, so is $J(-a)_Q$, so the first two properties are clear.

For the third property, note that $\check{C}J = J$ by Lemma 3.2. Thus in particular,

$$\begin{aligned} J_\beta &= (\check{C}J)_\beta &&= \operatorname{Hom}_R(R_{Q+\beta}, J) \\ \text{and} \quad J_{a+\beta} &= (\check{C}J)_{a+\beta} &&= \operatorname{Hom}_R(R_{Q+a+\beta}, J). \end{aligned}$$

The hypothesis in part 3 says that $R_{Q+\beta} = R_{Q+a+\beta}$, whence multiplication by $\mathbf{x}^a$ is an isomorphism $J_\beta \to J_{a+\beta}$, as required. □

The third property of Proposition 5.4 motivates the following definition:

Definition 5.5. An *essential point for* Q is an element $\varepsilon \in Q^{\mathrm{gp}}$ such that if $a \in Q$ and $(\varepsilon+Q)\cap Q = (a+\varepsilon+Q)\cap Q$, then a is a unit. The *essential set* $\mathcal{E}$ is the Q-set generated by the essential points; i.e., $\mathcal{E}$ is the union $\bigcup(Q+\varepsilon)$ over essential points ε.

Now we derive the main theorem of this section, Theorem 5.8, from combinatorial properties of the essential set, which we develop in detail in Section 6. The key results from that section which we use below are Lemma 6.6, which is an existence result for essential points, and Proposition 6.13, which allows us to control the size of the essential set when Q has simplicial saturation.

Proposition 5.6. *If M is straight, then $M_{\mathcal{E}}$ is an essential submodule of M.*

Proof. Suppose $0 \neq x \in M_\alpha$, and choose an essential point ε with $\varepsilon - \alpha \in Q$ and $(\varepsilon+Q)\cap Q = (\alpha+Q)\cap Q$, using Lemma 6.6. Proposition 5.4 shows that multiplication by $\mathbf{x}^{\varepsilon-\alpha}$ is an isomorphism $M_\alpha \to M_\varepsilon$. Thus any submodule of M containing x contains $\mathbf{x}^{\varepsilon-\alpha}x \in M_{\mathcal{E}}$. □

Observe that the results in this section so far have used no extra hypotheses on the affine semigroup Q. Since our goal involves *simplicial* semigroups, this will change starting now.

Proposition 5.7. *Let Q be an affine semigroup whose saturation is simplicial modulo units, and M a straight module over $k[Q]$. Then the Bass numbers of M are finite.*

Proof. Taking a as in Proposition 6.13, we find that $M_{\mathcal{E}} \subset M_{Q-a}$, so $M_{\mathcal{E}}(-a) \subset (M(-a))_Q$. Since $M(-a)$ is straight, $M(-a)_Q$ is finitely generated. Thus $M_{\mathcal{E}}$ is finitely generated, so M has a finitely generated essential submodule by Proposition 5.6. In particular, its Bass numbers in cohomological degree zero are finite. Moreover, if $J^\bullet$ is a minimal injective resolution of M, then $J^0(-a)$ is straight because it has a Q-graded essential submodule $M(-a)_Q$. Therefore $\mathrm{coker}(M(-a) \to J^0(-a))$ is straight, whence the result follows by induction on the cohomological degree. □

The Bass numbers of such an M at ungraded primes are also finite, by results of [**GW78**].

Theorem 5.8. *Let Q be an affine semigroup whose saturation is simplicial modulo units. If M is a finitely generated Q^{gp}-graded $k[Q]$-module, then the Bass numbers of $H^i_I(M)$ are finite, for any Q-graded ideal I.*

Proof. This is immediate from Proposition 5.7 and Corollary 5.3. □

Corollary 5.9. *Suppose Q is affine, with Q^{sat} simplicial modulo units, and let $M \in \mathcal{M}$ be finitely generated over $R = k[Q]$. Then there exist $\beta \in Q^{\text{gp}}$ and $n \in \mathbb{N}$ such that*

$$H^i_I(M)(-\beta) \cong \check{C}\underline{\text{Ext}}^i_R(R/I^n, M(-\beta)).$$

Proof. Since the Bass numbers of $H^i_I(M)$ are finite, by Remark 3.7 there exists β such that $H^i_I(M)(-\beta)$ is fixed by the Čech hull. Then $H^i_I(M)(-\beta)$ is the Cech hull of its Q-graded part and the result follows from Proposition 4.3. □

6. The essential set.

The essential set, introduced in Definition 5.5, fleshes out in some detail the combinatorics hidden in an affine semigroup. Since we believe this combinatorics is of independent interest, we determine in Theorem 6.2, Proposition 6.4, and the comments in between, the structure of the essential set in the saturated case, along with its relation to Hilbert bases, monomial modules, irrelevant ideals of toric varieties, and Alexander duality. The rest of the section we devote to providing the necessary relations between essential sets of unsaturated semigroups and those of their saturations, including the results already applied in Section 5.

We need a bit of notation. Associated to an affine semigroup Q are its *facets* $F_1, \ldots, F_r$; these are the degrees of homogeneous elements outside of the r codimension-one Q-graded primes of $k[Q]$. There are unique primitive integer-valued linear functionals $\{\tau_1, \ldots, \tau_r\}$ on Q^{gp}, nonnegative on Q, such that $F_i = \{b \in Q \mid \tau_i(b) = 0\}$. Given $\alpha \in Q^{\text{gp}}$, define $\tau(\alpha) \in \mathbb{Z}^r$ to be the vector $(\tau_1(\alpha), \ldots, \tau_r(\alpha))$, and let $\tau(\alpha)^+$ be the vector obtained from $\tau(\alpha)$ by replacing its negative entries with zeros.

Lemma 6.1. *Suppose Q is saturated. Then $(\alpha + Q) \cap Q = (\beta + Q) \cap Q$ if and only if $\tau(\alpha)^+ = \tau(\beta)^+$. In particular, ε is an essential point if and only if $\tau(\varepsilon)^+ \neq \tau(a + \varepsilon)^+$ for all nonunits a in some generating set for Q.*

Proof. Since Q is saturated, $(\alpha + Q) \cap Q$ is the set of lattice points $\gamma \in Q^{\text{gp}}$ inside the polyhderon defined by $\{\tau_i(\gamma) \geq 0 \text{ and } \tau_i(\gamma) \geq \tau_i(\alpha) \mid i = 1, \ldots, r\}$. This is the polyhedron defined by the inequalities $\{\tau_i(\gamma) \geq \tau(\alpha)^+_i \mid i = 1, \ldots, r\}$, and the first claim follows easily.

The map $\tau : Q^{\text{gp}} \to \mathbb{Z}^r$ takes Q to the semigroup $\tau(Q)$ isomorphic to the quotient of Q by its group of units. As a consequence, the definition of essential point translates to: ε is an essential point if and only if $\tau(\varepsilon)^+ \neq \tau(a + \varepsilon)^+$ for all nonunits $a \in Q$. But since τ_i is nonnegative on Q for all i, the second statement follows. □

The lack of nontrivial units in $\tau(Q)$ endows it with a unique minimal set $\mathcal{H}$ of semigroup generators, called the *Hilbert basis of* $\tau(Q)$. Each element of $\mathcal{H}$ imposes a condition that α must satisfy to be essential. To express this condition, define, for $h \in \mathbb{N}^r$, the set $\langle h \rangle = \{\zeta \in \mathbb{Z}^r \mid \zeta_i > -h_i \text{ for some } i \text{ such that } h_i > 0\}$. Observe that $\langle h \rangle$ is a union of half-spaces, and is defined in such a way that $\tau^{-1}(\langle \tau(a) \rangle) = \{\varepsilon \in Q^{\mathrm{gp}} \mid \tau(\varepsilon)^+ \neq \tau(\varepsilon + a)^+\}$.

Theorem 6.2. *If Q is saturated then the essential set $\mathcal{E}$ consists entirely of essential points. Furthermore, $\mathcal{E} = \tau^{-1}(\bigcap_{h \in \mathcal{H}} \langle h \rangle) = \bigcap_{h \in \mathcal{H}} \tau^{-1}(\langle h \rangle)$.*

Proof. The second sentence follows from Lemma 6.1 and the remarks following it because $\tau^{-1}(\mathcal{H})$ generates Q. Since $\langle h \rangle$ is stable under the action of Q, $\tau^{-1}(\langle h \rangle)$ is Q-stable, too. Thus the essential points already form a Q-set, which therefore equals $\mathcal{E}$. □

Example 6.3. We consider once again the semigroup Q generated by three elements x, y, z such that $x + y = 2z$. τ embeds Q in $\mathbb{Z}^2$ by sending x to $(2, 0)$, y to $(0, 2)$, and z to $(1, 1)$; the image of Q^{gp} in $\mathbb{Z}^2$ is the sublattice of index 2 consisting of pairs (u, v) such that u and v have the same parity.

The points $(0, 2)$, $(1, 1)$ and $(2, 0)$ form a Hilbert basis for $\tau(Q)$. $\langle (0, 2) \rangle$ consists of those points (u, v) in $\mathbb{Z}^2$ with $v \geq -1$; similarly $\langle (2, 0) \rangle$ consists of those points with $u \geq -1$. Finally $\langle (1, 1) \rangle$ consists of points with either u or v nonnegative. Thus the essential points are those points $\alpha \in Q^{\mathrm{gp}}$ such that $\tau(\alpha)$ has one of the following forms:

1. $(-1, v)$ for some odd positive v.
2. $(u, -1)$ for some odd positive u.
3. (u, v) for u, v nonnegative and with the same parity.

Figure 1 shows the essential set embedded in $\mathbb{Z}^2$ via τ. The spots (both hollow and solid) represent elements of Q^{gp}; solid spots are essential points. The regions in which $\alpha + Q \cap Q$ remain constant are enclosed by dotted lines. Note in particular that there is an essential point in every region, and that the essential points form a Q-set, as predicted by the theorem.

If we refer back to Example 2.4, we see that the socles of $R^p \check{C}$ computed there lie within the essential set, as Proposition 5.6 and Proposition 5.2 predict. Also note that $z + \mathcal{E} \subset Q$. One will be able to translate $\mathcal{E}$ so that it lies in Q precisely when Q is simplicial; this is the content of Proposition 6.13, which is the central goal of this section.

The essential set $\mathcal{E}$ is related to a number of other notions already playing roles in the study of semigroup algebras and toric varieties. For instance, the subset $\langle \mathcal{H} \rangle := \bigcap_{h \in \mathcal{H}} \langle h \rangle$ of $\mathbb{Z}^r$ is a *monomial module* [**BS98**], so $\mathcal{E}$ might be called a *skew monomial module* inside the lattice Q^{gp}. To get a better picture, $\langle \mathcal{H} \rangle \subset \mathbb{Z}^r$ is a "fuzzy neighborhood" of a certain union $\mathcal{U}$ of orthants, in the sense that there is a vector $z \in \mathbb{N}^r$ such that $\mathcal{U} \subseteq \langle \mathcal{H} \rangle \subseteq \mathcal{U} - z$. In fact,

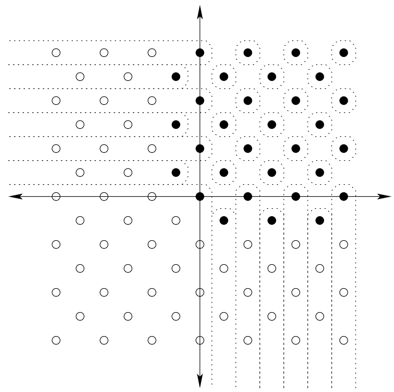


Figure 1. The essential set of the semigroup generated by $(2,0)$, $(1,1)$, and $(0,2)$.

each set $\langle h \rangle$ contains and approximates the union $U_h = \bigcup_{h_i>0}\{\zeta \in \mathbb{Z}^r \mid \zeta_i \geq 0\}$ of half-spaces, and $\mathcal{U} = \bigcap_{h\in\mathcal{H}} U_h$; our z can be any vector with $z_i > h_i$ for all $h \in \mathcal{H}$ and all i.

We can get an even better handle on $\mathcal{U}$ in the case where Q is the cone over $\overline{Q}$, an integral polytope in $\mathbb{Z}^{d-1} \times \{1\} \subset \mathbb{Z}^{d-1} \times \mathbb{Z} = \mathbb{Z}^d$. For each face F of Q, let $\mathrm{van}(F) \subseteq \{1,\ldots,r\}$ be the indices of functionals vanishing on F; similarly, for $z \in \mathbb{N}^r$, let $\mathrm{van}(z) = \{i \in \{1,\ldots,r\} \mid z_i = 0\}$. For instance, if $F = F_i$ is a facet then $\mathrm{van}(F_i) = \{i\}$, and $\mathrm{van}(z) = \mathrm{van}(F)$ if and only if $z_i = 0$ and $z_j > 0$ for all $j \neq i$. The polynomial ring $k[\mathbb{N}^r]$ is the *Cox homogeneous coordinate ring* [**Cox95**] of the projective toric variety X whose isomorphism class and embedding in projective space are determined by $\overline{Q}$. The Cox ring comes equipped with the *irrelevant ideal* $B = \langle \mathbf{x}^z \mid z \in \mathbb{N}^r \text{ and } \mathrm{van}(z) \subseteq \mathrm{van}(F) \text{ for some face } F \text{ of } Q\rangle$.

Proposition 6.4. *If $\check{C}$ is the Čech hull over $\mathbb{N}^r$, then $-\zeta \in \mathcal{U}$ if and only if $\mathbf{x}^\zeta \notin \check{C}(B)$. Equivalently, the $k[\mathbb{N}^r]$-submodule $\langle \mathbf{x}^\zeta \mid \zeta \in \mathcal{U}\rangle \subset k[\mathbb{Z}^r]$ is the shift by $(1,\ldots,1) \in \mathbb{Z}^r$ of the Čech hull $\check{C}(B^\star)$ of the ideal $B^\star$* Alexander dual *to B* [**ER98**], [**MP01**, Lecture VI].

In the case where $\overline{Q}$ is a simple polytope, so the corresponding projective toric variety is simplicial, $B^\star$ is the Stanley-Reisner ideal for the simplicial polytope polar to $\overline{Q}$.

Proof. The equivalence of the two statements is [**Mil98**, Lemma 2.11]. Note that $\check{C}(B)$ is, a priori, a submodule of $k[\mathbb{Z}^r]$ since the latter is the injective hull of B and the former is an essential extension. Now $\mathbf{x}^\zeta \in \check{C}(B)$ if and only if $\mathrm{van}(\zeta^+) \subseteq \mathrm{van}(F)$ for some face F by Remark 2.3. On the other hand, $-\zeta \in U_h$ precisely when $\zeta_i \leq 0$ for some i with $h_i > 0$; that is, when $\mathrm{van}(\zeta^+) \not\subseteq \mathrm{van}(h)$. Therefore, $-\zeta \notin \mathcal{U}$ if and only if $\mathrm{van}(\zeta^+) \subseteq \mathrm{van}(h)$ for some $h \in \mathcal{H}$. This occurs if and only if $\mathrm{van}(\zeta^+) \subseteq \mathrm{van}(F)$ for some face F, because: $\mathrm{van}(F) \subseteq \mathrm{van}(h)$ for all $h \in F$; and each $h \in \mathcal{H}$ lies in the relative interior of some F, so $\mathrm{van}(h) = \mathrm{van}(F)$ for this F. We conclude that $\mathbf{x}^\zeta \in \check{C}(B)$ if and only if $-\zeta \notin \mathcal{U}$. □

Example 6.5. We illustrate this for the semigroup Q of Example 4.10; that is, the semigroup on four generators $\{x, y, u, v\}$ with the relation $x+u = y+v$. If we order the four facets appropriately, τ embeds Q^{gp} in $\mathbb{Z}^4$ by sending x to $(1,0,0,1)$, y to $(0,1,0,1)$, u to $(0,1,1,0)$ and v to $(1,0,1,0)$. The image $\tau(Q^{\mathrm{gp}})$ is the lattice consisting of points (a,b,c,d) with $a+c = b+d$.

Now $\langle(1,0,0,1)\rangle$ is the set of (a,b,c,d) in $\mathbb{Z}^4$ such that either a or d is nonnegative. Thus, if $\alpha \in Q^{\mathrm{gp}}$ is an essential point, with $\tau(\alpha) = (a,b,c,d)$, then either a or d is nonnegative. Similarly, using the other elements of the Hilbert basis, we find that:

- Either b or d is nonnegative.
- Either b or c is nonnegative.
- Either a or c is nonnegative.

Note that in this example, the irrelevant ideal B of Proposition 6.4 is generated by the elements $\mathbf{x}^{(1,0,0,1)}, \mathbf{x}^{(0,1,0,1)}, \mathbf{x}^{(0,1,1,0)}$, and $\mathbf{x}^{(1,0,1,0)}$ in the polynomial ring $k[\mathbb{N}^4]$. The Čech hull of this ideal is thus supported precisely on those $(a,b,c,d) \in \mathbb{Z}^4$ such that at least one of the pairs $\{a,d\}$, $\{b,d\}$, $\{b,c\}$, $\{a,c\}$ consists of strictly positive integers. Therefore, an element of $\mathbb{Z}^4$ fails to be in this support if and only if its negative satisfies the above four conditions. To summarize, those elements α of Q^{gp} such that $-\tau(\alpha)$ is not in the support of $\check{C}B$ are essential points, as Proposition 6.4 predicts.

The conditions on $\tau(\alpha)$ given above, together with the fact that $(a,b,c,d) = \tau(\alpha)$ (and therefore $a+b = c+d$), imply that α is an essential point if (and only if) at most one of $\{a,b,c,d\}$ is negative. Note that no finite shift will take all of the essential points inside of Q, since one has essential points α whose negative coordinate is $-n$ for any natural number n. In particular, the degrees of the socle elements of $H^2_{(x,u)}(\omega_R)$ produced in Example 4.10 are a set of essential points whose negative coordinates are unbounded below.

More generally, the fact that we cannot shift the essential set into Q means that we cannot rule out the possibility of local cohomology having infinite Bass numbers. In fact, we will construct a local cohomology module with infinite Bass numbers whenever the essential set cannot be shifted into Q (Corollary 7.5).

During our proof of Theorem 5.8, we needed certain results about the structure of $\mathcal{E}$. In order to obtain them in the generality we used in Section 5, we no longer assume that Q is saturated. We begin with a lemma used in the proof of Proposition 5.6.

Lemma 6.6. *Given $\alpha \in Q^{\mathrm{gp}}$, there exists some essential point ε with $(\varepsilon + Q) \cap Q = (\alpha + Q) \cap Q$.*

Proof. Any element $\beta \in Q^{\mathrm{gp}}$ satisfying $(\beta + Q) \cap Q = (\alpha + Q) \cap Q$ must also satisfy $\tau(\beta) \preceq \tau(\gamma)$ for all $\gamma \in (\alpha + Q) \cap Q$, where $\preceq$ is the partial order by componentwise comparison. The set of possibilities for $\tau(\beta) \in \mathbb{Z}^r$ satisfying this condition is bounded above, and thus has a maximal element $\tau(\varepsilon)$. Moreover, $\tau(\beta) = \tau(\varepsilon)$ for some $\beta \in Q^{\mathrm{gp}}$ if and only if $\beta - \varepsilon$ is a unit of Q. This proves that ε is an essential point. □

The major combinatorial result used in the previous section is the fact that if Q^{sat} is (modulo its units) simplicial, then $\mathcal{E}$ can be shifted inside of Q. Therefore, we want an analog to Lemma 6.1 which holds even for semigroups which are not saturated. The key tool relating the combinatorics of a semigroup to the combinatorics of its normalization is provided by the next lemma. Recall that a *face* of Q is the set of degrees of elements outside a prime ideal of $k[Q]$.

Lemma 6.7. *Let F be a face of Q. There exists $a_F \in F$ such that $a_F + Q^F \subset Q$, where $Q^F := (Q + F^{\mathrm{gp}}) \cap Q^{\mathrm{sat}}$ is the* partial saturation of Q at F.

Proof. Let $R' = k[Q^F]$ and $\widetilde{R} = k[Q^{\mathrm{sat}}]$. Then, letting $\mathfrak{p} \subset R = k[Q]$ be the prime ideal such that $R/\mathfrak{p} = k[F]$, the R-algebra R' is the intersection $R_{(\mathfrak{p})} \cap \widetilde{R}$ of the homogeneous localization at $\mathfrak{p}$ with the normalization. The Lemma calls for a homogeneous element outside of $\mathfrak{p}$ to be in the *conductor ideal*

$$\mathrm{ann}_R(R'/R) = \{x \in R \mid xR' \subset R\}.$$

Such an element exists precisely when $\mathrm{ann}_R(R'/R)_{(\mathfrak{p})} = R$; i.e., when the localizations $R_{(\mathfrak{p})}$ and $R' \otimes_R R_{(\mathfrak{p})}$ are equal. But

$$\begin{aligned} R' \otimes_R R_{(\mathfrak{p})} &= R_{(\mathfrak{p})} \cap (\widetilde{R} \otimes_R R_{(\mathfrak{p})}) \\ &= R_{(\mathfrak{p})} \cap \widetilde{R}_{(\mathfrak{p}\widetilde{R})} \\ &= R_{(\mathfrak{p})} \end{aligned}$$

because $R_{(\mathfrak{p})} \subseteq \widetilde{R}_{(\mathfrak{p}\widetilde{R})}$. □

Remark 6.8. When $F = Q$, then $Q^F = Q^{\mathrm{sat}}$, and this is the well-known fact that every semigroup Q contains an element a with $a + Q^{\mathrm{sat}} \subset Q$.

With Lemma 6.7 in hand, we can now find a sufficient condition under which $(\beta + Q) \cap Q = (a + \beta + Q) \cap Q$. Let $\mathrm{van}_\tau(F) \subseteq \{\tau_1, \ldots, \tau_r\}$ be the subset consisting of functionals vanishing on the face F of Q.

Lemma 6.9. *Let $a \in F$, and suppose we have $\beta \in Q^{\mathrm{gp}}$ such that $\tau_i(a + a_F + \beta) \leq 0$ for all $\tau_i \notin \mathrm{van}_\tau(F)$. Then $(\beta + Q) \cap Q = (a + \beta + Q) \cap Q$.*

Proof. Noting that $\tau_j(a) = \tau_j(a_F) = 0$ for $\tau_j \in \mathrm{van}_\tau(F)$, the hypothesis on $a + a_F + \beta$ implies that the intersections with Q are contained in $(a + a_F + \beta + Q^{\mathrm{sat}})$. Therefore, it is enough to show that

$$(\beta + Q) \cap (a + a_F + \beta + Q^{\mathrm{sat}}) = (a + \beta + Q) \cap (a + a_F + \beta + Q^{\mathrm{sat}}).$$

This follows by adding β or $a+\beta$ to both sides of the equality in Lemma 6.10, below, and setting respectively $b = a + a_F$ or $b = a_F$. □

Lemma 6.10. *If $b \in F$ and $b + Q^F \subseteq Q$, then $Q \cap (b + Q^{\mathrm{sat}}) = b + Q^F$.*

Proof. We show $Q \cap (b+Q^{\mathrm{sat}}) = (b+Q^F) \cap (b+Q^{\mathrm{sat}})$, which obviously equals $b + Q^F$. Now $Q \cap (b + Q^{\mathrm{sat}}) \supseteq (b + Q^F) \cap (b + Q^{\mathrm{sat}})$, because Q contains $b+Q^F$; and $Q \cap (b+Q^{\mathrm{sat}}) \subseteq (b+Q^F) \cap (b+Q^{\mathrm{sat}})$, because $a - b \in Q^F$ when $a \in Q \cap (b + Q^{\mathrm{sat}})$, by definition of Q^F. □

We are now in a position to state and prove the unsaturated analog of Lemma 6.1.

Proposition 6.11. *Choose a_Q so that $\tau_i(a_Q) \geq \tau_i(a_F)$ for all i and F. Suppose $a \in Q$, $\beta \in Q^{\mathrm{gp}}$, and $\tau(a_Q + \beta)^+ = \tau(a + a_Q + \beta)^+$. Then $(\beta + Q) \cap Q = (a + \beta + Q) \cap Q$.*

Proof. Let F be the smallest face of Q containing a. Then for all τ_i not vanishing on F, we have $\tau_i(a) > 0$, so $\tau_i(a + a_Q + \beta) \leq 0$ (as otherwise the i^{th} coordinates of $\tau(a_Q + \beta)^+$ and $\tau(a + a_Q + \beta)^+$ are unequal). Thus $\tau_i(a + \beta) \leq \tau_i(-a_Q) \leq -\tau_i(a_F)$, and by Lemma 6.9 we have $(\beta + Q) \cap Q = (a + \beta + Q) \cap Q$, as required. □

The approximation to Theorem 6.2 in the unsaturated case is as follows:

Corollary 6.12. *Let $\mathcal{H}$ be the Hilbert basis for $\tau(Q)$, and a_Q be as in Proposition* 6.11*. Then $\mathcal{E} + a_Q \subseteq \bigcap_{h\in\mathcal{H}} \tau^{-1}(\langle h\rangle)$.*

Proof. Pick, for each $h \in \mathcal{H}$, an element $q_h \in Q$ with $\tau(q_h) = h$. Suppose ε is an essential point. Setting $\beta = \varepsilon$ and $a = q_h$ in Proposition 6.11, we have $\tau(\varepsilon + a_Q)^+ \neq \tau(\varepsilon + a_Q + q_h)^+$. Just as before Theorem 6.2, we have $\tau(\varepsilon + a_Q) \in \langle h\rangle$, and this holds for all $h \in \mathcal{H}$. □

Proposition 6.13. *Suppose Q^{sat} is simplicial* (*modulo units*)*. Then there exists $a \in Q$ such that $a + \mathcal{E} \subset Q$.*

Proof. The hypothesis on Q means precisely that for each $i = 1, \ldots, r$, the image $\tau(Q)$ contains an element $h^i \in \mathcal{H}$ in its Hilbert basis whose unique nonzero coordinate is $h^i_i > 0$. Observe that $\langle h^i\rangle = \{\zeta \in \mathbb{Z}^r \mid \zeta_i > -h^i_i\}$ is a half-space by definition. Setting $\mathbf{h} = (h^1_1, \ldots, h^r_r) \in \mathbb{Z}^r$, we find that $\mathbf{h}+\bigcap_{h\in\mathcal{H}}\langle h\rangle \subseteq \mathbf{h}+\bigcap_{i=1}^r\langle h^i\rangle \subseteq \mathbb{N}^r$. By Corollary 6.12 we may take $a = a_Q+\widetilde{h}$ for any $\widetilde{h} \in Q$ satisfying $\tau(\widetilde{h}) = \mathbf{h}$. □

7. Infinite-dimensional socles.

In this section we prove our principal result concerning semigroup rings, Theorem 7.1, by combining Theorem 5.8 with its converse, namely that if Q^{sat} is not simplicial then one can always find an ideal I for which local cohomology is not well-behaved. We avoid dealing with nontrivial units here, since they add nothing to the content, but obscure the statement.

Theorem 7.1. *Let Q be an affine semigroup of dimension d with trivial unit group (but not necessarily saturated). The following are equivalent:*

1. *The saturation Q^{sat} is simplicial.*
2. *For every Q-graded ideal I and every finitely generated Q-graded $k[Q]$-module M, the Bass numbers of $H^i_I(M)$ are finite.*
3. *For every Q-graded prime $\mathfrak{p}$ of dimension 2, $H^{d-1}_{\mathfrak{p}}(\omega_{k[Q^{\mathrm{sat}}]})$ has finitely generated socle.*

This theorem provides a proof and generalization of Example 4.10. The key to our argument is Yanagawa's computation of the local cohomology of the canonical module $\omega_{k[Q]}$ over a normal semigroup ring $k[Q]$ [**Yan01**]. To state it, let $\tau_1, \ldots, \tau_r$ be linear functionals which vanish on the facets $F_1, \ldots, F_r$ of Q and take nonnegative integer values on Q, as in the previous sections. For the sake of simplicity we assume that Q has no nonzero units. Choose a hyperplane H transverse to the real cone $\mathbb{R}_+Q$ generated by Q, so that $\overline{Q} = (\mathbb{R}_+Q) \cap H$ is a polytope of dimension $d-1 = \dim(k[Q]) - 1$ whose faces (including the empty face $\varnothing$) correspond to the primes of $k[Q]$.

Definition 7.2. Let $F \in Q$ correspond to $\overline{F} \in \overline{Q}$ (so $\mathbf{0} \in Q$ corresponds to $\varnothing \in \overline{Q}$, for example). Define the polyhedral cell subcomplex

$$\overline{F}(\alpha) = \{\overline{F'} \in \overline{F} \mid (\alpha + \mathbb{R}_+Q) \cap F' = \varnothing\}$$

of $\overline{F}$ for any face $\overline{F} \subseteq \overline{Q}$ and $\alpha \in Q^{\mathrm{gp}}$.

Theorem 7.3 ([**Yan01**, Theorem 6.1])**.** *Let Q be saturated and $\mathfrak{p}$ be a graded prime of $k[Q]$, corresponding to a face $\overline{F}$ of $\overline{Q}$. Then $H^{d-i}_{\mathfrak{p}}(\omega_{k[Q]})_\alpha \cong \widetilde{H}_{i-1}(\overline{F}, \overline{F}(\alpha))$ for all $\alpha \in Q^{\mathrm{gp}}$.*

We will apply this when $\mathfrak{p}$ has dimension 2; that is, when $\overline{F}$ is an edge of $\overline{Q}$.

Proposition 7.4. *If $\mathfrak{p}$ corresponds to the edge $\overline{F}$ and Q is saturated, then:*

1. $H^{d-1}_{\mathfrak{p}}(\omega_{k[Q]})_\alpha = 0$ *if $\tau_i(\alpha) > 0$ for some i such that $\overline{F}_i \cap \overline{F} \neq \varnothing$.*
2. $H^{d-1}_{\mathfrak{p}}(\omega_{k[Q]})_\alpha = 0$ *if $\tau_i(\alpha) \leq 0$ for all i such that $\overline{F}_i \cap \overline{F} = \varnothing$.*
3. $H^{d-1}_{\mathfrak{p}}(\omega_{k[Q]})_\alpha = k$ *if neither of the above conditions holds.*

Proof. Suppose the first condition holds. If $\overline{F}_i$ contains $\overline{F}$, then $\alpha + \mathbb{R}_+Q$ misses F entirely, so $\overline{F}(\alpha) = \overline{F}$, and the zeroth relative homology is zero.

Otherwise, $\overline{F}_i \cap \overline{F}$ is a vertex of the edge $\overline{F}$, and $\overline{F}(\alpha)$ contains at least that vertex. Thus the relative homology is again zero.

If the second condition holds, but the first does not, then $\tau_i(\alpha) \leq 0$ for all i. This implies $\overline{F}(\alpha)$ is the void complex—not even $\varnothing \in \overline{F}(\alpha)$, so the zeroth relative homology is still zero.

In the third case, $\overline{F}(\alpha)$ consists of just the empty face $\varnothing$, and the zeroth relative homology is the number of connected components of $\overline{F}$. ☐

Corollary 7.5. *Let $\overline{F}$ be an edge of $\overline{Q}$ such that there exists a facet $\overline{F}_j$ of $\overline{Q}$ with $\overline{F}_j \cap \overline{F} = \varnothing$. Let $\mathfrak{p}$ be the prime of $k[Q]$ corresponding to $\overline{F}$. Then if Q is saturated, $H^{d-1}_{\mathfrak{p}}(\omega_{k[Q]})$ has an infinite-dimensional socle.*

Proof. Every nonzero element $x \in H^{d-1}_{\mathfrak{p}}(\omega_{k[Q]})$ is annihilated by some power of the maximal ideal of $k[Q]$. To see this, suppose x is homogeneous of degree α, and assume that for some $\beta \in Q$, we had $\mathbf{x}^{n\beta}x \neq 0$ for all $n \in \mathbb{N}$. Then, by Proposition 7.4, $\tau_i(n\beta + \alpha) \leq 0$ for all n and all i such that $\overline{F}_i \cap \overline{F} \neq \varnothing$. Thus $\tau_i(\beta) = 0$ for all such i, so $\beta \in \overline{F}_i$ for all such i. But the intersection of all such $\overline{F}_i$ is empty, so $\beta = \mathbf{0}$.

If $\overline{F}_j \cap \overline{F} = \varnothing$, then choose $\alpha \in Q^{\mathrm{gp}}$ such that $\tau_j(\alpha) > 0$ and $\tau_i(\alpha) \leq 0$ for $i \neq j$. Proposition 7.4 implies that $H^{d-1}_{\mathfrak{p}}(\omega_{k[Q]})_\alpha$ is nonzero and killed by some power of the maximal ideal, so $H^{d-1}_{\mathfrak{p}}(\omega_{k[Q]})$ has nontrivial socle. Suppose its socle were finite-dimensional. Then there would exist $\beta \in Q^{\mathrm{gp}}$ such that $\tau_j(\beta)$ is maximal among the socle degrees in Q^{gp}. But since $\tau_j(\alpha) > 0$, we have $\tau_j(\beta) > 0$, so $\tau_j(2\beta) > \tau_j(\beta)$. Moreover the local cohomology is nontrivial in degree 2β. Taking a nonzero element of $H^{d-1}_{\mathfrak{p}}(\omega_{k[Q]})_{2\beta}$ and multiplying it by a sufficiently large power of the maximal ideal then yields a socle element in a degree γ with $\tau_j(\gamma) > \tau_j(\beta)$, which is a contradiction. ☐

Proof of Theorem 7.1. $1 \Rightarrow 2$ is Theorem 5.8, and $2 \Rightarrow 3$ is because $\omega_{k[Q^{\mathrm{sat}}]}$ is finitely generated over $k[Q]$. For $3 \Rightarrow 1$, the unsaturated case follows from the saturated case. Indeed, any Q^{sat}-graded ideal $I \subset k[Q^{\mathrm{sat}}]$ is generated up to radical by elements $\mathbf{y} = (y_1, \ldots, y_s)$ in $k[Q]$ (high powers of any homogeneous generating set for I will do). If M is any $k[Q^{\mathrm{sat}}]$-module, the cohomology of the Čech complex $C^\bullet(\mathbf{y}; M)$ on these generators is therefore a module over both $k[Q^{\mathrm{sat}}]$ and $k[Q]$. As such, it is simultaneously the local cohomology of M over $k[Q^{\mathrm{sat}}]$ with support on $I \subset k[Q^{\mathrm{sat}}]$ and over $k[Q]$ with support on $I \cap k[Q]$. Furthermore, any socle element of a $k[Q^{\mathrm{sat}}]$-module is also a socle element over $k[Q]$, since the maximal ideal of $k[Q]$ is contained in the maximal ideal of $k[Q^{\mathrm{sat}}]$.

Thus by Corollary 7.5, it suffices to produce, for any polytope $\overline{Q} \neq$ simplex, an edge $\overline{F}$ of $\overline{Q}$ that misses some facet. Equivalently, it suffices to show that if $\overline{Q}$ is a polytope in which every edge meets every facet then $\overline{Q}$ is a simplex. Let $\overline{F} \in \overline{Q}$ be a facet, and τ a linear functional supporting $\overline{F}$, nonnegative on $\overline{Q}$. Suppose τ takes a minimal nonzero value at a vertex

$v \notin \overline{F}$. If more than one vertex of $\overline{Q}$ lies off of $\overline{F}$, there is an edge (necessarily missing $\overline{F}$) connecting v to some vertex at which $\tau > 0$. Thus, if every edge meets every facet, there can be only one vertex of $\overline{Q}$ lying off of each facet, and $\overline{Q}$ must be a simplex. □

8. Open problems.

It has been seen above that affine semigroup rings provide a wealth of examples and counterexamples to general questions about local cohomology in singular varieties. In particular, they shed some light on some of the general questions posed by Huneke on local cohomology [**Hun92**]:

1. When is $H^i_I(M)$ zero?
2. When is $H^i_I(M)$ finitely generated?
3. When is $H^i_I(M)$ artinian?
4. When is the number of associated primes of $H^i_I(M)$ finite?

Although the answer to the fourth is trivially "always" in the cases discussed in this paper, the above examples provide clues as to how to refine the first three, given a grading.

Section 4 provides a possibility for answering Question 1: Relate the vanishing of local cohomology in a given cohomological degree to the vanishing of Ext modules in that cohomological degree and lower. Theorem 4.7 establishes this link for graded modules over semigroup rings; we believe that such a connection exists in significantly more generality, but we are unaware of how to relate infinitely generated modules to finitely generated ones without resorting to a grading. The key concept is that of a certain kind of "constancy", provided here by the Čech hull. This type of constancy is reminiscent of the characteristic 0 regular local case, in which the modules in question are treated as D-modules [**Lyu93**]. Perhaps the right generalization of D-module to the singular setting will provide the appropriate notion of constancy to bridge finitely generated Ext modules and local cohomology.

A partial answer to Question 2 in the general (local, ungraded) case concerns numerical criteria on the heights of primes and cohomological degrees involved [**Hun92**]. In the semigroup-graded case, finite generation can be viewed as a convex-geometric problem, dealing with Q-graded degrees in which the summands in a minimal injective resolution are nonzero. We expect in the Q-graded case for these considerations to yield geometric and combinatorial criteria in addition to the general numerical criteria. For the canonical module of a normal semigroup ring, for instance, local cohomology at a graded prime ideal $\mathfrak{p}$ of R is finitely generated if and only if it vanishes, since Proposition 4.9 expresses such cohomology in terms of derived functors of $\check{C}$, which are never finitely generated if they are nonzero, or in terms of $\check{C}\omega_{R/\mathfrak{p}}$, which is also never finitely generated.

As pointed out by Huneke [**Hun92**], Question 3 has two parts, namely:

3a. When is the maximal ideal the only associated prime of $H^i_I(M)$?
3b. When are the Bass numbers of $H^i_I(M)$ finite?

Both 3a and 3b should have concrete combinatorial answers in the semigroup case, at least when M is a canonical module. In fact, we expect the essential set to play a pivotal role in answering these and the following refinement of 3a: For which cohomological degrees i and graded ideals I is a given prime of $k[Q]$ associated to $H^i_I(\omega_{k[Q]})$?

As for Question 3b, it seems to be connected with the kinds of singularities which appear in the normalization of the ring R. Whether this holds in more generality than simply for semigroup rings is an interesting question. For instance, one can try classifying the singularities of a ring R or ideals I for which the modules $H^i_I(M)$ can have infinite Bass numbers (or, for that matter, which primes can appear with infinite Bass number). Even in the case of a semigroup ring, we do not have satisfactory answers to these last questions.

Finally, is there a global version of the Čech hull that works for toric varieties, and if so, what is its relation to the Čech hull over the Cox homogeneous coordinate ring [**Cox95**]? More generally, for varieties with a torus action, can a global Čech hull give information about cohomology with support on subvarieties fixed pointwise by subgroups of the acting torus? In the toric case, properties of a global Čech hull will be governed by the group of Weil divisors modulo Cartier divisors, introduced by Thompson to control resolutions of singularities [**Tho01**].

Acknowledgements. The authors wish to thank Bernd Sturmfels and Kohji Yanagawa for their useful suggestions. Both authors were partially supported by the National Science Foundation, and the second was partially supported by the Alfred P. Sloan Foundation.

References

[BS98] D. Bayer and B. Sturmfels, *Cellular resolutions of monomial modules*, J. Reine Angew. Math., **502** (1998), 123-140, MR 99g:13018, Zbl 0909.13011.

[BH93] W. Bruns and J. Herzog, *Cohen-Macaulay Rings*, Cambridge Studies in Advanced Mathematics, **39**, Cambridge University Press, Cambridge, 1993, MR 95h:13020, Zbl 0788.13005.

[Cox95] D. Cox, *The homogeneous coordinate ring of a toric variety*, J. Algebraic Geom., **4** (1995), 17-50, MR 95i:14046, Zbl 0846.14032.

[ER98] J.A. Eagon and V. Reiner, *Resolutions of Stanley-Reisner rings and Alexander duality*, J. Pure Appl. Algebra, **130**(**3**) (1998), 265-275, MR 99h:13017, Zbl 0941.13016.

[EMS00] D. Eisenbud, M. Mustaţă and M. Stillman, *Cohomology on toric varieties and local cohomology with monomial supports*, J. Symbolic Comput., **29**(**4-5**) (2000), 583-600, Symbolic computation in algebra, analysis, and geometry (Berkeley, CA, 1998), MR 2002a:14059.

[GW78] S. Goto and K. Watanabe, *On graded rings,* II ($\mathbb{Z}^n$*-graded rings*), Tokyo J. Math., **1**(**2**) (1978), 237-261, MR 81m:13022, Zbl 0406.13007.

[Har70] R. Hartshorne, *Affine duality and cofiniteness*, Invent. Math., **9** (1970), 145-164, MR 41 #1750, Zbl 0196.24301.

[Hun92] C. Huneke, *Problems on local cohomology*, in 'Free Resolutions in Commutative Algebra and Algebraic Geometry' (Sundance, UT, 1990), Jones and Bartlett, Boston, MA, 1992, 93-108, MR 93f:13010, Zbl 0782.13015.

[HS93] C. Huneke and R. Sharp, *Bass numbers of local cohomology modules*, Trans. Amer. Math Soc., **339** (1993), 765-779, MR 93m:13008, Zbl 0785.13005.

[Ish88] M.-N. Ishida, *The local cohomology groups of an affine semigroup ring*, in 'Algebraic Geometry and Commutative Algebra in Honor of Masayaoshi Nagata', Vol. I, Kinokuniya, Tokyo, 1988, 141-153, MR 90a:13029, Zbl 0687.14002.

[Lyu93] G. Lyubeznik, *Finiteness properties of local cohomology modules* (*an application of D-modules to commutative algebra*), Invent. Math., **113**(**1**) (1993), 41-55, MR 94e:13032, Zbl 0795.13004.

[Mil98] E. Miller, *Alexander duality for monomial ideals and their resolutions.* Preprint math.AG/9812095, 1998.

[Mil00] ———, *The Alexander duality functors and local duality with monomial support*, J. Algebra, **231**(**1**) (2000), 180-234, MR 2001k:13028, Zbl 0968.13009.

[MP01] E. Miller and D. Perkinson, *Eight lectures on monomial ideals*, in 'COCOA VI: Proceedings of the International School', Villa Gualino, May-June, 1999 (A.V. Geramita, ed.), Queens Papers in Pure and Applied Mathematics, **120**, Queen's Univ., Kingston, ON, 2001, 3-105.

[Mus00] M. Mustaţă, *Local cohomology at monomial ideals*, J. Symbolic Comput., **29**(**4-5**) (2000), 709-720, MR 2001i:13020, Zbl 0966.13010.

[Ter99] N. Terai, *Local cohomology modules with respect to monomial ideals.* Preprint, 1999.

[Tho01] H. Thompson, *On Toric Log Schemes*, Ph.D. Thesis, University of California at Berkeley, 2002.

[Wei94] C.A. Weibel, *An Introduction to Homological Algebra*, **38**, Cambridge University Press, Cambridge, 1994, MR 95f:18001, Zbl 0797.18001.

[Yan00] K. Yanagawa, *Bass numbers of local cohomology modules with supports in monomial ideals*, Math. Proc. Cambridge Phil. Soc., **131**(**1**) (2001), 45-60, MR 2002c:13037.

[Yan01] ———, *Sheaves on finite posets and modules over normal semigroup rings*, J. Pure Appl. Algebra, **161**(**3**) (2001), 341-366, MR 2002k:13033, Zbl 0992.13005.

Received September 14, 2001 and revised November 8, 2001.

Department of Mathematics
University of California
Berkeley CA 94720

E-mail address: dhelm@math.berkeley.edu

Mathematical Sciences Research Institute
1000 Centennial Drive
Berkeley CA 94720
E-mail address: emiller@msri.org

PACIFIC JOURNAL OF MATHEMATICS
Vol. 209, No. 1, 2003

EXAMPLES OF BIREDUCIBLE DEHN FILLINGS

James A. Hoffman and Daniel Matignon

If an irreducible manifold M admits two Dehn fillings along distinct slopes each filling resulting in a reducible manifold, then we call these *bireducible* Dehn fillings. The first example of bireducible Dehn fillings is due to Gordon and Litherland. More recently, Eudave-Muñoz and Wu presented the first infinite family of manifolds which admit bireducible Dehn fillings. We present another infinite family of hyperbolic manifolds which admit bireducible Dehn fillings. The manifolds obtained by the fillings are always the connect sum of two lens spaces.

0. Introduction.

Let M be an orientable 3-manifold with toroidal boundary T. Given a slope r on T, the *Dehn filling* of M along r, denoted by $M(r)$, is the manifold obtained by identifying T with the boundary of a solid torus V so that r bounds a meridian disk in V.

In this paper, we are especially interested in those Dehn fillings which produce reducible manifolds. Recall that a manifold is *reducible* if it contains an essential 2-sphere, that is, a 2-sphere which does not bound a 3-ball. If an irreducible manifold M admits two Dehn fillings along distinct slopes each filling resulting in a reducible manifold, then we call these *bireducible* Dehn fillings.

The first example of bireducible Dehn fillings is due to Gordon and Litherland [**GLi**]. More recently, Eudave-Muñoz and Wu [**EW**] presented the first infinite family of manifolds with admit bireducible Dehn fillings. They show that for each $p \neq 0$ there is a hyperbolic manifold M_p such that $M_P(\infty) \cong Q(2,-2)\#RP^3$ and $M_P(0) \cong Q(2p,-2p)\#RP^3$, where $Q(r,s)$ is the double branched cover of a Montesinos tangle $T[r,s]$.

We present another infinite family of hyperbolic manifolds which admit bireducible Dehn fillings, all with exactly one toroidal boundary component. They represent counterexamples to the generalization of the Cabling-Conjecture [**GS**] since they are hyperbolic. Notice that examples with a single boundary component can be constructed from the examples given in [**EW**] (for more details, see the end of Section 4).

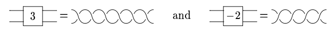
Theorem 1. *There exists an infinite family of hyperbolic manifolds which admit bireducible Dehn fillings. More precisely, there exist families of hyperbolic manifolds M_2^t, M_3^t and M_4^t, parameterized by an integer t, such that:*

a) $M_2^t(\infty) \cong L(-2,1)\#L(-4,1)$ *and*
$M_2^t(t) \cong L(2,1) \#L(t^2-2t+1, t-2)$ *for* $t \neq 0,1,2$,
b) $M_3^t(\infty) \cong L(-3,1)\#L(-3,1)$ *and*
$M_3^t(t) \cong L(3,1) \#L(t^2-t+1, t-1)$ *for* $t \neq 0,1$,
c) $M_4^t(\infty) \cong L(-4,1)\#L(-2,1)$ *and*
$M_4^t(t) \cong L(4,1) \#L(t^2+1, t)$ *for* $t \neq 0$.

Note that in each instance, the manifold resulting from the Dehn filling is the connect sum of two lens spaces. The lens space $L(p,q)$ is the manifold obtained by performing p/q-Dehn surgery on the unknot. The restrictions on the parameter t are there to account for cases where either the resultant manifold is not reducible (i.e., one of the summands is $L(1,n) \cong S^3$), or one summand of the resultant manifold is not a lens space. The latter occurs in Case (a) with $t = 1$. Here we get the summand $L(0,-1) \cong S^2 \times S^1$. This case is also uninteresting since the manifold (before the Dehn filling) is reducible.

1. Surgery instructions.

In this section, we show how to construct a family of manifolds with one toroidal boundary component. Begin with the five component link $L \subset S^3$ shown in Figure 1 having components A, B, C, D and K. The box labeled t represents t full twists. Positive values represent right-handed twists; and, negative values represent left-handed twists. For example,



Define $M^t_{(a,b,c,d)}$ to be the manifold obtained by removing a regular neighborhood of K and performing Dehn surgery on the components A, B, C and D along the respective slopes a, b, c and d. The parameter t represents the number of twists between the components K and C (as shown in Figure 1). In particular, we are interested in the families of manifolds $M^t_{(-1,-2,1,2)}$, $M^t_{(-1,-2,2,1)}$, $M^t_{(-2,-1,1,2)}$ and $M^t_{(-2,-1,2,1)}$ parametrized by the nonzero integer t.

Also, let $T = \partial M^t_{(a,b,c,d)}$ be the boundary torus. If r is a slope in T, then define $M^t_{(a,b,c,d)}(r)$ to be the manifold obtained by performing a Dehn filling along r.

2. Bireducibility.

Here we show that there are two slopes, namely $\infty = \frac{1}{0}$ and t, in the boundary of $M^t_{(a,b,c,d)}$ such that the Dehn fillings, $M^t_{(a,b,c,d)}(\infty)$ and $M^t_{(a,b,c,d)}(t)$, are both reducible manifolds. Moreover, the resulting summands are all lens spaces.

The following proofs use the link-calculus of 3-manifolds as described in Chapter 9H of [**R**]. The proofs consist of a series of link diagrams with accompanying surgery coefficients. Each transition between diagrams is either an isotopy or a twisting about an unknotted component. In order to simplify the statements of the following claims, we will consider $L(1,n) \cong S^3$ and $L(0,\pm 1) \cong S^2 \times S^1$ as lens spaces. This inclusion applies only to this section of the paper. Let us first consider the manifolds for which $a = -1$, $b = -2$.

Claim 1. $M^t_{(-1,-2,c,d)}(t) \cong L(t^2 + (1-c)t + 1, t + 1 - c)\# L(d+2, 1)$.

Proof. We refer the reader to Figure 2. In the first transition, component D is isotoped so that components B and D "pass through" component A in like fashion. This is done to facilitate the twisting of component A in the second transition.

In transition 2, we perform a single positive twist about component A. Since A now has a trivial surgery coefficient, it is removed from the diagram. Note that the surgery coefficients increase for components B and D as they link component A. In transition 3, we perform a single positive twist about component B. Again, the component is removed as it has a trivial surgery coefficient; and, the surgery coefficients for components C and D are increased.

Transition 4 introduces a new component F linking components C and K, and gives it $-t-1$ twists (to unwind the twisting of components C and K). Thus the surgery coefficient of F is $\frac{1}{-t-1}$. This new component is temporary and simplifies the diagram for the next transition. Moreover, we achieve the desired effect that the surgery coefficient of K is now -1. At this point, component C and D have, respectively, surgery coefficients $c - t$ and $d + 2$.

In transition 5, we perform a single positive twist about component K, giving it a trivial surgery coefficient. We will keep K in the diagram so that we may see how it lies in the resulting manifold. Note that the surgery coefficient of component C is unchanged as it has linking number zero with component K. The surgery coefficient of F is increased by the twist.

Transition 6 is an isotopy of component C. In transition 7, components D, F and K are isotoped to facilitate a twisting about component C.

In transition 8, we perform a single negative twist about component C. This is done so the coefficient of component F once again represents $-t-1$ twists. Transitions 9 and 10 are isotopies of component D in an attempt to separate components C and D. In transition 9, the lower arc of D is flyped to the top of the diagram. Also, a lower loop of D is untwisted. In transition 10, component D is pulled taut at the expense of twisting component K.

Transition 11 is an isotopy of components F and C. Component C is pulled down. We also localize component F so that, in transition 12, we may perform $t+1$ twists about F.

The final diagram shows us the knot K with trivial surgery coefficient. If we disregard K, we see two unknotted and unlinked components. Each component, C and D, represents a lens space summand (possibly S^3 or $S^2 \times S^1$) of $M^t_{(-1,-2,c,d)}(t)$, thus proving the claim. □

Next, consider the manifolds for which $a = -2$ and $b = -1$.

Claim 2. $M^t_{(-2,-1,c,d)}(t) \cong L(t^2 - ct + 1, t - c)\#L(d + 1, 1)$.

Proof. We refer the reader to Figure 3. The proof of this claim is nearly identical to the proof of Claim 1. Transition 1 is an isotopy of component D. In transition 2, we perform a single positive twist on component B. This gives B a trivial surgery coefficient; so it is removed. The surgery coefficient of components A and C increase to -1 and $c+1$, respectively. In transition 3, we perform a single positive twist on component A. This gives A a trivial surgery coefficient; so it too is removed. The surgery coefficients of components C and D each increase by 1. The remaining transitions are identical to those in Claim 1, the only difference being the surgery coefficients of components C and D. □

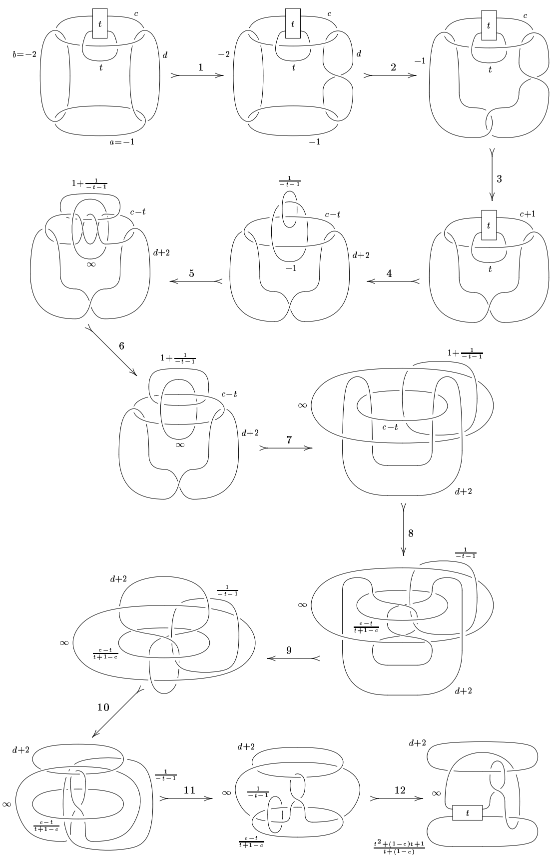


Figure 2. The equivalence of $M_{(-1,-2,c,d)}(t)$ and $L(t^2+(1-c)t+1, t+1-c)\#L(d+2,1)$.

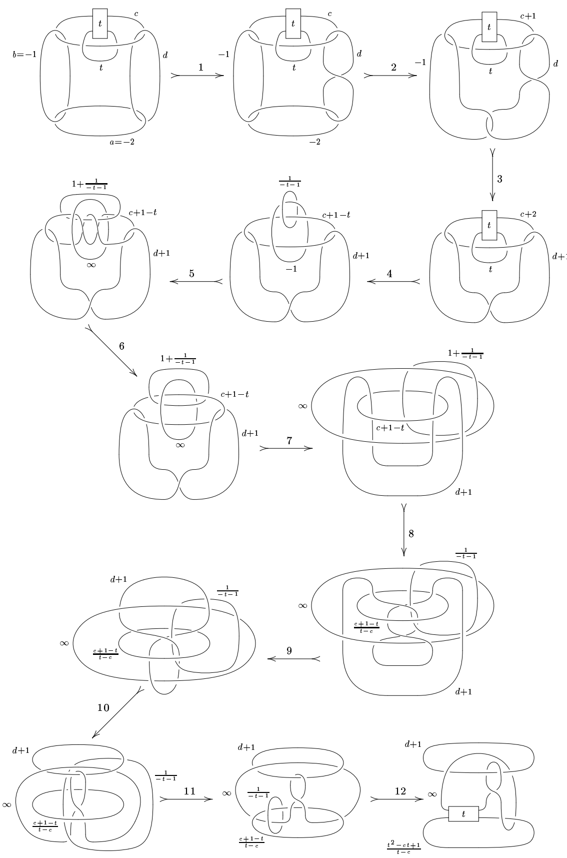


Figure 3. The equivalence of $M_{(-2,-1,c,d)}(t)$ and $L(t^2-ct+1, t-c)\#L(d+1,1)$.

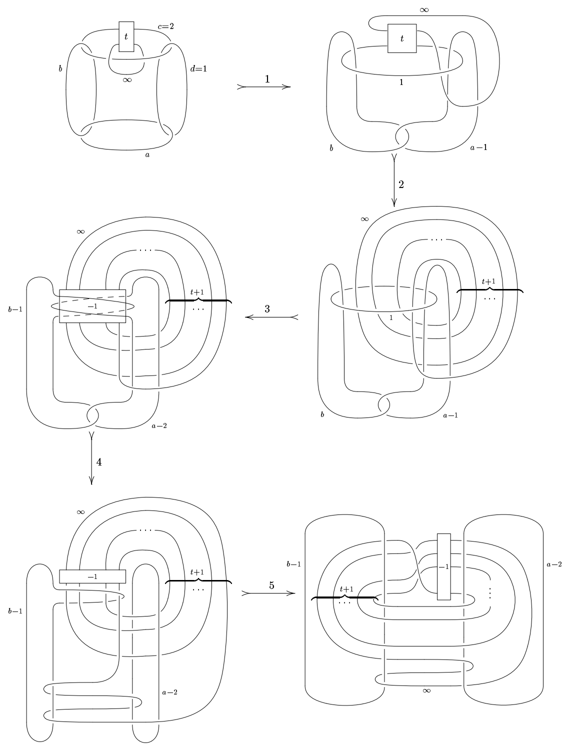


Figure 4. The equivalence of $M^t_{(a,b,2,1)}(\infty)$ and $L(a-2,1)\#L(b-1,1)$ with $t>0$.

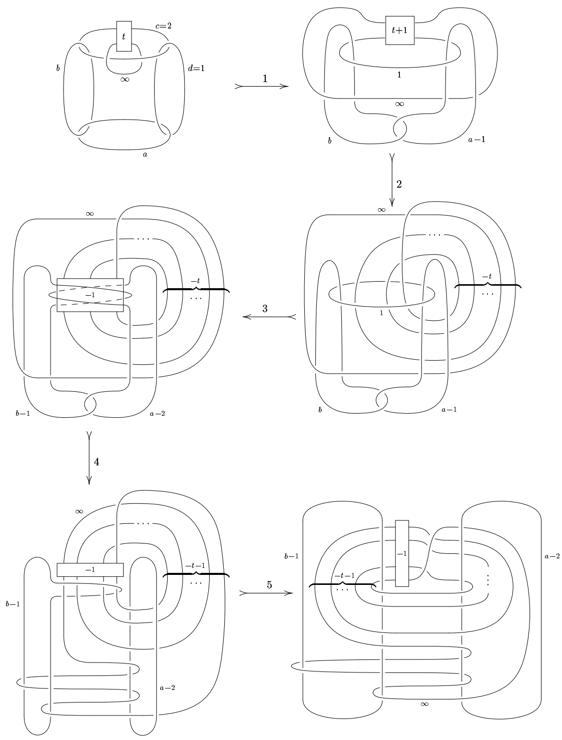


Figure 5. The equivalence of $M^t_{(a,b,2,1)}(\infty)$ and $L(a-2,1)\#L(b-1,1)$ with $t<0$.

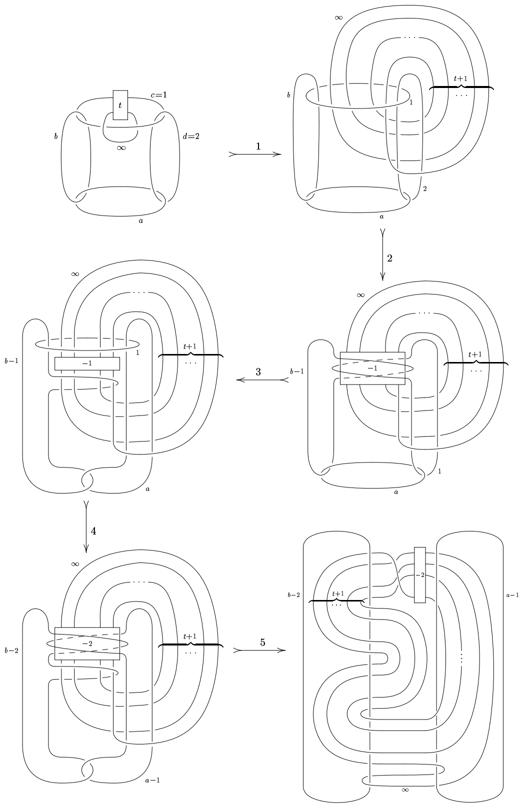


Figure 6. The equivalence of $M^t_{(a,b,1,2)}(\infty)$ and $L(a-1,1)\#L(b-2,1)$ with $t > 0$.

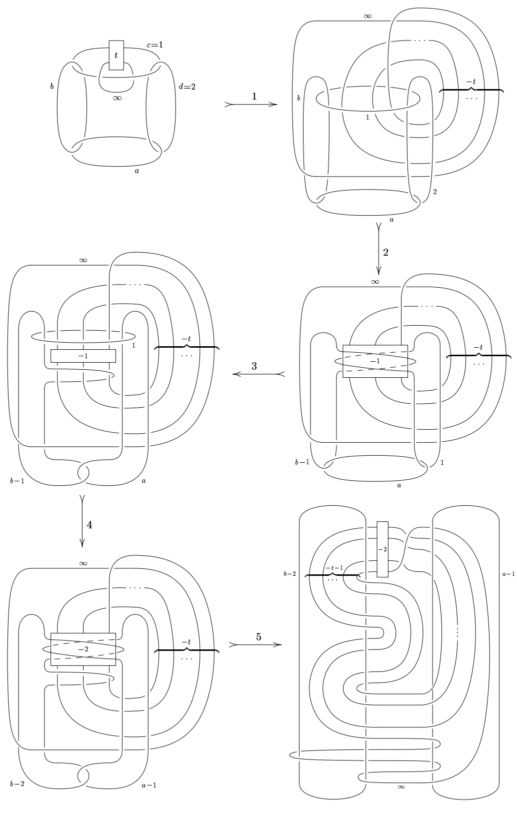


Figure 7. The equivalence of $M^t_{(a,b,1,2)}(\infty)$ and $L(a-1,1)\#L(b-2,1)$ with $t < 0$.

In our third claim, we consider the manifolds for which $c = 2$ and $d = 1$.

Claim 3. $M^t_{(a,b,2,1)}(\infty) \cong L(a-2,1)\#L(b-1,1)$.

Proof. There are two cases, $t > 0$ and $t < 0$, in the proof of this claim. The two cases are nearly identical. We refer the reader to Figures 4 and 5. In transition 1, we perform a single negative twist about component D giving it a trivial surgery coefficient. Component D thus is removed from the diagram. The surgery coefficients of components A and C are each reduced by 1. We also see in transitions 1 and 2 a rather involved isotopy of component K in which the twisting of components K and C is replaced by the looping of K around C. This is done to facilitate transition 3, where we perform a single negative twist about component C. This gives C a trivial surgery coefficient; so it is removed from the diagram. The surgery coefficients of components A and B are reduced to $a-2$ and $b-1$, respectively. Note that K keeps its trivial surgery coefficient.

Transition 4 is the start of another involved isotopy. This isotopy separates components A and B. The reader should first isotop the foremost (and leftmost) arc of component A so that it moves: In front of the diagram, through the interior of the disk bounded by component B, and behind the diagram. During this isotopy, A will snag one of the strands of component K. In transition 5, we deform component B to an oval. We note that components A and B are unlinked and unknotted. Thus each component, A and B, represents a lens space summand of $M^t_{(a,b,2,1)}(\infty)$ proving the claim. □

Finally, we consider the manifolds for which $c = 1$ and $d = 2$.

Claim 4. $M^t_{(a,b,1,2)}(\infty) \cong L(a-1,1)\#L(b-2,1)$.

Proof. Again there are two nearly identical cases to consider, $t > 0$ and $t < 0$. We refer the reader to Figures 6 and 7. This proof is similar to that of Claim 3. Transition 1 is an isotopy of component K in which the twisting of K and C is replaced by the looping of K around C. This is done to facilitate transition 2, where we perform a single negative twist about component C. This gives C a trivial surgery coefficient; so it is removed from the diagram. The surgery coefficients of components B and D are each reduced by one.

Transition 3 is an isotopy which shrinks component D and stretches component A. In transition 4, we perform a single negative twist about component D giving it a trivial surgery coefficient, and so it is removed from the diagram. The surgery coefficients of components A and B are reduced to $a-1$ and $b-2$, respectively.

We separate components A and B with an isotopy in transition 5. The reader should refer to the transitions 4 and 5 in the proof of Claim 3 for clarification. We note that components A and B are unlinked and unknotted. Thus each component, A and B, represents a lens space summand of $M^t_{(a,b,1,2)}(\infty)$ proving the claim. □

These four claims give us the following corollary. Part (a) of the corollary follows from Claims 2 and 3. Part (c) follows from Claims 1 and 4. And Part (b) follows from either Claims 2 and 4 or from Claims 1 and 3.

Corollary 2.1. *If t is an integer, then*

a) $M^t_{(-2,-1,2,1)}(\infty) \cong L(-4,1)\#L(-2,1)$ *and*
$M^t_{(-2,-1,2,1)}(t) \cong L(t^2-2t+1,t-2)\#L(2,1)$.
b) $M^t_{(-2,-1,1,2)}(\infty) \cong L(-3,1)\#L(-3,1)$ *and*
$M^t_{(-2,-1,1,2)}(t) \cong L(t^2-t+1,t-1)\#L(3,1)$, *also*
$M^t_{(-1,-2,2,1)}(\infty) \cong L(-3,1)\#L(-3,1)$ *and*
$M^t_{(-1,-2,2,1)}(t) \cong L(t^2-t+1,t-1)\#L(3,1)$,
c) $M^t_{(-1,-2,1,2)}(\infty) \cong L(-2,1)\#L(-4,1)$ *and*
$M^t_{(-1,-2,1,2)}(t) \cong L(t^2+1,t)\#L(4,1)$.

For the remainder of this paper, we wish to restrict our results to the cases which yield true lens spaces (thus excluding S^3 and $S^2 \times S^1$). In Case (a), we must exclude the values $t = 0, 1$ and 2. Likewise, in Case (b) we exclude $t = 0$ and 1, and in Case (c) we exclude $t = 0$. All other integral values of t are *admissible*.

We should remark, at this point, that these excluded cases are genuinely uninteresting. If the manifolds, before Dehn filling, were hyperbolic, then we might be able to claim a counterexample to a generalized cabling conjecture. But alas, the excluded manifolds are not hyperbolic.

3. Hyperbolicity.

Let M be any one of the manifolds $M^t_{(-1,-2,1,2)}$, $M^t_{(-1,-2,2,1)}$, $M^t_{(-2,-1,1,2)}$ and $M^t_{(-2,-1,2,1)}$ where t is an admissible integer. In this section, we prove that M is hyperbolic (i.e., the interior of M admits a hyperbolic structure). Since each manifold M has toroidal boundary $T = \partial M$, we need to show M is irreducible, ∂-irreducible, atoroidal, and not Seifert fibered [**T**].

Lemma 3.1. *If S is a separating reducing 2-sphere in $M(\infty)$ or $M(t)$, then S may not be isotoped so that $S \cap T = \emptyset$.*

Proof. It suffices to prove the claim for $M(t)$ since if a reducing 2-sphere S could be made disjoint from T in $M(\infty)$, then either S would also be a reducing 2-sphere disjoint from T in $M(t)$ or S would be inessential in $M(t)$. In the latter case, both summands of $M(t)$ would have to appear in $M(\infty)$.

First note that, for the connect sum of two irreducible manifolds, there is only one isotopy class of reducing (essential) 2-spheres. By Corollary 2.1, $M(t)$ is homeomorphic to the connect sum of two lens spaces. Thus any reducing 2-sphere in $M(t)$ can be isotoped to a "standard" sphere S which misses the cores of the lens space summands. That is, we may assume S separates to the two link components with the nontrivial surgery coefficients shown in the final diagrams in Figures 2 and 3. Since S is reducing, it must separate the two cores and intersect the component K. Recall that T is the torus boundary of a regular neighborhood of K. Now the problem can be restated by claiming K cannot be isotoped in $M(t)$ to miss S.

Again, we refer the reader to the final diagrams in Figures 2 and 3. Note that the link diagrams are identical, only the surgery coefficients differ. For this link diagram, with $t \neq 0$, the Alexander polynomial is given by $\Delta(a) = a^2 - 4a + 6 - 4a^{-1} + a^{-2}$ when all components are given a clockwise orientation. Recall that if L is a split link, then the Alexander polynomial is zero for that link. So we conclude that the three components in this diagram are indeed linked.

Every arc of K in $M(t) - S$ links with the core of the lens space summands. So for any product neighborhood $S \times I$ of S, where $S = S \times \{0\}$, such that $S \times \{1\}$ and K intersect transversaly, we have $|(S \times \{1\}) \cap K| \geq |S \cap K|$. Therefore, from Proposition 1.1 of [**E**], $|S \cap K|$ is minimal. So, it is impossible for K to be isotoped to miss S. □

Lemma 3.2. *M is irreducible and ∂-irreducible.*

Proof. Suppose that M is reducible with S a reducing sphere in M. If S is nonseparating in M, then S is nonseparating in $M(r)$. But by Corollary 2.1, there are slopes r for which $M(r)$ is the connect sum of two lens spaces. And these manifolds contain no nonseparating spheres.

So assume that S is a separating sphere in M. Then $M = X \# Y$ where $\partial X = \partial M$ and $Y \neq S^3$. Thus $M(r) = X(r) \# Y$. In particular, $M(\infty) = X(\infty) \# Y \cong L_1 \# L_2$ where L_1 and L_2 are lens spaces. By the uniqueness of decomposition, we can assume $X(\infty) \cong L_1$ and $Y \cong L_2$. But this contradicts Lemma 3.1.

If M is ∂-reducible, since M is irreducible and ∂M is a torus, then M must be a solid torus. But this is impossible since the fillings of a solid torus are well-known and do not correspond to the results of Corollary 2.1. □

Bireducible manifolds are not Seifert fibered. Thus we have the following lemma:

Lemma 3.3. *M is not Seifert fibered.*

Proof. If M is Seifert fibered, then $M(r)$ is Seifert fibered for all but one slope r for which M is reducible [**H**]. But by Corollary 2.1, we have two slopes for which Dehn filling produces reducible manifolds. □

The following lemma is proved in [**EW**]:

Lemma 3.4 (Eudave-Muñoz, Wu)**.** *Let W be an irreducible and ∂-irreducible 3-manifold. If both $W(r_1)$ and $W(r_2)$ are reducible and ∂-reducible, then $r_1 = r_2$.*

Next, we show that M does not contain an essential torus.

Lemma 3.5. *M is atoroidal.*

Proof. Suppose that M contains an essential torus F. Then F must be separating. Otherwise, $M(\infty)$ would contain a nonseparating torus or sphere, contradicting Corollary 2.1. Let W and W' be the two components of M cut along F, where W contains ∂M. Using Lemma 3.2, we may conclude that W is both irreducible and ∂-irreducible (as is W'). By Corollary 2.1, both $M(\infty)$ and $M(t)$ are atoroidal and reducible. Thus F must be compressible in both $W(\infty)$ and $W(t)$.

If both $W(\infty)$ and $W(t)$ are reducible, then this contradicts Lemma 3.4. Thus, we may assume that one of them is irreducible (i.e., it is a solid torus).

Let $x \in \{\infty, t\}$ such that $W(x) \cong S^1 \times D^2$. Let $y \in \{\infty, t\}$ and $x \neq y$. Let K_x be the core of the Dehn filling, and $V_x = N(K_x)$. Then $W = W(x) - \operatorname{int} V_x$. Note that $W' = M - W$ and $W' = M(\gamma) - W(\gamma)$, for all slopes $\gamma \in \partial M$. Recall that $F = \partial W' \cong \partial W(\gamma)$. Therefore $W' = M(x) - W(x) \cong \big(L(a,b)\#L(p,q)\big) - S^1 \times D^2$. Let r be a slope in F which corresponds to a meridian of $W(x) \cong S^1 \times D^2$. Then $W'(r)$ is reducible.

Now we examine $W' = M(y) - W(y)$. Let s be the slope in F, which corresponds to a meridian in $\partial W(y)$. So, r is the meridian slope in $\partial W(x)$ and s is the slope of the new meridian in $\partial W(y) \cong \partial W(x)$ after performing surgery on K_x along y.

We consider two cases, according to whether $W(y)$ is reducible or not.

Case 1: $W(y)$ is reducible (i.e., $W(y) \cong S^1 \times D^2 \# L(p,q)$).

It follows from theorems of Gabai and Scharlemann ([**Ga**] and [**S**]) that W is a cable space.

Case 2: $W(y)$ is irreducible.

From Gabai's Theorem 1.1 [**Ga**], W is the exterior of a braid in a solid torus. So, in both cases, we can apply Gordon's Lemma 3.3 [**Go**] (here W and K_x take place of respectively Y and J). Thus, $\Delta(r,s) = \left|\dfrac{nw^2}{(w,m)}\right| \geq |w|$, where w is the winding number of K_x in the solid torus $W(x)$.

In the first case, K_x is (p,q)-cable knot (hence $q \geq 2$). So $|w| > 1$. In the second case, K_x is a braid, then again $| w | > 1$. Consequently, $\Delta(r,s) > 1$. Note that this result also follows from [**B**] Theorem 2.5.

Now, in the first case $W'(s)$ is a lens space, which contradicts [**BZ**]. In the second case, $W'(s)$ is also reducible, which contradicts [**GLu**]. □

The results of this section show that any one of the manifolds $M^t_{(-1,-2,1,2)}$, $M^t_{(-1,-2,2,1)}$, $M^t_{(-2,-1,1,2)}$ and $M^t_{(-2,-1,2,1)}$, where t is an admissible integer, is a hyperbolic manifold. This fact and Corollary 2.1 suffice to prove Theorem 1.

4. Comments and questions.

The use of both positive and negative values in the parameter t produces redundancy in the list of manifolds up to homeomorphism. This redundancy is made explicit in the correspondences shown in the next theorem.

Theorem 2. *The following manifolds are homeomorphic:*

a) $M^t_{(-2,-1,2,1)}(t) \cong M^{2-t}_{(-2,-1,2,1)}(2-t)$ *for all* $t \leq -1$.
b) $M^t_{(-1,-2,2,1)}(t) \cong M^{1-t}_{(-1,-2,2,1)}(1-t)$ *for all* $t \leq -1$.
c) $M^t_{(-1,-2,1,2)}(t) \cong M^{-t}_{(-1,-2,1,2)}(-t)$ *for all* $t \leq -1$.

Proof. This proof is based on the fact that two lens spaces $L(p,q)$ and $L(p,q')$ are of the same homeomorphism type if and only if $\pm qq' \cong 1 (\operatorname{mod} p)$ [**R**]. We only prove Case (a), as the other two cases are similar. By Corollary 2.1, $M^t_{(-2,-1,2,1)}(t) \cong L(t^2-2t+1, t-2)$ and $M^{2-t}_{(-2,-1,2,1)}(2-t) \cong L(t^2-2t+1, -t)$. The homeomorphism of the two follows since $-(2-t)(-t) = 2t - t^2 = 1 - (t^2 - 2t + 1)$. □

We would like to point out that all known examples of bireducible fillings result in a summand which is homeomorphic to one of the lens spaces $L(2,1), L(3,1)$, or $L(4,1)$. The Eudave-Muñoz and Wu examples [**EW**] and the Gordon and Litherland example [**GLi**] always have an $L(2,1)$ summand. This begs the question: Does there exist an example in which no summand is homeomorphic to either $L(2,1), L(3,1)$, or $L(4,1)$?

Second, we would like the reader to note that there are no known examples of "trireducible" manifolds. According to Gordon and Luecke [**GLu**], if two fillings on an irreducible manifold with torus boundary produce reducible manifolds, then the slopes of the fillings must have a minimal geometric intersection of one. This means an irreducible manifold can have at most three slopes for which Dehn filling produces a reducible manifold. Does there exist a manifold which is trireducible?

Finally, in all known examples, the minimum number of times one of the reducing spheres meets the core of the Dehn filling is bounded by four. In our examples and the Eudave-Muñoz and Wu examples [**EW**], the other reducing sphere in each family meets the core an arbitrarily large number of times. Does there exist an example in which both minimal intersections are larger than four? And if so, is there a family of examples in which both minimal intersections are unbounded?

For the readers' convenience, here is a way to construct examples of hyperbolic manifolds with a single boundary manifold which have bireducible Dehn fillings, from the examples given in [**EW**].

It is shown in [**EW**] Theorem 3.6 that there is a collection of hyperbolic manifolds, denoted M_p, with two toroidal boundary components T_0 and T_1, such that T_0 has two reducing slopes. If r is a slope in T_0, denote the Dehn filling along r by $M_p(r)$; and, if r is in T_1, denote the Dehn filling by $(r)M_p$.

It follows from [**EW**] Lemma 3.1 that for the slopes ∞ and 0 in T_0, both $M_p(\infty)$ and $M_p(0)$ are reducible. Now, from [**EW**] Figure 3.1, $(\infty)M_p$ and $(0)M_p$ are also both reducible, where ∞ and 0 are slopes in T_1. Then, from [**EW**] Table 1.1, if r is a slope in T_1 such that $\Delta(r, \infty) > 3$ or $\Delta(r, 0) > 3$ then $(r)M_p$ is hyperbolic.

So, for almost all r (for all except at most 16 values of r) $(r)M_p$ is a hyperbolic manifold, with a single boundary component, and with bireducible Dehn fillings. Furthermore, $(r)M_p(\infty) = \mathrm{R}\,P^3 \# S_1$ and $(r)M_p(0) = \mathrm{R}\,P^3 \# S_2$, where S_1 and S_2 are small Seifert fibered spaces (in a few cases they are lens spaces).

Acknowledgements. We would very much like to thank the referee for some very insightful comments and Jeff Weeks and A.C. Manoharan for SnapPea PC which helped tremendously in verifying our work.

References

[B] J. Berge, *The knots in $D^2 \times S^1$ which have nontrivial Dehn surgeries that yield $D^2 \times S^1$*, Topology Appl., **38** (1991), 1-19, MR 92d:57005, Zbl 0725.57001.

[BZ] S. Boyer and X. Zhang, *On Culler-Shalen seminorms and Dehn filling*, Annals of Math., **148** (1998), 1-66, MR 2000d:57028.

[E] M. Eudave-Muñoz, 4-*Punctured tori in the exterior of knots*, J. Knot Theory Ramifications, **6** (1997), 659-676, MR 98f:57022, Zbl 0887.57014.

[EW] M. Eudave-Muñoz and Y.-Q. Wu, *Nonhyperbolic Dehn fillings on hyperbolic* 3-*manifolds*, Pacific J. Math., **190** (1999), 261-275, MR 2000j:57035.

[Ga] D. Gabai, *Surgery on knots in solid tori*, Topology, **28** (1989), 1-6, MR 90h:57005, Zbl 0678.57004.

[GS] F. González-Acuña and H. Short, *Knot surgery and primeness*, Math. Proc. Camb. Phil. Soc., **99** (1986), 89-102, MR 87c:57003, Zbl 0591.57002.

[Go] C. McA. Gordon, *Dehn surgery and satellite knots*, Trans. Amer. Math. Soc., **275** (1983), 687-708, MR 84d:57003, Zbl 0519.57005.

[GLi] C. McA. Gordon and R.A. Litherland, *Incompressible planar surfaces in* 3-*manifolds*, Topology Appl., **18** (1984), 121-144, MR 86e:57013, Zbl 0554.57010.

[GLu] C. McA. Gordon and J. Luecke, *Reducible manifolds and Dehn surgery*, Topology, **35** (1996), 385-409, MR 97b:57013, Zbl 0859.57016.

[H] Heil, *Elementary surgery on Seifert fiber spaces*, Yokohama Math. J., **22** (1974), 135-139, MR 51 #11515, Zbl 0297.57006.

[R] D. Rolfsen, *Knots and Links*, Math. Lect. Ser., **7**, Publish or Perish, Berkeley, California, 1976, MR 58 #24236, Zbl 0339.55004.

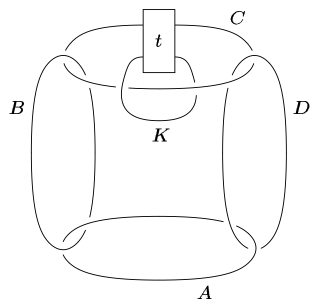
[S] M. Scharlemann, *Producing reducible* 3*-manifolds by surgery on a knot*, Topology, **29** (1990), 481-500, MR 91i:57003, Zbl 0727.57015.

[T] W. Thurston, *Three dimensional manifolds, Kleinian groups and hyperbolic geometry*, Bull. Amer. Math. Soc., **6** (1982), 357-381, MR 83h:57019, Zbl 0496.57005.

Received April 18, 2000 and revised February 23, 2002.

Analytical Mechanics Associates Inc.
303 Butler Farm Road
Hampton VA 23666
E-mail address: hoffman@ama-inc.com

Université de Provence
C.M.I. 39
rue Joliot Curie
F-13453 Marseille Cedex 13
France
E-mail address: matignon@gyptis.univ-mrs.fr


t
C
B
D
K
A

PACIFIC JOURNAL OF MATHEMATICS
Vol. 209, No. 1, 2003

HOMOTOPY MINIMAL PERIODS FOR MAPS OF THREE DIMENSIONAL NILMANIFOLDS

Jerzy Jezierski and Wacław Marzantowicz

A natural number m is called the homotopy minimal period of a map $f : X \to X$ if it is a minimal period for every map g homotopic to f. The set $\mathrm{HPer}(f)$ of all minimal homotopy periods is an invariant of the dynamics of f which is the same for a small perturbation of f. In this paper we give a complete description of the sets of homotopy minimal periods of self-maps of nonabelian three dimensional nilmanifold which is a counterpart of the corresponding characterization for three dimensional torus proved by Jiang and Llibre. As a corollary we show that if $2 \in \mathrm{HPer}(f)$ then $\mathrm{HPer}(f) = \mathbb{N}$ for such a map.

0. Introduction.

One of the natural problems in dynamical systems is the study of the homotopy minimal periods of self-map $f : X \to X$ i.e., these periods which are also minimal periods for every map g homotopic to f. An aim is to give a complete characterization, of the set $\mathrm{HPer}(f)$ of all homotopy minimal periods, in terms of the homological information on f. Since the homotopy minimal period preserves under a small perturbation of a manifold map, one can say that the set of all homotopy minimal periods describes the rigid part of dynamics of f. A description of the set of all homotopy minimal periods of a map is difficult in general, however here are some results for the mappings of compact homogenous spaces of Lie groups by a discrete subgroup.

After the case of maps of the circle in [**4**] (Block, Guckenheimer, Misiurewicz and Young) in the second instance maps of two-dimensional torus ($X = T^2$) have been investigated in a series of papers [**1**] and [**2**] by Alsedá, Baldwin, Llibre, Swanson and Szlenk. In our notion they gave a complete description of the set of all homotopy minimal periods of a map of the circle or two torus respectively. The answer is given in terms of the linearization of map f, i.e., an integral matrix of the linear map induced by f. In the work of Jiang and Llibre [**12**] the qualitative description of this set was successfully studied for maps of r-dimensional torus, for an arbitrary $r \geq 1$. All of them use the Nielsen theory, which for the torus maps has very nice algebraic description ([**5**]) and prepossessing geometric properties ([**12**], [**17**] and [**18**]).

Using the general result of [**12**] Jiang and Llibre gave also a complete description of the set of all homotopy minimal periods (called them the minimal set of periods) of a map of the three torus. It can be done with relatively easy handling using algebraic integers of degree equal or less than three.

Recently the authors extended the main theorem of [**12**] onto the case of a map f of an arbitrary compact nilmanifold X with the similar qualitative statement ([**10**] Thm. A). The crucial step of the mentioned fact was a proof that $NP_n(f) = 0$ implies that $f \sim g$, where g has no periodic points of the minimal period n. Basing also on this theorem we give here a complete description of the set of minimal homotopy periods of a compact nonabelian three dimensional nilmanifold (Theorem 3.1). A preliminary version of this theorem has been presented already in [**10**] (Thm. C) but that statement does not contain all restrictions on the sets of homotopy minimal periods that appear in the discussed case. Here we make use of the classification of compact three dimensional nilmanifolds and the fact that every such nilmanifold X forms a fibration with S^1 as the fiber and T^2 as the base (cf. [**6**]). Moreover every self-map of X is homotopic to a fiber map of this fibration due the Fadell-Husseini theorem (cf. [**6**]). This means that the integral 3×3 matrix A corresponding to f is a direct sum of one-dimensional and two-dimensional summand which yields that its characteristic polynomial is the multiple of a two polynomials of degree one and two, corresponding to the fiber mapf_1 and the base map $\bar{f}$ respectively. It lets us to derive the set of homotopy minimal periods of f from the corresponding sets of the factors f_1, $\bar{f}$ (Theorem 3.1) by use of a formula (Theorem 3.5, Corollary 3.6). Due to this factorization we can use the previous classification done in [**2**] and [**4**], and do not need to cope with algebra. The main necessary topological ingredient, with except the mentioned Thm. A of [**10**], is a description of the form of automorphism of any nilpotent nonabelian group of rank 3 (Proposition 2.12). In particular this yields that the degree of base map $\bar{f}$ is equal to the degree of fiber map f_1 (Corollary 2.13).

As an application we specify our theorem to the case of a homeomorphism of such a nilmanifold (Theorem 4.1).

Here is the scheme of the paper. In Section 1 we recall the formula for the homotopy minimal periods of self-maps of S^1 and $\mathbb{T}^2$ ([**4**], [**1**] and [**2**]). In Section 2 the necessary information about nilmanifolds and theorem on $\mathrm{HPer}\, f$ for the self-maps of nilmanifolds of [**10**] are recalled. Also general form of an automorphism of any nilpotent nonabelian group of rank 3 is given. This gives a necessary and sufficient condition on 3×3 matrix to be the linearization of a self-map of such a manifold. Then in Section 3 we show how to reduce the 3-nilmanifold case to S^1 and $\mathbb{T}^2$. This let us to prove the main result (Theorem 3.1). As an application we present a theorem of Šarkovskii type (Corollary 3.9) that says that for a self-map of nonabelian three nilmanifold the existence of homotopy period 2 implies the existence of all homotopy minimal periods. Finally we show that for a homeomorphism f of such manifold if $\mathrm{HPer}\,(f) \neq \emptyset$ then $\mathrm{HPer}\,(f) = \mathbb{N}$ with except two special cases when $\mathrm{HPer}\,(f) = \mathbb{N} \setminus 2\mathbb{N}$.

1. Homotopy minimal periods of self-maps of S^1 and T^2.

In this section we recall the explicit formulae of the homotopy minimal periods of self-maps of S^1 and T^2 presented in [**4**], [**1**] and [**2**]. First we recall the basic definitions used in [**12**] and [**10**]. Remaining a standard terminology, let $f : X \to X$ be a self-map of a compact connected polyhedron X, and n be a natural number. Let $\mathrm{Fix}(f)$ be the fixed point set of f, $P^m(f) := \mathrm{Fix}(f^m)$ and let

$$P_m(f) \;:=\; P^m(f) \setminus \bigcup_{n|m,\, n<m} P^n(f),$$

denote the set of periodic points with least period m.

Recall that $\mathrm{Per}(f)$ denotes the set of all minimal periods of f i.e.,

$$\mathrm{Per}(f) := \{m \in \mathbb{N};\; P_m(f) \neq \emptyset\}.$$

When a map $g : X \to X$ is homotopic to f, we shall write $g \simeq f$. Define the *set of homotopy minimal periods* to be the set

$$\mathrm{HPer}\,(f) := \bigcap_{g \simeq f} \mathrm{Per}(g). \tag{1.1}$$

Boju Jiang and Llibre use the name “the minimal set of periods” but we hope that what we use here more emphasizes that $n \in \mathrm{HPer}\,(f)$ iff n is a minimal period for every g homotopic to f.

We begin with $X = S^1$ which was studied by Block and co-authors in [**4**]. The meaning of letters (E), (F), (G) as well as the definition of matrix A and the set $T_A \subset \mathbb{N}$ in the theorem given below are given in the next section (Theorem 2.3).

Theorem 1.2 ([**4**]). *Let $f : S^1 \to S^1$ be a map of the circle and $d \in \mathbb{Z} = \mathcal{M}_{1\times 1}(\mathbb{Z})$ be the matrix corresponding to f i.e., the degree of f.*

There are three types for the minimal homotopy periods of f:

- (E) $\mathrm{HPer}\,(f) = \emptyset$ *if and only if* $d = 1$.
- (F) $\mathrm{HPer}\,(f)$ *is nonempty and finite if and only if* $d = -1$ *or* $d = 0$. *We have* $\mathrm{HPer}\,(f) = \{1\}$ *then. Moreover the sets* T_A *are equal to* $\mathbb{N} \setminus 2\mathbb{N}$ *and* $\mathbb{N}$ *correspondingly.*
- (G) $\mathrm{HPer}\,(f)$ *is equal to* $\mathbb{N}$ *for the remaining* d, *i.e.,* $|d| > 1$, *with the exception of one special case* $d = -2$ *where* $T_A = \mathbb{N}$ *but* $\mathrm{HPer}\,(f) = \mathbb{N} \setminus \{2\}$.

The case $X = T^2$ had been completely described by Alseda and co-authors in [**1**] and [**2**]. A reformulation of it is the following:

Theorem 1.3 ([**2**]). *Let $f : T^2 \to T^2$ be a map of the torus, $A \in \mathcal{M}_{2\times 2}(\mathbb{Z})$ the linearization of f, and $\chi_A(t) = t^2 - at + b$ be its characteristic polynomial.*

There are three types for the minimal homotopy periods of f:

(E) $\mathrm{HPer}\,(f) = \emptyset$ *if and only if* $-a + b + 1 = 0$.

(F) $\mathrm{HPer}\,(f)$ *is nonempty and finite for* 6 *cases corresponding to one of the six pairs* (a, b) *listed below*

$$(0, 0),\ (-1, 0),\ (-2, 1),\ (0, 1),\ (-1, 1),\ (1, 1).$$

We have $\mathrm{HPer}\,(f) \subset \{1, 2, 3\}$ *then. Moreover the sets T_A and* $\mathrm{HPer}\,(f)$ *are the following:*

	Cases of Type (F)	
$(a,\ b)$	T_A	$\mathrm{HPer}\,(f)$
$(0,\ 0)$	$\mathbb{N}$	$\{1\}$
$(0,\ 1)$	$\mathbb{N} \setminus 4\mathbb{N}$	$\{1, 2\}$
$(-1,\ 0)$	$\mathbb{N} \setminus 2\mathbb{N}$	$\{1\}$
$(-1,\ 1)$	$\mathbb{N} \setminus 3\mathbb{N}$	$\{1\}$
$(-2,\ 1)$	$\mathbb{N} \setminus 2\mathbb{N}$	$\{1\}$
$(1,\ 1)$	$\mathbb{N} \setminus 6\mathbb{N}$	$\{1, 2, 3\}$

(G) $\mathrm{HPer}\,(f)$ *is infinite for the remaining a, and b. Furthermore,* $\mathrm{HPer}\,(f)$ *is equal to $\mathbb{N}$ for all pairs $(a, b) \in \mathbb{Z}^2$ with the exception of the following special cases listed below. We say that a pair $(a, b) \in \mathbb{Z}^2$ satisfies condition*

1^0 *if $a \neq 0$ and $a + b + 1 = 0$,*

2^0 *if $a + b = 0$,*

3^0 *if $a + b + 2 = 0$ respectively,*

and (a, b) is not one of the pairs of case (E) *and* (F).

We have the following table of special cases:

	Special Cases of Type (G)	
$(a,\ b)$	T_A	$\mathrm{HPer}\,(f)$
$(-2,\ 2)$	$\mathbb{N}$	$\mathbb{N} \setminus \{2, 3\}$
$(-1,\ 2)$	$\mathbb{N}$	$\mathbb{N} \setminus \{3\}$
$(0,\ 2)$	$\mathbb{N}$	$\mathbb{N} \setminus \{4\}$
$(a,\ b)$, (a, b) satisfies 1^0	$\mathbb{N} \setminus 2\mathbb{N}$	$\mathbb{N} \setminus 2\mathbb{N}$
$(a,\ b)$, (a, b) satisfies 2^0	$\mathbb{N}$	$\mathbb{N} \setminus \{2\}$
$(a,\ b)$, (a, b) satisfies 3^0	$\mathbb{N}$	$\mathbb{N} \setminus \{2\}$

2. Nilmanifolds.

A compact manifold M is a nilmanifold iff it is of the form G/Γ where G is a simply connected nilpotent Lie group of dimension r and Γ is a lattice of rank r of G i.e., a discrete, torsion free, subgroup of G of rank r ([**14**] and [**16**]). Then the fundamental group of M is Γ and Γ uniquely determines M up to homeomorphism.

Fadell and Husseini in [**6**] show that every self map on M can be inductively fibered on an orientable fibration into a map on torus and a map on a lower dimensional nilmanifold ([**6**, Thm. 3.3]). This enables the proof of the following theorem (cf. [**9**] and [**13**], see also [**10**] for an exposition of it):

Theorem 2.1. *Let $f : X \to X$ be a map of a compact nilmanifold X of dimension r. Then there exists an $r \times r$ matrix A with integral coefficients such that*

$$L(f^m) = \det(\mathrm{I} - A^m)$$

for every $m \in \mathbb{N}$.

The integral matrix A is the basic object in study minimal and homotopy minimal periods of a self-map $f : X \to X$. Note that if $X = T^r$ is the torus then A is the unique homomorphism of $\Gamma = Z^r$ which corresponds to f and A is called the linearization of f (cf. [**9**] and [**13**]). As matter of fact the spectrum of matrix A, or equivalently the characteristic polynomial $\chi_A(t) \in \mathbb{Z}[t]$ determines the set $\mathrm{HPer}(f)$. For given $A \in \mathcal{M}_{r\times r}(\mathbb{Z})$ we set

$$T_A := \{\, n \in \mathbb{N} \mid \det(\mathrm{I} - A^n) \neq 0\}. \tag{2.2}$$

In the case if $A = A_f$ is the matrix associated to a self-map $f : X \to X$ of a compact nilmanifold X we call T_A *the set of algebraic periods of* f. The main result of [**9**] says the following:

Theorem 2.3 ([**10**], Thm. A)**.** *Let $f : X \to X$ be a map of a compact nilmanifold X of dimension r, A the matrix associated with f and $T_A \subset \mathbb{N}$ the set of algebraic periods of f.*

Then $\mathrm{HPer}(f) \subset T_A$ *and it is in one of the following three* (*mutually exclusive*) *types, where the letters E, F, and G are chosen to represent "empty", "finite" and "generic" respectively:*

- (E) $\mathrm{HPer}(f)$ *is empty if and only if* $N(f) = L(f) = 0$, *i.e., if and only if* 1 *is an eigenvalue of* A;
- (F) $\mathrm{HPer}(f)$ *is nonempty but finite if and only if all the eigenvalues of* A *are either zero or roots of unity different from* 1;
- (G) $\mathrm{HPer}(f)$ *is infinite and* $T_A \setminus \mathrm{HPer}(f)$ *is finite.*

Moreover, for every dimension r of X, there are finite sets $P(r)$, $Q(r)$ of integers such that $\mathrm{HPer}(f) \subset P(r)$ *in Type F and* $T_A \setminus \mathrm{HPer}(f) \subset Q(r)$ *in Type* (G).

Theorem 2.3 generalizes the corresponding result of Boju Jiang and Llibre ([**12**, Thm. B]) from the torus map onto the case of any compact nilmanifold. The last was used by Jiang and Llibre to give a complete description of all homotopy minimal periods of an arbitrary map of three torus in the terms of characteristic polynomial $\chi_A(t)$ of its linearization ([**12**, Thm. C]). The corresponding result for three dimensional nonabelian nilmanifolds was given in a not complete form in [**10**] (Thm. C). Now we would like to present a complete version of this description. To do this we need a little bit more information about three nilmanifolds.

We would like to remind the reader that the simplest nontrivial examples of compact nilmanifolds are *Iwasawa manifolds* $\mathbb{N}_n(\mathbb{R})/\mathbb{N}_n(\mathbb{Z})$ and $\mathbb{N}_n(\mathbb{C})/\mathbb{N}_n(\mathbb{Z}[\mathbf{i}])$, where $\mathbb{Z}[\mathbf{i}]$ is the ring of Gaussian integers and for any ring $\mathcal{R}$ with unity $\mathbb{N}_n(\mathcal{R})$ denotes the group of all unipotent upper triangular matrices whose entries are elements of the ring $\mathcal{R}$. The Iwasawa 3-manifold $\mathbb{N}_3(\mathbb{R})/\mathbb{N}_3(\mathbb{Z})$, called also "Baby Nil" is the simplest example of compact three dimensional nonabelian nilmanifold, since $\mathbb{N}_3(\mathbb{Z}) \neq \mathbb{Z}^3$. Generalizations of the Iwasawa manifolds are compact nilmanifolds $\mathbb{N}_3(\mathbb{R})/\Gamma_{p,q,r}$, where the subgroup $\Gamma_{p,q,r}$, with fixed p, q, $r \in \mathbb{N}$ consists of all matrices of the form

$$\begin{bmatrix} 1 & \frac{k}{p} & \frac{m}{p\cdot q\cdot r} \\ 0 & 1 & \frac{l}{q} \\ 0 & 0 & 1 \end{bmatrix}, \quad \text{where} \quad k,\, l,\, m \in \mathbb{Z}. \tag{2.4}$$

Since the group $\mathbb{N}_3(\mathbb{R})$ are named the Heisenberg group, the nilmanifolds $\mathbb{N}_3(\mathbb{R})/\Gamma_{p,q,r}$, are also called the Heisenberg nilmanifolds. The groups $\pi_1(X)$ for all compact nilmanifolds X are precisely all finitely generated torsion-free nilpotent groups (see [**3**], [**7**], [**14**] and [**16**]).

This leads to the following well-known classification theorem ([**7**, 4.1, Cor. 2]):

Theorem 2.5. *Let X be a compact nilmanifold of dimension* 3*. Then X is diffeomorphic to T^3 or to $\mathbb{N}_3(\mathbb{R})/\Gamma_{1,1,r}$ with some $r \in \mathbb{N}$.*

Proof. The point is that any finitely generated nilpotent group of rank 3 is isomorphic to Z^3 or to the group $\Gamma_{1,1,r}$ with some $r \in \mathbb{N}$ (cf. [**7**, 4.1, Cor. 2]).

We explain it briefly. In fact the correspondence

$$\begin{bmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix} \mapsto \begin{bmatrix} 1 & px & pqz \\ 0 & 1 & qy \\ 0 & 0 & 1 \end{bmatrix}$$

is an isomorphism of $\mathbb{N}_3(\mathbb{R})$ sending $\Gamma_{p,q,r}$ onto $\Gamma_{1,1,r}$. It is sufficient to show [!] that any discrete subgroup of rank 3 of $\mathbb{R}^3$, or $\mathbb{N}_3(\mathbb{R})$ is equal, up to isomorphism, to $\mathbb{Z}^3$, or $\Gamma_{p,q,r}$ respectively.

Since a nilmanifold is the quotient of a simply connected nilpotent Lie group by its uniform (hence discrete) subgroup, it remains to know that any three dimensional simply connected non-commutative nilpotent Lie group is isomorphic to the Heisenberg group $\mathbb{N}_3(\mathbb{R})$. The last follows from the fact that there is one non-commutative nilpotent Lie algebra of dimension three, up to isomorphism.

Then a three nilmanifold different than torus is of the form $\mathbb{N}_3(\mathbb{R})/\Gamma$ where Γ is a uniform subgroup in $\mathbb{N}_3(\mathbb{R})$ hence $\Gamma = \Gamma_{p,q,r}$ for some $p, q, r \in \mathbb{N}$. We notice that $\mathbb{N}_3(\mathbb{R})/\Gamma_{p,q,r} = \mathbb{N}_3(\mathbb{R})/\Gamma_{1,1,r}$. □

Next we point out that the Fadell-Husseini toral fibration of a three-dimensional compact nilmanifold has a special form. Since the commutator

$$G_1 = \left\langle [\mathbb{N}_3(\mathbb{R}), \mathbb{N}_3(\mathbb{R})] = \begin{bmatrix} 1 & 0 & z \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \,:\, z \in \mathbb{R} \right\rangle \tag{2.6}$$

is one dimensional, the quotient space $G_1/\Gamma \cap G_1 \approx S^1$. By the dimensional reasons the base space must be 2-torus and the fibration becomes $S^1 \subset \mathbb{N}_{1,1,r} \to T^2$.

The above gives the following statement:

Proposition 2.7. *Let $f : X \to X$ be a map of compact nilmanifold X of dimension 3 not diffeomorphic to T^3, and $f_1 : S^1 \to S^1$, $\bar{f} : T^2 \to T^2$ a pair of maps associated with f considered as a fiber map. Then the matrix A corresponding to f by Theorem 2.1 has the form*

$$\begin{bmatrix} d & 0 \\ 0 & A \end{bmatrix} = A_1 \oplus \overline{A},$$

where $A_1 = [d]$, with $d := \deg(f_1)$ the degree of the fiber map f_1, and $\overline{A} \in \mathcal{M}_{2\times 2}(\mathbb{Z})$ is the matrix corresponding to the map $\bar{f}$ of base T^2.

Consequently the characteristic polynomial of f is equal to $\chi_A(t) = \chi_{A_1}(t)\cdot \chi_{\overline{A}}(t) = (t-d)(t^2 - at + b)$, where $d \in \mathbb{Z}$, $t - d = \chi_{A_1}(t)$, a, $b \in \mathbb{Z}$ and $t^2 - at + b = \chi_{\overline{A}}(t)$ is the characteristic polynomial of $\overline{A}$. Moreover $a = \operatorname{tr} \overline{A}$ and $b = \det \overline{A} = \deg(\bar{f})$.

Proof. All with the exception of the last equality are obvious algebraically. The equality $\det \overline{A} = \deg(\bar{f})$ is well-known for the torus map induced by an integral matrix. □

The above proposition gives a natural restriction on an integral 3×3 matrix of the linearization any map of such a manifold. Now we formulate next algebraic restriction that comes from the geometry of the discussed spaces. First we recall a more general fact:

Proposition 2.8. *Let $\Gamma = \pi_1(X)$ be the fundamental group of a compact nilmanifold $X = G/\Gamma$. Then every map $f : X \to X$ is homotopic to a* *map given by a homomorphism* $\Phi : G \to G$ *and the induced homomorphism* $\pi_1(\Phi) : \Gamma \to \Gamma$ *is equal to* $\Phi_{|\Gamma}$.

Inversely, for every homomorphism $\phi : \Gamma \to \Gamma$ *there exist a map* $f : X \to X$ *such that* $\pi_1(f) = \phi$.

Proof. The statement follows from the fact that X is the $K(\Gamma, 1)$-space (cf. [**6**]), the fact that every endomorphism ϕ of Γ has a unique extension to an endomorphism Φ of G (cf. [**16**]), and that for a map $f : X \to X$ given by a homomorphism Φ of G the induced map of the fundamental group $\pi_1(f)$ is equal to $\Phi|\Gamma$. □

With respect to Theorem 2.5 and Proposition 2.8 and the below it is enough to determine the set of matrices of linearization of all endomorphisms of $\Gamma = \Gamma_{1,1,r}$. We begin with a description of $\Gamma_{1,1,r}$. Then we give a description of all endomorphisms of $\Gamma_{1,1,r}$. We follow the approach of [**8**] where the case of $\Gamma_{1,1,1}$ was discussed.

Assigning to any matrix

$$\begin{bmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix}, \quad \text{where} \quad x, y, z \in \mathbb{R} \quad \text{the vector} \quad (x, y, z)$$

we get the homeomorphism between $\mathbb{N}_3(\mathbb{R})$ and $\mathbb{R}^3$. In these coordinates the multiplication has form

$$(x, y, z) * (x', y', z') = (x + x', y + y', z + z' + xy').$$

Using the coordinates we see that $\Gamma_{1,1,r} \subset \mathbb{N}_3(\mathbb{R})$ is generated by the matrices

$$a := (1, 0, 0), \;\; b := (0, 1, 0), \;\; c := (0, 0, 1/r),$$

since $(m, p, q/r) = a^m b^p c^{q-mp}$. Moreover the only relations are

$$aba^{-1}b^{-1} = c^r, \;\; aca^{-1}c^{-1} = e, \;\; bcb^{-1}c^{-1} = e. \tag{2.9}$$

Let $\phi : \mathbb{R}^3 \to \mathbb{R}^3$ be a map and let

$$\phi(a) = (\alpha_1, \beta_1, \gamma_1), \;\; \phi(b) = (\alpha_2, \beta_2, \gamma_2), \;\; \phi(c) = (\alpha_3, \beta_3, \gamma_3).$$

We look for a necessary and sufficient condition on ϕ to extend to homomorphism of $\Gamma_{1,1,r}$. Suppose that ϕ extends to a such homomorphism. Then for some integer k

$$\phi(c) = c^k, \tag{2.10}$$

because the cyclic group generated by c is equal to the center of $\Gamma_{1,1,r}$, consequently $\alpha_3 = 0$ $\beta_3 = 0$ $\gamma_3 = k$. Using the first equality of (2.9) and (2.10), deriving $\phi(a)\phi(b)\phi(a)^{-1}\phi(b)^{-1}$, and comparing the coordinates we get

$$k = \alpha_1\beta_2 - \alpha_2\beta_1. \tag{2.11}$$

Note that $\phi(c) = c^k$ implies that the second and third relations of (2.9) are preserved, because $\phi(c)$ is in the center of $\Gamma_{1,1,r}$. Notice that γ_1 , γ_2 may be arbitrary. Since (2.9) are the only relations we get the following fact:

Proposition 2.12. *A map $\phi : \Gamma_{1,1,r} \to \Gamma_{1,1,r}$ defined in the coordinate system by its values on the generators a, b, c as*

$$\phi(a) = (\alpha_1, \beta_1, \gamma_1),\ \phi(b) = (\alpha_2, \beta_2, \gamma_2),\ \phi(c) = (\alpha_3, \beta_3, \gamma_3)$$

extends to an automorphism of $\Gamma_{1,1,r}$ iff $\alpha_3 = \beta_3 = 0$, and $\gamma_3 = \alpha_1\beta_2 - \alpha_2\beta_1$.

Consequently a 3×3 integral matrix A is the linearization matrix of a map of X given by an endomorphism of $\Gamma_{1,1,r}$ iff it is of the form

$$A = A_1 \oplus \overline{A} = \begin{bmatrix} k & 0 & 0 \\ 0 & \alpha_1 & \beta_1 \\ 0 & \alpha_2 & \beta_2 \end{bmatrix}$$

where $\det \overline{A} = k$.

Finally we formulate a topological consequence of Proposition 2.12:

Corollary 2.13. *Let $X \to X$ be a map of three dimensional nilmanifold not diffeomorphic to the torus.*

Then there exists $k \in \mathbb{Z}$ such that $\deg f = k^2$. *In particular if* $\deg f \neq 0$ *then f preserves the orientation.*

Proof. Note that for a fiber-map $f = (f_1, \bar{f})$ we have $\deg f = \deg f_1 \deg \bar{f}$. On the other hand we have just shown that for a map induced by a homomorphism, thus for every map, we have $\deg f_1 = d = \det \overline{A} = \deg \bar{f}$, by Proposition 2.12. □

3. The main theorem.

Theorem 3.1. *Let $f : X \to X$ be a map of three-dimensional compact nilmanifold X not diffeomorphic to T^3. Let $A = A_1 \oplus \overline{A} \in \mathcal{M}_{3\times 3}(\mathbb{Z})$ be the matrix induced by the fibre map $f = (f_1, \bar{f})$ (Theorem 2.1) and $\chi_A(t) = \chi_{A_1}(t) \cdot \chi_{\overline{A}}(t) = (t - d)(t^2 - at + b)$ be its characteristic polynomial. Then $d = b$ and there are three types for the minimal homotopy periods of f:*

(E) HPer $(f) = \emptyset$ *if and only if or $d = 1$ or $-a + d + 1 = 0$.*

(F) HPer (f) *is nonempty and finite only for* 2 *cases corresponding to*

$$d = 0$$

combined with one of the two pairs (a, b)

$$(0, 0), \quad \textit{and} \quad (-1, 0).$$

We have HPer $(f) = \{1\}$ *then. Moreover the sets T_A and* HPer (f) *are the following:*

Map	Cases of Type (F)	
$(d, \quad a, \quad b)$	T_A	$\mathrm{HPer}\,(f)$
$(0, \quad 0, \quad 0)$	$\mathbb{N}$	$\{1\}$
$(0, \ -1, \quad 0)$	$\mathbb{N}\setminus 2\mathbb{N}$	$\{1\}$

(G) $\mathrm{HPer}\,(f)$ *is infinite for the remaining* $(d, a, b = d)$. *Furthermore,* $\mathrm{HPer}\,(f)$ *is equal to* $\mathbb{N}$ *for all triples* $(d, a, b = d) \in \mathbb{Z}^3$ *with the exception of the following special cases listed below:*

	Special Cases of Type (G)	
$(\ d, \quad a, \quad b)$	T_A	$\mathrm{HPer}\,(f)$
$a + d + 1 = 0$, with $a \neq 0$, and $d \notin \{-2, -1, 0, 1\}$	$\mathbb{N}\setminus 2\mathbb{N}$	$\mathbb{N}\setminus 2\mathbb{N}$
$(\ 0, -2, \quad 0)$	$\mathbb{N}$	$\mathbb{N}\setminus\{2\}$
$(-1, \quad 1, -1)$	$\mathbb{N}\setminus 2\mathbb{N}$	$\mathbb{N}\setminus 2\mathbb{N}$
$(-1, -1, -1)$	$\mathbb{N}\setminus 2\mathbb{N}$	$\mathbb{N}\setminus 2\mathbb{N}$
$(-2, \quad 1, -2)$	$\mathbb{N}\setminus 2\mathbb{N}$	$\mathbb{N}\setminus 2\mathbb{N}$
$(-2, \quad 0, -2)$	$\mathbb{N}$	$\mathbb{N}\setminus\{2\}$
$(-2, \quad 2, -2)$	$\mathbb{N}$	$\mathbb{N}\setminus\{2\}$

Moreover for every pair subset $\mathcal{S}_1 \subset \mathcal{S}_2 \subset \mathbb{N}$, *appearing as* $\mathrm{HPer}\,(f)$ *and* T_A *listed above there exists a map* $f : X \to X$ *such that* $\mathrm{HPer}\,(f) = \mathcal{S}_1$ *and* $T_A = \mathcal{S}_2$.

To prove Theorem 3.1 we show an algorithm which allows us to express the homotopy minimal periods of a given self-map of a (nontrivial) 3-nilmanifold by the corresponding data of self-maps on S^1 and T^2. This will be obtained as a consequence of formulas deriving the set T_A of algebraic periods of $A = \oplus_1^l A_i$ from the sets T_{A_i} and analogously $\mathrm{HPer}\,(f)$ from $\mathrm{HPer}\,(f_i)$ for a map $f : X \to X$ where the sequence of torus maps $\{f_i\}_1^l$ come from consecutive applications of the Fadell-Husseini fibrations.

Let us start with some more general remarks. Recall that a matrix $A \in \mathcal{M}_{r\times r}(\mathbb{Z})$ of a self-map of compact nilmanifold X according to Theorem 2.1 is given by the following procedure: Suppose that a self-map $f : X \to X$, $\dim X = r$ is a fiber map given by the Fadell-Husseini theorem i.e., a map such that the diagram

$$\begin{array}{ccc} T^{s_1} & \xrightarrow{f_1} & T^{s_1} \\ {\scriptstyle \iota}\big\downarrow & & {\scriptstyle \iota}\big\downarrow \\ X & \xrightarrow{f} & X \\ {\scriptstyle p}\big\downarrow & & {\scriptstyle p}\big\downarrow \\ \overline{X} & \xrightarrow{\bar{f}} & \overline{X} \end{array}$$

commutes. Then $f_1 : T^{s_1} \to T^{s_1}$ is induced, up to homotopy, by a matrix $A_1 \in \mathcal{M}_{s_1 \times s_1}(\mathbb{Z})$. By induction on the dimension, we can assume that with $\bar{f}$ is assigned a matrix $\overline{A} = \oplus_2^l A_j$ $\overline{A} \in \mathcal{M}_{(r-s_1)\times(r-s_1)}(\mathbb{Z})$, $A_j \in \mathcal{M}_{s_j \times s_j}(\mathbb{Z})$.

Put

$$A := A_1 \oplus \overline{A} = \oplus_{i=1}^l A_j, \tag{3.1}$$

where l is the length of the given tower of consecutive Fadell-Husseini fibrations.

We begin with the following consequence of Theorem 2.1:

Proposition 3.2. *For a given map f of a compact nilmanifold X the matrix A and consequently its characteristic polynomial*

$$\chi_A(t) = \prod_1^l \chi_{A_j}(t) \in \mathbb{Z}[t]$$

depends only on the homotopy class of f.

Definition 3.3. For a given map $f : X \to X$, let $(f_1, f_2, \ldots, f_l)$ be a tower of torus maps given by the described above procedure. The number

$$s := \max_{1\le j\le l} s_j = \max_{1\le j\le l} \deg \chi_{A_j}(t)$$

we call *the size* of this tower.

Now we have to formulate a criterion to determine whether a natural number is a homotopy minimal period of a given map of nilmanifold (cf. [**8**] and [**12**] for the torus case, or [**9**] and [**10**] for the nilmanifold case).

Theorem 3.4. *Let $f : X \to X$ be a map of a compact nilmanifold X. Then $m \notin \mathrm{HPer}\,(f)$ if and only if either $N(f) = 0$ or $N(f^m) = N(f^{m/p})$ for some prime factor p of m.*

Consequently $m \in \mathrm{HPer}\,(f)$ if and only if:

a) $N(f^m) = |L(f^m)| = |\det(\mathrm{I} - A^m)| = |\chi_A^m(1)| \neq 0,$ *and*

b) *for every prime $p|m$ we have $N(f^m) > N(f^{m/p})$.*

Proof. Recall that $m \notin \mathrm{HPer}\,(f) \Leftrightarrow NP_m(f) = 0$ (by [**17**] and [**18**] for tori, by [**10**] for nilmanifolds). On the other hand $NP_m(f) = 0 \Leftrightarrow N(f) = 0$ or $N(f^m) = N(f^{m/p})$ for some prime factor p of m ([**12**] for tori, [**10**] for nilmanifolds). □

We are in position to formulate the formula which allows us to derive the sets T_A and $\mathrm{HPer}\,(f)$ of a map f with a given Fadell-Husseini tower $(f_1, f_2, \ldots, f_l)$.

Theorem 3.5. *Let $f : X \to X$ be a map of a compact nilmanifold X of dimension r. Let next $(f_1, \ldots, , f_l)$ be the tower of consecutive torus maps given by the Fadell-Husseini fibrations and $(A_1, \ldots, A_l)$ the sequence of their linearizations and $A = \oplus_1^l A_j$ the matrix corresponding to f.*

Then

$$T_A = \cap_1^l T_{A_j} \qquad \text{and}$$

$$T_A \cap (\cup_1^l \mathrm{HPer}\,(f_j)) \subset \mathrm{HPer}\,(f).$$

Proof. By the definition, $m \in T_A$ iff $\det(\mathrm{I} - A^m) = \chi_{A^m}(1) \neq 0$. But $\chi_A(1) = \prod_1^l \chi_{A_j}(1)$ which proves the first equality.

To prove the second formula, first note that $|\chi_{A^n}(1)|$ divides $|\chi_{A^m}(1)|$ if $n|m$ (provided $\chi_{A^n}(1) \neq 0$) for every integral matrix A. Consequently, by Theorem 3.4 it follows that $m \in \mathrm{HPer}\,(f)$ if $m \in T_A$ and there exists $1 \leq j_0 \leq l$ such that $|\chi_{A_{j_0}^m}(1)| > |\chi_{A_{j_0}^{m/p}}(1)|$ for every prime $p|m$, since $|\chi_{A_j^m}(1)| \geq |\chi_{A_j^{m/p}}(1)|$ for the remaining j. This shows the statement. □

As a consequence of the above theorem we have the following fact:

Corollary 3.6. *Let $f : X \to X$ be as in Theorem 3.5 and $m = p^a$ a prime power. Then $m \in \mathrm{HPer}\,(f)$ if and only if $m \in T_A \cap (\cup_1^l \mathrm{HPer}\,(f_j))$.*

Proof. By the argument of Theorem 3.5, since p is the only prime dividing m there exists $1 \leq j \leq l$ such that $|\chi_{A_j^m}(1)| > |\chi_{A_j^{m/p}}(1)|$. But this means that $m = p^a \in \mathrm{HPer}\,(f_j)$ in respect of Theorem 3.4. □

The next theorem reduces the computation of $\mathrm{HPer}\,(f)$ to $\mathrm{HPer}\,(\bar{f})$ and T_A which are given by Theorem 3.1.

Theorem 3.7. *Let X be a three dimensional compact nilmanifold different from a torus. Let $f : X \to X$ induces the pair of $(f_1, \bar{f})$ in the resulting*

Fadell-Husseini $S^1 \subset X \to T^2$ *for* X. *Let* $d = \deg f_1$. *Then*

$$\mathrm{HPer}\,(f) = \begin{cases} T_{\overline{A}} & \textit{for} \quad d \notin \{0, -1, +1, -2\} \\ \emptyset & \textit{for} \quad d = 1 \\ \mathrm{HPer}\,(\bar{f}) & \textit{for} \quad d = 0 \\ \mathrm{HPer}\,(\bar{f}) \setminus 2N & \textit{for} \quad d = -1 \\ T_{\overline{A}} \setminus \{2\} & \textit{for} \quad d = -2 \quad \textit{and} \quad 2 \notin \mathrm{HPer}\,(\bar{f}) \\ T_{\overline{A}} & \textit{for} \quad d = -2 \quad \textit{and} \quad 2 \in \mathrm{HPer}\,(\bar{f}). \end{cases}$$

Proof. We consider the following cases:

1. Let $d \notin \{0, -1, +1, -2\}$. Then $\mathrm{HPer}\,(f_1) = T_{A_1} = \mathbb{N}$. We will show that $\mathrm{HPer}\,(f) = T_{\overline{A}}$. $\subset$ is evident since $\mathrm{HPer}\,(f) \subseteq T_A = T_{\overline{A}} \cap T_{A_1} = T_{\overline{A}}$. On the other hand, Theorem 3.5 implies $\mathrm{HPer}\,(f) \supset T_A \cap \{\mathrm{HPer}\,(\bar{f}) \cup \mathrm{HPer}\,(f_1)\} = T_A = T_{\overline{A}}$, which gives $\supset$.

2. Let $d = 1$. Then $\mathrm{HPer}\,(f) = \emptyset$ by Theorem 2.3.

3. Let $d = 0$. Then $\chi_k(t) = t\overline{\chi}_k(t)$ gives $N(f^k) = |\chi_k(1)| = |\overline{\chi}_k(1)| = N(\bar{f}^k)$ which implies

$$\mathrm{HPer}\,(f) = \mathrm{HPer}\,(\bar{f}).$$

4. Let $d = -1$. We will show that $\mathrm{HPer}\,(f) = \mathrm{HPer}\,(\bar{f}) \setminus 2\mathbb{N}$. We notice that $T_{A_1} = \mathbb{N} \setminus 2\mathbb{N}$ and $\mathrm{HPer}\,(f_1) = \{1\}$. Now $\chi_k(t) = (t - (-1)^k)\overline{\chi}_k(t)$ gives

$$N(f^k) = \begin{cases} 2N(\bar{f}^k) & \text{for k odd,} \\ 0 & \text{for k even.} \end{cases}$$

By Theorem 3.4 notice that no even number k belongs to $\mathrm{HPer}\,(f)$ since $N(f^k) = 0$ and that any odd number k will either belong to both of $\mathrm{HPer}\,(\bar{f})$ and $\mathrm{HPer}\,(f)$ or neither of these since $N(f^k) = 2N(\bar{f}^k)$. Consequently

$$\mathrm{HPer}\,(f) = \mathrm{HPer}\,(\bar{f}) \setminus 2\mathbb{N}$$

in this case.

5. Let $d = -2$. Now by definition $T_{A_1} = \mathbb{N}$ and from Theorem 1.2 $\mathrm{HPer}\,(f_1) = \mathbb{N} \setminus \{2\}$. Then $T_A = T_{\overline{A}}$, and by Theorem 3.5 $\mathrm{HPer}\,(f) \supset T_A \cap \{\mathrm{HPer}\,(f_1) \cup \mathrm{HPer}\,(\bar{f})\} = T_{\overline{A}} \cap \{(\mathbb{N} \setminus \{2\}) \cup \mathrm{HPer}\,(\bar{f})\}$.

Consequently we have

$$\mathrm{HPer}\,(f) = \begin{cases} T_{\overline{A}} \setminus \{2\} & \text{for } 2 \notin \mathrm{HPer}\,(\bar{f}) \\ T_{\overline{A}} & \text{for } 2 \in \mathrm{HPer}\,(\bar{f}). \end{cases}$$

□

Proof of Theorem 3.1. We shall use Theorems 1.2, 1.3, 3.7, and Proposition 2.12. From Proposition 2.12 it follows that $d = b$. At first we notice that

$$\begin{aligned}\mathrm{HPer}\,(f) = \emptyset &\iff \mathrm{HPer}\,(\bar f) = \emptyset \quad \text{or} \quad \mathrm{HPer}\,(f_1) = \emptyset \\ &\iff \det(\overline{A}) = 0 \quad \text{or} \quad d = 1 \\ &\iff 1 - a + d = 0 \quad \text{or} \quad d = 1.\end{aligned}$$

We will assume now that $\mathrm{HPer}\,(f) \neq \emptyset$. Suppose first that $d \notin \{-2, -1, 0, 1\}$. From Theorem 1.3 it follows that $T_{A_1} = \mathbb{N}$ and consequently $\mathrm{HPer}\,(f) = T_A = T_{\overline{A}}$ by Theorem 3.7. Now we look for the case $T_{\overline{A}} \neq \mathbb{N}$ and $d \notin \{-2, -1, 0, 1\}$ in the tables of Theorem 1.3. The second condition does not hold if $\bar f \in$ (F). On the other hand if $\bar f \in$ (G) then the first condition holds iff (a, b) satisfies 1^0) i.e., $a \neq 0$, and $a + d + 1 = 0$. $\mathrm{HPer}\,(f) = T_A = T_{\overline{A}} = \mathbb{N} \setminus 2\mathbb{N}$ then. This gives the first row of the table of special cases (G) of the statement.

Let $d = 0$. Then $\mathrm{HPer}\,(f) = \mathrm{HPer}\,(\bar f)$ by Theorem 3.7. Looking at the tables of Theorem 1.3 we get the two triples which gives the case (F) of the statement. Moreover, substituting $b = d = 0$ to the special cases 1^0, 2^0, 3^0 of (G) of Theorem 1.3, deriving a, and excluding pairs (a, b) that have been already listed we get the second row of the case (G) of the statement.

Let $d = -1$. Then $T_{A_1} = \mathbb{N} \setminus 2\mathbb{N}$ and $\mathrm{HPer}\,(f_1) = \{1\}$. Thus $T_A = T_{\overline{A}} \setminus 2\mathbb{N}$. On the other hand $\mathrm{HPer}\,(f) \supset T_A \cap \mathrm{HPer}\,(\bar f) = \mathrm{HPer}\,(\bar f) \setminus 2\mathbb{N}$. Now looking at the tables of Theorem 1.3 we notice that $b = d = -1$ may occur only in (G). But even then $\mathrm{HPer}\,(\bar f) \supset \mathbb{N} \setminus 2\mathbb{N}$. Thus $\mathbb{N} \setminus 2\mathbb{N} = T_A \supset \mathrm{HPer}\,(f) \supset \mathrm{HPer}\,(\bar f) \setminus 2\mathbb{N} \supset \mathbb{N} \setminus 2\mathbb{N}$ implies $\mathrm{HPer}\,(f) = \mathbb{N} \setminus 2\mathbb{N}$. On the other hand we notice that this occurs exactly for $(1, -1)$ and $(-1, -1)$. This gives the third and the fourth line of the exceptional cases in 3.1 (G).

Let $d = -2$. By Theorem 3.7

$$\mathrm{HPer}\,(f) = \begin{cases} T_{\overline{A}} \setminus \{2\} & \text{for } 2 \notin \mathrm{HPer}\,(\bar f) \\ T_{\overline{A}} & \text{for } 2 \in \mathrm{HPer}\,(\bar f). \end{cases}$$

Lemma 3.8 shows that $2 \notin \mathrm{HPer}\,(\bar f)$ iff $a = 0, 1, 2$. In all remaining cases ($d = -2, \mathrm{HPer}\,(\bar f) \neq \emptyset$) $\mathrm{HPer}\,(f) = T_{\overline{A}}$. In Theorem 1.3 we look for the cases $T_{\overline{A}} \neq \mathbb{N}$ (with $d = -2$). This is possible only for $(a, d) = (1, -2), (2, -2), (0, -2)$ the three exceptional cases discussed in Lemma 3.8. In all remaining cases $\mathrm{HPer}\,(f) = T_{\overline{A}} = \mathbb{N}$. Now the three last lines in the table in Theorem 3.1 (G) follow from Lemma 3.8.

We are left with the task to prove that for every pair of sets listed as (T_A, and $\mathrm{HPer}\,(f)$) in the statement of Theorem 3.1 there exists a map $f : \mathbb{N}_3(\mathbb{R})/\Gamma_{p,q,r} \to \mathbb{N}_3(\mathbb{R})/\Gamma_{p,q,r}$ which gives this pair. Fix a, $b = d \in \mathbb{Z}$. For given a, b we define an integral matrix

$$A = \begin{bmatrix} b & 0 & 0 \\ 0 & a & b \\ 0 & -1 & 0 \end{bmatrix}.$$

By Proposition 2.12 A defines a homomorphisms of $\Gamma_{p,q,r}$ and hence a map of $\mathbb{N}_3(\mathbb{R})/\Gamma_{p,q,r}$ whose linearization is equal $A = A_1 \oplus \overline{A}$. We have $\det \overline{A} = b$, and $\operatorname{tr} \overline{A} = a$, which proves the theorem. □

An elementary consideration gives the lemma below, which verification is left to the reader.

Lemma 3.8. *If $d = -2$ and $\mathrm{HPer}\,(f) \neq \emptyset$ then $2 \notin \mathrm{HPer}\,(f) \iff a = 0, 1, 2$. Moreover:*

- *$T_A = \mathrm{HPer}\,(f) = \mathbb{N} \setminus 2\mathbb{N}$ for $a = 1$,*
- *$T_A = \mathbb{N}$, $\mathrm{HPer}\,(f) = \mathbb{N} \setminus \{2\}$ for $a = 0$ or $a = 2$.*

As a consequence of Theorem 3.1 we get the following:

Corollary 3.9. *If a self map of a 3-nilmanifold different than 3-torus is such that $3 \in \mathrm{HPer}\,(f)$ then $\mathbb{N} \setminus 2\mathbb{N} \subset \mathrm{HPer}\,(f) \subset \mathrm{Per}(f)$. If $2 \in \mathrm{HPer}\,(f)$ then $\mathbb{N} = \mathrm{HPer}\,(f) = \mathrm{Per}(f)$. In particular, the first assumption is satisfied if $L(f^3) \neq L(f)$ and the second if $L(f^2) \neq L(f)$.*

Proof. By Theorem 3.1, $\mathrm{HPer}\,(f)$ finite implies $\mathrm{HPer}\,(f) \subset \{1\}$. Thus $3 \in \mathrm{HPer}\,(f)$ implies case (G) hence $\mathrm{HPer}\,(f) \supset \mathbb{N} \setminus 2\mathbb{N}$. If $2 \in \mathrm{HPer}\,(f)$ then the special cases in Theorem 3.1 are excluded hence $\mathrm{HPer}\, f = \mathbb{N}$. □

Remark 3.10. It is easy to note that one may modify Theorem 3.1 to a nilmanifold of any dimension provided the size of its Fadell-Husseini tower is less or equal to two. If the size of tower is less or equal to three these approach should still work due to the complete description of the homotopy minimal periods of the three torus maps done by Jiang and Llibre in [**12**].

Remark 3.11. Roughly speaking Corollary 3.9 is a Šarkovskii type theorem. Instead of the existence of an orbit of a given length (here 2 or 3) we need a stronger assumption 2 or $3 \in \mathrm{HPer}\,(f)$. However the conclusion is also stronger, because it states the existence of homotopy minimal periods.

Remark 3.12. The next natural and possible to achieve case is a description of minimal homotopy periods of maps of some low dimensional compact solvmanifolds especially of dimension 4. The latter needs a slight modifications of theorems of [**10**] and some facts already proved in [**9**]. The possibility of non Nielsen number fibre uniformity on the associated Mostow fibrations for solvmanifolds makes the study more complicated.

4. Homeomorphisms of 3-nilmanifolds.

We will formulate a version of the last section for homeomorphisms of three dimensional nilmanifolds (see [**12**] for the corresponding theorem for a homeomorphism of the three dimensional torus).

Theorem 4.1. *Let $f : X \to X$ be a homeomorphism of three-dimensional compact nilmanifold X not diffeomorphic to T^3. Let $A = A_1 \oplus \bar{A} \in \mathcal{M}_{3\times 3}(\mathbb{Z})$ be the matrix induced by the fibre map $f = (f_1, \bar{f})$ (Proposition 2.7) and $\chi_A(t) = \chi_{A_1}(t) \cdot \chi_{\bar{A}}(t) = (t-d)(t^2 - at + b)$ its characteristic polynomial. Then $d = b = \pm 1$ and consequently $\mathrm{HPer}\,(f) = \emptyset$ iff $d = 1$ or ($d = -1$ and $a = 0$). For $d = -1$ and the remaining a we have $\mathrm{HPer}\,(f) = \mathbb{N}$ with the only two exceptions being when $a = 1$ or $a = -1$. For these special cases $T_A = \mathrm{HPer}\,(f) = \mathbb{N} \setminus 2\mathbb{N}$.*

Proof. The statement follows from Theorem 3.1 and the fact that $d = \pm 1$. ☐

As a direct consequence we get the following analog of the Šarkovskii type for a homeomorphisms of nonabelian three nilmanifolds:

Corollary 4.2. *Let $f : X \to X$ be a homeomorphism, or more general a homotopy equivalence, of a compact three dimensional nilmanifold X not diffeomorphic to the torus. If $\mathrm{HPer}\,(f) \neq \emptyset$ then $\mathbb{N}\setminus 2\mathbb{N} \subset \mathrm{HPer}\,(f)$. Moreover if $2 \in \mathrm{HPer}\,(f)$, e.g., if $L(f^2) \neq L(f)$, (or if any $2k \in \mathrm{HPer}\,(f)$) then $\mathrm{HPer}\,(f) = \mathbb{N}$.*

Acknowledgements. The authors wish to express their thanks to the referee for several helpful comments concerning the subject and form of the paper.

References

[1] L. Alsedà, S. Baldwin, J. Llibre, R. Swanson and W. Szlenk, *Minimal sets of periods for torus maps via Nielsen numbers*, Pacific J. Math., **169**(**1**) (1995), 1-32, MR 96f:55001, Zbl 0843.55004.

[2] ———, *Torus maps and Nielsen numbers*, in 'Nielsen Theory and Dynamical Systems' (Ch. McCord, ed.), Contemp. Math., **152**, Amer. Math. Soc., Providence, 1993, 1-7, CMP 1 243 466, Zbl 0804.54032.

[3] L. Auslander, *An exposition of the structure of solvmanifolds*, Bull. Amer. Math. Soc., **79**(**3**) (1973), 227-261, MR 58 #6066a, Zbl 0265.22016.

[4] L. Block, J. Guckenheimer, M. Misiurewicz and L.S. Young, *Periodic points end topological entropy of one-dimensional maps*, Lectures Notes in Mathematics, **819**, Springer Verlag, Berlin, Heidelberg-New York, 1983, 18-24, MR 82j:58097, Zbl 0447.58028.

[5] R. Brooks, R. Brown, J. Pak and D. Taylor, *The Nielsen number of maps of tori*, Proc. Amer. Math. Soc., **52** (1975), 348-400, MR 51 #11483, Zbl 0309.55005.

[6] E. Fadell and S. Husseini, *On a theorem of Anosov on Nielsen numbers for nilmanifolds*, in 'Nonlinear Functional Analysis and its Applications' (S.P. Singh, ed.), D. Reidel Publishing Company, 1986, 47-53, MR 87j:55004, Zbl 0596.58035.

[7] V.V. Gorbatsevich, A.L. Onishchik and E.B. Vinberg, *Foundations of Lie Theory and Lie Transformation Groups*, Reprint of the 1993 translation (*Lie groups and Lie algebras.* I), Encyclopaedia Math. Sci., **20**, Springer, Berlin, 1993; Springer-Verlag, Berlin, 1997, MR 99c:22009.

[8] B. Halpern, *Periodic points on tori*, Pacific J. Math., **83** (1979), 117-133, MR 81a:55001, Zbl 0438.58023.

[9] P. Heath and E. Keppelmann, *Fibre techniques in Nielsen periodic point theory on nil and solvmanifolds* I, Topology Appl., **76** (1997), 217-247, MR 98e:55004, Zbl 0881.55002.

[10] J. Jezierski and W. Marzantowicz, *Homotopy minimal periods for nilmanifold maps*, Math. Z., **239** (2002), 381-414, CMP 1 888 231.

[11] B. Jiang, *Lectures on Nielsen Fixed Point Theory*, Contemp. Math., **14**, Amer. Math. Soc., Providence, 1983, MR 84f:55002, Zbl 0512.55003.

[12] B. Jiang and J. Llibre, *Minimal sets of periods for torus maps*, Discrete Contin. Dynam. Systems, **4**(**2**) (1998), 301-320, MR 99j:58159, Zbl 0965.37019.

[13] E.C. Keppelmann and C.K. McCord, *The Anosov theorem for exponential solvmanifolds*, Pacific J. Math., **170**(**1**) (1995), 143-159, MR 96j:55005, Zbl 0856.55003.

[14] A.I. Malcev, *On a class of homogenous spaces*, Izv. Akad. Nauk SSSR, Ser Mat., **13** (1949), 9-22.

[15] C.K. McCord, *Estimating Nielsen numbers on infrasolvmanifolds*, Pacific J. Math., **147**(**1**) (1992), 345-368, MR 93d:55001, Zbl 0766.55002.

[16] M.S. Raghunathan, *Discrete Subgroups of Lie Groups*, Springer-Verlag, 1972, MR 58 #22394a, Zbl 0254.22005.

[17] C.Y. You, *The least number of periodic points on tori*, Adv. in Math. (China), **24**(**2**) (1995), 155-160, MR 96d:57040, Zbl 0833.55003.

[18] ———, *A note on periodic points on tori*, Beijing Math., **1** (1995), 224-230.

Received July 12, 2001 and revised October 3, 2001. This research was supported by KBN grant nr 2 PO3A 03315.

Institute of Applied Mathematics
University of Agriculture
ul. Nowoursynowska 166
02-787 Warszawa
Poland
E-mail address: jezierski@alpha.sggw.waw.pl

Faculty of Mathematics and Computer Science
Adam Mickiewicz University of Poznań
ul. Matejki 48/49
60-769 Poznań
Poland
E-mail address: marzan@math.amu.edu.pl

PACIFIC JOURNAL OF MATHEMATICS
Vol. 209, No. 1, 2003

SOME EISENSTEIN SERIES IDENTITIES RELATED TO MODULAR EQUATIONS OF THE SEVENTH ORDER

Zhi-Guo Liu

Dedicated to my friend Richard Lewis

In this paper we will use one well-known modular equation of seventh order, one theta function identity of S. McCullough and L.-C. Shen, 1994, and the complex variable theory of elliptic functions to prove some new septic identities for theta functions. Then we use these identities to provide new proofs of some Eisenstein series identities in Ramanujan's notebooks or "lost" notebook. We also derive a new identity for Eisenstein series and some curious trigonometric identities.

1. Introduction.

Suppose throughout that $q = \exp(2\pi i\tau)$, where τ has positive imaginary part, and set

$$(z;q)_\infty = \prod_{n=0}^{\infty}(1 - zq^n). \tag{1.1}$$

The Dedekind eta-function is defined by

$$\eta(\tau) = q^{\frac{1}{24}}(q;q)_\infty = e^{\frac{\pi i\tau}{12}}\prod_{n=1}^{\infty}(1 - e^{2\pi i n\tau}). \tag{1.2}$$

For brevity, we define

$$h(\tau) = \frac{\eta^4(7\tau)}{\eta^4(\tau)}, \quad k(\tau) = \frac{\eta^7(\tau)}{\eta(7\tau)}, \quad \text{and} \quad \rho(\tau) = 7\frac{\eta(49\tau)}{\eta(\tau)}. \tag{1.3}$$

Throughout this article we will use $\left(\frac{n}{7}\right)$ to denote the Legendre symbol.

The Eisenstein series $T(\tau), L(\tau), M(\tau)$, and $N(\tau)$ are defined by

$$T(\tau) = 1 + 2\sum_{n=1}^{\infty}\left(\frac{n}{7}\right)\frac{q^n}{1-q^n} = 1 + 2\sum_{n=1}^{\infty}\left(\frac{n}{7}\right)\frac{e^{2\pi i n\tau}}{1-e^{2\pi i n\tau}}, \tag{1.4}$$

$$L(\tau) = 1 - 24\sum_{n=1}^{\infty}\frac{nq^n}{1-q^n} = 1 - 24\sum_{n=1}^{\infty}\frac{ne^{2\pi i n\tau}}{1-e^{2\pi i n\tau}}, \tag{1.5}$$

$$M(\tau) = 1 + 240 \sum_{n=1}^{\infty} \frac{n^3 q^n}{1-q^n} = 1 + 240 \sum_{n=1}^{\infty} \frac{n^3 e^{2\pi i n\tau}}{1-e^{2\pi i n\tau}}, \tag{1.6}$$

and

$$N(\tau) = 1 - 504 \sum_{n=1}^{\infty} \frac{n^5 q^n}{1-q^n} = 1 - 504 \sum_{n=1}^{\infty} \frac{n^5 e^{2\pi i n\tau}}{1-e^{2\pi i n\tau}}. \tag{1.7}$$

In his lost notebook [**17**, p, 53], S. Ramanujan recorded without proofs formulas for $T(r\tau), L(r\tau), M(r\tau)$, and $N(r\tau)$, for certain positive integers r, as sums of quotients of Dedekind eta-functions. These particular quotients (called Hauptmoduls) frequently arise in the theory and applications of modular forms and elliptic functions. In particular, Ramanujan claimed that:

Theorem 1. *Let* $k(\tau), h(\tau), M(\tau)$ *and* $N(\tau)$ *defined by* (1.3), (1.6), *and* (1.7), *respectively. Then we have*

$$\begin{aligned} M(\tau) = k(\tau)^{4/3} &\left(1 + 245h(\tau) + 2401h^2(\tau)\right) \\ &\cdot \left(1 + 13h(\tau) + 49h^2(\tau)\right)^{1/3}, \end{aligned} \tag{1.8}$$

$$M(7\tau) = k(\tau)^{4/3} \left(1 + 5h(\tau) + h^2(\tau)\right) \left(1 + 13h(\tau) + 49h^2(\tau)\right)^{1/3}, \tag{1.9}$$

$$\begin{aligned} N(\tau) = k(\tau)^2 &\Big(1 - 7^2(5+2\sqrt{7})h(\tau) - 7^3(21+8\sqrt{7})h^2(\tau)\Big) \\ &\cdot \Big(1 - 7^2(5-2\sqrt{7})h(\tau) - 7^3(21-8\sqrt{7})h^2(\tau)\Big), \end{aligned} \tag{1.10}$$

and

$$\begin{aligned} N(7\tau) = k(\tau)^2 &\Big(1 + (7+2\sqrt{7})h(\tau) + (21+8\sqrt{7})h^2(\tau)\Big) \\ &\cdot \Big(1 + (7-2\sqrt{7})h(\tau) + (21-8\sqrt{7})h^2(\tau)\Big). \end{aligned} \tag{1.11}$$

These identities reveal deep connections between Eisenstein series and Dedekind eta-functions. The first published proofs of (1.8)-(1.11) are due to S. Raghavan and S.S. Rangachari [**16**], who used the theory of modular forms with which Ramanujan was unfamiliar. These proofs give a uniform explanation of the existence of these identities but do not provide any insight into how Ramanujan discovered the identities. These proofs are essentially verifications. It is desirable to find more natural proofs of the aforementioned identities without employing the theory of modular forms. B.C. Berndt, H.H. Chan, J. Sohn, and S.H. Son [**3**] recently found proofs of (1.8)-(1.11) based entirely on results found in Ramanujan's notebooks [**18**]. In fact, their proofs depend upon some modular equations of the seventh order of Ramanujan.

In the present paper, we present a quite different approach. Our main tools are the following three Lemmas:

Lemma 2. *The sum of all the residues of an elliptic function at the poles inside a period-parallelogram is zero.*

Lemma 3. *Let* $\theta_1(z|q)$ *be Jacobi theta function defined by* (2.1) *below. Then:*

$$(q;q)_\infty \frac{\theta_1(2z|q)}{\theta_1(z|q)} = 2\sum_{n=-\infty}^{\infty}(-1)^n q^{\frac{1}{2}(3n^2+n)}\cos(6n+1)z, \tag{1.12}$$

$$\begin{aligned}\frac{\theta_1'}{\theta_1}(x|q)+\frac{\theta_1'}{\theta_1}(y|q)+\frac{\theta_1'}{\theta_1}(z|q)-\frac{\theta_1'}{\theta_1}(x+y+z|q)\\ = \theta_1'(0|q)\frac{\theta_1(x+y|q)\theta_1(y+z|q)\theta_1(z+x|q)}{\theta_1(x|q)\theta_1(y|q)\theta_1(z|q)\theta_1(x+y+z|q)}.\end{aligned} \tag{1.13}$$

Lemma 4. *Let* $h(\tau)$ *and* $\rho(\tau)$ *be defined by* (1.3)*. Then*

$$\begin{aligned}7\rho^3(\tau)+35\rho^2(\tau)+49\rho(\tau)+(\rho^2(\tau)+7\rho(\tau)+7)\\ \cdot\sqrt{4\rho^3(\tau)+21\rho^2(\tau)+28\rho(\tau)} = 98h(\tau).\end{aligned} \tag{1.14}$$

Lemma 2 is a fundamental theorem of elliptic functions and can be found in [**5**, p. 22]. Recently, in [**9**, **10**, **11**, **12**, **13**], we have used Lemma 2 to set up many important theta function identities. Identity (1.12) is the well-known quintuple identity [**6**, **7**, **8**, **21**]. For an interesting account of this identity, one can consult [**2**, p. 83]. Identity (1.13) was derived by S. McCullough and L.-C. Shen in their remarkable paper [**14**], in which they used the properties of theta functions to study the Sezgö kernel of an annulus. Identity (1.14) is [**22**, p. 117, Equation (4.5)]. It plays a pivotal role in the study of the modular equations of degree 7.

It should be emphasized that our method is constructive and can be used to derive theta function identities and Eisenstein series identities, rather than just to verify previously derived identities. This method provides deeper insight into the theory of theta function identities and Eisenstein series identities.

In this paper we will also prove the following identities:

Theorem 5. *Let* $k(\tau), h(\tau)$*, and* $T(\tau)$ *be defined by* (1.3) *and* (1.4)*, respectively. Then we have*

$$8-7\sum_{n=1}^{\infty}\left(\frac{n}{7}\right)\frac{n^2q^n}{1-q^n} = k(\tau)(8+49h(\tau)), \tag{1.15}$$

$$T(\tau) = k(\tau)^{1/3}\left(1+13h(\tau)+49h^2(\tau)\right)^{1/3}, \tag{1.16}$$

$$A(\tau):=1+4\sum_{n=1}^{\infty}\frac{nq^n}{1-q^n}-28\sum_{n=1}^{\infty}\frac{nq^{7n}}{1-q^{7n}} \tag{1.17}$$
$$=T^2(\tau)=k(\tau)^{2/3}\left(1+13h(\tau)+49h^2(\tau)\right)^{2/3}$$

and

$$16+\sum_{n=1}^{\infty}\left(\frac{n}{7}\right)\frac{n^4q^n}{1-q^n}=k(\tau)^{5/3}(16+49h(\tau)) \tag{1.18}$$
$$\cdot\left(1+13h(\tau)+49h^2(\tau)\right)^{2/3}.$$

Equation (1.15) can also be found in [**17**, p. 53] and the first published proof of (1.15) are due to S. Raghavan [**15**], who used the theory of modular forms. Equations (1.16) and (1.17) are contained in Entry 5 (i) of Chapter 21 of Ramanujan's second notebook [**18**]. In [**2**, p, 467-473], B.C. Berndt has given proofs of (1.8) and (1.9) by using some modular equations of the seventh order. Many wonderful applications of (1.16) have been given in [**10**]. To the author's best knowledge (1.18) is a new identity.

In the course of our investigations, we obtain the following intriguing identities of theta functions:

Theorem 6. *If* $k(\tau), h(\tau)$ *and* $\rho(\tau)$ *are defined by* (1.3)*. Then we have*

$$\frac{\theta_1(\frac{2\pi}{7}|q)}{\theta_1(\frac{\pi}{7}|q)}-\frac{\theta_1(\frac{3\pi}{7}|q)}{\theta_1(\frac{2\pi}{7}|q)}+\frac{\theta_1(\frac{\pi}{7}|q)}{\theta_1(\frac{3\pi}{7}|q)}=1+\rho(\tau), \tag{1.19}$$

$$\frac{\theta_1(\frac{\pi}{7}|q)}{\theta_1(\frac{2\pi}{7}|q)}-\frac{\theta_1(\frac{2\pi}{7}|q)}{\theta_1(\frac{3\pi}{7}|q)}+\frac{\theta_1(\frac{3\pi}{7}|q)}{\theta_1(\frac{\pi}{7}|q)} \tag{1.20}$$
$$=\frac{1}{2}(3\rho(\tau)+4)+\frac{1}{2}\sqrt{4\rho^3(\tau)+21\rho^2(\tau)+28\rho(\tau)},$$

$$\frac{\theta_1^2(\frac{\pi}{7}|q)}{\theta_1(\frac{3\pi}{7}|q)}-\frac{\theta_1^2(\frac{2\pi}{7}|q)}{\theta_1(\frac{\pi}{7}|q)}+\frac{\theta_1^2(\frac{3\pi}{7}|q)}{\theta_1(\frac{2\pi}{7}|q)}=0, \tag{1.21}$$

$$\frac{\theta_1(\frac{2\pi}{7}|q)}{\theta_1^4(\frac{\pi}{7}|q)}-\frac{\theta_1(\frac{\pi}{7}|q)}{\theta_1^4(\frac{3\pi}{7}|q)}+\frac{\theta_1(\frac{3\pi}{7}|q)}{\theta_1^4(\frac{2\pi}{7}|q)}=\frac{1}{\sqrt{7}}\eta^{-2}(\tau)\eta^{-1}(7\tau)\left(8+49h(\tau)\right), \tag{1.22}$$

$$\frac{\theta_1^4(\frac{3\pi}{7}|q)}{\theta_1(\frac{\pi}{7}|q)}-\frac{\theta_1^4(\frac{\pi}{7}|q)}{\theta_1(\frac{2\pi}{7}|q)}-\frac{\theta_1^4(\frac{2\pi}{7}|q)}{\theta_1(\frac{3\pi}{7}|q)}=\sqrt{7}\eta^2(\tau)\eta(7\tau)\left(5+49h(\tau)\right), \tag{1.23}$$

$$\frac{\theta_1^7(\frac{2\pi}{7}|q)}{\theta_1^7(\frac{\pi}{7}|q)}-\frac{\theta_1^7(\frac{3\pi}{7}|q)}{\theta_1^7(\frac{2\pi}{7}|q)}+\frac{\theta_1^7(\frac{\pi}{7}|q)}{\theta_1^7(\frac{3\pi}{7}|q)}=57+2\times 7^3h(\tau)+7^4h^2(\tau), \tag{1.24}$$

$$\begin{aligned}(1.25)\qquad & \frac{\theta_1^7(\frac{\pi}{7}|q)}{\theta_1^7(\frac{2\pi}{7}|q)}-\frac{\theta_1^7(\frac{2\pi}{7}|q)}{\theta_1^7(\frac{3\pi}{7}|q)}+\frac{\theta_1^7(\frac{3\pi}{7}|q)}{\theta_1^7(\frac{\pi}{7}|q)}\\ &=289+18\times 7^3h(\tau)+19\times 7^4h^2(\tau)+7^6h^3(\tau),\end{aligned}$$

$$\begin{aligned}(1.26)\qquad & \frac{\theta_1^3(\frac{3\pi}{7}|q)}{\theta_1^6(\frac{\pi}{7}|q)}-\frac{\theta_1^3(\frac{\pi}{7}|q)}{\theta_1^6(\frac{2\pi}{7}|q)}+\frac{\theta_1^3(\frac{2\pi}{7}|q)}{\theta_1^6(\frac{3\pi}{7}|q)}\\ &=\frac{1}{\sqrt{7}}\eta^{-2}(\tau)\eta^{-1}(7\tau)\left(46+637h(\tau)+49^2h^2(\tau)\right),\end{aligned}$$

$$\begin{aligned}(1.27)\qquad & \left(\frac{\theta_1(\frac{3\pi}{7}|q)}{\theta_1^2(\frac{\pi}{7}|q)}-\frac{\theta_1(\frac{\pi}{7}|q)}{\theta_1^2(\frac{2\pi}{7}|q)}+\frac{\theta_1(\frac{2\pi}{7}|q)}{\theta_1^2(\frac{3\pi}{7}|q)}\right)^3\\ &=7\sqrt{7}\eta^{-2}(\tau)\eta^{-1}(7\tau)\left(1+13h(\tau)+49h^2(\tau)\right)\end{aligned}$$

and

$$\begin{aligned}(1.28)\qquad & \theta_1^{-7}\left(\frac{\pi}{7}\,|q\right)-\theta_1^{-7}\left(\frac{2\pi}{7}\,|q\right)-\theta_1^{-7}\left(\frac{3\pi}{7}\,|q\right)\\ &=\sqrt{7}\eta^{-14}(\tau)\eta^{-7}(7\tau)\left(1+7h(\tau)\right)\\ &\qquad\cdot\left(1+13h(\tau)+49h^2(\tau)\right)^{1/3}.\end{aligned}$$

Using the product representation of $\theta_1(z|q)$ given by (2.2) and letting $q\to 0$ in (1.19)-(1.28), we readily find the following curious trigonometric identities:

Corollary 7. *We have:*

$$(1.29)\qquad \frac{\sin(2\pi/7)}{\sin(\pi/7)}-\frac{\sin(3\pi/7)}{\sin(2\pi/7)}+\frac{\sin(\pi/7)}{\sin(3\pi/7)}=1,$$

$$(1.30)\qquad \frac{\sin(\pi/7)}{\sin(2\pi/7)}-\frac{\sin(2\pi/7)}{\sin(3\pi/7)}+\frac{\sin(3\pi/7)}{\sin(\pi/7)}=2,$$

$$(1.31)\qquad \frac{\sin^2(\pi/7)}{\sin(3\pi/7)}-\frac{\sin^2(2\pi/7)}{\sin(\pi/7)}+\frac{\sin^2(3\pi/7)}{\sin(2\pi/7)}=0,$$

$$(1.32)\qquad \frac{\sin(2\pi/7)}{\sin^4(\pi/7)}-\frac{\sin(\pi/7)}{\sin^4(3\pi/7)}+\frac{\sin(3\pi/7)}{\sin^4(2\pi/7)}=\frac{64}{7}\sqrt{7},$$

$$(1.33)\qquad \frac{\sin^4(3\pi/7)}{\sin(\pi/7)}-\frac{\sin^4(\pi/7)}{\sin(2\pi/7)}-\frac{\sin^4(2\pi/7)}{\sin(3\pi/7)}=\frac{5}{8}\sqrt{7},$$

$$(1.34)\qquad \frac{\sin^7(2\pi/7)}{\sin^7(\pi/7)}-\frac{\sin^7(3\pi/7)}{\sin^7(2\pi/7)}+\frac{\sin^7(\pi/7)}{\sin^7(3\pi/7)}=57,$$

$$(1.35)\qquad \frac{\sin^7(\pi/7)}{\sin^7(2\pi/7)}-\frac{\sin^7(2\pi/7)}{\sin^7(3\pi/7)}+\frac{\sin^7(3\pi/7)}{\sin^7(\pi/7)}=289,$$

$$\frac{\sin^3(3\pi/7)}{\sin^6(\pi/7)}-\frac{\sin^3(\pi/7)}{\sin^6(2\pi/7)}+\frac{\sin^3(2\pi/7)}{\sin^6(3\pi/7)} = \frac{368}{\sqrt{7}}, \tag{1.36}$$

$$\frac{\sin(2\pi/7)}{\sin^2(3\pi/7)}-\frac{\sin(\pi/7)}{\sin^2(2\pi/7)}+\frac{\sin(3\pi/7)}{\sin^2(\pi/7)} = 2\sqrt{7}, \tag{1.37}$$

$$\csc^7\left(\frac{\pi}{7}\right)-\csc^7\left(\frac{2\pi}{7}\right)-\csc^7\left(\frac{3\pi}{7}\right) = 2^7\sqrt{7}. \tag{1.38}$$

Equations (1.31) and (1.37) have been found by Berndt and Zhang [**4**].

The rest of the article is organized as follows: In Section 2 we introduce some basic facts about theta function $\theta_1(z|q)$. In Section 3 we prove (1.19) using the quintuple product identity. Section 4 is devoted to the proofs of (1.20) and (1.21). In Section 5 we derive (1.22) and (1.23). Sections 6 and 7 are devoted to the proofs of (1.24)-(1.28). In Section 8 we prove (1.15), (1.16), and (1.17). In Sections 9 and 10 we derive (1.8)-(1.11). Lastly, in Section 11 we prove (1.18).

2. Some basic facts about $\theta_1(z|\tau)$.

We begin with the definition of the classical theta function $\theta_1(z|q)$ [**23**, p. 464]

$$\begin{aligned}\theta_1(z|q) &= -iq^{\frac{1}{8}}\sum_{n=-\infty}^{\infty}(-1)^n q^{\frac{1}{2}n(n+1)}e^{(2n+1)iz} \\ &= 2q^{\frac{1}{8}}\sum_{n=0}^{\infty}(-1)^n q^{\frac{1}{2}n(n+1)}\sin(2n+1)z.\end{aligned} \tag{2.1}$$

Using the Jacobi triple product formula we have [**23**, p. 470]

$$\theta_1(z|q) = 2q^{\frac{1}{8}}(\sin z)(q;q)_\infty(qe^{2iz};q)_\infty(qe^{-2iz};q)_\infty. \tag{2.2}$$

Differentiating the above equation with respect to z and then putting $z=0$ we find that

$$\theta_1'(0|q) = 2q^{\frac{1}{8}}(q;q)_\infty^3 = 2\eta^3(\tau), \tag{2.3}$$

where and throughout this paper the prime means the partial derivative with respective to z.

From the definition of $\theta_1(z|q)$, the functional equations

$$\theta_1(z+\pi|q) = -\theta_1(z|q), \quad \theta_1(z+\pi\tau|q) = -q^{-1/2}e^{-2\pi iz}\theta_1(z|q) \tag{2.4}$$

can be easily verified. Differentiating the above equations with respect to z, and then setting $z=0$, we find that

$$\theta_1'(\pi|q) = -\theta_1'(0|q), \quad \theta_1'(\pi\tau|q) = -q^{-1/2}\theta_1'(0|q). \tag{2.5}$$

Taking $z = \frac{\pi}{7}, \frac{2\pi}{7}$, and $\frac{3\pi}{7}$, respectively in (2.2) and then multiplying the three resulting equations together we find that

(2.6)
$$\theta_1\left(\frac{\pi}{7}\,|q\right)\theta_1\left(\frac{2\pi}{7}\,|q\right)\theta_1\left(\frac{3\pi}{7}\,|q\right) = \sqrt{7}q^{\frac{3}{8}}(q;q)_\infty^2(q^7;q^7)_\infty = \sqrt{7}\eta^2(\tau)\eta(7\tau).$$

The Fourier series expansion for the logarithmic derivatives of $\theta_1(z|q)$ [**23**, p. 489] is

$$\frac{\theta_1'}{\theta_1}(z|q) = \cot z + 4\sum_{n=1}^{\infty}\frac{q^n}{1-q^n}\sin 2nz. \tag{2.7}$$

Substituting

$$\cot z = \frac{1}{z} - \frac{z}{3} - \frac{z^3}{45} - \frac{2z^5}{945} - \frac{z^7}{4725} + \cdots \tag{2.8}$$

and

$$\sin z = z - \frac{1}{3!}z^3 + \frac{1}{5!}z^5 - \frac{1}{7!}z^7 + \cdots \tag{2.9}$$

into (2.7) gives

$$\begin{aligned}\frac{\theta_1'}{\theta_1}(z|q) &= \frac{1}{z} - \frac{1}{3}L(\tau)z - \frac{1}{45}M(\tau)z^3 - \frac{2}{945}N(\tau)z^5 \\ &\quad - \frac{1}{4725}\left(1 + 480\sum_{n=1}^{\infty}\frac{n^7q^n}{1-q^n}\right)z^7 + \cdots .\end{aligned} \tag{2.10}$$

By the infinite products expansion for $\theta_1(z|q)$ and direct computation, we find that

$$\theta_1(7z|q^7) = -\frac{(q^7;q^7)_\infty}{(q;q)_\infty^7}\theta_1(z|q)\prod_{r=1}^{3}\theta_1\left(z - \frac{r\pi}{7}\,|q\right)\theta_1\left(z + \frac{r\pi}{7}\,|q\right). \tag{2.11}$$

We now take the logarithmic derivative of this equation and obtain

$$\sum_{r=1}^{3}\frac{\theta_1'}{\theta_1}\left(z - \frac{r\pi}{7}\,|q\right) + \sum_{r=1}^{3}\frac{\theta_1'}{\theta_1}\left(z + \frac{r\pi}{7}\,|q\right) = 7\frac{\theta_1'}{\theta_1}(7z|q^7) - \frac{\theta_1'}{\theta_1}(z|\tau). \tag{2.12}$$

Using (2.10) on the right-hand side of (2.12) yields

$$\begin{aligned}&\sum_{r=1}^{3}\frac{\theta_1'}{\theta_1}\left(z - \frac{r\pi}{7}\,|q\right) + \sum_{r=1}^{3}\frac{\theta_1'}{\theta_1}\left(z + \frac{r\pi}{7}\,|q\right) \\ &= \frac{1}{3}\left(L(\tau) - 7^2L(7\tau)\right)z + \frac{1}{45}\left(M(\tau) - 7^4M(7\tau)\right)z^3 + O(z^5).\end{aligned} \tag{2.13}$$

Differentiating with repect to z and then setting $z=0$ gives

(2.14)
$$\left(\frac{\theta_1'}{\theta_1}\right)'\left(\frac{\pi}{7}|q\right)+\left(\frac{\theta_1'}{\theta_1}\right)'\left(\frac{2\pi}{7}|q\right)+\left(\frac{\theta_1'}{\theta_1}\right)'\left(\frac{3\pi}{7}|q\right)=\frac{1}{6}\left(L(\tau)-7^2L(7\tau)\right).$$

Differentiating (2.13) with repect to z, three times, and then setting $z=0$ we obtain

$$\begin{aligned}(2.15)\qquad &\left(\frac{\theta_1'}{\theta_1}\right)'''\left(\frac{\pi}{7}|q\right)+\left(\frac{\theta_1'}{\theta_1}\right)'''\left(\frac{2\pi}{7}|q\right)+\left(\frac{\theta_1'}{\theta_1}\right)'''\left(\frac{3\pi}{7}|q\right)\\ &=\frac{1}{15}\left(M(\tau)-7^4M(7\tau)\right).\end{aligned}$$

3. The proof of (1.19).

We recall the quintuple product identity (see Lemma 3)

$$(3.1)\qquad (q;q)_\infty\frac{\theta_1(2z|q)}{\theta_1(z|q)}=2\sum_{n=-\infty}^{\infty}(-1)^nq^{\frac{1}{2}(3n^2+n)}\cos(6n+1)z.$$

When $z=0$, (3.1) reduces to the Euler identity

$$(3.2)\qquad (q;q)_\infty=\sum_{n=-\infty}^{\infty}(-1)^nq^{\frac{1}{2}(3n^2+n)}.$$

Denote

$$(3.3)\qquad s(n):=\cos\frac{(6n+1)\pi}{7}-\cos\frac{2(6n+1)\pi}{7}+\cos\frac{3(6n+1)\pi}{7}.$$

By taking $z=\frac{\pi}{7}$, $z=-\frac{2\pi}{7}$, and $z=\frac{3\pi}{7}$, respectively, in (3.1) and then adding the resulting equations we obtain

$$\begin{aligned}(3.4)\qquad &(q;q)_\infty\left\{\frac{\theta_1(\frac{2\pi}{7}|q)}{\theta_1(\frac{\pi}{7}|q)}-\frac{\theta_1(\frac{3\pi}{7}|q)}{\theta_1(\frac{2\pi}{7}|q)}+\frac{\theta_1(\frac{\pi}{7}|q)}{\theta_1(\frac{3\pi}{7}|q)}\right\}\\ &=2\sum_{n=-\infty}^{\infty}(-1)^nq^{\frac{1}{2}(3n^2+n)}s(n).\end{aligned}$$

From the following easily verified elementary trigonometric facts:

$$(3.5)\qquad s(n)=\begin{cases}-3, & n\equiv 1 \pmod 7\\ \frac{1}{2}, & n\not\equiv 1 \pmod 7),\end{cases}$$

we have the evaluation

(3.6)

$$
\begin{aligned}
&2\sum_{n=-\infty}^{\infty}(-1)^n q^{\frac{1}{2}(3n^2+n)}s(n)\\
&=2\sum_{\substack{n=-\infty\\ n\not\equiv 1 \pmod 7}}^{\infty}(-1)^n q^{\frac{1}{2}(3n^2+n)}s(n)+2\sum_{\substack{n=-\infty\\ n\equiv 1 \pmod 7}}^{\infty}(-1)^n q^{\frac{1}{2}(3n^2+n)}s(n)\\
&=\sum_{\substack{n=-\infty\\ n\not\equiv 1 \pmod 7}}^{\infty}(-1)^n q^{\frac{1}{2}(3n^2+n)}-6\sum_{\substack{n=-\infty\\ n\equiv 1 \pmod 7}}^{\infty}(-1)^n q^{\frac{1}{2}(3n^2+n)}\\
&=\sum_{n=-\infty}^{\infty}(-1)^n q^{\frac{1}{2}(3n^2+n)}+7q^2\sum_{n=-\infty}^{\infty}(-1)^n q^{\frac{1}{2}(147n^2+49n)}\\
&=(q;q)_\infty+7q^2(q^{49};q^{49})_\infty.
\end{aligned}
$$

In the last step we have used Euler's identity (3.2). Substituting the above equation into (3.4) we obtain (1.19). This completes the proof of (1.19).

4. The proofs of (1.20) and (1.21).

We first prove (1.21) and then prove (1.20).

Let

$$f(z)=\frac{\theta_1^3(z|q)}{\theta_1(z-\frac{\pi}{7}|q)\theta_1(z-\frac{2\pi}{7}|q)\theta_1(z-\frac{4\pi}{7}|q)}. \tag{4.1}$$

Using (2.4) we can easily show that $f(z)$ is an elliptic functions with periods π and $\pi\tau$. It has three simple poles $\frac{\pi}{7}$, $\frac{2\pi}{7}$, and $\frac{4\pi}{7}$ and no other poles.

Let $\operatorname{res}(f;x)$ denote the residue of $f(z)$ at x. We have the following evaluations:

$$
\begin{aligned}
\operatorname{res}\left(f;\frac{\pi}{7}\right)&=\lim_{z\to\frac{\pi}{7}}\left(z-\frac{\pi}{7}\right)f(z)\\
&=\lim_{z\to\frac{\pi}{7}}\frac{(z-\frac{\pi}{7})}{\theta_1(z-\frac{\pi}{7}|q)}\times\lim_{z\to\frac{\pi}{7}}\frac{\theta_1^3(z|q)}{\theta_1(z-\frac{2\pi}{7}|q)\theta_1(z-\frac{4\pi}{7}|q)}.
\end{aligned} \tag{4.2}
$$

By L'Hôpital's rule,

$$\lim_{z\to\frac{\pi}{7}}\frac{(z-\frac{\pi}{7})}{\theta_1(z-\frac{\pi}{7}|q)}=\frac{1}{\theta_1'(0|q)}. \tag{4.3}$$

It is plain that

$$\lim_{z\to\frac{\pi}{7}}\frac{\theta_1^3(z|q)}{\theta_1(z-\frac{2\pi}{7}|q)\theta_1(z-\frac{4\pi}{7}|q)}=\frac{\theta_1^2(\frac{\pi}{7}|q)}{\theta_1(\frac{3\pi}{7}|q)}. \tag{4.4}$$

Therefore we have

$$\operatorname{res}\left(f;\frac{\pi}{7}\right)=\frac{\theta_1^2(\frac{\pi}{7}|q)}{\theta_1'(0|q)\theta_1(\frac{3\pi}{7}|q)}. \tag{4.5}$$

In the same way we find that

$$\operatorname{res}\left(f;\frac{2\pi}{7}\right)=-\frac{\theta_1^2(\frac{2\pi}{7}|q)}{\theta_1'(0|q)\theta_1(\frac{\pi}{7}|q)}, \tag{4.6}$$

$$\operatorname{res}\left(f;\frac{4\pi}{7}\right)=\frac{\theta_1^2(\frac{3\pi}{7}|q)}{\theta_1'(0|q)\theta_1(\frac{2\pi}{7}|q)}. \tag{4.7}$$

On the other hand, Lemma 2 gives

$$\operatorname{res}\left(f;\frac{\pi}{7}\right)+\operatorname{res}\left(f;\frac{2\pi}{7}\right)+\operatorname{res}\left(f;\frac{4\pi}{7}\right)=0. \tag{4.8}$$

Substituting (4.5)-(4.7) into the above equation we obtain (1.21).

We are now ready to prove (1.20). Letting

$$a:=\frac{\theta_1(\frac{2\pi}{7}|q)}{\theta_1(\frac{\pi}{7}|q)},\quad b:=-\frac{\theta_1(\frac{3\pi}{7}|q)}{\theta_1(\frac{2\pi}{7}|q)},\quad c:=\frac{\theta_1(\frac{\pi}{7}|q)}{\theta_1(\frac{3\pi}{7}|q)}, \tag{4.9}$$

and recalling (1.3), we find that (1.19) can be rewritten as

$$a+b+c=1+\rho(\tau). \tag{4.10}$$

Using (4.4) we find that (1.21) can be written as

$$ab^2-a^2+c=0. \tag{4.11}$$

It is obvious that

$$abc=-1. \tag{4.12}$$

Multiplying (4.11) by a^{-1} and c, respectively, and then using (4.12) in the resulting equations we find that

$$bc^2-b^2+a=0, \tag{4.13}$$

$$ca^2-c^2+b=0. \tag{4.14}$$

Denote

$$Q:=ab+bc+ca,\quad P:=a+b+c=1+\rho(\tau),\quad R:=abc=-1. \tag{4.15}$$

Multiplying (4.11) by a, (4.13) by b, and (4.14) by c and then adding the resulting equations we find that

$$(a^2b^2+b^2c^2+c^2a^2)-(a^3+b^3+c^3)+ab+bc+ca=0. \tag{4.16}$$

Using the theory of elementary symmetric polynomials, we readily find that the above equation can be rewritten as

$$Q^2+(3\rho(\tau)+4)Q-(\rho^3(\tau)+3\rho^2(\tau)+\rho(\tau)-4)=0. \tag{4.17}$$

Solving the above equation for Q, we obtain

$$Q = -\frac{1}{2}(3\rho(\tau) + 4) - \frac{1}{2}\sqrt{4\rho^3(\tau) + 21\rho^2(\tau) + 28\rho(\tau)}. \tag{4.18}$$

Noting the definitions of a, b, and c, (4.9), we find that (4.18) is (1.20).

5. The proofs of (1.22) and (1.23).

Using (2.6) and (4.9) we readily find that

$$y_1 := a^3 b = -\sqrt{7}\eta^2(\tau)\eta(7\tau)\frac{\theta_1(\frac{2\pi}{7}|q)}{\theta_1^4(\frac{\pi}{7}|q)}, \tag{5.1}$$

$$y_2 := b^3 c = -\sqrt{7}\eta^2(\tau)\eta(7\tau)\frac{\theta_1(\frac{3\pi}{7}|q)}{\theta_1^4(\frac{2\pi}{7}|q)}, \tag{5.2}$$

$$y_3 := c^3 a = \sqrt{7}\eta^2(\tau)\eta(7\tau)\frac{\theta_1(\frac{\pi}{7}|q)}{\theta_1^4(\frac{3\pi}{7}|q)}. \tag{5.3}$$

From (4.11)-(4.14) and some straightforward evaluations we find that

$$y_1 y_2 = -y_1 - 1, \tag{5.4}$$

$$y_2 y_3 = -y_2 - 1, \tag{5.5}$$

$$y_3 y_1 = -y_3 - 1, \tag{5.6}$$

$$y_1 y_2 y_3 = 1. \tag{5.7}$$

We now compute $y_1 + y_2 + y_3$ and $y_1y_2 + y_2y_3 + y_3y_1$. Noting (4.12) and (4.15), we have the evaluation

$$\begin{aligned} PQ &= (a+b+c)(ab+bc+ca) \\ &= ac^2 + cb^2 + ba^2 + ab^2 + bc^2 + ca^2 - 3. \end{aligned} \tag{5.8}$$

Adding (4.11), (4.13), and (4.14), we find that

$$\begin{aligned} ab^2 + bc^2 + ca^2 &= a^2 + b^2 + c^2 - a - b - c \\ &= (a+b+c)^2 - 2(ab+bc+ca) - a - b - c \\ &= P^2 - 2Q - P. \end{aligned} \tag{5.9}$$

Substituting the above equation into (5.8), we find that

$$ac^2 + cb^2 + ba^2 = -P^2 + PQ + P + 2Q + 3. \tag{5.10}$$

Using (4.11), (4.13), (4.14), and the above equation, we readily find that

$$\begin{aligned} ab^3 + bc^3 + ca^3 &= a(c^2 - b) + b(a^2 - c) + c(b^2 - a) \\ &= ac^2 + cb^2 + ba^2 - ab - bc - ca \\ &= -P^2 + PQ + P + Q + 3. \end{aligned} \tag{5.11}$$

Employing (4.11), (4.12), (4.13), (4.14), and the above equation, we find that

$$\begin{aligned}(5.12)\qquad a^3b+b^3c+c^3a &= (a^2+b^2+c^2)(ab+bc+ca)\\ &\quad -ab^3-bc^3-ca^3+a+b+c\\ &= (P^2-2Q)Q+P^2-PQ-P-Q-3+P\\ &= P^2Q+P^2-2Q^2-PQ-Q-3.\end{aligned}$$

Therefore, by using Lemma 4, (4.10), (4.18), and the definitions of y_1, y_2, and y_3, we obtain

(5.13)

$$\begin{aligned}&y_1+y_2+y_3 = a^3b+b^3c+c^3a\\ &= P^2Q+P^2-PQ-2Q^2-Q-3\\ &= (\rho^2(\tau)+7\rho(\tau)+7)Q-2\rho^3(\tau)-5\rho^2(\tau)+6\\ &= -\frac{1}{2}(\rho^2(\tau)+7\rho(\tau)+7)\left(3\rho(\tau)+4+\sqrt{4\rho^3(\tau)+21\rho^2(\tau)+28\rho(\tau)}\right)\\ &\quad -2\rho^3(\tau)-5\rho^2(\tau)+6\\ &= -\frac{1}{2}\left(\rho^2(\tau)+7\rho(\tau)+7\right)\sqrt{4\rho^3(\tau)+21\rho^2(\tau)+28\rho(\tau)}\\ &\quad -\frac{1}{2}\left(7\rho^3(\tau)+35\rho^2(\tau)+49\rho(\tau)\right)-8\\ &= -8-49\frac{\eta^4(7\tau)}{\eta^4(\tau)} = -8-49h(\tau).\end{aligned}$$

The above equation is equivalent to (1.22).

Adding (5.4), (5.5), and (5.6) and then using the above equation we immediately have

$$\begin{aligned}(5.14)\qquad y_1y_2+y_2y_3+y_3y_1 &= -(y_1+y_2+y_3)-3\\ &= 5+49\frac{\eta^4(7\tau)}{\eta^4(\tau)} = 5+49h(\tau).\end{aligned}$$

The above equation is equivalent to (1.23).

6. The proofs of (1.24) and (1.25).

Multiplying (4.11) by ab, (4.13) by bc, (4.14) by ac, and noting the definitions of y_1, y_2, and y_3, we find that

$$(6.1)\qquad a^2b^3 = y_1+1,\quad b^2c^3 = y_2+1,\quad c^2a^3 = y_3+1.$$

Multiplying (4.11) by b^3, (4.13) by c^3, (4.14) by a^3, and using the definitions of y_1, y_2, and y_3, we obtain

$$ab^5 = a^2b^3 - y_2, \quad bc^5 = b^2c^3 - y_3, \quad ca^5 = c^2a^3 - y_1. \tag{6.2}$$

Combining (6.1) and (6.2) we have

$$ab^5 = y_1 - y_2 + 1, \quad bc^5 = y_2 - y_3 + 1, \quad ca^5 = y_3 - y_1 + 1. \tag{6.3}$$

Multiplying (4.11) by a^5, (4.13) by b^5 and (4.14) by c^5, we find that

$$a^7 = a^5c + y_1^2, \quad b^7 = b^5a + y_2^2, \quad c^7 = c^5b + y_3^2. \tag{6.4}$$

From (6.3) and (6.4) we find the following relations:

$$a^7 = y_1^2 - y_1 + y_3 + 1, \quad b^7 = y_2^2 - y_2 + y_1 + 1, \quad c^7 = y_3^2 - y_3 + y_2 + 1. \tag{6.5}$$

Using the above relations, (5.13), and (5.14), we immediately have

$$\begin{aligned} a^7 + b^7 + c^7 &= y_1^2 + y_2^2 + y_3^2 + 3 \\ &= (y_1 + y_2 + y_3)^2 - 2(y_1y_2 + y_2y_3 + y_3y_1) + 3 \\ &= (8 + 49h(\tau))^2 - 2(5 + 49h(\tau)) + 3 \\ &= 57 + 2 \times 7^3h(\tau) + 7^4h^2(\tau). \end{aligned} \tag{6.6}$$

The above equation is equivalent to (1.24).

By using (6.5) and (5.4)-(5.7) we find that

$$a^7b^7 = y_1(y_1 + 1)^2, \quad b^7c^7 = y_2(y_2 + 1)^2, \quad c^7a^7 = y_3(y_3 + 1)^2. \tag{6.7}$$

Adding the three equations together in (6.7) and then using (5.4)-(5.7), (5.13), and (5.14), we obtain

$$\begin{aligned} a^7b^7 + b^7c^7 + c^7a^7 &= y_1(y_1 + 1)^2 + y_2(y_2 + 1)^2 + y_3(y_3 + 1)^2 \\ &= (y_1 + y_2 + y_3)^3 - 3(y_1 + y_2 + y_3)(y_1y_2 + y_2y_3 + y_3y_1) \\ &\quad + 3y_1y_2y_3 + 2(y_1 + y_2 + y_3)^2 - 4(y_1y_2 + y_2y_3 + y_3y_1) \\ &\quad\quad + y_1 + y_2 + y_3 \\ &= (y_1 + y_2 + y_3)^3 + 5(y_1 + y_2 + y_3)^2 + 14(y_1 + y_2 + y_3) + 15 \\ &= -289 - 18 \times 7^3h(\tau) - 19 \times 7^4h^2(\tau) - 7^6h^3(\tau). \end{aligned} \tag{6.8}$$

The above equation is equivalent to (1.25).

7. The proofs of (1.26), (1.27) and (1.28).

Multiplying (4.11) by a^4b^2, (4.13) by b^4c^2, (4.14) by a^2c^4, and using (5.1)-(5.4), we find that

$$a^5b^4 = y_1^2 + y_1, \quad b^5c^4 = y_2^2 + y_2, \quad c^5a^4 = y_3^2 + y_3. \tag{7.1}$$

Therefore we have

$$\begin{aligned} a^5b^4 + b^5c^4 + c^5a^4 &= y_1^2 + y_1 + y_2^2 + y_2 + y_3^2 + y_3 \\ &= (y_1 + y_2 + y_3)^2 - 2(y_1y_2 + y_2y_3 + y_3y_1) \\ &\quad + y_1 + y_2 + y_3. \end{aligned} \tag{7.2}$$

Substituting (5.13) and (5.14) into the above equation we obtain

$$a^5b^4 + b^5c^4 + c^5a^4 = 46 + 13 \times 49h(\tau) + 49^2h^2(\tau). \tag{7.3}$$

The above equation is the same as (1.26).

Now we prove (1.27). By a direct evaluation,

$$\begin{aligned} &(x_1 + x_2 + x_3)^3 \\ &= x_1^3 + x_2^3 + x_3^3 + 6x_1x_2x_3 \\ &\quad + 3x_1^2x_2 + 3x_1^2x_3 + 3x_2^2x_1 + 3x_2^2x_3 + 3x_3^2x_1 + 3x_3^3x_2. \end{aligned} \tag{7.4}$$

Taking $x_1 = \sqrt[3]{y_1^2y_2}, x_2 = \sqrt[3]{y_2^2y_3}$, and $x_3 = \sqrt[3]{y_3^2y_1}$ and using (5.4)-(5.7), we obtain

$$\begin{aligned} &\left(\sqrt[3]{y_1^2y_2} + \sqrt[3]{y_2^2y_3} + \sqrt[3]{y_3^2y_1}\right)^3 \\ &= y_1^2y_2 + y_2^2y_3 + y_3^2y_1 + 3(y_1 + y_2 + y_3) \\ &\quad + 3(y_1y_2 + y_2y_3 + y_3y_4) + 6 \\ &= -y_1(y_1 + 1) - y_2(y_2 + 1) - y_3(y_3 + 1) \\ &\quad + 3(y_1 + y_2 + y_3) + 3(y_1y_2 + y_2y_3 + y_3y_4) + 6 \\ &= -y_1^2 - y_2^2 - y_3^2 + 2(y_1 + y_2 + y_3) \\ &\quad + 3(y_1y_2 + y_2y_3 + y_3y_4) + 6 \\ &= -(y_1 + y_2 + y_3)^2 - 3(y_1 + y_2 + y_3) - 9 \\ &= -49\left(1 + 13h(\tau) + 49h^2(\tau)\right). \end{aligned} \tag{7.5}$$

Noting the definitions of y_1, y_2, and y_3, we find that the above equation is equivalent to (1.27).

Finally we prove (1.28). Denote

$$\Delta := -8 - 49h(\tau). \tag{7.6}$$

Then (5.13) and (5.14) can be written in the following forms, respectively:

$$y_1 + y_2 + y_3 = \Delta \tag{7.7}$$

$$y_1y_2 + y_2y_3 + y_4y_5 = -\Delta - 3. \tag{7.8}$$

By (5.4)-(5.7), (7.7), and (7.8),

$$y_1^2 + y_2^2 + y_3^2 = \Delta^2 + 2\Delta + 6, \tag{7.9}$$

$$y_1^3 + y_2^3 + y_3^3 = \Delta^3 + 3\Delta^2 + 9\Delta + 3, , \tag{7.10}$$

$$y_1^4 + y_2^4 + y_3^4 = \Delta^4 + 4\Delta^3 + 14\Delta^2 + 16\Delta + 18, \tag{7.11}$$

$$y_1^5 + y_2^5 + y_3^5 = \Delta^5 + 5\Delta^4 + 20\Delta^3 + 35\Delta^2 + 50\Delta + 15. \tag{7.12}$$

Taking $x_1 = \sqrt[3]{y_1^5y_2}, x_2 = \sqrt[3]{y_2^5y_3}$, and $x_3 = \sqrt[3]{y_3^5y_1}$ in (7.4) and using (5.4)-(5.7), we obtain

$$\begin{aligned}
&\left(\sqrt[3]{y_1^5y_2} + \sqrt[3]{y_2^5y_3} + \sqrt[3]{y_3^5y_1}\right)^3 \\
&= y_1^5y_2 + y_2^5y_3 + y_3^5y_1 + 3(y_1^3y_3 + y_3^3y_2 + y_2^3y_1) \\
&\quad + 3(y_1^3y_2^2 + y_2^3y_3^2 + y_3^3y_1^2) + 6 \\
&= -(y_1^5 + y_2^5 + y_3^5) - (y_1^4 + y_2^4 + y_3^4) \\
&\quad + 3(y_1^3 + y_2^3 + y_3^3) + 3(y_1^2 + y_2^2 + y_3^2) + 3(y_1 + y_2 + y_3) - 3 \\
&= -(\Delta^2 + 3\Delta + 9)(\Delta + 1)^3 \\
&= 7^5(1 + 13h(\tau) + 49h^2(\tau))(1 + 7h(\tau)).
\end{aligned} \tag{7.13}$$

Substituting (5.1)-(5.3) and (7.6) into the above equation, we obtain (1.28).

8. The proofs of (1.15), (1.16) and (1.17).

We recall the following identity (see, for example, [**20**]):

$$\begin{aligned}
&\cot^2 y - \cot^2 x + 8\sum_{n=1}^{\infty} \frac{nq^n}{1-q^n}(\cos 2nx - \cos 2ny) \\
&= \theta_1'(0|q)^2 \frac{\theta_1(x-y|q)\theta_1(x+y|q)}{\theta_1^2(x|q)\theta_1^2(x|q)}.
\end{aligned} \tag{8.1}$$

Dividing both sides of this equation by $x - y$ and then letting $y \to x$, we get

$$2\cot x(1 + \cot^2 x) - 16\sum_{n=1}^{\infty} \frac{n^2q^n}{1-q^n}\sin 2nx = \theta_1'(0|q)^3 \frac{\theta_1(2x|q)}{\theta_1^4(x|q)}. \tag{8.2}$$

Taking $x=\frac{\pi}{7}, \frac{2\pi}{7}$, and $-\frac{3\pi}{7}$, respectively, in the above equation and then adding the resulting equations we get

$$s-16\sum_{n=1}^{\infty} s(n)\frac{n^2q^n}{1-q^n}=\theta_1'(0|q)^3\left(\frac{\theta_1(\frac{2\pi}{7}|q)}{\theta_1^4(\frac{\pi}{7}|q)}-\frac{\theta_1(\frac{\pi}{7}|q)}{\theta_1^4(\frac{3\pi}{7}|q)}+\frac{\theta_1(\frac{3\pi}{7}|q)}{\theta_1^4(\frac{2\pi}{7}|q)}\right). \tag{8.3}$$

Here

$$\begin{aligned} s=&\,2\cot\frac{\pi}{7}\left(1+\cot^3\frac{\pi}{7}\right)+2\cot\frac{2\pi}{7}\left(1+\cot^3\frac{2\pi}{7}\right)\\ &-2\cot\frac{3\pi}{7}\left(1+\cot^3\frac{3\pi}{7}\right), \end{aligned} \tag{8.4}$$

$$s(n)=\sin\frac{2n\pi}{7}+\sin\frac{4n\pi}{7}-\sin\frac{6n\pi}{7}. \tag{8.5}$$

Setting $q=0$ in (8.3) and then using (1.32) we have

$$s=\frac{\sin(2\pi/7)}{\sin^4(\pi/7)}-\frac{\sin(\pi/7)}{\sin^4(3\pi/7)}+\frac{\sin(3\pi/7)}{\sin^4(2\pi/7)}=\frac{64}{7}\sqrt{7}. \tag{8.6}$$

From [**13**, p. 145, Equation (7.18)] we know that

$$s(n)=\sin\frac{2n\pi}{7}+\sin\frac{4n\pi}{7}-\sin\frac{6n\pi}{7}=\frac{\sqrt{7}}{2}\left(\frac{n}{7}\right). \tag{8.7}$$

Substituting (8.6) and (8.7) into (8.3) and then using (1.22) in the resuting equation we obtain (1.15).

To prove (1.16), we recall the identity of Mccullogh and L.-C. Shen (see Lemma 3)

$$\begin{aligned} &\frac{\theta_1'}{\theta_1}(x|q)+\frac{\theta_1'}{\theta_1}(y|q)+\frac{\theta_1'}{\theta_1}(z|q)-\frac{\theta_1'}{\theta_1}(x+y+z|q)\\ &=\theta_1'(0|q)\frac{\theta_1(x+y|q)\theta_1(y+z|q)\theta_1(z+x|q)}{\theta_1(x|q)\theta_1(y|q)\theta_1(z|q)\theta_1(x+y+z|q)}. \end{aligned} \tag{8.8}$$

Taking $(x,y,z)=(\frac{\pi}{7},-\frac{3\pi}{7},-\frac{3\pi}{7}),(\frac{\pi}{7},-\frac{2\pi}{7},-\frac{2\pi}{7})$, and $(\frac{\pi}{7},\frac{\pi}{7},\frac{2\pi}{7})$, respectively, in the above equation we obtain

$$\frac{\theta_1'}{\theta_1}\left(\frac{\pi}{7}|q\right)-\frac{\theta_1'}{\theta_1}\left(\frac{2\pi}{7}|q\right)-2\frac{\theta_1'}{\theta_1}\left(\frac{3\pi}{7}|q\right)=\theta_1'(0|q)\frac{\theta_1(\frac{2\pi}{7}|q)}{\theta_1^2(\frac{3\pi}{7}|q)}, \tag{8.9}$$

$$\frac{\theta_1'}{\theta_1}\left(\frac{\pi}{7}|q\right)-2\frac{\theta_1'}{\theta_1}\left(\frac{2\pi}{7}|q\right)+\frac{\theta_1'}{\theta_1}\left(\frac{3\pi}{7}|q\right)=\theta_1'(0|q)\frac{\theta_1(\frac{\pi}{7}|q)}{\theta_1^2(\frac{2\pi}{7}|q)}, \tag{8.10}$$

$$2\frac{\theta_1'}{\theta_1}\left(\frac{\pi}{7}|q\right)+\frac{\theta_1'}{\theta_1}\left(\frac{2\pi}{7}|q\right)+\frac{\theta_1'}{\theta_1}\left(\frac{3\pi}{7}|q\right)=\theta_1'(0|q)\frac{\theta_1(\frac{3\pi}{7}|q)}{\theta_1^2(\frac{\pi}{7}|q)}. \tag{8.11}$$

Adding (8.9), (8.10), and (8.11) gives

$$\begin{aligned}(8.12)\qquad &2\left(\cot\frac{\pi}{7}+\cot\frac{2\pi}{7}-\cot\frac{3\pi}{7}\right)\\ &+8\sum_{n=1}^{\infty}\frac{q^n}{1-q^n}\left(\sin\frac{2n\pi}{7}+\sin\frac{4n\pi}{7}-\sin\frac{6n\pi}{7}\right)\\ &=\theta_1'(0|q)\left(\frac{\theta_1(\frac{3\pi}{7}|q)}{\theta_1^2(\frac{\pi}{7}|q)}-\frac{\theta_1(\frac{\pi}{7}|q)}{\theta_1^2(\frac{2\pi}{7}|q)}+\frac{\theta_1(\frac{2\pi}{7}|q)}{\theta_1^2(\frac{3\pi}{7}|q)}\right).\end{aligned}$$

Setting $q=0$ and then using (1.35), we obtain

$$(8.13)\qquad \cot\frac{\pi}{7}+\cot\frac{2\pi}{7}-\cot\frac{3\pi}{7}=\sqrt{7}.$$

Substituting (8.7), (8.13), and (1.27) into the above equation we obtain (1.16).

To prove (1.14), we construct the following elliptic function:

$$(8.14)\qquad f(z):=\frac{\theta_1(z+\frac{\pi}{7}|q)\theta_1(z+\frac{2\pi}{7}|q)\theta_1(z-\frac{3\pi}{7}|q)}{\theta_1^3(z|q)}.$$

By using (2.4), it is easy to check that $f(z)$ is an elliptic function with periods π and $\pi\tau$. Also, $f(z)$ has only one pole at 0, and its order is 3. We now compute $\operatorname{res}(f;0)$.

It is plain that

$$(8.15)\qquad \operatorname{res}(f;0)=\frac{1}{2}\left[\frac{d^2(z^3f(z))}{d^2z}\right]_{z=0}.$$

Set

$$(8.16)\qquad F(z):=z^3f(z),\quad \phi(z)=\frac{F'(z)}{F(z)}.$$

By logarithmic differentitation we easily find that

$$(8.17)\qquad \operatorname{res}(f;0)=\frac{1}{2}\left[\frac{d^2(z^3f(z))}{d^2z}\right]_{z=0}=\frac{1}{2}F(0)\left(\phi(0)^2+\phi'(0)\right).$$

Using (2.10) we find that

$$\begin{aligned}(8.18)\qquad \phi(z)&=\frac{z}{3}-3\frac{\theta_1'}{\theta_1}(z|q)+\frac{\theta_1'}{\theta_1}\left(z+\frac{\pi}{7}\,|q\right)\\ &\quad+\frac{\theta_1'}{\theta_1}\left(z+\frac{2\pi}{7}\,|q\right)+\frac{\theta_1'}{\theta_1}\left(z-\frac{3\pi}{7}\,|q\right)\\ &=L(\tau)z+\frac{\theta_1'}{\theta_1}\left(z+\frac{\pi}{7}\,|q\right)\\ &\quad+\frac{\theta_1'}{\theta_1}\left(z+\frac{2\pi}{7}\,|q\right)+\frac{\theta_1'}{\theta_1}\left(z-\frac{3\pi}{7}\,|q\right)+O(z^3).\end{aligned}$$

Setting $z=0$ and then using (8.7) and (8.13), we obtain

$$\begin{aligned}\phi(0) &= \frac{\theta_1'}{\theta_1}\left(\frac{\pi}{7}\,|q\right)+\frac{\theta_1'}{\theta_1}\left(\frac{2\pi}{7}\,|q\right)-\frac{\theta_1'}{\theta_1}\left(\frac{3\pi}{7}\,|q\right) \\ &= \left(\cot\frac{\pi}{7}+\cot\frac{2\pi}{7}-\cot\frac{3\pi}{7}\right) \\ &\quad +4\sum_{n=1}^{\infty}\frac{q^n}{1-q^n}\left(\sin\frac{2n\pi}{7}+\sin\frac{4n\pi}{7}-\sin\frac{6n\pi}{7}\right) \\ &= \sqrt{7}\left(1+2\sum_{n=1}^{\infty}\left(\frac{n}{7}\right)\frac{q^n}{1-q^n}\right). \end{aligned} \tag{8.19}$$

Differentiating (8.18) with respect to z, setting $z=0$, and using (2.14), we find that

$$\begin{aligned}\phi'(0) &= L(\tau)+\left(\frac{\theta_1'}{\theta_1}\right)'\left(\frac{\pi}{7}\,|q\right)+\left(\frac{\theta_1'}{\theta_1}\right)'\left(\frac{2\pi}{7}\,|q\right)+\left(\frac{\theta_1'}{\theta_1}\right)'\left(\frac{3\pi}{7}\,|q\right) \\ &= -7\left(1+4\sum_{n=1}^{\infty}\frac{nq^n}{1-q^n}-28\sum_{n=1}^{\infty}\frac{nq^{7n}}{1-q^{7n}}\right). \end{aligned} \tag{8.20}$$

Note that

$$F(0) = -\frac{\theta_1(\frac{\pi}{7}|q)\theta_1(\frac{2\pi}{7}|q)\theta_1(\frac{3\pi}{7}|q)}{\theta_1'(0|q)^3} \neq 0. \tag{8.21}$$

Substituting (8.19) and (8.20) into (8.17) and using Lemma 2, we find that

$$1+4\sum_{n=1}^{\infty}\frac{nq^n}{1-q^n}-28\sum_{n=1}^{\infty}\frac{nq^{7n}}{1-q^{7n}} = \left(1+2\sum_{n=1}^{\infty}\left(\frac{n}{7}\right)\frac{q^n}{1-q^n}\right)^2. \tag{8.22}$$

Combining (1.16) and (8.22) we obtain (1.17).

9. The proofs of (1.8) and (1.9).

To prove (1.8) and (1.9), we introduce the function

$$f(z) = \frac{\theta_1(2z|q)\theta_1(3z|q)}{\theta_1^6(z|q)\theta_1(7z|q^7)}. \tag{9.1}$$

By using (2.4) we readily verify that $f(z)$ is an elliptic function with periods π and $\pi\tau$. The poles of $f(z)$ are 0 and $\frac{\pi}{7}, \frac{2\pi}{7}, \ldots, \frac{6\pi}{7}$. Furthermore, $\frac{\pi}{7}, \frac{2\pi}{7}, \ldots, \frac{6\pi}{7}$ are simple poles and 0 is a pole of order 5.

From Lemma 2, we have

$$\operatorname{res}(f;0)+\sum_{k=1}^{6}\operatorname{res}\left(f;\frac{k\pi}{7}\right)=0. \tag{9.2}$$

Now,

$$\text{res}\left(f;\frac{\pi}{7}\right)=\lim_{z\to\frac{\pi}{7}}\left(z-\frac{\pi}{7}\right)f(z) \tag{9.3}$$
$$=-\frac{\theta_1(\frac{2\pi}{7}|q)\theta_1\left(\frac{3\pi}{7}\,|q\right)}{7\theta_1'(0|q^7)\theta_1^6(\frac{\pi}{7}|q)}=-\frac{1}{2\sqrt{7}}\frac{\eta^2(\tau)}{\eta^2(7\tau)}\theta_1^{-7}\left(\frac{\pi}{7}\,|q\right),$$

and we also find that

$$\text{res}\left(f;\frac{6\pi}{7}\right)=\text{res}\left(f;\frac{\pi}{7}\right)=-\frac{1}{2\sqrt{7}}\frac{\eta^2(\tau)}{\eta^2(7\tau)}\theta_1^{-7}\left(\frac{\pi}{7}\,|q\right). \tag{9.4}$$

In the same way we find that

$$\text{res}\left(f;\frac{2\pi}{7}\right)=\text{res}\left(f;\frac{5\pi}{7}\right)=\frac{1}{2\sqrt{7}}\frac{\eta^2(\tau)}{\eta^2(7\tau)}\theta_1^{-7}\left(\frac{2\pi}{7}\,|q\right), \tag{9.5}$$

$$\text{res}\left(f;\frac{3\pi}{7}\right)=\text{res}\left(f;\frac{4\pi}{7}\right)=\frac{1}{2\sqrt{7}}\frac{\eta^2(\tau)}{\eta^2(7\tau)}\theta_1^{-7}\left(\frac{3\pi}{7}\,|q\right). \tag{9.6}$$

To compute $\text{res}(f;0)$, we define

$$F(z):=z^5f(z),\qquad \phi(z):=\frac{F'(z)}{F(z)}. \tag{9.7}$$

It is plain that

$$F(0)=\frac{6}{7\theta_1'(0|q^7)\theta_1'(0|q)^4}=\frac{3}{112\eta^3(7\tau)\eta^{12}(\tau)}. \tag{9.8}$$

By an elementary calculation,

$$\text{res}(f;0)=\frac{1}{24}\left[F^{(4)}(z)\right]_{z=0} \tag{9.9}$$
$$=\frac{F(0)}{24}\left(\phi^4(0)+6\phi^2(0)\phi'(0)+4\phi(0)\phi''(0)+3\phi'(0)^2+\phi'''(0)\right).$$

Using (2.10), we find that

$$\phi(z)=\frac{5}{z}-6\frac{\theta_1'}{\theta_1}(z|q)+2\frac{\theta_1'}{\theta_1}(2z|q)+3\frac{\theta_1'}{\theta_1}(3z|q)-7\frac{\theta_1'}{\theta_1}(7z|q^7) \tag{9.10}$$
$$=\frac{7}{3}\left(7L(7\tau)-L(\tau)\right)z$$
$$+\frac{7}{45}\left(343M(7\tau)-13M(\tau)\right)z^3+O(z^5).$$

This yields

$$\phi'(0)=\frac{7}{3}\left(7L(7\tau)-L(\tau)\right)=14A(\tau),\quad \phi(0)=0,\quad \phi''(0)=0, \tag{9.11}$$
$$\phi'''(0)=\frac{14}{15}\left(343M(7\tau)-13M(\tau)\right).$$

Substituting the above equations into (9.9) we arrive at

$$\operatorname{res}(f;0) = \frac{1}{960}\eta^{-3}(7\tau)\eta^{-12}(\tau) \cdot \left(630A^2(\tau) + 343M(7\tau) - 13M(\tau)\right). \tag{9.12}$$

Substituting (9.3)-(9.6) and (9.13) into (9.2) we obtain

$$\begin{aligned} &630A^2(\tau) + 343M(7\tau) - 13M(\tau) \\ &= \frac{960}{\sqrt{7}}\eta^{14}(\tau)\eta(7\tau)\left(\theta_1^{-7}\left(\frac{\pi}{7}\,|q\right) - \theta_1^{-7}\left(\frac{2\pi}{7}\,|q\right) - \theta_1^{-7}\left(\frac{3\pi}{7}\,|q\right)\right). \end{aligned} \tag{9.13}$$

Substituting (1.28) into (9.13) we obtain the following interesting result:

Lemma 8. *We have*

$$\begin{aligned} &630A^2(\tau) + 343M(7\tau) - 13M(\tau) \\ &= 960k(\tau)^{4/3}\left(1 + 7h(\tau)\right)\left(1 + 13h(\tau) + 49h^2(\tau)\right)^{1/3}. \end{aligned} \tag{9.14}$$

From [**1**, pp. 24, 48, 69] we know that

$$\eta(-1/\tau) = \sqrt{-i\tau}\eta(\tau), \tag{9.15}$$

$$L(-1/\tau) = -\frac{6\tau i}{\pi} + \tau^2 L(\tau), \tag{9.16}$$

$$M(-1/\tau) = \tau^4 M(\tau), \tag{9.17}$$

$$N(-1/\tau) = \tau^6 N(\tau). \tag{9.18}$$

It follows that

$$\eta(-1/7\tau) = \sqrt{-7i\tau}\eta(7\tau), \tag{9.19}$$

$$A(-1/7\tau) = -7\tau^2 A(\tau), \tag{9.20}$$

$$M(-1/7\tau) = (7\tau)^4 M(7\tau), \tag{9.21}$$

$$N(-1/7\tau) = (7\tau)^6 N(7\tau), \tag{9.22}$$

$$h(-1/7\tau) = 7^{-2}h^{-1}(\tau). \tag{9.23}$$

Replacing τ by $-1/7\tau$ in (9.14) and then using (9.20), (9.21), and (9.23) in the resulting equation we deduce that:

Lemma 9. *We have*

$$\begin{aligned} &90A^2(\tau) - 91M(7\tau) + M(\tau) \\ &= 960k(\tau)^{4/3}\left(7h(\tau) + h^2(\tau)\right)\left(1 + 13h(\tau) + 49h^2(\tau)\right)^{1/3}. \end{aligned} \tag{9.24}$$

By solving the linear system of equations, (9.14) and (9.24), for $M(\tau)$ and $M(7\tau)$ we deduce the following theorem:

Theorem 10. *We have*

$$7M(7\tau) = 15A^2(\tau) - 8k(\tau)^{4/3}\left(1 + 20h(\tau) + 91h^2(\tau)\right) \cdot \left(1 + 13h(\tau) + 49h^2(\tau)\right)^{1/3} \tag{9.25}$$

$$M(\tau) = 105A^2(\tau) - 8k(\tau)^{4/3}\left(13 + 140h(\tau) + 343h^2(\tau)\right) \cdot \left(1 + 13h(\tau) + 49h^2(\tau)\right)^{1/3}. \tag{9.26}$$

Substituting (1.17) into the above equations, respectively, we obtain (1.8) and (1.9).

10. The proofs of (1.10) and (1.11).

Let

$$f(z) = \frac{\theta_1(z|q)\theta_1(2z|q^7)}{\theta_1^{11}(z|q^7)}. \tag{10.1}$$

It is easy to check that $f(z)$ is an elliptic function with periods π and $7\pi\tau$. Also, $f(z)$ has only one pole at 0, and its order is 9. From lemma 2 we have

$$\mathrm{res}(f;0) = 0. \tag{10.2}$$

Set

$$F(z) := z^9 f(z), \qquad \phi(z) := \frac{F'(z)}{F(z)}. \tag{10.3}$$

Using (2.10) we find that

$$\begin{aligned}\phi(z) &= \frac{9}{z} + \frac{\theta_1'}{\theta_1}(z|q) - 11\frac{\theta_1'}{\theta_1}(z|q^7) + 2\frac{\theta_1'}{\theta_1}(2z|q^7) \\ &= 2z - \frac{2}{15}z^3 - \frac{4}{35}z^5 - \frac{246}{4725}z^7 + 4\sum_{n=1}^{\infty}\frac{q^n}{1-q^n}\sin 2nz \\ &\quad + 4\sum_{n=1}^{\infty}\frac{q^{7n}}{1-q^{7n}}(2\sin 4nz - 11\sin 2nz) + O(z^9).\end{aligned} \tag{10.4}$$

It follows that

$$\phi'(0) = 2A(\tau), \tag{10.5}$$

$$\phi'''(0) = -\frac{2}{15}\left(M(\tau) + 5M(7\tau)\right), \tag{10.6}$$

$$\phi^{(5)}(0) = -\frac{16}{63}\left(N(\tau) + 53N(7\tau)\right), \tag{10.7}$$

$$\phi(0) = \phi''(0) = \phi^{(4)}(0) = \phi^{(6)}(0) = 0, \tag{10.8}$$

and

$$\phi^{(7)}(0) = -\frac{16}{15}\left(1 + 480\sum_{n=1}^{\infty}\frac{n^7q^n}{1-q^n} + 245 + 245 \times 480\frac{n^7q^{7n}}{1-q^{7n}}\right). \tag{10.9}$$

Employing the identity [**1**, p. 199], [**19**]

$$M^2(\tau) = 1 + 480\sum_{n=1}^{\infty}\frac{n^7q^n}{1-q^n}, \tag{10.10}$$

Equation (10.9) can be written as

$$\phi^{(7)}(0) = -\frac{16}{15}\left(M^2(\tau) + 245M^2(7\tau)\right). \tag{10.11}$$

Using the fact that $\phi(0) = \phi''(0) = \phi^{(4)}(0) = \phi^{(6)}(0) = 0$, we find that by a direct computation,

$$\begin{aligned}\operatorname{res}(f;0) = \frac{1}{8!}F(0)\Big(&105\phi'(0)^4 + 210\phi'(0)^2\phi'''(0)\\ &+ 28\phi'(0)\phi^{(5)}(0) + 35\phi'''(0)^2 + \phi^{(7)}(0)\Big).\end{aligned} \tag{10.12}$$

Substituting (10.5), (10.6), (10.7), and (10.11) into (10.12) and then using (10.1) yields

$$\begin{aligned}&N(\tau) + 53N(7\tau)\\ &= \frac{63}{8}A^2(\tau)\left(15A(\tau) - M(\tau) - 5M(7\tau)\right)\\ &\quad - \frac{1}{32A(\tau)}\left(M^2(\tau) - 14M(\tau)M(7\tau) + 553M^2(7\tau)\right).\end{aligned} \tag{10.13}$$

Replacing τ by $-1/7\tau$ in the above equation and then applying (9.20)-(9.23) in the resulting equation, we deduce that

$$\begin{aligned}&53N(\tau) + 7^6N(7\tau)\\ &= -\frac{441}{8}A(\tau)\left(15 \times 7^2A^2(\tau) - 5M(\tau) - 7^4M(7\tau)\right)\\ &\quad + \frac{1}{32A(\tau)}\left(79M^2(\tau) - 2 \times 7^4M(\tau)M(7\tau) + 7^7M^2(7\tau)\right).\end{aligned} \tag{10.14}$$

Solving the above two equations for $N(\tau)$ and $N(7\tau)$ we obtain the following lemma:

Lemma 11. *We have*

$$N(\tau)=\frac{49}{2320}A(\tau)\left(135\times 7^2A^2(\tau)-2\times 7^4M(7\tau)-388M(\tau)\right)\\ -\frac{1}{27840A(\tau)}\Big(7^7M^2(7\tau)-6\times 7^4M(\tau)M(7\tau)\\ +923M^2(\tau)\Big), \tag{10.15}$$

$$N(7\tau)=\frac{7}{2320}A(\tau)\left(-135A^2(\tau)+2M(\tau)+388M(\tau)\right)\\ +\frac{1}{27840A(\tau)}\Big(M^2(\tau)-42M(\tau)M(7\tau)\\ +6461M^2(7\tau)\Big). \tag{10.16}$$

Substituting (1.8), (1.9), and (1.16) into the above two equations, respectively, we obtain (1.10) and (1.11).

11. The proof of (1.18).

In this section we first evaluate some elementary trigonometric sums. Let ω =exp($\frac{2\pi i}{7}$). It is well-known that

$$(1-x)\prod_{r=1}^{6}(1-x\omega^r)=1-x^7. \tag{11.1}$$

It follows that for $x\neq 1$,

$$\left(1-2x\cos\frac{2\pi}{7}+x^2\right)\left(1-2x\cos\frac{4\pi}{7}+x^2\right)\left(1-2x\cos\frac{6\pi}{7}+x^2\right)\\ =\frac{1-x^7}{1-x}. \tag{11.2}$$

Letting $x\to 1$ gives

$$2^6\sin^2\frac{\pi}{7}\sin^2\frac{2\pi}{7}\sin^2\frac{3\pi}{7}=7, \tag{11.3}$$

and from this we obtain

$$\sin\frac{\pi}{7}\sin\frac{2\pi}{7}\sin\frac{3\pi}{7}=\frac{1}{8}\sqrt{7}. \tag{11.4}$$

Similarly, setting $x=-1$ in (11.2), we have

$$\cos\frac{\pi}{7}\cos\frac{2\pi}{7}\cos\frac{3\pi}{7}=\frac{1}{8}. \tag{11.5}$$

Combining the above two equations we obtain

$$\cot\frac{\pi}{7}\cot\frac{2\pi}{7}\cot\frac{3\pi}{7}=\frac{1}{\sqrt{7}}. \tag{11.6}$$

We recall the identity (see (8.13))

$$\cot\frac{\pi}{7}+\cot\frac{2\pi}{7}-\cot\frac{3\pi}{7}=\sqrt{7}. \tag{11.7}$$

Taking $q=0$ in (2.14), we obtain

$$\cot^2\frac{\pi}{7}+\cot^2\frac{2\pi}{7}+\cot^2\frac{3\pi}{7}=5. \tag{11.8}$$

From (11.6), (11.7), and (11.8), we readily find that

$$\cot\frac{\pi}{7},\quad \cot\frac{2\pi}{7},\quad \text{and}\quad -\cot\frac{3\pi}{7} \tag{11.9}$$

are the roots of cubic equation

$$x^3-\sqrt{7}x^2+x+\frac{1}{\sqrt{7}}=0. \tag{11.10}$$

Let

$$s_n=\cot^n\frac{\pi}{7}+\cot^n\frac{2\pi}{7}+(-1)^n\cot^n\frac{3\pi}{7}. \tag{11.11}$$

Then from (11.10) we obtain the following recurrence formula:

$$s_{n+3}=\sqrt{7}s_{n+2}-s_{n+1}-\frac{1}{\sqrt{7}}s_n,\quad s_0=3,\quad s_1=\sqrt{7},\quad s_2=5. \tag{11.12}$$

It follows that

$$s_3=\frac{25}{\sqrt{7}},\quad s_4=19,\quad s_5=\frac{103}{\sqrt{7}}. \tag{11.13}$$

It can be easily verified that

$$\cot^{(4)}x=16\cot x+40\cot^3 x+24\cot^5 x. \tag{11.14}$$

Therefore we have

$$\cot^{(4)}\frac{\pi}{7}+\cot^{(4)}\frac{2\pi}{7}+\cot^{(4)}\frac{3\pi}{7}=16s_1+40s_3+24s_5=\frac{3584}{\sqrt{7}}. \tag{11.15}$$

Now we begin to prove (1.18). Using (2.4) we can verify that

$$f(z)=\frac{\theta_1(2z|q)\theta_1(z+\frac{\pi}{7}|q)\theta_1(z+\frac{2\pi}{7}|q)\theta_1(z-\frac{3\pi}{7}|q)}{\theta_1^7(z|q)} \tag{11.16}$$

is an elliptic function with only one pole, namely, at 0 with order 6.

Set

$$F(z):=z^6f(z),\qquad \phi(z):=\frac{F'(z)}{F(z)}. \tag{11.17}$$

We find that

$$\begin{aligned}(11.18)\qquad \phi(z) &= \frac{6}{z} - 7\frac{\theta_1'}{\theta_1}(z|q) + 2\frac{\theta_1'}{\theta_1}(z|q)\\ &\quad + \frac{\theta_1'}{\theta_1}\left(z+\frac{\pi}{7}\,|q\right) + \frac{\theta_1'}{\theta_1}\left(z+\frac{2\pi}{7}\,|q\right) + \frac{\theta_1'}{\theta_1}\left(z-\frac{3\pi}{7}\,|q\right)\\ &= L(\tau)z - \frac{z^3}{5}M(\tau) + \frac{\theta_1'}{\theta_1}\left(z+\frac{\pi}{7}\,|q\right)\\ &\quad + \frac{\theta_1'}{\theta_1}\left(z+\frac{2\pi}{7}\,|q\right) + \frac{\theta_1'}{\theta_1}\left(z-\frac{3\pi}{7}\,|q\right) + O(z^5).\end{aligned}$$

Setting $z=0$ and then using (8.19), we find that

$$\begin{aligned}(11.19)\qquad \phi(0) &= \frac{\theta_1'}{\theta_1}\left(\frac{\pi}{7}\,|q\right) + \frac{\theta_1'}{\theta_1}\left(\frac{2\pi}{7}\,|q\right) - \frac{\theta_1'}{\theta_1}\left(\frac{3\pi}{7}\,|q\right)\\ &= \sqrt{7}\left(1 + 2\sum_{n=1}^{\infty}\left(\frac{n}{7}\right)\frac{q^n}{1-q^n}\right).\end{aligned}$$

Differentiating (11.18) with repect to z and then setting $z=0$ and finally using (2.14), we obtain

$$\begin{aligned}(11.20)\quad \phi'(0) &= L(\tau) + \left(\frac{\theta_1'}{\theta_1}\right)'\left(\frac{\pi}{7}\,|q\right) + \left(\frac{\theta_1'}{\theta_1}\right)'\left(\frac{2\pi}{7}\,|q\right) + \left(\frac{\theta_1'}{\theta_1}\right)'\left(\frac{3\pi}{7}\,|q\right)\\ &= -7A(\tau).\end{aligned}$$

Differentiating (11.18) twice with repect to z , seting $z=0$, and using (8.6) and (8.7), we obtain

$$\begin{aligned}(11.21)\qquad \phi''(0) &= \left(\frac{\theta_1'}{\theta_1}\right)''\left(\frac{\pi}{7}\,|q\right) + \left(\frac{\theta_1'}{\theta_1}\right)''\left(\frac{2\pi}{7}\,|q\right) - \left(\frac{\theta_1'}{\theta_1}\right)''\left(\frac{3\pi}{7}\,|q\right)\\ &= \frac{8}{\sqrt{7}}\left(8 - 7\sum_{n=1}^{\infty}\left(\frac{n}{7}\right)\frac{n^2q^n}{1-q^n}\right).\end{aligned}$$

Using (2.15), we find that

$$\begin{aligned}(11.22)\qquad \phi'''(0) &= -\frac{6}{5}M(\tau) + \left(\frac{\theta_1'}{\theta_1}\right)'''\left(\frac{\pi}{7}\,|q\right)\\ &\quad + \left(\frac{\theta_1'}{\theta_1}\right)'''\left(\frac{2\pi}{7}\,|q\right) + \left(\frac{\theta_1'}{\theta_1}\right)'''\left(\frac{3\pi}{7}\,|q\right)\\ &= -\frac{1}{15}\left(7M(\tau) + 2401M(7\tau)\right).\end{aligned}$$

From (11.16) and (8.7), we have

$$
\begin{aligned}
(11.23)\quad \phi^{(4)}(0) &= \left(\frac{\theta_1'}{\theta_1}\right)^{(4)}\left(\frac{\pi}{7}\,|q\right) + \left(\frac{\theta_1'}{\theta_1}\right)^{(4)}\left(\frac{2\pi}{7}\,|q\right) - \left(\frac{\theta_1'}{\theta_1}\right)^{(4)}\left(\frac{3\pi}{7}\,|q\right) \\
&= \cot^{(4)}\frac{\pi}{7} + \cot^{(4)}\frac{2\pi}{7} + \cot^{(4)}\frac{3\pi}{7} \\
&\quad + 64\sum_{n=1}^{\infty}\frac{n^4q^n}{1-q^n}\left(\sin\frac{2n\pi}{7} + \sin\frac{4n\pi}{7} - \sin\frac{6n\pi}{7}\right) \\
&= 32\sqrt{7}\left(16 + \sum_{n=1}^{\infty}\left(\frac{n}{7}\right)\frac{n^4q^n}{1-q^n}\right).
\end{aligned}
$$

By logarithmic differentiation we find that

$$
\begin{aligned}
(11.24)\qquad \operatorname{res}(f;0) = \frac{1}{120}F(0)\Big(&\phi(0)^5 + 10\phi(0)^3\phi'(0) + 5\phi(0)\phi'''(0) \\
&+ 10\phi(0)^2\phi''(0) + 15\phi(0)\phi'(0)^2 \\
&\quad + 10\phi'(0)\phi''(0) + \phi^{(4)}(0)\Big).
\end{aligned}
$$

Substituting (11.19)-(11.23) into the above equation and then using (8.19) in the resulting equation and finally using the fact that $\operatorname{res}(f;0) = 0$, we obtain

$$
\begin{aligned}
(11.25)\qquad & 96\left(16 + \sum_{n=1}^{\infty}\left(\frac{n}{7}\right)\frac{n^4q^n}{1-q^n}\right) \\
&= \left(1 + 2\sum_{n=1}^{\infty}\left(\frac{n}{7}\right)\frac{q^n}{1-q^n}\right)(17M(\tau) + 2401M(7\tau)) \\
&\quad - 882\left(1 + 2\sum_{n=1}^{\infty}\left(\frac{n}{7}\right)\frac{q^n}{1-q^n}\right)^5.
\end{aligned}
$$

Substituting (1.8), (1.9), and (1.16) into the above equation we immediately obtain (1.18).

Acknowledgements. This work is supported, in part, by Academic Research Fund R146000027112 from the National University of Singapore. The author is grateful to the referee for his many helpful suggestions leading to an improvement of an earlier version of this paper. He would also like to thank Heng Huat Chan, Li-Chien Shen, and Yifan Yang for their comments.

References

[1] T.M. Apostol, *Modular Functions and Dirichlet Series in Number Theory*, Springer-Verlag, New York, 1976, MR 54 #10149, Zbl 0332.10017.

[2] B.C. Berndt, *Ramanujan's Notebooks*, Part III, Springer-Verlag, New York, 1991, MR 92j:01069, Zbl 0733.11001.

[3] B.C. Berndt, H.H. Chan, J. Sohn and S.H. Son, *Eisenstein series in Ramanujan's lost notebook*, Ramanujan J., **4** (2000), 81-114, MR 2001j:11018.

[4] B.C. Berndt and L.-C. Zhang, *A new class of theta-function identities originating in Ramanujan's notebooks*, J. Number Theory, **48** (1994), 224-242, MR 95m:11046, Zbl 0812.11028.

[5] K. Chandrasekharan, *Elliptic Functions*, Springer-Verlag, New York, 1985, MR 87e:11058, Zbl 0575.33001.

[6] R.J. Evans, *Theta function identities*, J. Math. Anal. Appl., **147** (1990), 97-121, MR 91d:11045, Zbl 0707.11033.

[7] F.G. Garvan, *A simple proof of Watson's partition congruences for powers of* 7, J. Austral. Math. Soc. (Series A), **36** (1984), 316-334, MR 85f:11072, Zbl 0533.10006.

[8] M.D. Hirschhorn and D.C. Hunt, *A simple proof of the Ramanujan conjecture for powers of* 5, J. Reine Angew. Math., **326** (1981), 1-17, MR 82m:10023, Zbl 0452.10015.

[9] Z.-G. Liu, *The Borweins' cubic theta function identity and some cubic modular identities of Ramanujan*, Ramanujan J., **4** (2000), 43-50, MR 2001f:33026.

[10] ———, *On certain identities of Ramanujan*, J. Number Theory, **83** (2000), 59-75, MR 2001f:11066, Zbl 0981.11016.

[11] ———, *Some Eisenstein series identities*, J. Number Theory, **85** (2000), 231-252, MR 2001k:11075, Zbl 0973.11053.

[12] ———, *Some theta function identities associated with the modular equations of degree* 5, Integers: Electronic Journal of Combinatorial Number Theory, **1** (2001), A#03, 14pp., MR 2002c:33017.

[13] ———, *Residue theorem and theta function identities*, Ramanujan J., **5** (2001), 129-151, MR 2002g:11056.

[14] S. McCullough and L.-C. Shen, *On the Szgö kernel of an annulus*, Proc. Amer. Math. Soc., **121** (1994), 1110-1120, MR 94j:30005, Zbl 0807.30005.

[15] S. Raghavan, *On certain identities due to Ramanujan*, Quart. J. Math. Oxford(2), **37** (1986), 221-229, MR 87j:11113, Zbl 0585.10033.

[16] S. Raghavan and S.S. Rangachari, *Ramanujan's elliptic integrals and modular identities*, in 'Number Theory and Related Topics', Oxford University Press, Bombay, 1989, 119-149, MR 98b:11045, Zbl 0748.33013.

[17] S. Ramanujan, *The Lost Notebook and Other Unpublished Papers*, Narosa, New Delhi, 1988, MR 89j:01078, Zbl 0639.01023.

[18] ———, *Notebooks* (2 volumes), Tata Institute of Fundamental Research, Bombay, 1957, MR 20 #6340, Zbl 0138.24201.

[19] ———, *On certain arithmetical functions*, Trans. Cambridge Philos. Soc., **22** (1916), 159-184.

[20] L.-C. Shen, *On the additive formula of the theta functions and a collection of Lambert series pertaining to the modular equations of degree* 5, Trans. Amer. Math. Soc., **345** (1994), 323-345, MR 95a:33038, Zbl 0808.33014.

[21] ———, *On the products of three theta functions*, Ramanujan J., **3** (1999), 343-357, MR 2000k:33038.

[22] G.N. Watson, *Ramanujan Vermutung üeber Zerfällungsanzahlen*, J. Reine Angew. Math., **179** (1938), 97-128.

[23] E.T. Whittaker and G.N. Watson, *A Course of Modern Analysis*, 4th ed. Cambridge Univ. Press, Cambridge, 1966, MR 97k:01072, Zbl 0951.30002.

Received August 28, 2001 and revised October 22, 2001.

Department of Mathematics
East China Normal University
Shanghai 200062
P. R. of China
E-mail address: liuzg18@hotmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 209, No. 1, 2003

IRREDUCIBLE NUMERICAL SEMIGROUPS

J.C. Rosales and M.B. Branco

We give a characterization for irreducible numerical semigroups. From this characterization we obtain that every irreducible numerical semigroup is either a symmetric or pseudo-symmetric numerical semigroup. We study the minimal presentations of an irreducible numerical semigroup. Separately, we deal with the cases of maximal embedding dimension and multiplicity 3 and 4.

1. Introduction and basic concepts.

A **numerical semigroup** is a subset S of $\mathbb{N}$ closed under addition, $0 \in S$ and generates $\mathbb{Z}$ as a group (here $\mathbb{N}$ and $\mathbb{Z}$ denote the set of nonnegative integers and the set of the integers, respectively). From this definition we obtain (see [**2**] and [**13**]) the following results.

(1) The set $\mathbb{N} \setminus S$ is finite, we refer to the greatest integer not belonging to S as the **Frobenius number** of S and denote it by $g(S)$.
(2) The semigroup S has a unique minimal system of generators $\{n_0 < n_1 < \cdots < n_p\}$. We refer to the numbers n_0 and $p+1$ as the **multiplicity** and **embedding dimension** of S and denote them by $m(S)$ and $\mu(S)$, respectively.

Let $F = \{a_0X_0 + \cdots + a_pX_p : a_0, \ldots, a_p \in \mathbb{N}\}$ be the free monoid generated by $\{X_0, \ldots, X_p\}$ and let $\varphi : F \to S$ be the monoid epimorphism defined by

$$\varphi(a_0X_0 + \cdots + a_pX_p) = a_0n_0 + \cdots + a_pn_p.$$

It is well-known that if σ is the kernel congruence of φ (that is, $x\sigma y$ if $\varphi(x) = \varphi(y)$), then S is isomorphic to the quotient monoid F/σ (see [**13**]). Rédei shows in [**9**] that the congruence σ is finitely generated and therefore there exists

$$\rho = \{(x_1, y_1), \ldots, (x_t, y_t)\} \subseteq F \times F$$

such that σ is the smallest congruence on F that contains ρ. The set ρ is called a **presentation** for the numerical semigroup S. We say that ρ is **minimal presentation** if no proper subset of ρ generates σ. In [**10**] it is shown that the concepts of minimal presentation and presentation with the lowest cardinality coincide for a numerical semigroup.

Numerical semigroups have been widely studied in the literature not only from the semigroupist point of view but also to give us a series of examples in ring theory through the concept of the semigroup ring associated to a numerical semigroup (see for instance [**7**], [**4**], [**5**], [**8**], [**15**]). Along this line, if K is a field, $K[S]$ is the finite type K-algebra associated to S and $K[X_0, \ldots, X_p]$ is the polynomial ring in $p+1$ indeterminates, the K-algebras epimorphism $\lambda : K[X] \to K[S]$ such that $X_i \mapsto t^i$ is a S-graded ring homomorphism with degree zero. Therefore, the prime ideal $P = \text{kernel}(\lambda)$ (called the ideal associated to the semigroup) is homogeneous and defines a monomial curve in the $(p+1)$-dimensional affine space on K. Herzog proves in [**7**] that finding a system of generators for P is equivalent to finding a presentation for S. Let us also notice that Kunz in [**8**] proves that $K[S]$ is Gorestein if and only if S is **symmetric** and Barucci-Dobbs-Fontana prove in [**2**] that $K[S]$ is Kunz if and only if S is **pseudo-symmetric**.

We say that a numerical semigroup is **irreducible** if it can not be expressed as an intersection of two numerical semigroups containing it properly. In Theorem 1, we see that S is irreducible if and only if S is maximal in the set of all numerical semigroups with Frobenius number $g(S)$. From [**6**] and [**2**] we deduce that the class of irreducible semigroups with odd (respectively even) Frobenius number is the same as the class of symmetric (respectively pseudo-symmetric) numerical semigroups. Moreover, [**11**] provides a study of the irreducible numerical semigroups with odd Frobenius number. Our aim in this paper is to generalize these results for irreducible numerical semigroups in general (that is, with Frobenius number even or odd).

The contents in this paper are organized as follows. In Section 2 we characterize the irreducible numerical semigroups giving special attention to their Apéry sets. In Section 3 we study the irreducible numerical semigroups with multiplicity 3 and 4. We explicitly give the family of irreducible numerical semigroups of this kind. The aim of Section 4 is to give an upper bound for the cardinal of a minimal presentation for an irreducible numerical semigroup in function of their multiplicity and embedding dimension. Finally, in Section 5 we study those irreducible numerical semigroups having multiplicity greater than or equal to five and embedding dimension equal to its multiplicity minus one.

2. Characterization of irreducible numerical semigroups.

Throughout this section S denotes a numerical semigroup, such that $S \neq \mathbb{N}$. It is well-known (see for instance [**12**]) that $S \cup \{g(S)\}$ is also a numerical semigroup.

Theorem 1. *The following conditions are equivalent:*

1) *S is irreducible,*

2) *S is maximal in the set of all numerical semigroups with Frobenius number $g(S)$,*
3) *S is maximal in the set of all numerical semigroups that do not contain $g(S)$.*

Proof. 1) $\Rightarrow$ 2) Let $\overline{S}$ be a numerical semigroup such that $S \subseteq \overline{S}$ and $g(\overline{S}) = g(S)$. Then $S = (S \cup \{g(S)\}) \cap \overline{S}$. Since S is irreducible, we deduce that $S = \overline{S}$.

2) $\Rightarrow$ 3) Let $\overline{S}$ be a numerical semigroup such that $S \subseteq \overline{S}$ and $g(S) \notin \overline{S}$. Then $\overline{S} \cup \{g(S)+1, g(S)+2, \dots\}$ is a numerical semigroup that contains S with Frobenius number $g(S)$. Therefore, $S = \overline{S} \cup \{g(S)+1, g(S)+2, \dots\}$ and so $S = \overline{S}$.

3) $\Rightarrow$ 1) Let S_1 and S_2 be two numerical semigroups that contain S properly. Then, by hypothesis, $g(S) \in S_1$ and $g(S) \in S_2$. Therefore $S \neq S_1 \cap S_2$ and so S is irreducible. □

From [**6**] and [**2**] we deduce the next result.

Proposition 2.

1) *If $g(S)$ is odd, then S is irreducible if and only if for all $h, h^{'} \in \mathbb{Z}$, such that $h + h^{'} = g(S)$, we have that either $h \in S$ or $h^{'} \in S$ (that is, S is symmetric).*
2) *If $g(S)$ is even, then S is irreducible if and only if for all $h, h' \in \mathbb{Z} \setminus \{\frac{g(S)}{2}\}$, such that $h + h' = g(S)$, we have that either $h \in S$ or $h' \in S$ (that is, S is pseudo-symmetric).*

Let $n \in S \setminus \{0\}$. Denote by $0 = w(1) < w(2) \cdots < w(n)$ the smallest elements in S in respective congruence classes mod n. We denote by $\mathrm{Ap}(S, n)$, the **Apéry set** of n in S (see [**1**]), the set $\{0 = w(1) < w(2) < \cdots < w(n)\}$. It is well-known (see [**13**]) that $\mathrm{Ap}(S, n) = \{x \in S : \ x - n \notin S\}$ and that $w(n) = g(S) + n$.

The following result is also well-known (see [**1**], [**4**] or [**11**]):

Proposition 3. *Let $n \in S \setminus \{0\}$. Then S is irreducible with an odd Frobenius number (that is, S is symmetric) if and only if $w(i)+w(n-i+1) = w(n)$ for all $i \in \{1, \dots, n\}$.*

Now we see how is the $\mathrm{Ap}(S, n)$ when S is irreducible with an even Frobenius number.

Lemma 4. *If S is irreducible with an even Frobenius number and $n \in S \setminus \{0\}$, then $\frac{g(S)}{2} + n \in \mathrm{Ap}(S, n)$.*

Proof. It is enough to prove that $\frac{g(S)}{2} + n \in S$, since $\frac{g(S)}{2} \notin S$, but this follows from Proposition 2 $((\frac{g(S)}{2} + n) + (\frac{g(S)}{2} - n) = g(S))$. □

Proposition 5. *Let S be a numerical semigroup with an even Frobenius number and $n \in S \setminus \{0\}$. Then S is irreducible if and only if*

$$\mathrm{Ap}(S,n) = \{0 = w(1) < w(2) < \dots < w(n-1) = g(S) + n\} \cup \{g(S)/2 + n\}$$

and $w(i) + w(n-i) = w(n-1)$ for all $i \in \{1, \dots, n-1\}$.

Proof. First note that if $g(S)$ is even, then $\frac{g(S)}{2}+n \in \mathrm{Ap}(S,n)$ and $\frac{g(S)}{2}+n < \max \mathrm{Ap}(S,n)$. If $i \in \{1, \dots, n-1\}$, then $w(i) - n \notin S$ and $w(i) - n \neq \frac{g(S)}{2}$. By Proposition 2, we have that $g(S) - (w(i) - n) \in S$ and thus $w(n-1) - w(i) = g(S) + n - w(i) \in S$. Since $w(n-1) \in \mathrm{Ap}(S,n)$ we deduce that $w(n-1) - w(i) \in \mathrm{Ap}(S,n)$. Furthermore $w(n-1) - w(i) \neq \frac{g(S)}{2} + n$ because otherwise we would have $w(i) = \frac{g(S)}{2}$. Hence the reader can check that $w(i) + w(n-i) = w(n-1)$.

Conversely, let x be an integer such that $x \neq \frac{g(S)}{2}$ and $x \notin S$. Let us show that $g(S) - x \in S$. Take $w \in \mathrm{Ap}(S,n)$ such that $w \equiv x (\mathrm{mod}\ n)$. Then $x = w - kn$ for some $k \in \mathbb{N} \setminus \{0\}$. We distinguish two cases:

(1) If $w = \frac{g(S)}{2}+n$, then $g(S)-x = g(S)-(\frac{g(S)}{2}+n-kn) = \frac{g(S)}{2}+(k-1)n$. Besides, $x \neq \frac{g(S)}{2}$ leads to $k \neq 1$ and therefore $k \geq 2$. Hence we can assert that $g(S) - x \in S$.

(2) If $w \neq \frac{g(S)}{2} + n$, then $g(S) - x = g(S) - (w - kn) = g(S) + n - w + (k-1)n = w(n-1) - w + (k-1)n \in S$, since $w(n-1) - w \in S$ by hypothesis.

□

Note that if S has embedding dimension two, then S is irreducible with odd Frobenius number (i.e., S is symmetric); in fact S is a **complete intersection** (see [**3**, **7**]).

Observe also that $\mu(S) \leq m(S)$ for every numerical semigroup S. The semigroups with $\mu(S) = m(S)$ have been widely studied in the literature (see for instance [**2**, **12**, **15**]) and are called **MED-semigroups** (numerical semigroups with maximal embedding dimension).

Proposition 6. *Let S be an irreducible numerical semigroup.*

1) *If $g(S)$ is odd and $m(S) \geq 3$, then $\mu(S) \leq m(S) - 1$.*
2) *If $g(S)$ is even and $m(S) \geq 4$, then $\mu(S) \leq m(S) - 1$.*

Proof. 1. See Section 2 of [**11**].

2. It is enough to prove that $\mu(S) \neq m(S)$. If $\mu(S) = m(S)$, then S is minimally generated by $\{m(S), n_1, \dots, n_{m(S)-1}\}$ and therefore $\mathrm{Ap}(S,n)$ is of the form

$$\mathrm{Ap}(S,n) = \{0 < n_2 < \cdots < n_{m(S)-1}\} \cup \left\{ n_1 = \frac{g(S)}{2} + m(S) \right\}.$$

Since $m(S) - 1 \geq 3$ then $n_1 \neq n_2 \neq n_{m(S)-1}$. By Proposition 5 we deduce that $n_{m(S)-1} - n_2 \in S$, which contradicts the fact that $\{m(S), n_1, \ldots, n_{m(S)-1}\}$ is a minimal system of generators for S. □

Note that $S = \langle 3, 7, 11 \rangle$ is an irreducible numerical semigroup with Frobenius number $g(S) = 8$ (it is easy to see that 8 belongs to every numerical semigroup that properly contains S). That is why in 2) of the above proposition we need that $m(S) \geq 4$ instead of $m(S) \geq 3$.

Using 1) and 2) of the above proposition we can assert that if S is an irreducible numerical semigroup with $m(S) \geq 4$, then $\mu(S) \leq m(S) - 1$.

3. Irreducible numerical semigroups with multiplicity 3 and 4.

In this section we study the irreducible numerical semigroups with multiplicity 3 and 4. By the remark made after Proposition 5, we know that if $\mu(S) = 2$, then S is irreducible. Recall also, that from Proposition 6, if $m(S) = 4$ and S is irreducible then $\mu(S) \leq 3$.

Therefore, we focus our study in the cases:

1) S is irreducible with $m(S) = \mu(S) = 3$,
2) S is irreducible with $m(S) = 4$ and $\mu(S) = 3$.

The following result is an immediate consequence of [**2**, Theorems I.4.2, I.4.4]. Here we offer an alternative proof by using Apéry sets.

Theorem 7. *The following conditions are equivalent:*

1) *S is an irreducible numerical semigroup with $m(S) = \mu(S) = 3$,*
2) *S is generated by $\{3, x+3, 2x+3\}$ with x not a multiple of 3.*

Proof. 1) ⇒ 2) If $m(S) = \mu(S) = 3$, then $\{3, n_1, n_2\}$ is a minimal system of generators for S. From Proposition 6 we deduce that $g(S)$ is even and by Proposition 5 we have that

$$\mathrm{Ap}(S, 3) = \left\{0, n_1 = \frac{g(S)}{2} + 3, n_2 = g(S) + 3\right\}.$$

Taking $x = \frac{g(S)}{2}$ we have that $n_1 = x + 3$ and $n_2 = 2x + 3$. Furthermore, $x = \frac{g(S)}{2} \notin S$ and thus x is not a multiple of 3.

2) ⇒ 1) Clearly $\{3, x+3, 2x+3\}$ is a minimal system of generators for S and thus $m(S) = \mu(S) = 3$. We have that $\mathrm{Ap}(S, 3) = \{0, x+3, 2x+3\}$. Hence $2x + 3 = g(S) + 3$ and therefore $\frac{g(S)}{2} + 3 = x + 3$. From Proposition 5 we deduce that S is irreducible. □

$S = \langle 3, 3+x, 2x+3 \rangle$ is a MED-semigroup. Applying the results obtained in [**12**] we deduce that a minimal presentation for S is:

$$\rho = \{(2X_1, X_0 + X_2), (2X_2, xX_0 + X_1), ((x+1)X_0, X_1 + X_2)\}.$$

Now we study the irreducible numerical semigroups with multiplicity 4. We distinguish two cases taking into account that the Frobenius number is odd (a symmetric semigroup) or even (a pseudo-symmetric semigroup).

Herzog proves in [**7**] that a numerical semigroup S with minimal system of generators $\{n_0, n_1, n_2\}$ is irreducible with an odd Frobenius number (i.e., symmetric) if and only if it is a complete intersection. Applying the results obtained in [**5**] this occurs if and only if $n_i \in \left\langle \frac{n_j}{(n_j,n_k)}, \frac{n_k}{(n_j,n_k)} \right\rangle$ for some $\{i, j, k\} = \{0, 1, 2\}$, where (n_j, n_k) denotes the greatest common divisor (gcd for short) of n_j, n_k.

Theorem 8. *The following conditions are equivalent:*

1) *S is an irreducible numerical semigroup, $g(S)$ is odd, $m(S) = 4$ and $\mu(S) = 3$,*
2) *S is a numerical semigroup generated by $\{4, 2x, x+2y\}$ with $y \in \mathbb{N}\backslash\{0\}$ and x an odd integer greater than or equal to 3.*

Proof. 1) $\Rightarrow$ 2) If $m(S) = 4$ and $\mu(S) = 3$, then $\{4, n_1, n_2\}$ is a minimal system of generators for S. From the previous remark we only have two cases:

a) Assume that $d = \gcd\{4, n_1\}$ and $n_2 \in \langle \frac{4}{d}, \frac{n_1}{d} \rangle$. Notice that $d = 2$ and $n_1 = 2x$ with x an odd number greater than or equal to 3. Furthermore $1 = \gcd\{4, n_1, n_2\}$, then n_2 is an odd number and $n_2 \in \langle 2, x\rangle$ thus $n_2 = x + 2y$ (because all odd numbers in $\langle 2, x\rangle$ are of this kind).
b) Assume that $d = \gcd\{n_1, n_2\}$ and $4 \in \langle \frac{n_1}{d}, \frac{n_2}{d} \rangle$. From here we deduce that $n_1 = 2d$, $n_2 = k_2 d$ with k_2 odd and d an odd integer greater than or equal to 3. Therefore, $n_2 = d+(k_2-1)d$ with $(k_2-1)d$ even. Taking $x = d$ and $y = \frac{(k_2-1)d}{2}$ we obtain the desired result.

2) $\Rightarrow$ 1) Clearly, $2 = \gcd\{4, 2x\}$ and $x+2y \in \langle \frac{4}{2}, \frac{2x}{2} \rangle$. By the remark made before this theorem we have that S is an irreducible numerical semigroup with an odd Frobenius number. Now, we need to show that $\{4, 2x, x + 2y\}$ is a minimal system of generators for S, but this is clear because:

1) $x + 2y \notin \langle 4, 2x\rangle$, since $x + 2y$ is odd,
2) $2x \notin \langle 4, x + 2y\rangle$, since if $2x = a4 + b(x + 2y)$ with $a, b \in \mathbb{N}$, then applying that $2x$ is an even integer not a multiple of 4 and that $x+2y$ is odd, we deduce that $b \geq 2$, contradicting that $2(x + 2y) > 2x$.

□

The semigroup $S = \langle 4, 2x, x + 2y\rangle$ has Frobenius number $g(S) = 3x + 2y - 4$, furthermore using that it is a complete intersection we deduce that a minimal presentation for S is:

$$\rho = \{(2X_1, xX_0), (2X_2, yX_0 + X_1)\}.$$

Finally, we study the irreducible numerical semigroups such that $g(S)$ is even, $m(S) = 4$ and $\mu(S) = 3$.

Theorem 9. *The following conditions are equivalent:*

1) *S is an irreducible numerical semigroup, $g(S)$ is even, $m(S) = 4$ and $\mu(S) = 3$,*
2) *S is generated by $\{4, x+2, x+4\}$ with x an odd integer greater than or equal to* 3.

Proof. 1) $\Rightarrow$ 2) If $m(S) = 4$ and $\mu(S) = 3$, then $\{4, n_1, n_2\}$ is a minimal system of generators for S. From Lemma 4 we know that $\frac{g(S)}{2} + 4 \in \mathrm{Ap}(S, 4)$. We distinguish two cases:

a) If $\frac{g(S)}{2} + 4$ is a minimal generator then, by Proposition 5, we deduce that
$$\mathrm{Ap}(S,4) = \left\{0, n_1 = \frac{g(S)}{2} + 4, n_2, 2n_2 = g(S) + 4\right\}.$$
Taking $x = \frac{g(S)}{2}$, then $n_1 = x + 4$ and $n_2 = x + 2$. Furthermore $g(S) \notin S$ and therefore x is odd.

b) If $\frac{g(S)}{2} + 4$ is not a minimal generator, then
$$\mathrm{Ap}(S,4) = \left\{0, n_1, n_2, \frac{g(S)}{2} + 4\right\}.$$
Hence $g(S) + 4 = n_1$ or $g(S) + 4 = n_2$. Suppose that $g(S) + 4 = n_1$ then, by Proposition 5, we deduce that $n_1 - n_2 \in S$, contradicting that $\{4, n_1, n_2\}$ is a minimal system of generators.

2) $\Rightarrow$ 1) Clearly, $\{4, x+2, x+4\}$ is a minimal system of generators of S, whence $m(S) = 4$ and $\mu(S) = 3$. The reader can check that
$$\mathrm{Ap}(S,4) = \{0, x+2, x+4, 2x+4\}.$$
Therefore $g(S) = 2x$ and then
$$\mathrm{Ap}(S,4) = \left\{0, \frac{g(S)}{2} + 4, \frac{g(S)+4}{2}, g(S) + 4\right\}.$$
Using Proposition 5 we obtain that S is irreducible. □

Note that $S = \langle 4, x+2, x+4 \rangle$ has Frobenius number $2x$. Applying [**7**] and that this semigroup is not symmetric (therefore it is not a complete intersection), we can deduce that a minimal presentation for S is:
$$\rho = \{(2X_2, X_0 + 2X_1), (3X_1, kX_0 + X_2), (tX_0, X_1 + X_2)\}$$
with $k = \frac{3(x+2)-(x+4)}{4}$ and $t = \frac{(x+4)+(x+2)}{4}$. Observe that $3(x+2) - (x+4)$ is a multiple of 4 if and only if x is odd, and $(x+4) + (x+2)$ is a multiple of 4 if and only if x is odd.

4. An upper bound of the cardinality of a minimal presentation for an irreducible numerical semigroup.

Let S be a numerical semigroup with minimal system of generators $\{n_0 < n_1 < \cdots < n_p\}$. In [**12**] it is shown the following result ($\#\mathrm{MR}S$ denotes the cardinality of a minimal presentation for S).

Proposition 10. *Let S be a numerical semigroup. Then*

$$\#\mathrm{MR}S \leq \frac{n_0(n_0-1)}{2} - 2(n_0 - 1 - p).$$

In [**11**] this bound is improved in the case S is symmetric. In fact, the following result is given there:

Proposition 11. *If S is symmetric, $n_0 \geq 3$ and $p \geq 2$, then*

$$\#\mathrm{MR}S \leq \frac{(n_0-2)(n_0-1)}{2} - 1 + (p + 2 - n_0).$$

Our aim in this section is to prove the analogue to this result for S an irreducible semigroup with even Frobenius number.

Throughout this section S is an irreducible numerical semigroup with $g(S)$ even and $p \geq 3$.

For $n \in S$ define the graph $G_n = (V_n, E_n)$, as

$$V_n = \{n_i \in \{n_0, \ldots, n_p\} : n - n_i \in S\},$$
$$E_n = \{[n_i, n_j] : n - (n_i + n_j) \in S, i, j \in \{0, \ldots, p\}, i \neq j\}.$$

From [**12**] we can deduce the following result.

Proposition 12. *If $\{n_0, n_1, \ldots, n_p, g(S)\}$ is a minimal system of generators for $S' = S \cup \{g(S)\}$, $g(S) > n_0$ and n_i and n_0 are in the same connected component of $G_{g(S)+n_0+n_i}$ for all $i \in \{1, \ldots, p\}$, then*

$$\#\mathrm{MR}S + p + 2 = \#\mathrm{MR}S'.$$

Applying Proposition 5 and using that $p \geq 3$ we deduce that $g(S) + n_o \geq n_i + n_j$ for some $i, j \in \{1, \ldots, p\}$ and therefore $g(S) > n_0$. Furthermore, $\{n_0, n_1, \ldots, n_p, g(S)\}$ is a minimal system of generators for $S' = S \cup \{g(S)\}$, since otherwise we would deduce from [**12**] that $n_p = g(S) + n_0$, which contradicts Proposition 5 for $p \geq 3$.

Lemma 13. *If $i \in \{1, \ldots, p\}$, $w \in \mathrm{Ap}(S, n_0)$ and n_0 and n_i are in two different connected components of G_{w+n_i}, then for all $w' \in \mathrm{Ap}(S, n_0)$ such that $w - w' \in S \setminus \{0\}$ we have that $w' + n_i \in \mathrm{Ap}(S, n_0)$*

Proof. Suppose that $w' + n_i \notin \mathrm{Ap}(S, n_0)$, then $w' + n_i - n_0 \in S$. Let $s \in S \setminus \{0\}$ be such that $w = w' + s$ and $j \in \{0, \ldots, p\}$ such that $s - n_j \in S$. Then, $w + n_i - (n_i + n_j) \in S$ and $w + n_i - (n_j + n_0) \in S$. Therefore $[n_i, n_j]$, $[n_j, n_0] \in E_n$ and so n_i and n_0 are in the same connected component of G_{w+n_i}. ∎

Lemma 14. *If $i \in \{1, \dots, p\}$, then n_0 and n_i are in the same connected component of $G_{g(S)+n_0+n_i}$.*

Proof. Suppose that n_0 and n_i are in two different connected components of $G_{g(S)+n_0+n_i}$. Let $j \in \{1, \dots, p\}$ be such that $n_j \neq \frac{g(S)}{2} + n_0$ and $n_i \neq n_j$ (this is possible because $p \geq 3$). By Lemma 13 and Proposition 5 we deduce that $g(S) + n_0 - n_j + n_i \in \mathrm{Ap}(S, n_0)$.

Observe that $g(S) + n_0 - n_j + n_i = \frac{g(S)}{2} + n_0$, since otherwise using Proposition 5 we would obtain that $g(S) + n_0 - (g(S) + n_0 - n_j + n_i) \in S$ and therefore $n_j - n_i \in S$, contradicting that $\{n_0, \dots, n_p\}$ is a minimal system of generators for S.

Let us observe that $n_i \neq \frac{g(S)}{2} + n_0$ because otherwise we would deduce from $g(S)+n_0-n_j+n_i = \frac{g(S)}{2}+n_0$, that $n_j = g(S)+n_0$ and applying Proposition 5 we can assert that $S = \langle n_0, \frac{g(S)}{2} + n_0, g(S) + n_0\rangle$, which contradicts that $p \geq 3$.

Now assume that $\mathrm{Ap}(S, n_0) = \{0 = w(1) < \cdots < w(n_0-1)\} \cup \left\{\frac{g(S)}{2} + n_0\right\}$. We distinguish two cases:

1) If $\frac{g(S)}{2} + n_0 \in \{n_1, \dots, n_p\}$, then from Proposition 5 and Lemma 13 we have that

$$w(1) + n_i = w(2), w(2) + n_i = w(3), \dots, w(n_0 - 2) + n_i = w(n_0 - 1).$$

Hence,

$$\mathrm{Ap}(S, n_0) = \{0, n_i, 2n_i, \dots, (n_0 - 2)n_i\} \cup \left\{\frac{g(S)}{2} + n_0\right\}$$

and thus $S = \langle n_0, n_i, \frac{g(S)}{2} + n_0\rangle$, a contradiction because $p \geq 3$.

2) If $\frac{g(S)}{2} + n_0 \notin \{n_1, \dots, n_p\}$, then again from Proposition 5 and Lemma 13 we obtain that

$$\mathrm{Ap}(S, n_0) = \left\{0, n_i, \dots, kn_i = \frac{g(S)}{2} + n_0, n_j, n_j + n_i, \dots, n_j + tn_i = g(S) + n_0\right\}$$

for some $k, t \in \mathbb{N}$. Therefore, $S = \langle n_0, n_i, n_j\rangle$, in contradiction again with $p \geq 3$.

□

Proposition 15.

$$\#\mathrm{MR}S \leq \frac{(n_0 - 2)(n_0 - 1)}{2} - 1 + (p + 2 - n_0).$$

Proof. Applying Lemma 14 and Proposition 12 we deduce that $\#\mathrm{MR}S = \#\mathrm{MR}(S \cup \{g(S)\}) - (p+2)$. From Proposition 10 we have that

$$\#\mathrm{MR}(S \cup \{g(S)\}) \leq \frac{n_0(n_0-1)}{2} - 2(n_0 - 1 - p - 1).$$

Hence,

$$\#\mathrm{MR}S \leq \frac{(n_0-2)(n_0-1)}{2} - 1 + (p + 2 - n_0).$$

□

From Propositions 15 and 11 we can obtain the following result.

Theorem 16. *If S is an irreducible numerical semigroup with $\mu(S) \geq 4$, then*

$$\#\mathrm{MR}S \leq \frac{(m(S)-2)(m(S)-1)}{2} - 1 + (\mu(S) + 1 - m(S)).$$

Note that if $\mu(S) = 2$, then $\#\mathrm{MR}S = 1$ and if $\mu(S) = 3$, then $\#\mathrm{MR}S = 2$ or 3 depending on the parity of $g(S)$ (see[**7**]).

5. Irreducible numerical semigroups with maximal embedding dimension.

A **MEDI-semigroup** is an irreducible semigroup with multiplicity $m \geq 5$ and embedding dimension $m - 1$. Remember from Proposition 6 that if S is irreducible and $m(S) \geq 5$, then $\mu(S) \leq m(S) - 1$ and this is why we use the name MEDI-semigroup to indicate that it is an irreducible numerical semigroup with the maximal possible embedding dimension.

If $S = \langle m(S), n_1, \ldots, n_{m(S)-2} \rangle$ is a MEDI-semigroup, then

$$\mathrm{Ap}(S, m(S)) = \{0, n_1, \ldots, n_{m(S)-2}, g(S) + m(S)\}.$$

Moreover, from Propositions 3 and 5 we can deduce that $g(S) + m(S) = n_i + n_j$ with $i, j \in \{1, \ldots, m(S) - 2\}$ and $i \neq j$. Applying now [**14**, Theorem 1] we get that

$$\#\mathrm{MR}S = \frac{(m(S)-2)(m(S)-1)}{2} - 1.$$

Note that for $m(S) \in \{3, 4\}$, the previous formula is not true (for this reason in the definition of MEDI-semigroup we need that $m(S) \geq 5$). In fact, for $m(S) = 3$ applying the previous formula, we have $\#\mathrm{MR}S = 0$ but we know that a minimal presentation for $\langle 3, n_1 \rangle$ has cardinal 1. For $m(S) = 4$ applying the previous formula, we have $\#\mathrm{MR}S = 2$ and we know that in this class there are semigroups with minimal presentation of cardinal 3 (see the remark after Theorem 9).

If S is a MEDI-semigroup with $g(S)$ odd, then S is a **MEDSY-semigroup** according to the terminology used in [**11**].

Theorem 17. *If S is an irreducible numerical semigroup with $\mu(S) \geq 5$, then the following conditions are equivalent:*

1) *S is a* MEDI-*semigroup, and*
2) $\#\mathrm{MR}S = \frac{(m(S)-2)(m(S)-1)}{2} - 1$.

Proof. 2) ⇒ 1) Since $\mu(S) \geq 4$, by Theorem 16 we know that

$$\#\mathrm{MR}S \leq \frac{(m(S)-2)(m(S)-1)}{2} - 1 + (\mu(S) + 1 - m(S)).$$

Since

$$\#\mathrm{MR}S = \frac{(m(S)-2)(m(S)-1)}{2} - 1,$$

we get that $\mu(S) = m(S) - 1$ and therefore S is a MEDI-semigroup.

1) ⇒ 2) Proved already (see the beginning of this section). □

The next result appears in [**11**].

Lemma 18. *Let $A = \{0 = w(1), w(2), \ldots, w(m)\}$ be a subset of $\mathbb{N}$ such that $w(i) \not\equiv w(j) (\mathrm{mod}\ m)$ for all $1 \leq i < j \leq m$, and let S be a numerical semigroup generated by $A \cup \{m\}$. Then $\mathrm{Ap}(S, m) = A$ if and only if for all $1 \leq i, j \leq m$ there exist $1 \leq k \leq m$ and $t \in \mathbb{N}$ such that $w(i) + w(j) = w(k) + tm$.*

Proposition 19. *If S is an irreducible numerical semigroup with $m(S) \geq 5$ and*

$$\mathrm{Ap}(S, m(S)) = \{0 = w(1) < w(2) < \cdots < w(m(S))\},$$

then the semigroup S' generated by

$$\{m(S), w(2) + m(S), \ldots, w(m(S) - 1) + m(S)\}$$

is a MEDI-*semigroup.*

Proof. In [**11**, Proposition 2.4] it is proved that $\{m(S), w(2) + m(S), \ldots, w(m(S) - 1) + m(S)\}$ is a minimal system of generators for S'. Furthermore, in that proposition, it is also shown that if S is symmetric, then S' is MEDSY-semigroup. Therefore it is enough to prove that if S is irreducible with $g(S)$ even, then S' is irreducible. From Lemma 18 we obtain that

$$\mathrm{Ap}(S', m(S)) = \{0 < w(2) + m(S) < \cdots < w(m(S) - 1) + m(S) < w(m(S)) + 2m(S)\}$$

and, by Proposition 5, we get that S' is irreducible. □

As a consequence of the previous proof we have that $g(S') = g(S) + 2m(S)$.

Proposition 20. *If S is a* MEDI-*semigroup with a minimal system of generators $\{m(S) < n_1 < \cdots < n_{m(S)-2}\}$, then the semigroup S' generated by $\{m(S), n_1 - m(S), \ldots, n_{m(S)-2} - m(S)\}$ is irreducible.*

Proof. In [**11**, Propositon 2.5] it is proved that if S is a MEDSY-semigroup then S' is symmetric. Therefore, it is enough to prove that if S is a MEDI-semigroup with $g(S)$ even, then S' is irreducible.

Assume that $n_j = \frac{g(S)}{2} + m(S)$ and

$$\mathrm{Ap}(S, m(S)) = \{0, n_1, \ldots, n_{m(S)-2}, g(S) + m(S) = n_1 + n_{m(S)-2}\}.$$

Using Lemma 18 it is easy to prove that

$$\mathrm{Ap}(S', m(S)) = \{0, n_1 - m(S), \ldots, n_{m(S)-2} - m(S), g(S) - m(S)\}.$$

From Proposition 5 we conclude that S' is irreducible (note that $g(S') = g(S) - 2m(S)$ and $n_j - m(S) = \frac{g(S')}{2} + m(S)$). □

Applying Propositions 19 and 20 and a similar reasoning to the one used in the proof of [**11**, Theorem 2.6] we obtain the following result:

Theorem 21. *There is one to one correspondence between the irreducible semigroups with Frobenius number g and multiplicity $m \geq 5$ and the* MEDI-*semigroups with Frobenius number $g + 2m$, multiplicity m and the rest of minimal generators greater than $2m$.*

Acknowledgements. Special thanks to P. A. García-Sánchez and the referee for their comments and suggestions.

References

[1] R. Apéry, *Sur les branches superlinéaires de courbes algébriques*, C.R. Acad. Sci. Paris, **222** (1946), 1198-2000, MR 8,221a, Zbl 0061.35404.

[2] V. Barucci, D.E. Dobbs and M. Fontana, *Maximality properties in numerical semigroups and applications to one-dimensional analytically irreducible local domains*, Memoirs Amer. Math. Soc., **125**(**598**) (1997), MR 97g:13039, Zbl 0868.13003.

[3] A. Brauer, *On a problem of partitions*, Amer. J. Math., **64** (1942), 299-312, MR 3,270d, Zbl 0061.06801.

[4] H. Bresinsky, *On prime ideals with generic zero $x_i = t^{ni}$*, Proc. Amer. Soc., **47** (1975), 329-332, MR 52 #10741, Zbl 0296.13007.

[5] C. Delorme, *Sous-monoides d'intersion complete de* ℕ, Ann. Scient. Ec. Norm. Sup., 4-série, **9** (1976), 145-154, MR 53 #10821, Zbl 0325.20065.

[6] R. Fröberg, G. Gottlieb and R. Häggkvist, *On numerical semigroups*, Semigroup Forum, **35** (1987), 63-83, MR 88d:20092, Zbl 0614.10046.

[7] J. Herzog, *Generators and relations of abelian semigroups and semigroup rings*, Manuscripta Math., **3** (1970), 175-193, MR 42 #4657, Zbl 0211.33801.

[8] E. Kunz, *The value semigroup of a one dimensional Gorenstein ring*, Proc. Amer. Math. Soc., **25** (1970), 748-751, MR 42 #263, Zbl 0197.31401.

[9] L. Rédei, *The Theory of Finitely Generated Commutative Semigroups*, Pergamon, Oxford-Edinburgh, New York, 1965, MR 32 #5761, Zbl 0133.27904.

[10] J.C. Rosales, *An algorithm for determining a minimal relation associated to a numerical semigroup*, Int. J. Algebra and Comput., **6** (1996), 441-455, MR 97f:20080, Zbl 0863.20026.

[11] ______, *On symmetric numerical semigroups*, J. Algebra, **182** (1996), 422-434, MR 98h:20099, Zbl 0856.20043.

[12] ______, *On numerical semigroups*, Semigroup Forum, **52** (1996), 307-318, MR 96m:20092, Zbl 0853.20041.

[13] J.C. Rosales and P.A. García-Sánchez, *Finitely Generated Commutative Monoids*, Nova Science Publishers, New York, 1999, MR 2000d:20074, Zbl 0966.20028.

[14] ______, *On numerical semigroups with high embedding dimension*, J. Algebra, **203** (1998), 567-578, MR 99g:20108, Zbl 0906.20042.

[15] J. Sally, *Cohen-Macaulay local rings of maximal embedding dimension*, J. Algebra, **56** (1979), 168-183, MR 80e:14022, Zbl 0401.13016.

Received September 26, 2001 and revised February 26, 2002. The first author has been supported by the project BFM2000-1469. The second author has been supported by the University of Évora.

Departamento de Álgebra
Universidad de Granada
E-18071 Granada
Spain
E-mail address: jrosales@ugr.es

Departamento de Matemática
Universidade de Évora
7000 Évora
Portugal
E-mail address: mbb@dmat.uevora.pt

PACIFIC JOURNAL OF MATHEMATICS
Vol. 209, No. 1, 2003

THE SCHRÖDINGER EQUATION ON SPHERES

Michael Taylor

It is shown that the fundamental solution to the Schrödinger equation on a d-dimensional sphere has an explicit description at times that are rational multiples of π. This leads to sharp L^p estimates on the solution operator at those times. Analogous, though less explicit, results are obtained when spheres are replaced by Zoll manifolds, and when potentials are added.

1. Introduction.

Let Δ denote the Laplace-Beltrami operator on the d-dimensional sphere S^d, with its standard metric. The fundamental solution to the Schrödinger equation:

$$i\frac{\partial u}{\partial t} = \Delta u, \quad u(0,x) = \delta_p(x), \tag{1.1}$$

is a distribution on $\mathbb{R} \times S^d$ with fairly nasty behavior; its singular support is all of $\mathbb{R} \times S^d$. However, J. Rauch pointed out to me that when $d = 1$ and t is a rational multiple of π (we say $t \in \pi\mathbb{Q}$), then $e^{-it\Delta}\delta(x) \in \mathcal{D}'(S^1)$ is a finite sum of delta functions on S^1. Hence for such $t \in \pi\mathbb{Q}$, $e^{-it\Delta}$ is bounded on $L^p(S^1)$ for each $p \in [1,\infty]$. Here we work out an equally precise description of $e^{-it\Delta}$ on $\mathcal{D}'(S^d)$, for each $t \in \pi\mathbb{Q}$. From this follows a precise account of the L^p-Sobolev mapping properties of $e^{-it\Delta}$, for such t.

In §2 we will derive the basic identities for $e^{-it\Delta}$ on $\mathcal{D}'(S^d)$ when $t \in \pi\mathbb{Q}$. For such t we express $e^{-it\Delta}$ in terms of solution operators to a *wave* equation. This leads to the sharp L^p-Sobolev estimates. Establishing sharpness is simply a matter of showing that certain coefficients in the formula for $e^{-it\Delta}$ do not vanish. This issue is settled in §3.

In §4 we discuss various extensions of these results. It is mentioned that another extension of the S^1 case is to d-dimensional tori, and that formulas there have a number-theoretical significance. We also discuss extensions to Zoll surfaces and to situations where a potential is added to the Laplace operator.

2. Basic identities.

For $d = 1$, we have the Fourier series representation for $S(t, x) = e^{-it\Delta}\delta(x)$ at $t = 2\pi m/n$:

$$S(2\pi m/n, x) = \frac{1}{2\pi} \sum_{\nu=-\infty}^{\infty} e^{2\pi i\nu^2 m/n}\, e^{i\nu x}. \tag{2.1}$$

If we set $\nu = nj + \ell$ and produce a double sum over $j \in \mathbb{Z}$, $\ell \in \{0, \dots, n-1\}$, the sum over j becomes

$$\sum_{j=-\infty}^{\infty} e^{injx} = \frac{2\pi}{n} \sum_{j=0}^{n-1} \delta_{2\pi j/n}, \tag{2.2}$$

and we obtain the distribution on $S^1 = \mathbb{R}/(2\pi\mathbb{Z})$:

$$S(2\pi m/n, x) = \frac{1}{n} \sum_{j=0}^{n-1} G(m, n, j)\, \delta_{2\pi j/n}(x), \tag{2.3}$$

with

$$G(m, n, j) = \sum_{\ell=0}^{n-1} e^{2\pi i(\ell^2 m + \ell j)/n}. \tag{2.4}$$

Let us note that the sum is really over $\ell \in \mathbb{Z}/(n)$. In particular we can replace ℓ by $-\ell$ and hence see that $G(m, n, j)$ is even in j. We also note that our formulas implicitly assume $n > 0$, but we need make no restriction on the sign of m.

The following alternative presentation of (2.3) has some advantages: Set

$$\Gamma(m, k, j) = \frac{1}{2k} G(m, 2k, j) = \frac{1}{2\pi} \sum_{\ell=0}^{2k-1} e^{\pi i(\ell^2 m + \ell j)/k}. \tag{2.5}$$

If we set $n = 2k$ in (2.3), we have

$$S(\pi m/k, x) = \sum_{j=0}^{2k-1} \Gamma(m, k, j)\, \delta_{\pi j/k}(x). \tag{2.6}$$

These formulas will prove useful in our analysis of the higher-dimensional case.

We now consider $e^{-it\Delta}$ on S^d. As is common in analysis on spheres, we take

$$A = \sqrt{-\Delta + \Big(\frac{d-1}{2}\Big)^2}. \tag{2.7}$$

It is well-known that

$$\operatorname{Spec} A = \Big\{\frac{d-1}{2} + j : j = 0, 1, 2, \dots\Big\}. \tag{2.8}$$

In particular $\operatorname{Spec} A \subset \mathbb{Z}$ if d is odd and $\operatorname{Spec} A \subset \mathbb{Z} + \frac{1}{2}$ if d is even.

Now when $\operatorname{Spec} A \subset \mathbb{Z}$ we can analyze functions of the self-adjoint operator A by the formula

$$\varphi(A) = \int_{S^1} \hat{\varphi}(t) \, \cos tA \, dt, \tag{2.9}$$

provided φ is even, with

$$\hat{\varphi}(t) = \frac{1}{2\pi} \sum_{\nu=-\infty}^{\infty} \varphi(\nu) \, e^{i\nu t}. \tag{2.10}$$

This is easily confirmed via the spectral theorem and Fourier inversion, if $\varphi : \mathbb{Z} \to \mathbb{C}$ is summable. Then a limiting argument gives it for any bounded (or even polynomially bounded) φ (yielding $\hat{\varphi} \in \mathcal{D}'(S^1)$), exploiting the smooth dependence on t of the family of operators $\cos tA$, acting on $\mathcal{D}'(S^d)$. In particular we can apply (2.9) with $\varphi(\nu) = \varphi_s(\nu) = e^{is\nu^2}$. If $s = m\pi/k$, then $\hat{\varphi}_s$ is given by (2.1), with $n = 2k$, so by (2.6) we have

$$e^{\pi i (m/k) A^2} = \sum_{j=0}^{2k-1} \Gamma(m,k,j) \, \cos \frac{\pi j}{k} A, \tag{2.11}$$

yielding a formula for

$$e^{-it\Delta} = e^{-it(d-1)^2/4} \, e^{itA^2} \tag{2.12}$$

at $t = m\pi/k$, on $\mathcal{D}'(S^d)$, when d is odd.

In case $\operatorname{Spec} A \subset \mathbb{Z} + \frac{1}{2}$, then $\operatorname{Spec} 2A \subset \mathbb{Z}$, and (2.9) applies with A replaced by $2A$. Taking $\varphi(\nu) = \varphi_s(\nu)$ as before, we have

$$e^{4\pi i (m/k) A^2} = \sum_{j=0}^{2k-1} \Gamma(m,k,j) \, \cos \frac{2\pi j}{k} A, \tag{2.13}$$

yielding a formula for $e^{-it\Delta}$ at $t = 4\pi m/k$ on $\mathcal{D}'(S^d)$, valid for d even (and also for d odd).

Note that $\cos tA$ is a solution operator for a hyperbolic PDE. In fact, given $f \in \mathcal{D}'(S^d)$, $u(t,x) = (\cos tA) f(x)$ solves

$$u_{tt} - \Big(\Delta - \Big(\frac{d-1}{2}\Big)^2\Big) u = 0, \quad u(0,x) = f(x), \ u_t(0,x) = 0. \tag{2.14}$$

In particular, for each t, $\cos tA$ is a Fourier integral operator of order zero. Its mapping properties on L^p Sobolev spaces $H^{s,p}$ are well-known (cf. [**S3**]), and by (2.11) and (2.13) they are shared by $e^{-it\Delta}$ for each $t \in \pi\mathbb{Q}$. Thus we have the following:

Proposition 2.1. *Given* $p \in (1,\infty)$, $s \in \mathbb{R}$, *we have*

$$e^{-\pi i(m/k)\Delta} : H^{s,p}(S^d) \longrightarrow H^{s-(d-1)|1/2-1/p|,p}(S^d). \tag{2.15}$$

Such estimates also hold in the endpoint cases $p = 1, \infty$, *with* L^1 *replaced by the local Hardy space* $\mathfrak{h}^1$ *and* L^∞ *replaced by* bmo.

It is well-known that such a mapping property cannot be improved for $\cos tA$ when $0 < t < \pi$. We aim to show that (2.15) cannot be improved, with some obvious exceptions, noted in §3. In view of (2.11) and (2.13) this merely amounts to examining whether cancellations can arise. In fact (2.11) and (2.13) can be collapsed somewhat, to involve $\cos t_\nu A$ with $\{t_\nu\}$ running over $[0,\pi]$, without multiplicities. We take up the task of doing this.

3. Further analysis of the coefficients.

First look at (2.11), and note that $\Gamma(m,k,j)$, given by (2.5), is even in j and periodic of period $2k$ in j, so $\Gamma(m,k,j) = \Gamma(m,k,2k-j)$. Also, when $\operatorname{Spec} A \subset \mathbb{Z}$, $\cos(2\pi - t)A = \cos tA$, so $\cos(\pi j/k)A$ is unchanged when j is replaced by $2k - j$. Thus we have on $\mathcal{D}'(S^d)$ for d odd:

$$e^{\pi i(m/k)A^2} = \Gamma(m,k,0)I + \Gamma(m,k,k)P + 2\sum_{j=1}^{k-1} \Gamma(m,k,j)\,\cos\frac{\pi j}{k}A. \tag{3.1}$$

Here I is the identity operator and $P = \cos\pi A$. When A is given by (2.7) on $\mathcal{D}'(S^d)$, with d odd, we have

$$Pf(x) = (-1)^{(d-1)/2} f(-x). \tag{3.2}$$

Looking at (2.13), we also see that

$$e^{4\pi i(m/k)A^2} = \Gamma(m,k,0)I + \Gamma(m,k,k)Q + 2\sum_{j=1}^{k-1} \Gamma(m,k,j)\,\cos\frac{2\pi j}{k}A. \tag{3.3}$$

In this case $Q = \cos 2\pi A$. From (2.8) we see that $Q = (-1)^{d-1}I$. In other words,

$$\begin{aligned} e^{4\pi i(m/k)A^2} &= \bigl\{\Gamma(m,k,0) + (-1)^{d-1}\Gamma(m,k,k)\bigr\}I \\ &\quad + 2\sum_{j=1}^{k-1} \Gamma(m,k,j)\,\cos\frac{2\pi j}{k}A. \end{aligned} \tag{3.4}$$

Now we have terms $\cos t_j A$, with $t_j \in (0, 2\pi)$, and we still want to cut the sum down. This time, use

$$\cos(2\pi - t)A = (\cos 2\pi A)\cos tA - (\sin 2\pi A)\sin tA. \tag{3.5}$$

As seen above, $\cos 2\pi A = (-1)^{d-1}I$, and meanwhile $\operatorname{Spec} 2A \subset \mathbb{Z} \Rightarrow \sin 2\pi A = 0$, so we have $\cos(2\pi - t)A = (-1)^{d-1}\cos tA = -\cos tA$ if d is even, which we assume from here to the end of formula (3.8) below. To work on the sum over j in (3.4), we consider separately two cases.

First suppose $k = 2\nu$ is even. Then we can write the sum over $1 \le j \le k-1$ in (3.4) as

$$\sum_{j=1}^{k/2-1} \big\{\Gamma(m,k,j) - \Gamma(m,k,k-j)\big\} \cos\frac{2\pi j}{k} A + \Gamma(m,k,\nu)\cos \pi A. \tag{3.6}$$

Recalling that $\operatorname{Spec} A \subset \mathbb{Z} + \frac{1}{2}$ for d even, we have $\cos \pi A = 0$, so in this case we have

$$\begin{aligned} e^{4\pi i(m/k)A^2} &= \big\{\Gamma(m,k,0) - \Gamma(m,k,k)\big\} I \\ &\quad + 2\sum_{j=1}^{k/2-1} \big\{\Gamma(m,k,j) - \Gamma(m,k,k-j)\big\} \cos\frac{2\pi j}{k} A. \end{aligned} \tag{3.7}$$

Next suppose $k = 2\nu + 1$ is odd. Then we have

$$\begin{aligned} e^{4\pi i(m/k)A^2} &= \big\{\Gamma(m,k,0) - \Gamma(m,k,k)\big\} I \\ &\quad + 2\sum_{j=1}^{(k-1)/2} \big\{\Gamma(m,k,j) - \Gamma(m,k,k-j)\big\} \cos\frac{2\pi j}{k} A. \end{aligned} \tag{3.8}$$

To reiterate, (3.7) and (3.8) hold on $\mathcal{D}'(S^d)$ for d even.

Now at this point there are no cancellations between terms in any of the formulas (3.1), (3.7), and (3.8). To be precise, suppose f is supported in a ball of radius $< 1/2k$ in S^d. Then, in each of these formulas, we have that the various terms $(\cos t_\nu A) f$ that arise have disjoint singular support. There remains the issue of whether the coefficients of these various terms $\cos t_\nu A$ might vanish. In fact, some do, and we now take up the question of exactly which coefficients vanish and which do not.

We mention some properties of $\Gamma(m,k,j)$, whose proofs are given in [**HB**], and also in [**T2**]. First, $\Gamma(m,k,j)$ can vanish sometimes. In fact

$$mk + j \ \text{ odd} \Longrightarrow \Gamma(m,k,j) = 0. \tag{3.9}$$

Now let us set $\Gamma(m,k) = \Gamma(m,k,0)$. There is the following result of [**HB**]:

Lemma 3.1. *Assume that m and k are relatively prime.*

(i) *If mk and j are even, then, with μ solving $\mu m = 1 \bmod k$,*

$$\Gamma(m,k,j) = e^{-\pi i (m/k)(j/2)^2 \mu^2}\, \Gamma(m,k). \tag{3.10}$$

(ii) *If mk and j are odd, then, with ν solving $4\nu m = 1 \bmod k$,*

$$\Gamma(m,k,j) = e^{-4\pi i (m/k)\nu^2 j^2}\, \Gamma(4m,k). \tag{3.11}$$

Furthermore, in Cases (i) *and* (ii), *respectively, we have*

$$|\Gamma(m,k)| = k^{-1/2}, \quad |\Gamma(4m,k)| = k^{-1/2}. \tag{3.12}$$

In particular, when m and k are relatively prime,

$$mk + j \quad \textit{even} \Longrightarrow \Gamma(m,k,j) \neq 0. \tag{3.13}$$

The results (3.9) and (3.13) specify precisely which coefficients arising in (3.1), for $e^{\pi i(m/k)A^2}$ on $\mathcal{D}'(S^d)$ with d odd, are nonvanishing, when m and k are relatively prime. It remains to look at the coefficients that arise in (3.7)-(3.8), describing $e^{4\pi i(m/k)A^2}$ on $\mathcal{D}'(S^d)$ for d even. Again we take m and k to be relatively prime. We consider three main cases, each having two subcases.

Case (I): m even, k odd.

$$\begin{aligned} j \ \text{ even} &\Longrightarrow \Gamma(m,k,j) \neq 0 \ \text{ and } \ \Gamma(m,k,k-j) = 0, \\ j \ \text{ odd} &\Longrightarrow \Gamma(m,k,j) = 0 \ \text{ and } \ \Gamma(m,k,k-j) \neq 0. \end{aligned}$$

In both subcases, $\Gamma(m,k,j) - \Gamma(m,k,k-j) \neq 0$.

Case (II): m odd, k even.

$$\begin{aligned} j \ \text{ odd} &\Longrightarrow \Gamma(m,k,j) = \Gamma(m,k,k-j) = 0, \\ j \ \text{ even} &\Longrightarrow \Gamma(m,k,j) - \Gamma(m,k,k-j) \\ &\qquad = \{e^{-\pi i(m/k)(j/2)^2\mu^2} - e^{-\pi i(m/k)(k/2-j/2)^2\mu^2}\}\Gamma(m,k). \end{aligned}$$

We take a closer look at this expression. Recall that $|\Gamma(m,k)| = k^{-1/2}$ in this case. Meanwhile the quantity in braces is equal to

$$e^{-\pi i(m/k)(j/2)^2\mu^2}\Big[1 - e^{-\pi i(m/k)(k^2/4-jk/2)\mu^2}\Big]. \tag{3.14}$$

We look at the exponent in the last exponential in (3.14). Say $\mu m = 1 + ak$, $a \in \mathbb{Z}$. Then $(\mu + bk)m = 1 + (a + bm)k = 1 + a_1 k$, and since m is odd we can arrange that a_1 be even. The quantity (3.10) is independent of the choice of μ mod k, so we can just assume a is even. Now the exponent mentioned above is seen to be the negative of

$$\pi i\mu\Big(\frac{k}{4} - \frac{j}{2}\Big) + \pi i a\mu\Big(\frac{k^2}{4} - \frac{jk}{2}\Big),$$

so the quantity (3.14) vanishes if and only if $k/4 - j/2$ is an even integer.

Case (III): m odd, k odd.

$$\begin{aligned} j \ \text{ even} &\Longrightarrow \Gamma(m,k,j) = 0 \ \text{ and } \ \Gamma(m,k,k-j) \neq 0, \\ j \ \text{ odd} &\Longrightarrow \Gamma(m,k,j) \neq 0 \ \text{ and } \ \Gamma(m,k,k-j) = 0. \end{aligned}$$

These considerations specify all the coefficients arising in (3.1), (3.7), and (3.8), except that (3.12) specifies $\Gamma(m,k)$ (for mk even) only up to phase. Further specification can be found in [**T2**].

Having these formulas for $e^{-it\Delta}$ when $t \in \pi\mathbb{Q}$, we note that there are explicit formulas for the action of $\cos tA$ on functions on S^d. We write them down here (at least for d odd). Demonstrations can be found in Chapter 8 of [**T1**]. If $d = 2\nu + 1$ is odd, we have

$$(\cos tA)f(x) = C_\nu(\sin t)\Big(\frac{1}{\sin t}\frac{\partial}{\partial t}\Big)^\nu \big(\sin^{2\nu-1} t\ \overline{f}_x(t)\big), \tag{3.15}$$

for $0 < t < \pi$, where $\overline{f}_x(t)$ is the mean value of f over the shell

$$\Sigma_x(t) = \{y \in S^d : \mathrm{dist}(x,y) = t\}, \tag{3.16}$$

and where $\mathrm{dist}(x,y)$ denotes the spherical distance. Here $C_\nu = 1/(2\nu-1)!!$, where $(2\nu-1)!! = 3\cdot 5\ldots(2\nu-1)$. Note that the strong Huygens principle holds here; $(\cos tA)\delta_p$ is supported on the shell $\Sigma_p(t)$. If $d = 2\nu$ is even, there is a formula in similar analogy to the solution to the wave equation on $\mathbb{R}\times\mathbb{R}^d$ for d even; we refer to [**T1**] for details. Of course, in this case the strong Huygens principle fails, but the singular support of $(\cos tA)\delta_p$ lies on $\Sigma_p(t)$.

We apply the results of this section to demonstrate the sharpness of the operator regularity stated in (2.15), with a discrete set of exceptions. Some of the exceptions are apparent from the fact that $\mathrm{Spec}(-\Delta) = \{j(j+d-1) : j = 0,1,2,\ldots\}$, hence consists of integers (even integers if d is even). Thus $e^{-\pi it\Delta} = I$ on functions on S^d whenever d is odd and t is an even integer, and whenever d is even and t is an integer. We check this observation against the formulas (3.1) and (3.7)-(3.8). Doing so will produce another discrete set of ts for which $e^{-\pi it\Delta}$ is bounded on all the spaces $H^{s,p}(S^d)$ and show that for all other $t \in \mathbb{Q}$ the operator mapping properties given in (2.15) cannot be improved.

First suppose d is odd, so (3.1) applies. Assume m and k are relatively prime. The "wave contribution" is

$$2\sum_{j=1}^{k-1}\Gamma(m,k,j)\,\cos\frac{\pi j}{k}A. \tag{3.17}$$

If $k = 1$, this sum is empty, and we have $e^{\pi imA^2} = \Gamma(m,1,0)I + \Gamma(m,1,1)P$. When m is even, $\Gamma(m,1,1) = 0$ and we recover the observation made in the previous paragraph. If $k = 2$ in (3.1), then the sum (3.17) has one term, involving $\Gamma(m,2,1)$, which vanishes, by (3.9), so $e^{\pi i(m/2)A^2} = \Gamma(m,2,0)I + \Gamma(m,2,2)P$. Hence $e^{-\pi it\Delta} : H^{s,p}(S^d) \to H^{s,p}(S^d)$ for all s,p if d is odd and $2t \in \mathbb{Z}$. If $k \geq 3$ in (3.1), then the sum (3.17) contains terms involving the coeffcients $\Gamma(m,k,1)$ and $\Gamma(m,k,2)$, and by (3.13) at least one of these terms is not zero, so (2.15) cannot be improved.

Next suppose d is even, so (3.7) applies to $e^{-\pi it\Delta}$ if $t = 4m/k$ with k even, and (3.8) applies if k is odd. Assume m and k are relatively prime. This time the "wave contribution" is

$$2 \sum_{j=1}^{[(k-1)/2]} \bigl\{\Gamma(m,k,j) - \Gamma(m,k,k-j)\bigr\} \cos \frac{2\pi j}{k} A. \tag{3.18}$$

If $k = 1$ or 2, this sum is empty. If $k = 4$, the sum (3.18) has one term, with coefficient $\Gamma(m,4,1) - \Gamma(m,4,3) = 0$. These results recover the observation that $e^{-\pi it\Delta} = I$ for $t \in \mathbb{Z}$. If $k \geq 3$ is odd, then Cases (I) and (III) show none of the terms in (3.18) vanish. If $k \geq 6$ is even, a check of Case (II) shows some of the terms in (3.18) can vanish, but in no cases do all of them vanish, so there are no further cases when (2.15) can be improved. In particular, for $k = 6$ one has $\Gamma(m,6,2) - \Gamma(m,6,4) \neq 0$ and for $k = 8$ one has $\Gamma(m,8,2) - \Gamma(m,8,6) \neq 0$.

4. Remarks and extensions.

Here we make several remarks on applications and extensions of the calculations made in §2.

Remark 1. The quantities $G(m,n,j)$ and $\Gamma(m,k,j)$ defined in (2.4)-(2.5), which arise in the calculation of $S(2\pi m/n, x)$, are Gauss sums, of interest in number theory. Another way to evaluate $S(2\pi m/n, x)$ is to take the free-space fundamental solution $S_0(t,x)$ giving $e^{-it\Delta}$ on $\mathbb{R} \times \mathbb{R}$ and sum its translates $S_0(t, x - 2\pi\nu)$, $\nu \in \mathbb{Z}$. Carrying this out yields a formula similar to (2.3), but with coefficients involving different Gauss sums. Comparing the calculations produces a straightforward and nifty proof of the reciprocity formula for Gauss sums:

(4.1)
$$\sum_{\ell=0}^{n-1} e^{2\pi i\ell^2 m/n}\, e^{2\pi i\ell j/n} = \frac{1+i}{2}\Bigl(\frac{n}{m}\Bigr)^{1/2} e^{-\pi i j^2/2mn} \sum_{\ell=0}^{2m-1} e^{-\pi i\ell^2 n/2m}\, e^{\pi i j\ell/m}.$$

This formula was first derived (by different means) by Landsberg and Schaar, in the 1890s. Specializing to $j = 0,\ m = 1$ gives the classical formula

$$\sum_{\ell=0}^{n-1} e^{2\pi i\ell^2/n} = \frac{1+i}{2}(1 + i^{-n})\sqrt{n}, \tag{4.2}$$

due to Gauss, used in one of his proofs of the quadratic reciprocity formula.

Remark 2. Another multidimensional extension of (2.3) involves $e^{itQ(D)}\delta(x)$ on the torus $\mathbb{T}^d = \mathbb{R}^d/(2\pi\mathbb{Z})^d$, where

$$Q(\xi) = \xi \cdot A\xi, \quad A^t = A \in G\ell(d,\mathbb{Z}), \quad \det A = \pm 1. \tag{4.3}$$

Then $e^{itQ(D)}\delta(x)$ can be evaluated for $t \in \pi\mathbb{Q}$. Again two methods work. One involves Fourier series on $\mathbb{T}^d$. The other involves writing down the fundamental solution $S_0^Q(t,x)$ giving $e^{itQ(D)}$ on $\mathbb{R}\times\mathbb{R}^d$ and summing $S_0^Q(t,x+2\pi\nu)$ over $\nu \in \mathbb{Z}^d$. Both produce finite linear combinations of delta functions supported on a lattice in $\mathbb{T}^d$, and comparing calculations produces reciprocity formulas for multivariate Gauss sums, obtained (by different means) in [**K**]. The reader can try this as an exercise, or see [**T2**] for details.

Remark 3. There are other extensions of the material presented above, arising from the fact that the identity (2.11) is valid whenever A is a self-adjoint operator with spectrum contained in $\mathbb{Z}$. This can be applied as follows: Let M be a d-dimensional Zoll manifold, a compact Riemannian manifold on which all geodesics have minimal period 2π, and consider $H = -\Delta + V$, with positive, real-valued $V \in C^\infty(M)$. As shown in [**CdV**] (cf. also §29.2 of [**H**]), there exists a positive, self-adjoint $A \in OPS^1(M)$ and $S \in OPS^{-1}$, commuting with each other (and with H) and $\alpha \in \mathbb{R}$, such that

$$\sqrt{-\Delta+V} = A + \alpha I + S, \quad \operatorname{Spec} A \subset \mathbb{Z}. \tag{4.4}$$

(Here $OPS^m(M)$ denotes the space of mth order pseudodifferential operators of classical type on M.) Then

$$e^{it(-\Delta+V)} = e^{it(\alpha^2+2AS+2\alpha S+S^2)}\, e^{2i\alpha tA}\, e^{itA^2}. \tag{4.5}$$

Noting that e^{itA} is a group of operators with the same Sobolev space mapping properties as used to prove Proposition 2.1, and that the first factor on the right side of (4.5) is a family of operators in $OPS^0(M)$, and applying (2.11) to e^{itA^2} for $t \in \pi\mathbb{Q}$, we see that

$$e^{\pi i(m/k)(-\Delta+V)} : H^{s,p}(M) \longrightarrow H^{s-(d-1)|1/2-1/p|,p}(M), \tag{4.6}$$

extending (2.15) to this context.

Remark 4. In [**GGR**] it is shown that, when $d=1$, $e^{-it\Delta}$ does not map $L^p(S^1)$ to itself for any $p \neq 2$ when $t \notin \pi\mathbb{Q}$. By contrast with such an indication that $S(t,\cdot) \in \mathcal{D}'(S^1)$ is less regular for $t \notin \pi\mathbb{Q}$ than for $t \in \pi\mathbb{Q}$, results in [**KR**] give a sense in which $S(t,\cdot) \in \mathcal{D}'(S^1)$ is *more regular* for (most) $t \notin \pi\mathbb{Q}$ than for $t \in \pi\mathbb{Q}$. Here regularity is measured in the scale of Besov spaces $B^s_{\infty,\infty}(S^1)$. It is clear from Formula (2.3) that $S(t,\cdot) \in B^{-1}_{\infty,\infty}(S^1)$ for $t \in \pi\mathbb{Q}$ and one cannot improve this. It is shown in [**KR**] that, for a.e. $t \in \mathbb{R}$,

$$S(t,\cdot) \in B^{-s}_{\infty,\infty}(S^1), \quad \forall\ s > \frac{1}{2}. \tag{4.7}$$

More precise results relate the best s to the continued fraction expansion of t/π; see [**KR**] for details.

It is tempting to speculate that $S(t,\cdot) \in \mathcal{D}'(S^1)$ is more regular in $B^{-s}_{p,p}(S^1)$ for rational than for irrational t/π when $p < 2$ and more regular for irrational than for rational t/π when $p > 2$. But at this point this is just a speculation.

Remark 5. We also mention the recent paper [**BGT**], dealing with the Schrödinger equation on a compact d-dimensional Riemannian manifold M. It is shown that, given $\varphi \in C_0^\infty(\mathbb{R})$, there exists $\alpha > 0$ such that, for $h \in (0,1]$,

$$\|e^{-it\Delta}\varphi(h^2\Delta)f\|_{L^\infty(M)} \le C|t|^{-d/2}\|f\|_{L^1(M)}, \quad |t| \le \alpha h. \tag{4.8}$$

This is used to produce Strichartz estimates, leading to solvability results for nonlinear Schrödinger equations. In [**BGT**], the sharpness of some of their Strichartz estimates is verified in particular for $M = S^d$, with its standard metric.

Acknowledgment. Thanks to Jeffrey Rauch for several interesting conversations related to this work. This work was performed while the author was at MSRI, and was partially supported by NSF grant DMS-9877077.

References

[BGT] N. Burq, P. Gérard and N. Tzvetkov, *The Schrödinger equation on a compact manifold: Strichartz estimates and applications*, Jour. 'Equations aux Derivées Partielles' (Plestin-les-Graves, 2001), Exposé No. 5, 18 pp., Univ. Nantes, 2001, MR 2003a:58055.

[CdV] Y. Colin de Verdière, *Sur le spectre des opérateurs elliptiques à bicharacteristiques toutes périodiques*, Comment. Math. Helv., **54** (1979), 508-522, MR 81a:58052, Zbl 0459.58014

[GGR] P. Gérard, F. Golse and J. Rauch, *unpublished manuscript*.

[HB] J. Hannay and M. Berry, *Quantization of linear maps on a torus — Fresnel diffraction by a periodic grating*, Physica, **1D** (1980), 267-290, MR 82g:81005.

[H] L. Hörmander, *The Analysis of Linear Partial Differential Operators*, Vol. 4, Springer-Verlag, New York, 1985, MR 98f:35002, Zbl 0612.35001.

[KR] L. Kapitanski and I. Rodnianski, *Does a quantum particle know the time?*, in 'Emerging Applications in Number Theory', 355-371, D. Hejhal et al. (eds.), Springer-Verlag, New York, 1999, MR 2000g:81044.

[K] A. Krazer, *Zur Theorie der mehrfachen Gausschen Summen*, H. Weber Fetschrift, Leipzig, 1912, 181.

[S3] A. Seeger, C. Sogge and E. Stein, *Regularity properties of Fourier integral operators*, Ann. of Math., **134** (1991), 231-251, MR 92g:35252, Zbl 0754.58037.

[T1] M. Taylor, *Partial Differential Equations*, Vol. 2, Springer-Verlag, New York, 1996, MR 98b:35003, Zbl 0869.35003.

[T2] ______, *Tidbits in Harmonic Analysis*, Lecture Notes, UNC, 1998.

Received December 10, 2001 and revised March 25, 2002.

Mathematics Department
University of North Carolina
Chapel Hill, NC 27599
E-mail address: met@math.unc.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 209, No. 1, 2003

THE MEROMORPHIC CONTINUATION OF THE RESOLVENT OF THE LAPLACIAN ON LINE BUNDLES OVER $\mathbb{C}H(n)$

Cynthia E. Will

Let $G = SU(n,1)$, $K = S(U(n) \times U(1))$, and for $l \in \mathbb{Z}$, let $\{\tau_l\}_{l\in\mathbb{Z}}$ be a one-dimensional K-type and let E_l the line bundle over G/K associated to τ_l. In this work we prove that the resolvent of the Laplacian, acting on C_c^∞-sections of E_l is given by convolution with a kernel which has a meromorphic continuation to $\mathbb{C}$. We prove that this extension has only simple poles and we identify the images of the corresponding residues with $(\mathfrak{g}, K)$-submodules of the principal series representations. We show that for certain values of the parameters these modules are holomorphic (or antiholomorphic) discrete series.

1. Introduction.

In [**9**] the meromorphic continuation of the resolvent kernel of the Laplacian acting on functions was studied in the case of the so called Damek-Ricci spaces. These include, in particular, symmetric spaces of strictly negative curvature. This meromorphic continuation has simple poles and the residues are finite rank operators whose images can be explicitly described and their dimensions determined. In the present paper, we shall prove similar results in the case of the action of the Laplacian on line bundles over $\mathbf{C}H^n$. We use work by Shimeno on the theory of spherical functions in this context. We prove that in a certain open half-plane of $\mathbb{C}$, the resolvent is given by convolution with an explicit kernel and this has a meromorphic continuation to $\mathbb{C}$. We prove that this continuation has simple poles located at parameters of reducibility of certain principal series representations of G. The corresponding residues are convolution operators and their images are isomorphic to $(\mathfrak{g}, K)$-submodules of the principal series representations. For some values of the parameters, these modules are finite dimensional and for others they are holomorphic, antiholomorphic or limits of discrete series representations, hence infinite dimensional. This is in contrast with the case of the trivial K-type, studied in [**9**].

An outline of the paper is as follows: In Sections 2 and 3 we introduce notation and describe some results due mainly to Shimeno, to be used in the rest of the paper. In Section 4 we study the meromorphic continuation of the resolvent kernel, and we describe the images of the residues as $(\mathfrak{g}, K)$-submodules of the principal series. We show in particular that any holomorphic or anti-holomorphic representation occurs as image of a residue, as well as any limit of discrete series and any finite dimensional representation whose contain a one dimensional K-type. Moreover, in the last case, we also prove that this module is the kernel of the standard intertwining operator of the principal series representation.

We finally discuss in Section 5, in detail, the case when $G = SU(1,1)$.

This paper is part of my thesis work. I am very grateful to my advisor, R. Miatello, for his guidance and constant support. I also wish to thank J. Vargas for helpful conversations.

2. Preliminaries.

2.1. Basic notation. We begin by introducing notation that will be used throughout this paper. As is customary, we will denote a Lie group by an upper case letter and its Lie algebra by the corresponding lower case gothic letter.

If $G = \mathrm{S}U(n,1)$, then the Lie algebra of G is given by $\mathfrak{g} = \{X \in \mathfrak{sl}(n+1,\mathbb{C}) \,:\, XJ + J\overline{X}^t = 0\}$, where $J = \left[\begin{smallmatrix} 0 & 0 & 1 \\ 0 & \mathrm{Id} & 0 \\ 1 & 0 & 0 \end{smallmatrix}\right]$.

Let $\mathfrak{g} = \mathfrak{k} + \mathfrak{p}$ be the Cartan decomposition associated to the Cartan involution $\theta(X) = \overline{X}^t$. Thus

$$\mathfrak{k} = \left\{ \begin{bmatrix} A & 0 \\ 0 & y \end{bmatrix} : A \in \mathfrak{u}(n),\, \mathrm{tr}(A) + y = 0 \right\} \qquad \text{and}$$

$$\mathfrak{p} = \left\{ \begin{bmatrix} 0 & b \\ \overline{b}^t & 0 \end{bmatrix} :\, b \in \mathbb{C}^n \right\}.$$

If we put $H_0 = \left[\begin{smallmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{smallmatrix}\right]$, it is easy to see that $\mathfrak{a} = \mathbb{R}H_0$ is a maximal abelian subalgebra of $\mathfrak{p}$ and $\mathfrak{z} = \mathbb{R}\left[\begin{smallmatrix} \frac{\mathbf{i}}{n}I & 0 \\ 0 & -\mathbf{i} \end{smallmatrix}\right]$ is the center of $\mathfrak{k}$, where $\mathbf{i} = \sqrt{-1}$. We have that $\mathfrak{k} = \mathfrak{k}_s + \mathfrak{z}$, where $\mathfrak{k}_s = [\mathfrak{k}, \mathfrak{k}]$ is the semisimple part of $\mathfrak{k}$. Let M be the centralizer of A in K, that is for $n > 1$

$$M = \left\{ \begin{bmatrix} e^{\mathbf{i}s} & 0 & 0 \\ 0 & U & 0 \\ 0 & 0 & e^{\mathbf{i}s} \end{bmatrix} :\, U \in \mathrm{U}(n-1),\, \det(U)e^{2\mathbf{i}s} = 1 \right\}.$$

If $\mathfrak{t}$ is the set of diagonal matrices of $\mathfrak{k}$, then $\mathfrak{t}_c$ is a Cartan subalgebra of $\mathfrak{g}_c$. The corresponding root system is

$$\Delta = \{\gamma_{i,j} = \epsilon_i - \epsilon_j \;\; : \; 1 \leq i \neq j \leq n+1\}$$

where $\varepsilon_i(\mathrm{Diag}(h_1, \dots, h_{n+1})) = h_i$. We choose an ordering in the dual space of $\mathbf{i}\mathfrak{t}$ such that the system of positive roots is $\Delta^+ = \{\gamma_{i,j} \;\; : \; i < j\}$. Let Δ_c

and Δ_n be the set of compact and noncompact roots respectively. We fix a bilinear form B on $\mathfrak{g}$, given by a multiple of the Killing form of $\mathfrak{g}$ such that $B(H_0, H_0) = 1$, and for $\gamma \in \mathfrak{t}_c^*$ we denote by H_γ the element of $\mathfrak{t}$ defined by $\gamma(H) = B(H, H_\gamma)$ for all $H \in \mathfrak{t}$. Denote by

$$\mathfrak{t}^- = \mathbb{R}H_{\gamma_{1,n+1}}, \quad \text{and} \quad \mathfrak{t}^+ = \{H \in \mathfrak{t} \ : \ \gamma_{1,n+1}(H) = 0\}.$$

Since $\{\gamma_{1,n+1}\}$ is a basis of Δ_n, we have that $\mathfrak{t} = \mathfrak{t}^- \oplus \mathfrak{t}^+$, and there exists an automorphism c of $\mathfrak{g}_c$, such that c maps $\mathbf{i}\mathfrak{t}^-$ bijectively to $\mathfrak{a}$, fixing $\mathfrak{t}^+$ (see [**10**, p. 281]). Therefore, $\mathfrak{h} = \mathfrak{t}^+ + \mathfrak{a}$ is a Cartan subalgebra of $\mathfrak{g}$, where

$$\mathfrak{h} = \left\{ H = \begin{bmatrix} \mathbf{i}u_1 & 0 & 0 & t \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & \mathbf{i}u_n & 0 \\ t & 0 & 0 & \mathbf{i}u_1 \end{bmatrix} \ : \ \begin{matrix} \sum_{j=2}^{n} u_j + 2u_1 = 0, \\ \\ t, u_j \in \mathbb{R} \end{matrix} \right\}.$$

Let ε_j be the linear funtional on $\mathfrak{a}_c^*$ defined by

$$\varepsilon_1(H) = \mathbf{i}u_1 + t, \ \ \varepsilon_{n+1}(H) = \mathbf{i}u_1 - t, \ \text{and} \ \ \varepsilon_j(H) = \mathbf{i}u_j \ \ (1 < j \le n).$$

Thus, with the natural ordering, the corresponding set of positive roots is

$$R^+ = \{\alpha_{i,j} = \varepsilon_i - \varepsilon_{j+1} : 1 \le i \le j \le n\}.$$

We denote by Σ the set of restricted roots of the pair $(\mathfrak{g}, \mathfrak{a})$, and we use a compatible ordering in the dual space of $\mathfrak{a}$. Hence, for $n > 1$, $\Sigma^+ = \{\alpha, \frac{1}{2}\alpha\}$ is the set of positive restricted roots, where α is the restriction of $\alpha_{1,n+1}$. The corresponding root spaces are given by

$$\mathfrak{g}_{\alpha/2} = \left\{ \begin{bmatrix} 0 & {}^t\overline{x} & 0 \\ -x & 0 & x \\ 0 & {}^t\overline{x} & 0 \end{bmatrix} ; x \in \mathbb{C}^{n-1} \right\} \ \text{and} \ \ \mathfrak{g}_\alpha = \mathbb{R} \begin{bmatrix} -\mathbf{i} & 0 & \mathbf{i} \\ 0 & 0 & 0 \\ -\mathbf{i} & 0 & \mathbf{i} \end{bmatrix},$$

and thus $m_\alpha = \dim \mathfrak{g}_\alpha = 1$ and $m_{\alpha/2} = \dim \mathfrak{g}_{\alpha/2} = 2(n-1)$.

We will identify the dual space $\mathfrak{a}_c^*$ with $\mathbb{C}$ under the correspondence $\nu = z\frac{1}{2}\alpha \mapsto z$. In other words, since $\alpha(H_0) = 2$, we are identifying ν with $\nu(H_0)$. As usual, let ρ be the linear functional on $\mathfrak{a}$ defined by $\rho(H) = \frac{1}{2}\sum_{\beta \in R^+} m_\beta \beta(H)$. Hence, under the above convention, ρ is identified with n.

We denote by W the Weyl group of G, and we note that in this case $W = \{\pm \mathrm{Id}\}$.

If $A^+ = \{\exp(tH_0) : \ t > 0\}$, then we have the Cartan decomposition of G, $G = K\mathrm{Cl}(A^+)K$. We take on A the measure $da = dt$, on K we use the Haar measure so that the total mass is one, and on G we use the Haar measure such that

$$\int_G f(g)dg = \int_{KA^+K} \delta(t) f(k_1 a k_2) dk_1 da dk_2$$

where $\delta(t) = 2^{2n-1}(\sinh t)^{2(n-1)} \sinh 2t$.

For any $g \in G$, let $g = \kappa(g) \exp H(g) n(g)$ be the Iwasawa decomposition of g.

2.2. Representations. We denote by $\hat{K}$ and $\hat{M}$ the set of irreducible unitary representations of K and M, respectively. For $l \in \mathbb{Z}$ let τ_l be the one-dimensional representation of K associated to the character $\chi_l\left(\left[\begin{smallmatrix} A & 0 \\ 0 & y \end{smallmatrix}\right]\right) = y^l$. We note that every one-dimensional representation of K is of this form. We set $\sigma_l = \tau_{l|M}$.

For each $l \in \mathbb{Z}$, we define $m_\alpha(l) = 1 - 2l$, $\; m_{\alpha/2}(l) = 2(n-1) + 2l$ and $\rho(l) = \frac{1}{2} \sum_{\beta \in R^+} m_\beta(l)\beta$. Thus, under the above identification $\rho(l) = n - l$ (see §2.1).

Let E_l denote the homogeneous line bundle over G/K associated with τ_l. We identify the space of C^∞-sections of E_l with the space $C^\infty(G/K; \tau_l)$ of C^∞-functions on G such that $f(xk) = \tau_l(k)^{-1} f(x)$ for any $x \in G$, $k \in K$. We denote by $D_l = D_l(G/K)$, the space of left invariant differential operators on G which leave $C^\infty(G/K; \tau_l)$ invariant. We note that for $l = 0$, τ_l is the trivial representation of K, and $D_0 = D(G/K)$. Recall that we have the isomorphism (see for instance [**12**, Thm. 2.1])

$$D_l \simeq U(\mathfrak{g})^K / U(\mathfrak{g})^K \cap U(\mathfrak{g})\mathfrak{k}_l$$

where $\mathfrak{k}_l = \{X + \tau_l(X) \mid X \in \mathfrak{z}\}$.

Since τ_l is a one-dimensional representation, $\tau_{l|M}$ is clearly multiplicity free (i.e., no constituent occurs twice), then by [**1**, Thm. 3] D_l is commutative.

Definition 2.1. A complex valued function f on G is said to be τ_l-*radial* if

$$f(k_1 x k_2) = \tau_l(k_1)^{-1} f(x) \tau_l(k_2)^{-1} \text{ for all } g \in G,\; k_1, k_2 \in K.$$

The space of τ_l-radial C^∞-functions on G will be denoted by $C_l^\infty(G)$. We note that $C_l^\infty(G)$ is an algebra with the following convolution product:

$$f \star g(x) = \int_G f(y^{-1}x) g(y) dy.$$

Let f^- denote the restriction to A^+ of a function $f \in C_l^\infty(G)$. It follows from the Cartan decomposition $G = K\mathrm{Cl}(A^+)K$ that $f \in C_l^\infty(G)$ is determined by f^-. For $D \in U(\mathfrak{g})$, we denote by $\Delta_l(D)$ the τ_l-radial component, that is, $\Delta_l(D)$ is a differential operator on A^+ satisfying

$$(Df)^- = \Delta_l(D)(f^-) \qquad \forall\, f \in C_l^\infty(G).$$

We will now recall some facts on the radial component of C, the Casimir operator of $\mathfrak{g}_c$ with respect to B. Let $X_1, \ldots, X_{2(n-1)}$ and X_0 be basis of $\mathfrak{g}_{\alpha/2}$ and $\mathfrak{g}_\alpha$ respectively, such that $-B(X_i, \theta(X_j)) = \delta_{i,j}$. Let $\{U_1, \ldots, U_r\}$ be an orthonormal basis of $\mathfrak{m}$ with respect to $-B_{|\mathfrak{m}}$.

Proposition 2.2. *If $f \in C_l^\infty(G)$ and $C_{\mathfrak{m}}$ denotes the Casimir element of $\mathfrak{m}$ with respect to $-B_{|\mathfrak{m}}$, then*

$$\Delta_l(C)f(a_t) = \left(\frac{d^2}{dt^2} - \tau_l(C_{\mathfrak{m}}) + ((2n-1)\coth t + 2\coth 2t)\frac{d}{dt} - \frac{l^2}{(\cosh t)^2}\right) f(a_t).$$

Proof. If we define, as usually, for $j = 0, \dots, 2(n-1)$

$$Z_j = 2^{-\frac{1}{2}}(X_j + \theta(X_j)), \quad Y_j = 2^{-\frac{1}{2}}(X_j - \theta(X_j)),$$

then, it is easy to see that

$$C = H_0^2 - C_{\mathfrak{m}} + \sum_{j=0}^{2(n-1)} Y_j^2 - \sum_{j=0}^{2(n-1)} Z_j^2.$$

Using arguments analogous to those in [**13**, p. 280] (see also [**3**, Lemma 22]), we can see that in our case we obtain for $f \in C_l^\infty(G/K)$:

$$\begin{aligned} Cf(a_t) = {} & \frac{d^2}{dt^2} f(a_t) - \tau_l(C_{\mathfrak{m}})f(a_t) + (2(n-1)\coth t + 2\coth 2t)\frac{d}{dt} f(a_t) \\ & + (\sinh t)^{-2} \sum_{j=1}^{2(n-1)} \tau_l(Z_j^2)f(a_t) + (\sinh 2t)^{-2}\tau_l(Z_0^2)f(a_t) \\ & + (\coth t)^2 \sum_{j=1}^{2(n-1)} f(a_t)\tau_l(Z_j^2) + (\coth 2t)^2 f(a_t)\tau_l(Z_0^2) \\ & - 2(\sinh t)^{-1}(\coth t) \sum_{j=1}^{2(n-1)} \tau_l(Z_j)f(a_t)\tau_l(Z_j) \\ & - 2(\sinh 2t)^{-1}(\coth 2t)\, \tau_l(Z_0)f(a_t)\tau_l(Z_0) - \sum_{j=0}^{2(n-1)} \tau_l(Z_j^2)f(a_t). \end{aligned}$$

Here, we have used that $\alpha(H_0) = 2$, and therefore $\alpha(\log(a_t)) = 2t$.

On the other hand, it is easy to see that if $X = \begin{bmatrix} 0 & {}^t\overline{x} & 0 \\ -x & 0 & x \\ 0 & {}^t\overline{x} & 0 \end{bmatrix} \in \mathfrak{g}_{\alpha/2}$, then

$$(X + \theta(X)) = \begin{bmatrix} 0 & 2{}^t\overline{x} & 0 \\ -2x & 0 & 0 \\ 0 & & 0 \end{bmatrix}.$$

Hence, by the definition of τ_l, $\tau_l(Z_j) = 0$ for $j = 1, \dots, 2(n-1)$. We also have that $Z_0 = -\mathbf{i}(E_{1,1} - E_{n+1,n+1})$, and then $\tau_l(Z_0) = \mathbf{i}l$. Therefore, the above equation becomes

$$\begin{aligned}Cf(a_t) = \frac{d^2}{dt^2}f(a_t) - \tau_l(C_{\mathfrak{m}})f(a_t) + (2(n-1)\coth t + 2\coth 2t)\frac{d}{dt}f(a_t)\\ - l^2\left((\sinh 2t)^{-2} + (\coth 2t)^2 - 2(\sinh 2t)^{-1}\coth 2t - 1\right)f(a_t)\end{aligned}$$

and the last term of the right-hand side of the above equation equals $\frac{l^2}{(\cosh t)^2}$, as was to be shown. □

Remark. In the case $l = 0$, a τ_0-radial function corresponds to a K-biinvariant function on G, and Proposition 2.2 generalizes the formula for the action of $\Delta(C)$ in this case given in [**8**, §1, p. 667].

3. Spherical functions.

Definition 3.1. If $l \in \mathbb{Z}$ and ϕ is a complex valued τ_l-radial continuous function on G, then ϕ is said to be a τ_l-*spherical* function if $\phi(e) = 1$ and $D\phi = \chi(D)\phi$ for each $D \in D_l$, with $\chi(D) \in \mathbb{C}$.

We have the following description of the τ_l-spherical functions (see for instance [**12**, Prop. 6.1]), as an Einsenstein Integral (see [**13**, 8.12.2]).

Proposition 3.2. *For $l \in \mathbb{Z}$, $g \in G$ and $\nu \in \mathfrak{a}^*$, define*

$$\Phi_{\nu,l}(g) = \int_K e^{-(\nu+\rho)(H(g^{-1}k))}\tau_l(k^{-1}\kappa(g^{-1}k))dk.$$

The function $\phi_{\nu,l}$ is a τ_l-spherical function on G and every τ_l-spherical function is of this form, for some $\nu \in \mathfrak{a}_c^*$. Furthermore, $\phi_{\nu,l} = \phi_{\mu,l}$ if and only if $\mu = s\nu$ for some $s \in W$, and the map $\nu \to \phi_{\nu,l}(g)$ is holomorphic for each $g \in G$.

In order to give another characterization of τ_l-spherical functions, we will recall some facts on the principal series representations of G. Let $P = MAN$ be the minimal parabolic subgroup of G. For $\nu \in \mathfrak{a}_c^*$ and $l \in \mathbb{Z}$, let $(\pi_{\nu,l}, H_P^{l,\nu})$ be the induced representation from P to G of the representation $(\pi^{l,\nu}, H_\nu)$ of P, given by $\pi^{l,\nu}(man)v = a^{(\nu+\rho)}\sigma_l(m)v$.

Since $[\tau_l : \sigma_l] = 1$, by Frobenius reciprocity, τ_l appears in the K-decomposition of $H_P^{l,\nu}$, and then we can define $1_{\nu,l} \in H_P^{l,\nu}$ such that $1_{\nu,l|K} = \tau_l$. Moreover. we have that

$$\begin{aligned}\langle \pi_{\nu,l}(g)1_{\nu,l}, 1_{\nu,l}\rangle &= \int_K 1_{\nu,l}(g^{-1}k)\overline{1_{\nu,l}(k)}dk\\ &= \int_K e^{-(\nu+\rho)H(g^{-1}k)}\tau_{-l}(k^{-1}\kappa(g^{-1}k))dk\end{aligned}$$

This means that $\phi_{\nu,-l}(g) = \langle \pi_{\nu,l}(g) 1_{\nu,l}, 1_{\nu,l} \rangle$, and so it can be shown that the restriction $\Phi^-_{\nu,l}$ of $\Phi_{\nu,l}$ to A^+ satisfies the differential equation:

$$\Delta_l(C)\Phi^-_{\nu,l} = \chi(\nu,l)\Phi^-_{\nu,l}$$

where $\chi(\nu,l) = \nu^2 - \rho^2 + \tau_{-l}(C_{\mathfrak{m}})$ ([**13**, p. 280]).

As in the case of the trivial K-type, these spherical functions are also related with hypergeometric functions, as we will now see.

Proposition 3.3 ([**11**, Prop. 2.6]). *Let $u(t) = 2\cosh t$, then we have that*

$$u(t)^l \cdot (\Delta_l(C) + \rho^2 - \tau_l(C_{\mathfrak{m}})) \cdot u(t)^{-l} = L(l) + \rho(l)^2$$

where $L(l) = \frac{d^2}{dt^2} + ((2n-1)\coth t + (1-2l)\tanh t)\frac{d}{dt}$.

Using this proposition one can see that the function

$$\psi(t) = u^l(t)\Phi^-_{\nu,l}(\exp(tH_0))$$

is an even smooth function on $(0,+\infty)$ satisfying $\psi(0) = 1$ and

$$L(l)\psi = \lambda(\nu,l)\psi \tag{1}$$

where $\lambda(\nu,l) = \nu^2 - \rho(l)^2$. Furthermore, it is known that the Jacobi function

$$\phi^{(n-1,-l)}_{i\nu} = {}_2F_1\left(\frac{n-l+\nu}{2}, \frac{n-l-\nu}{2}, n, -(\sinh t)^2\right)$$

is the unique solution satisfying these conditions (see [**6**, §2.1]). Therefore

$$\Phi_{\nu,l}(\exp tH_0) = (2\cosh t)^{-l}\phi^{(n-1,-l)}_{i\nu}(t).$$

It can also be seen (see [**6**, p. 7] and [**11**, p. 384]) that for $\nu \notin -\mathbb{N}$, a second solution of (1) in $(0,+\infty)$ is given by

(2)
$$\widetilde{Q}_{\nu,l}(t) = (2\cosh t)^{-(\nu+\rho(l))}\,{}_2F_1\left(\frac{n-l+\nu}{2}, \frac{n+l+\nu}{2}, 1+\nu, (\cosh t)^{-2}\right).$$

As a function of ν, $\widetilde{Q}_{\nu,l}$ is holomorphic on $\mathbb{C}\setminus\mathbb{N}$, and for $\nu \notin \mathbb{Z}$, $\widetilde{Q}_{\nu,l}$ and $\widetilde{Q}_{-\nu,l}$ are linearly independent, and so, as in the case $l = 0$, we can write

$$(2\cosh t)^l\Phi_{\nu,l}(\exp(tH_0)) = c(\nu,l)\widetilde{Q}_{-\nu,l}(t) + c(-\nu,l)\widetilde{Q}_{\nu,l}(t) \tag{3}$$

where

$$c(\nu,l) = \frac{2^{n-l-\nu}(n-1)!\,\Gamma(\nu)}{\Gamma(\frac{\nu+n+l}{2})\Gamma(\frac{\nu+n-l}{2})}, \tag{4}$$

and if $\operatorname{Re}\nu > 0$, the asymptotic behavior of $\Phi_{\nu,l}$, as $t \to \infty$, is given by:

$$\Phi_{\nu,l}(\exp(tH_0)) \sim c(\nu,l)\, e^{t(\nu-\rho)}. \tag{5}$$

We will also need the following fact (see [**4**, Prop. 2.2]):

$$\delta^{1/2}\cdot(\Delta_l(C)+\rho^2)\cdot\delta^{-1/2} = \frac{d^2}{dt^2}+\tau_{-l}(C_{\mathfrak{m}})+\sum_{\beta\in R^+}\frac{1}{4}m_{\frac{\beta}{2}}(l)(2-m_{\frac{\beta}{2}}(l)-2m_\beta(l))4\sinh(2t)^{-2}. \tag{6}$$

With all these elements in place, we can adapt most of the arguments in [**8**, **9**], to obtain generalizations of the results of [**9**, §3].

Theorem 3.4. *If $\nu\in\mathbf{C}$, $\nu\notin-\mathbb{N}$, then there exists a function $Q_{\nu,l}\in C^\infty_{-l}(G-K)$ with the following properties:*

(a) $\Delta_l(C)Q_{\nu,l}=\chi(\nu,l)Q_{\nu,l}$. *$Q_{\nu,l}(x)$ is holomorphic for $\nu\notin-\mathbb{N}$ and if $\nu\in-\mathbb{N}$, $Q_{\nu,l}(x)$ has at most a simple pole.*

(b) $\Phi^-_{\nu,l}=c(-\nu,l)Q^-_{\nu,l}+c(\nu,l)Q^-_{-\nu,l}$.

(c) *As $t\mapsto 0$, $Q_{\nu,l}(\exp(tH_0))\sim d(\nu)t^{-2(n-1)}|\log t|^{\delta_{n,1}}$, for some meromorphic function $d(\nu)$ on $\mathbb{C}$, holomorphic if $\nu\notin-\mathbb{N}$. Furthermore, if $\nu\in\mathbb{C}\setminus-\mathbb{N}$, then $Q_{\nu,l}(g)$ lies in $L^1_{\rm loc}(G)$, and if $\operatorname{Re}\nu>\rho$, $Q_{\nu,l}(g)\in L^1(G)$.*

(d) $\lim_{t\mapsto 0^+}\delta(t)\,\frac{d}{dt}Q_{\nu,l}(\exp(tH_0))=-2\nu c(\nu,l)$.

(e) *If $f\in C^\infty_c(G/K,\tau_l)$ and $\nu\notin-\mathbb{N}$ then for $\operatorname{Re}\nu>\rho$*

$$\int_G Q_{\nu,l}(x^{-1}y)(C-\lambda(\nu,l)\mathrm{Id})\mathrm{f(y)dy}=-2\nu\, c(\nu,l)f(x). \tag{7}$$

Proof. Let $Q_{\nu,l}(ka_tk')=\tau_l(k)u(t)^{-l}\widetilde{Q}_{\nu,l}(t)\tau_l(k')$. As we noted before, since $\widetilde{Q}_{\nu,l}$ is a solution of $L(l)g(t)=\lambda(\nu,l)g(t)$, $Q_{\nu,l}$ is a solution of $\Delta_l(C)f^-=\chi(\nu,l)f^-$.

It is clear from the definition that $Q_{\nu,l}\in C^\infty_{-l}(G\setminus K)$, and by the above observations, it satisfies (a). From our definition, it is also clear that (b) is equivalent to (3).

The proof of (c) is similar to that in the case of the trivial K-type, so it will be omitted (see [**9**]). Note that (2) implies that $\widetilde{Q}_{\nu,l}(t)\sim e^{-t(\nu+\rho(l))}$ when $t\mapsto\infty$, therefore $Q_{\nu,l}(\exp(tH_0))\sim e^{-(\nu+\rho)}$ when $t\mapsto\infty$. This fact allows us to prove (d) as in the case of $l=0$.

In order to see (e), we first note that since if $f\in C^\infty_c(G/K,\tau_l)$, so is $L_{x^{-1}}f$, then it suffices to see that

$$\int_G Q_{\nu,l}(y)(C-\lambda(\nu,l))f(y)\,dy=-2\nu c(\nu,l)f(e).$$

The left-hand side equals

$$\int_0^\infty Q_{\nu,l}(a_t)(C-\lambda(\nu,l))\int_K\tau_l(k)f(ka_t)\,dk\,\delta(t)\,dt.$$

If $f \in C_c^\infty(G/K, \tau_l)$, then $f^l(a_t) := \int_K \tau_l(k) f(ka_t)\, dk$ is a τ_l-radial function on G. Hence we can replace C by its radial part to obtain

$$\int_0^\infty \left(\delta^{\frac{1}{2}}(t) Q_{\nu,l}(a_t) \delta^{\frac{1}{2}}(t) \Delta_l(C) f^l(a_t) - \delta^{\frac{1}{2}}(t) \Delta_l(C) Q_{\nu,l}(a_t) \delta^{\frac{1}{2}}(t) f^l(a_t)\right) dt.$$

Now using the radialization (6) and arguing as in [**9**, pp. 1225], we obtain that the above is equal to

$$\int_0^\infty \frac{d}{dt}\left(\delta(t) Q_{\nu,l}(\exp(tH_0)) \frac{d}{dt} f^l(a_t) - \delta(t) \frac{d}{dt} Q_{\nu,l}(\exp(tH_0)) f^l(a_t)\right) dt.$$

Therefore, looking at the asymptotic behavior as $t \mapsto 0$, and as $t \mapsto \infty$, we obtain that the above integral equals $\lim_{t \mapsto 0^+} \delta(t) \frac{d}{dt} Q_{\nu,l}(\exp(tH_0)) f^l(e)$. Then, using (d) we are done. □

4. The residues of the resolvent kernel.

Let $\widetilde{R}(\lambda(\nu,l))$ denote the kernel operator with kernel $K_{\nu,l}(x,y) := -\frac{Q_{\nu,l}(x^{-1}y)}{2\nu c(\nu,l)}$ and let $R(\lambda(\nu,l))$ denote the resolvent of C acting on $L^2(G/K, \tau_l)$. By Theorem 3.4, if $\operatorname{Re}\nu > \rho$, then $\widetilde{R}(\lambda(\nu,l)) = R(\lambda(\nu,l))$.

Since $\nu \mapsto K_{\nu,l}$ is defined also for $\operatorname{Re}\nu \le \rho$, we are interested in $\widetilde{R}(\lambda(\nu,l))$ acting on $C_c^\infty(G/K,\tau_l)$ as a meromorphic continuation of $R(\lambda(\nu,l))$. In the next theorem, we will give a description of the singularities of $\widetilde{R}(\lambda(\nu,l))$.

Theorem 4.1. *$\widetilde{R}(\lambda(\nu,l))$ has simple poles lying at $\nu = \nu^\pm_{k,l}$ with $\nu^-_{k,l} = -|l| - n - 2k$, $k \in \mathbb{N}_0$, and $\nu^+_{k,l} = |l| - n - 2k$, for $k \in \mathbb{N}_0$ such that $|l| - n - 2k \ge 0$. If ν is a pole and we set, for $f \in C_c^\infty(G,K,\tau_l)$, $T_\nu(f) := \operatorname{Res}_{z=\nu} \widetilde{R}(\lambda(z,l))(f)$, then $T_\nu(f) = p(\nu)\, f * \check{\Phi}_{\nu,-l}$.*

Proof. We know that $\Phi_{\nu,l}(g)$ is everywhere holomorphic as a function of ν. Hence, using (a) and (b) of Theorem 3.4, we find that the poles of $K_{\nu,l}$ are precisely the zeros of $2\nu c(\nu,l)$.

Furthermore, it is easy to see from (4) that the zeros of $c(\nu,l)$ are at $\nu = \nu^\pm_{k,l}$, as in the statement of the theorem. On the other hand, we have that $|\nu^+_{k,l}| < |\nu^-_{0,l}|$, when $|l| > n$, so that $\nu^+_{k,l}$ is defined. Hence $\frac{Q_{-\nu,l}}{2\nu c(-\nu,l)}$ is analytic at $\nu^\pm_{k,l}$. Thus, for $f \in C_c^\infty(G/K,\tau_l)$, using Theorem 3.4(b), we get that if ν is a pole, then

$$\operatorname{Res}_{z=\nu} \widetilde{R}(\lambda(z,l))(f) = p(\nu)\, f * \check{\Phi}_{\nu,-l},$$

where $p(\nu) = -\operatorname{Res}_{z=\nu}(2\nu c(\nu,l) c(-\nu,l))^{-1}$. □

Now we want to study the image of these operators, and in order to do this, we will introduce certain irreducible representations of K, for $n > 1$.

For $p, q \in \mathbb{N}_0$, we denote by $V_{p,q}$, the set of harmonic polynomials in $z \in \mathbb{C}^n$ of bidegree (p, q), and define on this space the action of K given by

$$\tau_{l,p,q}\left(\left[\begin{array}{cc} A & 0 \\ 0 & y \end{array}\right]\right) f(z) = y^{q-p+l} f({}^t z A).$$

Proposition 4.2 ([**2**, § 2]). *Let τ be an arbitrary K-type which contains the M-type σ_l. Then there exist $p, q \in \mathbb{N}_0$ such that τ is equivalent with $V_{p,q}$.*

Actually, if we put $F_{p,q}(z) = z_1^p \bar{z}_1^q \; {}_2F_1(-p, -q, n-1, -(|z_2|^2 + \cdots + |z_n|^2)/|z_1|^2)$ then $F_{p,q} \in V_{p,q}$, and it is easy to see that $\tau_{l,p,q}(X) F_{p,q} = \sigma_l(X) F_{p,q}$ for $X \in M$.

For any $P \in \operatorname{Hom}_M(V_{p,q}, H_l)$ and $\nu \in \mathfrak{a}_c^*$ we can define a K-intertwining operator from $V_{p,q}$ to $H_P^{l,\nu}$ by $L(P, f, \nu)(g) = e^{-(\nu+\rho)H(g)} P(\tau_{l,p,q}(\kappa(g)^{-1}) f)$, for $f \in V_{p,q}$ (see [**13**, 8.11.4]). Furthermore, since $[\tau_{l,p,q} : \sigma_l] = 1$, we have that $f \mapsto L(P, f, \nu)$ is an injective K-intertwining operator.

Let P denote the linear map from $V_{p,q}$ to $\mathbb{C}$ defined by $P(f) = f(1, 0, \ldots, 0)$. It is clear that $P \in \operatorname{Hom}_M(V_{p,q}, H_l)$, and then we have the related homomorphism $L(P, f, \nu)$. Let $\widetilde{V}_{p,q} \subset H_P^{l,\nu}$ denote the image of $V_{p,q}$ under this homomorphism and let $A(w, l, \nu) : H_P^{\nu,l} \mapsto H_P^{-\nu,l}$ denote the standard intertwining operator, where $w = \operatorname{diag}(-1, -1, 1, \ldots, 1) \in K$ is a representative of the nontrivial element of W.

In particular, from [**2**, § 3] we have that

$$A(w, l, \nu) L(P, F_{p,q}, \nu) = (-1)^{p+q} c_{\tau_{l,p,q}}(\sigma_l, \nu) L(P, F_{p,q}, -\nu)$$

where $c_{\tau_{l,p,q}}(\sigma_l, \nu)$ is given by

$$c_{\tau_{l,p,q}}(\sigma_l, \nu) = \frac{k\Gamma(\nu) \prod_{j=0}^{p-1}(\nu - n + l - 2j) \prod_{j=0}^{q-1}(\nu - n - l - 2j)}{\Gamma(\frac{\nu+n-l+2p}{2}) \Gamma(\frac{\nu+n+l+2q}{2})}.$$

If $\nu \neq 0$, let $D_\nu^l = \{(p, q) \in \mathbb{N}_0^2 \;:\; c_{\tau_{l,p,q}}(\sigma_l, \nu) = 0\}$. If $(p, q) \in D_\nu^l$, it is clear that $L(P, F_{p,q}, \nu) \in \operatorname{Ker} A(w, l, \nu)$ which is a G-module. Hence $\widetilde{V}_{p,q} \subset \operatorname{Ker} A(w, l, \nu)$, and moreover, by Frobenius reciprocity and Proposition 4.2 we have that

$$\operatorname{Ker} A(w, l, \nu) = \oplus_{(p,q) \in D_\nu^l} \widetilde{V}_{p,q}.$$

It is easy to see that

$$D_{\nu_k^-}^l = \left\{(p, q) \in \mathbb{N}_0^2 \;:\; p \leq k + \frac{l + |l|}{2},\; q \leq k + \frac{|l| - l}{2}\right\},$$

and therefore $\operatorname{Ker} A(w, l, \nu_k^-) = \sum_{(p,q) \in D_{\nu_k^-}^l} \widetilde{V}_{p,q}$ is a finite dimensional $(\mathfrak{g}, K)$-module. It is clear that its restriction contains $\tau_l = \tau_{l,0,0}$.

For $\nu = 0$, since we know that $c(\nu, l)$ has a pole, we can consider the normalized intertwining operator $B(w, l, \nu) = \Gamma(\nu)^{-1} A(w, l, \nu)$; now, since

$\Gamma(\nu)^{-1}c_{\tau_{l,p,q}}(\sigma_l,\nu)$ is holomorphic at $\nu=0$, then as in the other cases, we have that $\operatorname{Ker} B(w,l,0)=\sum_{(p,q)\in D_0^l}\widetilde{V}_{p,q}$, where

$$D_0^l=\left\{(p,q)\in\mathbb{N}_0^2\ :\ \Gamma(\nu)^{-1}c_{\tau_{l,p,q}}(\sigma_l,\nu)|_{\nu=0}=0\right\}.$$

We note that since $\nu=0$ is a pole, then $|l|-n=2k$ with $k\in\mathbb{N}$ and one can verifies that

$$D_0^l=\left\{(p,q)\in\mathbb{N}_0^2\ :\ p\le k=\frac{l-n}{2}\right\},\quad \text{if } l>0,$$

$$D_0^l=\left\{(p,q)\in\mathbb{N}_0^2\ :\ q\le k=\frac{-l-n}{2}\right\},\quad \text{if } l<0.$$

It is well-known that $H_P^{l,\nu}$ is equivalent to $H_{\overline{P}}^{l,-\nu}$ and the intertwining operator is $R(w)$, where R is the right regular representation of G. It is also known that if $A(\overline{P},P,\sigma_l,\nu)$ denotes the standard intertwining operator from $H_P^{l,\nu}$ to $H_{\overline{P}}^{l,\nu}$, then $A(w,l,\nu)=R(w)A(\overline{P},P,\sigma_l,\nu)$ (see [**5**, VII §4]).

Let $V(\mu,l)$ denote the image of the residue of $\widetilde{R}(\lambda(\nu,l))$ at $\nu=\mu$, and $V(\mu,l)_K$ the space of K-finite vectors in $V(\mu,l)$. Now, using a generalization of Helgason's theorem ([**10**, §7]), we can give a very explicit description of $V(\mu,l)_K$.

Theorem 4.3. *If μ is a pole of $\widetilde{R}(\lambda(\nu,l))$, then $V(\mu,l)_K$ is a $(\mathfrak{g},K)$-module. This module is of finite dimension only in the case when $\mu=\nu_{k,l}^-$ for $k\in\mathbb{N}_0$. The modules corresponding to $\mu=\nu_{k,l}^+$ are equivalent, as $(\mathfrak{g},K)$-modules, to holomorphic discrete series representations. Moreover, in the case when $\mu=\nu_{k,l}^-$ and $\mu=0$ these $(\mathfrak{g},K)$-modules are isomorphic to* $\operatorname{Ker} A(w,l,\nu_{k,l}^-)$ *and* $\operatorname{Ker} B(w,l,0)$ *respectively.*

Proof. If $f\in C_c^\infty(G/K,l)$ and $x\in G$, by Theorem 4.1 we have that:

$$\begin{aligned}T_{\nu_{k,l}^\pm}(f)(x)&=p(\nu_{k,l}^\pm)\,f*\check{\phi}_{\nu_{k,l}^\pm,-l}(x)\\&=p(\nu_{k,l}^\pm)\,\left\langle\pi_{\nu_{k,l}^\pm}(x^{-1})\,\pi_{\nu_{k,l}^\pm,l}(f)1_{\nu_{k,l}^\pm,l},1_{\nu_{k,l}^\pm,l}\right\rangle.\end{aligned}$$

Hence, $V(\nu,l)_K$ is isomorphic to the $(\mathfrak{g},K)$-module generated by $1_{l,\nu}$, so we will now describe this module.

In order to do this, we will give a condition on ν for $1_{l,\nu}$ to generate a finite dimensional $(\mathfrak{g},K)$-submodule of $H_P^{l,\nu}$.

For $\lambda\in\mathfrak{h}_c^*$ we define

$$m_0=\lambda(\mathbf{i}X)\ \text{ and }\ m_1=\frac{2\langle\lambda,\alpha\rangle}{\langle\alpha,\alpha\rangle},$$

where as in [**10**, 4.4], $X = \begin{bmatrix} -1 & & \\ & \frac{2}{n-1}\mathrm{Id} & \\ & & -1 \end{bmatrix}$.

If $\lambda|_{\mathfrak{t}^+\cap\mathfrak{k}_1} = 0$, then by [**10**, Prop 7.1] , λ is dominant integral if and only if m_0 and m_1 are integers such that $|m_0| \leq m_1$ and $(-1)^{m_0} = (-1)^{m_1}$.

We note that $\{\alpha, \varepsilon_1, \ldots, \varepsilon_n\}$ is a basis of $\mathfrak{h}_c^*$, then straightforward calculation shows that if $\lambda = a_1\alpha + \sum_{i=2}^n a_i\varepsilon_i \in \mathfrak{h}_c^*$, then $\lambda|_{\mathfrak{t}^+\cap\mathfrak{k}_1} = 0$ if and only if $a_2 = a_3 = \cdots = a_n$. Hence, if we denote $\beta = \sum_{i=2}^n \varepsilon_i$, and $\lambda = a_1\alpha + a_0\beta$, then we have that $m_0 = -2a_0$ and $m_1 = 2a_1$.

On the other hand, by [**10**, Thm. 7.2], $\lambda = a_1\alpha + a_0\beta$, is a highest weight of a finite dimensional irreducible representation of G, whose restriction to K contains the one-dimensional K-type χ_{m_0} with multiplicity one.

Furthermore, in the proof of the theorem, we can see that this representation is equivalent to $\operatorname{Im} A(\overline{P}, P, \sigma_l, \mu)$, a subrepresentation of $H_{\overline{P}}^{l,\nu}$, where $\mu = \lambda_{|\mathfrak{a}} + \rho$.

Therefore, $1_{l,\nu}$ generates a finite dimensional G-submodule of $H_P^{l,\nu}$ (with highest weight λ) if and only if $\nu = -\lambda_{|\mathfrak{a}} - \rho$, where $\lambda = \frac{l}{2}\beta + (|l| + 2k)\frac{\alpha}{2}$.

We note that because of our identification, $\nu = \nu_{k,l}$, as we want to show.

On the other hand, we have proved that $\operatorname{Ker} A(w, l, \nu_{k,l}^-)$ is a finite dimensional submodule of $H_P^{l,\nu_{k,l}^-}$ which contains $1_{l,\nu_{k,l}^-}$. We also know that $H_P^{l,\nu_{k,l}^-}$ is equivalent to $H_{\overline{P}}^{l,-\nu_{k,l}^-}$, and it has only one irreducible representation ([**5**, p. 273]). Therefore, it is clear that the $(\mathfrak{g}, K)$-submodule of $H_P^{l,\nu_{k,l}^-}$ generated by $1_{l,\nu_{k,l}^-}$ is $\operatorname{Ker} A(w, l, \nu_{k,l}^-)$.

We will now study the case when $\nu = \nu_{k,l}^+$. We begin by observing that in the rank one case, [**11**, Thm. 5.1] states that if $\lambda \in \mathfrak{a}_c^*$, $\lambda = \nu.\frac{\alpha}{2}$, with $\nu \geq 0$, then $\Phi_{\lambda,l}$ belongs to $L^2(G/K, \tau_l)$ if and only if

$$\frac{\langle -\lambda - \rho(|l|), \alpha\rangle}{\langle \alpha, \alpha\rangle} \in \mathbb{N}. \tag{8}$$

In particular, there exists $\lambda \in \mathfrak{a}_c^*$ such that $\Phi_{\lambda,l}$ belongs to $L^2(G/K, \tau_l)$ if and only if $|l| > n$. We note that (8) means that $-\nu - (n - |l|) = 2k$ with $k \in \mathbb{N}$, or equivalently, $\nu = \nu_{k,l}^+$ for some k. Hence the $(\mathfrak{g}, K)$-module generated by $1_{l,\nu}$ in $H_P^{l,\nu}$ is infinitesimally equivalent to a discrete series representation if and only if $\nu = \nu_{k,l}^+$

Moreover, Shimeno proves in [**11**, Thm 5.10], that these are actually infinitesimally equivalent to holomorphic discrete series representations.

For $\nu = 0$, it is known that if $H_p^{l,0}$ is reducible, it is a sum of two inequivalent irreducible representations. These representations are called limits of discrete series. Since they are inequivalent and $B(w, l, 0)$ is an intertwining operator, it is easy to see that $\operatorname{Ker} B(w,l,0)$ is the $(\mathfrak{g},K)$-submodule of $H_P^{l,0}$ generated by $1_{l,0}$, concluding the proof. □

Remark 4.4. We wish to point out that every (irreducible) finite dimensional, discrete series, or limit of discrete series representation of G, containing a one-dimensional K-type can be seen as a residue of the resolvent kernel. That is, if (π, H_π) is a finite dimensional representation of G containing a one-dimensional K-type χ_m, then there exists a line bundle over G/K such that this representation is isomorphic to the residue of the meromorphic continuation of the resolvent of the Casimir operator acting on that line bundle. In fact, if λ is the highest weight of π, then by [**10**, Thm 7.2] $\lambda = a\alpha + b\beta$, where $a = |b| + k$. Then by the above, H_π is isomorphic to $V_{k,2b}$, the image of the residue of $R(\lambda(\nu, 2b))$ at $\nu = \nu^-_{k,2b}$.

In the case that (π, H_π) is a discrete series, this implies that $\Phi_{\lambda,m}$ belongs to $L^2(G/K, \tau_m)$, and then by [**11**, Thm 5.1] $\lambda = \nu^+_{k,m}$. Hence, H_π is isomorphic to $V(\nu^+_{k,m}, m)$.

Finally, if (π, H_π) is a limit of discrete series containing the one-dimensional K-type τ_m, it means that $H_P^{m,0}$ is reducible, and so $m \equiv n \quad (2)$ and $|m| > n$ ([**5**, p. 621]). Thus, H_π is isomorphic to the image of the residue of $R(\lambda(\nu, m))$ at $\nu = \nu^+_{\frac{|m|-n}{2},m}$.

Remark 4.5. We observe that $\nu = 0$ is not a pole of the resolvent kernel $\widetilde{R}_l(\lambda(\nu, l))$, in the case when $l = 0$.

We will now use the Weyl dimension formula to calculate the dimension of the representation $V(\nu^-_{k,l})$. The fundamental weights of $\mathfrak{g}_c$ are $\Lambda_j = \epsilon_1 + \cdots + \epsilon_j$, $1 \le j \le n$, hence $\alpha = \Lambda_1 + \Lambda_n$ and $\beta = \Lambda_n - \Lambda_1$. Hence, we are interested in the dimension of the $\mathfrak{g}_c$-module asociated to $\Lambda_{k,l} = \frac{l}{2}\beta + \left(\frac{|l|+2k}{2}\right)\alpha = \left(\frac{|l|-l}{2} + k\right)\Lambda_1 + \left(\frac{|l|+l}{2} + k\right)\Lambda_n$. Then we have that

$$\begin{aligned}\dim(V_{k,l}) &= \prod_{1\le i<j\le n+1} \frac{\langle \Lambda_{k,l} + \rho, \epsilon_i - \epsilon_j\rangle}{\langle \rho, \epsilon_i - \epsilon_j\rangle} \\ &= \prod_{1<j\le n} \frac{\frac{|l|-l}{2} + k + j - 1}{j-1} \cdot \prod_{1<i\le n} \frac{\frac{|l|+l}{2} + k + n + 1 - i}{n+1-i} \\ &\quad \cdot \frac{1}{n}\left(\frac{|l|+l}{2} + \frac{|l|-l}{2} + 2k + n\right)\end{aligned}$$

and so

$$\dim(V_{k,l}) = \binom{\frac{|l|-l}{2} + k + n - 1}{\frac{|l|-l}{2} + k} \cdot \binom{\frac{|l|+l}{2} + k + n - 1}{\frac{|l|+l}{2} + k} \cdot \frac{|l| + 2k + n}{n}.$$

5. The case $G = \mathrm{SU}(1,1)$.

We will consider now the case when $G = \mathrm{SU}(1,1)$. We shall see that the results will be entirely similar to those in the case of $\mathrm{SU}(n,1), n > 1$, but we shall analyze this case separately, because the notation and some of the definitions are different.

We have that $\theta(X) = \overline{X}^t$ and $\mathfrak{g} = \mathfrak{k} + \mathfrak{p}$, where

$$\mathfrak{k} = \left\{ \begin{bmatrix} \mathbf{i}t & 0 \\ 0 & -\mathbf{i}t \end{bmatrix} : t \in \mathbb{R} \right\} \text{ and } \mathfrak{p} = \left\{ \begin{bmatrix} 0 & b \\ \bar{b} & 0 \end{bmatrix} : b \in \mathbb{C} \right\}.$$

If $H_0 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ and $\mathfrak{a} = \mathbb{R}H_0$, then $M = \{\pm I\}$ and in this case, $\hat{K} = \{\tau_l : l \in \mathbb{Z}\}$ and $\hat{M} = \{1, \epsilon\}$, where ϵ denotes the nontrivial character of M. Therefore, for each $\nu \in \mathbb{C}$ we have two principal series representations, $H^{\nu,+}$ and $H^{\nu,-}$ corresponding to 1 and ϵ, respectively and $\tau_{l|\mathfrak{m}} = I$ if and only if $l \equiv 0$ (2), $\tau_{l|\mathfrak{m}} = \epsilon$, otherwise. Now, Proposition 2.2 may be stated as follows:

$$\Delta_l(C) = \frac{d^2}{dt^2} + 2\coth t \frac{d}{dt} + l^2(\cosh t)^{-2}. \tag{9}$$

Furthermore, Proposition 3.3 becomes

$$u(t)^l \circ (\Delta_l(C) + \rho^2) \circ u(t)^{-l} = \frac{d^2}{dt^2} + (\coth t + (1-2l)\tanh t)\frac{d}{dt} + \rho(l)^2,$$

where $u(t) = 2\cosh t$.

We now define a differential operator on $\mathbb{R}^+$, as in Proposition 3.3:

$$L(l) = \frac{d^2}{dt^2} + (\coth(t) + (1-2l)\tanh(t))\frac{d}{dt}.$$

As in the case when $n > 1$, one can relate the spherical functions $\Phi_{\nu,l}$ with the solutions of $L(l)f = (\nu^2 - \rho(l)^2)f$, where $\rho(l) = 1 - l$ (see §2.1), and find that they are given by

$$\Phi_{\nu,l}(\exp(tH_0)) = (2\cosh t)^{-l}\phi_{i\nu}^{(0,-l)}(t).$$

In the same way, we can see that if we take the solution in (2) of the above equation for $n = 1$ and $\nu \notin -\mathbb{N}$, we get the following eigenfunction of $\Delta_l(C)$ on A^+:

$$\begin{aligned} &Q_{\nu,l}(\exp tH_0) \\ &= (2\cosh(t))^{-(\nu+\rho(l))}\, {}_2F_1\left(\frac{1-l+\nu}{2}, \frac{1+l+\nu}{2}, 1+\nu, \cosh(t)^{-2}\right), \end{aligned}$$

which satisfies

$$(2\cosh t)^l \Phi_{\nu,l}(\exp(tH_0)) = c(\nu,l)Q_{-\nu,l}(a_t) + c(-\nu,l)Q_{\nu,l}(a_t)$$

where

$$c(\nu, l) = \frac{2^{1-l-\nu}\Gamma(\nu)}{\Gamma(\frac{\nu+1+l}{2})\Gamma(\frac{\nu+1-l}{2})}.$$

With all this in place, we can prove Theorem 3.4 in our case, obtaining in the same way the meromorphic continuation of the resolvent kernel $K_{\nu,l}(x,y) = -\frac{Q_{\nu,l}(x^{-1}y)}{2\nu c(\nu,l)}$, hence we have the following theorem, which gives results analogous to those in Theorem 4.1 and Theorem 4.3 in the present case.

Theorem 5.1. *$\widetilde{R}(\lambda(\nu))$ has simple poles lying at $\nu = \nu_{k,l}^{\pm}$, where $\nu_{k,l}^{-} = -|l| - 1 - 2k$, with $k \in \mathbb{N}_0$, and $\nu_{k,l}^{+} = |l| - 1 - 2k$, with $k \in \mathbb{N}$ such that $|l| - 1 - 2k \geq 0$. If ν is a pole, $\mathrm{Res}_{z=\nu}\, \widetilde{R}(\lambda(z))(f) = p(\nu)\, f * \check{\phi}_{\nu,-l}$.*

Moreover, the rank of the residue of $\widetilde{R}(\lambda(\nu))$ at $\nu = \nu_{k,l}^{-}$ is a finite dimensional $(\mathfrak{g}, K)$-module. The residues at $\nu = \nu_{k,l}^{+}$ are infinitesimally equivalent to holomorphic discrete series representations.

Proof. Since the proof can be done as in the general case, we will only prove the representation theory assertion. As in the general case, we have that $\phi_{\nu,-l}(g) = \langle \pi_{\nu,l} 1_{l,\nu}, 1_{l,\nu}\rangle$, where $1_{l,\nu}(kan) = a^{-(\nu+\rho)}\tau_l(k)^{-1}$ belongs to $H^{\nu,+}$ (resp. $H^{\nu,-}$) if l is even (resp. l odd).

If we denote $k(\theta) = \begin{bmatrix} e^{\mathrm{i}\theta} & 0 \\ 0 & e^{-\mathrm{i}\theta} \end{bmatrix}$, then $\tau_l(k(\theta)) = e^{-\mathrm{i}l\theta}$, and therefore $1_{\nu,l}(k(\theta)an) = a^{-(\nu+\rho)}e^{\mathrm{i}l\theta}$. On the other hand, since $\mathfrak{g}$ is isomorphic to $\widetilde{\mathfrak{g}} = sl(2,\mathbb{R})$, for each $\nu \in \mathbb{C}$, $1_{\nu,l}$ can be identified with the function of $SL(2,\mathbb{R})$ defined by

$$\phi_{-l}\left(\begin{bmatrix} e^{-\mathrm{i}t} & 0 \\ 0 & e^{-\mathrm{i}t} \end{bmatrix}\begin{bmatrix} e^{2\mathrm{i}t} & e^{2\mathrm{i}t}x \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}\right) = e^{(\nu+1)t}e^{-\mathrm{i}l\theta}.$$

This function belongs to $H(\nu) = \{f : Sl(2,\mathbb{R}) \mapsto \mathbb{C} \;:\; f(ank) = a^{\nu+1}f(k),\ f_{|K} \in L^2(K)\}$ (see [**7**, p. 116]) and the $(\mathfrak{g}, K)$-modules of $H^{\nu,\pm}$ generated by $1_{\nu,l}$ are isomorphic to the $(\widetilde{\mathfrak{g}}, K)$-modules of $H(\nu)$ generated by ϕ_{-l}.

We note that the difference in the sign (with [**7**]) is due to the different choices in the Iwasawa decompositions.

We thus have that $V_{\nu_{k,l}^{-},l} \simeq \displaystyle\sum_{j=1}^{|l|-1+2k} \langle \phi_{-(|l|+2(k-j))}\rangle$, and therefore $V_{\nu_{k,l}^{-},l}$ is finite dimensional. If $\nu_{k,l}^{+} \neq 0$, then we obtain the discrete series:

$$V_{\nu_{k,l}^{+},l} \;\simeq\; \begin{cases} \displaystyle\sum_{\substack{j\equiv l\,(2)\\ j\leq -l+2k}} \langle \phi_j\rangle & l > 0 \\ \displaystyle\sum_{\substack{j\equiv l\,(2)\\ j\geq -l-2k}} \langle \phi_j\rangle & l < 0. \end{cases}$$

Finally, we can see that if $\nu = 0$ is a pole then l is odd, and therefore we obtain the so called 'Mock discrete series' or limit of discrete series representations:

$$V_{0,l} \simeq \begin{cases} \sum\limits_{\substack{j\equiv l\,(2)\\ j\leq -1}} \langle \phi_j \rangle & l>0 \\ \sum\limits_{\substack{j\equiv l\,(2)\\ j\geq 1}} \langle \phi_j \rangle & l<0 \end{cases}$$

thus concluding the proof. □

References

[1] A. Deitmar, *Invariant operators on higher K-types*, J. Reine Angew. Math., **412** (1990), 97-107, MR 92b:22013, Zbl 0712.43006.

[2] M. Eguchi and S. Koizumi, *The explicit formula for the Harish-Chandra C-function of $SU(n,1)$ for arbitrary irreducible representation of K which contain one-dimensional K-types*, Proc. Japan Acad., Ser A, **70** (1994), 256-259, MR 95i:22019, Zbl 0838.43008.

[3] Harish-Chandra, *Differential equations and semisimple Lie groups*, manuscript, 1960, in 'Collected Papers', Vol. III, 57-120, Springer Verlag, New York, 1984, MR 85e:01061c.

[4] G.J. Heckman and E.M. Opdam, *Root system and hypergeometric functions* I, Compositio Mathematicae, **64** (1987), 329-352, MR 89b:58192a, Zbl 0656.17006.

[5] A.W. Knapp, *Representation Theory of Semisimple Groups*, Princeton Univ. Press, Princeton, New Jersey, 1986, MR 87j:22022, Zbl 0604.22001.

[6] T. Koornwinder, *Jacobi functions and analysis on noncompact semisimple Lie groups*, in 'Special Functions: Group Theorical Aspects and Applications', Reidel, Dordrecht, 1984, 1-85, MR 86m:33018, Zbl 0584.43010.

[7] S. Lang, $SL(2,\mathbb{R})$, Springer-Verlag, New York, 1985, Zbl 0583.22001.

[8] R.J. Miatello and N.R. Wallach, *The resolvent of the Laplacian on negatively curved locally symmetric spaces of finite volume*, J. Differential Geom., **36** (1992), 663-698, MR 93i:58160, Zbl 0766.53044.

[9] R. Miatello and C.E. Will, *The residues of the resolvent on Damek-Ricci spaces*, Proc. Amer. Math. Soc., **128** (2000), 1221-1229, MR 2001i:22011, Zbl 0941.22005.

[10] H. Schlichtkrull, *One dimensional K-types in finite dimensional representations of semisimple Lie groups: A generalization of Helgason's theorem*, Math. Scand., **54** (1984), 279-294, MR 85i:22014, Zbl 0545.22015.

[11] N. Shimeno, *The Plancherel formula for spherical functions with a one-dimensional K-type on a simply connected simple Lie group of Hermitian type*, J. Funct. Anal., **121** (1994), 330-388, MR 95b:22036, Zbl 0830.43018.

[12] ——, *Eigenspaces of invariant differential operators on a homogeneous line bundle on a Riemannian symmetric space*, J. Fac. Sci. Univ. Tokyo, Sect. I.A, Math., **37** (1990), 201-234, MR 91j:22008, Zbl 0713.58053.

[13] N. Wallach, *Harmonic Analysis on Homogeneous Spaces*, Pure and Appl. Math., **19**, Marcel Dekker Inc., New York, 1973, MR 58 #16978, Zbl 0265.22022.

Received June 8, 2000 and revised May 10, 2001. This research was supported by a fellowship from CONICET and research grants from CONICET and SeCyT UNC (Argentina).

Ciem, FaMAF
Universidad Nacional de Córdoba
5000 Córdoba
Argentina
E-mail address: cwill@mate.uncor.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 209, No. 1, 2003

THE HEAT KERNEL AND THE RIESZ TRANSFORMS ON THE QUATERNIONIC HEISENBERG GROUPS

Fuliu Zhu

In this paper we use the method of stochastic integral due to Gaveau to construct the heat kernel for the quaternionic Heisenberg groups, and then follow the line of Coulhon et al. to deduce the uniformly boundedness of the Riesz transforms on these nilponent Lie groups.

0. Introduction.

As the heat kernel plays an important role in many problems in harmonic analysis, an explicit usable expression is very much desirable.

An explicit expression for the heat kernel for the Heisenberg group $H^n = \mathbf{C}^n \times \mathbf{R}$ was obtained by Hulanicki [**9**] and by Gaveau [**7**]. Gaveau [**7**] also obtained the heat kernel for free nilpotent Lie groups of step two. Cygan [**4**] obtained the heat kernel for all nilpotent Lie groups of step two. But neither Gaveau's expression for free nilpotent Lie groups nor Cygan's expression for arbitrary nilpotent Lie groups of step two were as explicit as in the case of Heisenberg groups.

The Hulanicki-Gaveau's formula for the heat kernel for the Heisenberg group has many interesting applications: Hueber [**8**] et al. used it to describe the Martin boundary corresponding to the sublaplacian of the Heisenberg group, Garofalo [**6**] et al. used it to study the regularity of boundary points in the Dirichlet problem for the heat equation on the Heisenberg group, while Coulhon [**3**] et al. used it to show the uniform boundedness of Riesz transforms on the Heisenberg group. Although these applications are very impressive, they depend heavily on explicit expressions for the heat kernel. All of these works motivate the following question: Are there other nilpotent Lie groups for which the expressions for the heat kernel are as explicit as in the case of the Heisenberg group?

The first aim of this paper is to look for such formulae for the heat kernel for the quaternionic Heisenberg groups. These groups are defined by replacing the complex field $\mathbf{C}$ by the field of quaternions $\mathbf{H}$ in the definition of $\mathbf{H}^n$. More precisely, we make $\mathbf{H}^n \times \mathbf{R}^3$ into a nilpotent Lie group of step two by suitably defining the group operation. On this group there is a natural sublaplacian with an associated heat kernel. We use the method of Gaveau

[**7**], i.e., the stochastic integral, to calculate the heat kernel for the quaternionic Heisenberg group and obtain a closed form expression which closely resembles that of the heat kernel for Heisenberg groups. As we know, apart from the standard Heisenberg group, the quaternionic Heisenberg group is the only nilpotent Lie group on which an explicit formula for the heat kernel has been obtained up to now.

The second aim of this paper is to use the explicit formula for the heat kernel to study the uniform boundedness of the Riesz transforms on the quaternionic Heisenberg group. That is, the Riesz transforms are bounded on L^p spaces with norms independent of the dimension of the group. On the standard Heisenberg group this problem was addressed by Colhon [**3**] et al.. We apply their method to the quaternionic groups and by overcoming considerable difficulties in the process of calculation, finally prove the uniform boundedness of Riesz transforms on the quaternionic Heisenberg group.

We hope that we can use this explicit expression of the heat kernel to solve other problems in the harmonic analysis on the quaternionic Heisenberg group.

1. Prelimilaries.

We identify the division ring $\mathbf{H}$ of quaternions with $\mathbf{R}\times\mathbf{R}^3$. For $\mathbf{p} = (x_0, \mathbf{x}), \mathbf{q} = (y_0, \mathbf{y}) \in \mathbf{H}$, the quaternionic multiplication is defined as:

$$\mathbf{pq} = (x_0, \mathbf{x})(y_0, \mathbf{y}) = (x_0y_0 - \mathbf{x}.\mathbf{y}, x_0\mathbf{y} + y_0\mathbf{x} + \mathbf{x}\times\mathbf{y}),$$

where $\mathbf{x}.\mathbf{y}$ and $\mathbf{x}\times\mathbf{y}$ are the inner and exterior product of $\mathbf{x}$ and $\mathbf{y}$ respectively. For $\mathbf{p} = (x_0, \mathbf{x}) \in \mathbf{H}$, we use the notations $x_0 = \operatorname{Re}\mathbf{p}$, $\mathbf{x} = \operatorname{Im}\mathbf{p}$. The conjugate of $\mathbf{p}$ is denoted as $\overline{\mathbf{p}} = (x_0, -\mathbf{x})$ and $|\mathbf{p}| = (\mathbf{p}.\overline{\mathbf{p}})^{1/2}$ is the norm of $\mathbf{p}$.

The product space $\mathbf{H}^n\times\mathbf{R}^3$ together with the multiplication

$$\begin{aligned}&(\mathbf{p_1},\dots,\mathbf{p_n},\mathbf{u}).(\mathbf{q_1},\dots,\mathbf{q_n},\mathbf{v})\\ &= \left(\mathbf{p_1}+\mathbf{q_1},\dots,\mathbf{p_n}+\mathbf{q_n},\mathbf{u}+\mathbf{v}+2\sum_{r=1}^{n}\operatorname{Im}(\mathbf{q}_r.\overline{\mathbf{p}}_r)\right)\end{aligned}$$

constitutes a Lie group, called the quaternionic Heisenberg group, and denoted by $\mathbf{H}H^n$ (Allcock's notation [**1**]).

We know [**2**] that if

$$\mathbf{p}_r = (x_{r0}, \mathbf{x_r}) = (x_{r0}, (x_{r1}, x_{r2}, x_{r3})) \in \mathbf{H}^n,$$
$$\text{for } 1\le r\le n \text{ and } \mathbf{u} = (u_1, u_2, u_3) \in \mathbf{R}^3,$$

then the vector fields

$$\begin{aligned}
X_{r0} &= \frac{\partial}{\partial x_{r0}} - 2x_{r1}\frac{\partial}{\partial u_1} - 2x_{r2}\frac{\partial}{\partial u_2} - 2x_{r3}\frac{\partial}{\partial u_3},\\
X_{r1} &= \frac{\partial}{\partial x_{r1}} + 2x_{r0}\frac{\partial}{\partial u_1} + 2x_{r3}\frac{\partial}{\partial u_2} - 2x_{r2}\frac{\partial}{\partial u_3},\\
X_{r2} &= \frac{\partial}{\partial x_{r2}} - 2x_{r3}\frac{\partial}{\partial u_1} + 2x_{r0}\frac{\partial}{\partial u_2} + 2x_{r1}\frac{\partial}{\partial u_3},\\
X_{r3} &= \frac{\partial}{\partial x_{r3}} + 2x_{r2}\frac{\partial}{\partial u_1} - 2x_{r1}\frac{\partial}{\partial u_2} + 2x_{r0}\frac{\partial}{\partial u_3}
\end{aligned}$$

form a basis of the Lie algebra of $\mathbf{H}H^n$. The commutators of these vector fields satisfy

$$\begin{aligned}
[X_{r0}, X_{s1}] &= 4\delta_{rs}\frac{\partial}{\partial u_1} = [X_{r2}, X_{s3}],\\
[X_{r0}, X_{s2}] &= 4\delta_{rs}\frac{\partial}{\partial u_2} = [X_{r3}, X_{s1}],\\
[X_{r0}, X_{s3}] &= 4\delta_{rs}\frac{\partial}{\partial u_3} = [X_{r1}, X_{s2}]
\end{aligned}$$

with all other brackets equal to zero. So the quaternionic Heisenberg group is a nilpotent Lie group of step two.

Following the case of the Heisenberg groups [**5**], we introduce on $\mathbf{H}H^n$ the group $\{\delta_t : 0 < t < \infty\}$ of dilations defined by

$$\delta_t(\mathbf{p}, \mathbf{u}) = (t\mathbf{p}, t^2\mathbf{u}) = (t\mathbf{p}_1, \ldots, t\mathbf{p}_n, t^2\mathbf{u}).$$

These dilations satisfy the distributive law

$$\delta_t((\mathbf{p}, \mathbf{u}).(\mathbf{q}, \mathbf{v})) = \delta_t(\mathbf{p}, \mathbf{u}).\delta_t(\mathbf{q}, \mathbf{v}).$$

We also define the norm function on $\mathbf{H}H^n$ by

$$|(\mathbf{p}, \mathbf{u})| = (|\mathbf{p}|^4 + |\mathbf{u}|^2)^{1/4} = \left(\left(\sum_{r=1}^{n}\sum_{i=0}^{3}|x_{ri}|^2\right)^2 + \sum_{j=1}^{3}u_j^2\right)^{1/4},$$

which satisfies

$$|\delta_t(\mathbf{p}, \mathbf{u})| = t|(\mathbf{p}, \mathbf{u})|.$$

Let $e = (0, 0)$ be the identity element of the group $\mathbf{H}H^n$.

We know [**5**] that Kohn's sublaplace operator on the quaternionic Heisenberg group is defined as

$$\triangle = \sum_{r=1}^{n}\sum_{i=0}^{3} X_{ri}^2.$$

A simple calculation shows that

$$\triangle = \sum_{r=1}^{n}\sum_{i=0}^{3}\frac{\partial^2}{\partial x_{ri}^2} + 4\sum_{r=1}^{n}\sum_{i=1}^{3}\left(x_{r0}\frac{\partial^2}{\partial x_{ri}\partial u_i} - x_{ri}\frac{\partial^2}{\partial x_{r0}\partial u_i}\right)$$
$$+ 4\sum_{r=1}^{n}\sum_{i=0}^{3}x_{ri}^2\sum_{j=1}^{3}\frac{\partial^2}{\partial u_j^2} + 4\sum_{r=1}^{n}\sum_{(i,j,k)}x_{ri}\left(\frac{\partial^2}{\partial x_{rj}\partial u_k} - \frac{\partial^2}{\partial x_{rk}\partial u_j}\right),$$

where (i,j,k) means the cyclic permutation of $(1,2,3)$.

2. The heat kernels of the quaternionic Heisenberg groups.

In this section we shall use the method developed by Gaveau [**7**] and Hulanicki [**9**] to derive an explicit expression of the heat kernel of the quaternionic Heisenberg group. Firstly we have:

Lemma 2.1. *The diffusion of the infinitesimal generator* $\frac{1}{2}\triangle$ *starting at* e *is the process*

$$g(s) = (x_{ri}(s), u_j(s))_{1\le r\le n,\ 0\le i\le 3,\ 1\le j\le 3},$$

where $(x_{ri}(s))_{1\le r\le n,\ 0\le i\le 3}$ *are* $4n$ *standard Brownian motions, and*

$$u_j(s) = 2\sum_{r=1}^{n}\int_0^s x_{r0}dx_{rj}(t) - x_{rj}dx_{r0}(t) + x_{rk}dx_{ri}(t) - x_{ri}dx_{rk}(t).$$

Proof. As in [**7**], the projection on $\mathbf{H}^n = \mathbf{R}^{4n}$ of the diffusion to be found is the diffusion of the infinitesimal generator $\frac{1}{2}\sum_{r=1}^{n}\sum_{i=0}^{3}\frac{\partial^2}{\partial x_{ri}{}^2}$, which is given by $4n$ standard Brownian motions $(\mathbf{x}_1(s),\ldots,\mathbf{x}_n(s))$. Hence it is sufficient to compute the stochastic differentials $du_j (1\le j\le 3)$.

We observe that the matrix of principal symbols of $\triangle$ is given by

$$G = \begin{pmatrix} I_4 & \ldots & 0 & A_1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \ldots & I_4 & A_n \\ {}^tA_1 & \ldots & {}^tA_n & C \end{pmatrix}$$

where, for $1\le r\le n$,

$$A_r = \begin{pmatrix} -2x_{r1} & -2x_{r2} & -2x_{r3} \\ 2x_{r0} & 2x_{r3} & -2x_{r2} \\ -2x_{r3} & 2x_{r0} & 2x_{r1} \\ 2x_{r2} & -2x_{r1} & 2x_{r0} \end{pmatrix}$$

and

$$C = \left(4\sum_{r=1}^{n}|\mathbf{x_r}|^2\right) I_3.$$

In these expressions, I_k denotes the identity matrix of order k. Let

$$\Sigma = \begin{pmatrix} I_4 & \dots & 0 & A_1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & I_4 & A_n \end{pmatrix}.$$

It is obvious that

$${}^t\Sigma\Sigma = G,$$

and we know that [**7**] the matrix Σ gives stochastic differentials of the diffusion $\frac{1}{2}\triangle$:

$$(d\mathbf{x}_1, \dots, d\mathbf{x}_n, du_1, du_2, du_3) = (d\mathbf{x}_1, \dots, d\mathbf{x}_n) \begin{pmatrix} I_4 & \dots & 0 & A_1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & I_4 & A_n \end{pmatrix}.$$

It is easy to obtain

$$du_i = 2\sum_{r=1}^{n}(x_{r0}dx_{ri} - x_{ri}dx_{r0} + x_{rj}dx_{rk} - x_{rk}dx_{rj}).$$

Let

$$\frac{1}{2}\triangle = \frac{\partial}{\partial s}$$

be the equation of propagation of heat, where $s(\geq 0)$ denotes the time. Assume that $p_s(e, \mathbf{g})$, for $\mathbf{g} \in \mathbf{H}H^n$, be the heat kernel with pole at e. From the definition we have

$$p_s(e, \mathbf{g})d\mathbf{g} = \operatorname{Prob}(\mathbf{g}_\omega(s) \in d\mathbf{g}),$$

where $d\mathbf{g}$ is the left-invariant measure on $\mathbf{H}H^n$. Let

$$\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_n) = (y_{10}, \dots, y_{13}, \dots, y_{n0}, \dots, y_{n3})$$

and $\mathbf{v} = (v_1, v_2, v_3)$ be the dual variables of $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ and $\mathbf{u} = (u_1, u_2, u_3)$ respectively. We write $|\mathbf{x}|^2 = \sum\limits_{r=1}^{n}\sum\limits_{i=0}^{3} x_{ri}^2$ and

$$a = 4\begin{pmatrix} 0 & v_1 & v_2 & v_3 \\ -v_1 & 0 & v_3 & -v_2 \\ -v_2 & -v_3 & 0 & v_1 \\ -v_3 & v_2 & -v_1 & 0 \end{pmatrix},$$

$$A = \operatorname{diag}(\underbrace{a, \dots, a}_{n}).$$

If X is a skew-symmetric matrix of order $4n$ and $\mathbf{w} \in \mathbf{R}^{4n}$, we write

$$\psi_s(X, \mathbf{w}) = \exp\left[\frac{1}{s}\left(-|\mathbf{w}|^2 + {}^t\mathbf{w}\left(I_{4n} - \frac{s^2X^2}{4\pi^2}\right)^{-1}\mathbf{w}\right)\right]\det\left(I_{4n} - \frac{sX}{2\pi}\right)^{-1}$$

where ${}^t\mathbf{w}$ is the transpose of the vector $\mathbf{w}$.

With these preparations, we have:

Theorem 2.2. *The Euclidean Fourier transform of* $p_s(e, \mathbf{g})$ *is given by the formula*

$$\begin{aligned}&\hat{p}_s(e,.)(\mathbf{y},\mathbf{v})\\&= (2\pi s)^{-2n} \int\limits_{\mathbf{R}^{4n}} \exp\left(\sqrt{-1}\sum_{r=1}^{n}\sum_{i=0}^{3} y_{ri}x_{ri} - \frac{|\mathbf{x}|^2}{2s}\right) \prod_{m=1}^{\infty} \psi_s\left(\frac{A}{m}, \mathbf{x}\right) d\mathbf{x}.\end{aligned}$$

Proof. By the definition, we have

$$\begin{aligned}&\hat{p}_s(e,.)(\mathbf{y},\mathbf{v})\\
&= \int_{\mathbf{R}^{4n+3}} \exp\left(\sqrt{-1}\left(\sum_{r=1}^{n}\sum_{i=0}^{3} y_{ri}x_{ri} + \sum_{j=1}^{3} v_j u_j\right)\right) p_s(e,(\mathbf{x},\mathbf{u}))d\mathbf{x}d\mathbf{u}\\
&= E_0\left(\exp\sqrt{-1}\left(\sum_{r=0}^{n}\sum_{i=0}^{3} y_{ri}x_{ri}(s) + \sum_{j=1}^{3} v_j u_j(s)\right)\right)\\
&= E_0\left(\exp\sqrt{-1}\sum_{r=1}^{n}\sum_{i=0}^{3} y_{ri}x_{ri}(s)\right)\\
&\qquad \cdot E_0\left(\exp\sqrt{-1}\sum_{j=1}^{3} v_j u_j(s)|, x_{ri}(s) = x_{ri}\right),\end{aligned}$$

where $E_0(\ldots|x_{ri}(s) = x_{ri})$ denotes the conditional expectation given the $x_{ri}(s)$. So it is sufficient to evaluate

$$E_0\left(\exp\left(\sqrt{-1}\sum_{j=1}^{3} v_j u_j(s)\right)|x_{ri}(s) = x_{ri}\right).$$

We express the $4n$ standard real Brownian motions as the Fourier series with independent Gaussian variables as their coefficients, i.e., when $s \neq 2\pi$,

$$x_{ri}(s) = \frac{s\xi^{(ri)}}{\sqrt{2\pi}} + \sum_{m=1}^{\infty} \frac{1}{m\sqrt{\pi}}\left(\xi_m^{(ri)}(\cos ms - 1) - \xi_m^{\prime(ri)}\sin ms\right),$$

where $\xi_m^{(ri)}$ and $\xi_m^{\prime(ri)}$ are one-dimensional standard normal distributions which are independent of each other, and $x_{ri}(2\pi) = U_{ri}$.

From Lemma 2.1 we get immediately

$$u_i(2\pi) = 4\sum_{r=1}^{n} \frac{1}{m} \left[\xi_m^{(r0)} \left(\xi_m^{\prime(ri)} - \frac{U_{ri}}{\sqrt{\pi}}\right) - \xi_m^{(ri)} \left(\xi_m^{\prime(r0)} - \frac{U_{r0}}{\sqrt{\pi}}\right)\right.$$
$$\left. + \xi_m^{(rj)} \left(\xi_m^{\prime(rk)} - \frac{U_{rk}}{\sqrt{\pi}}\right) - \xi_m^{(rk)} \left(\xi_m^{\prime(rj)} - \frac{U_{rj}}{\sqrt{\pi}}\right)\right].$$

Thus

$$E_0 \left(\exp\left(\sqrt{-1}\sum_{j=1}^{3} v_j u_j(2\pi)\right) |x_{ri}(2\pi) = U_{ri}\right)$$
$$= \prod_{m=1}^{\infty} E_0\Bigg(\exp \frac{4\sqrt{-1}}{m}\Bigg\{ \sum_{i=1}^{3} v_i \sum_{r=1}^{n} \Bigg[\xi_m^{(r0)} \left(\xi_m^{\prime(ri)} - \frac{U_{ri}}{\sqrt{\pi}}\right)$$
$$- \xi_m^{(ri)} \left(\xi_m^{\prime(r0)} - \frac{U_{r0}}{\sqrt{\pi}}\right) + \xi_m^{(rj)} \left(\xi_m^{\prime(rk)} - \frac{U_{rk}}{\sqrt{\pi}}\right)$$
$$- \xi_m^{(rk)} \left(\xi_m^{\prime(rj)} - \frac{U_{rj}}{\sqrt{\pi}}\right)\Bigg]\Bigg\}\Bigg|x_{ri}(2\pi) = U_{ri}\Bigg).$$

In the last expression, all conditional information is exhausted, so the conditional expectation is actually reduced to the expectation. Hence it is sufficient to consider the terms

$$J_m = E\Bigg(\exp \frac{4\sqrt{-1}}{m}\Bigg\{ \sum_{i=1}^{3} v_i \sum_{r=1}^{n} \Bigg[\xi_m^{(r0)} \left(\xi_m^{\prime(ri)} - \frac{U_{ri}}{\sqrt{\pi}}\right) - \xi_m^{(ri)} \left(\xi_m^{\prime(r0)} - \frac{U_{r0}}{\sqrt{\pi}}\right)$$
$$+ \xi_m^{(rj)} \left(\xi_m^{\prime(rk)} - \frac{U_{rk}}{\sqrt{\pi}}\right) - \xi_m^{(rk)} \left(\xi_m^{\prime(rj)} - \frac{U_{rj}}{\sqrt{\pi}}\right)\Bigg]\Bigg\}\Bigg).$$

Setting $\xi_m^{"(ri)} = \xi_m^{\prime(ri)} - \frac{U_{ri}}{\sqrt{\pi}}$, we first integrate with respect to $\xi_m^{(ri)}$, which yields

$$J_m = E\left(\exp\left(-\frac{1}{2m^2}\sum_{r=1}^{n}\sum_{i=0}^{3} (\mu_m^{(ri)})^2\right)\right),$$

where

$$\mu_m^{(r)} = \begin{pmatrix} \mu_m^{(r0)} \\ \mu_m^{(r1)} \\ \mu_m^{(r2)} \\ \mu_m^{(r3)} \end{pmatrix} = 4 \begin{pmatrix} 0 & v_1 & v_2 & v_3 \\ -v_1 & 0 & v_3 & -v_2 \\ -v_2 & -v_3 & 0 & v_1 \\ -v_3 & v_2 & -v_1 & 0 \end{pmatrix} \begin{pmatrix} \xi_m^{"(r0)} \\ \xi_m^{"(r1)} \\ \xi_m^{"(r2)} \\ \xi_m^{"(r3)} \end{pmatrix} = a\xi_m^{"(r)},$$

written briefly as

$$\mu_m = A\xi_m^{"}.$$

The skew-symmetric matrix A is more simple than that in [**7**], which makes the calculation from now on easier than that in [**7**], and causes the heat kernel of the quaternionic Heisenberg groups to be more simple and more concrete than those of the general nilpotent Lie groups of step two.

Thus

$$\sum_{r=1}^{n}\sum_{i=0}^{3}(\mu_m^{(ri)})^2 = (A\mu_m'', A\mu_m'') = -(A^2\mu_m'', \mu_m'').$$

Let $B = \frac{1}{m^2}A^2$, as in [7]. Then we have

$$\begin{aligned} J_m &= \frac{1}{(2\pi)^{2n}}\int_{\mathbf{R}^{4n}} \exp\frac{1}{2}\left[(B\xi_m'', \xi_m'') - \left(\xi_m'' - \frac{U_m}{\sqrt{\pi}}, \xi_m'' - \frac{U_m}{\sqrt{\pi}}\right)\right] d\xi_m'' \\ &= \exp\left(-\frac{|U_m|^2}{2\pi} + \frac{1}{2\pi}{}^tU_m(I_{4n} - B)^{-1}U_m\right) \Big/ \sqrt{\det(I_{4n} - B)}. \end{aligned}$$

This yields

$$E_0\left(\exp\left(\sqrt{-1}\sum_{j=1}^{3} v_j \sum_{r=1}^{n}\sum_{i=0}^{3} x_{ri}(s)\right) |x_{ri}(s) = x_{ri}\right) = \prod_{m=1}^{\infty}\psi_s\left(\frac{1}{m}A, \mathbf{x}\right).$$

Thus the proof of Theorem 2.2 is finished.

We may explicitly give the diagonalization by 2×2 block matrices of the skew-symmetric matrix A as follows: Let $\rho = \sqrt{v_1^2 + v_2^2 + v_3^2}$, $\sigma = \sqrt{v_2^2 + v_3^2}$, and set

$$Y_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, Y_1 = \frac{1}{\rho}\begin{pmatrix} 0 \\ v_1 \\ v_2 \\ v_3 \end{pmatrix}, Y_2 = \frac{1}{\rho\sigma}\begin{pmatrix} 0 \\ -\sigma^2 \\ v_1v_2 \\ v_1v_3 \end{pmatrix}, Y_3 = \frac{1}{\sigma}\begin{pmatrix} 0 \\ 0 \\ -v_3 \\ v_2 \end{pmatrix}.$$

It is easy to see that these vectors are orthonormal and

$$aY_0 = -4\rho Y_1, aY_1 = 4\rho Y_0, aY_2 = -4\rho Y_3, aY_3 = 4\rho Y_2.$$

Now we introduce an orthogonal matrix $\omega = (Y_0, Y_1, Y_2, Y_3)$. It is readily seen that

$${}^t\omega a\omega = \operatorname{diag}\left(\begin{pmatrix} 0 & 4\rho \\ -4\rho & 0 \end{pmatrix}, \begin{pmatrix} 0 & 4\rho \\ -4\rho & 0 \end{pmatrix}\right) = p.$$

Furthermore, if we set

$$\Omega = \operatorname{diag}(\underbrace{\omega, \dots, \omega}_{n}),$$

then

$${}^t\Omega A\Omega = \operatorname{diag}\left(\underbrace{\begin{pmatrix} 0 & 4\rho \\ -4\rho & 0 \end{pmatrix}, \dots, \begin{pmatrix} 0 & 4\rho \\ -4\rho & 0 \end{pmatrix}}_{2n}\right) = P.$$

Now we can give the explicit expression of the heat kernel:

Theorem 2.3. *The heat kernel on the quaternionic Heisenberg group* $\mathbf{H}H^n$ *is given by*

$$\begin{aligned} p_s(\mathbf{x},\mathbf{u}) &= p_s(e,(\mathbf{x},\mathbf{u})) \\ &= (2\pi s)^{2n}(2\pi)^{-(4n+3)} \int_{\mathbf{R}^3} \exp(-(\sqrt{-1}\mathbf{u}.\mathbf{v} + |\mathbf{x}|^2 \rho \coth 2s\rho)) \\ &\quad \cdot \left(\frac{2s\rho}{\sinh 2s\rho}\right)^{2n} d\mathbf{v}. \end{aligned}$$

Proof. After the orthogonal transformation represented by matrix Ω, Theorem 2.2 can be rewritten as

$$\begin{aligned} (1)\qquad & \hat{p}_s(e,.)(\mathbf{y},\mathbf{v}) \\ &= (2\pi s)^{-2n} \int_{\mathbf{R}^{4n}} \exp\left(\sqrt{-1}\sum_{r=1}^{n}\sum_{i=0}^{3} y_{ri}x_{ri} - \frac{|\mathbf{x}|^2}{2s}\right) \\ &\quad \cdot \prod_{m=1}^{\infty} \exp\left[\frac{1}{s}\left(-|\mathbf{x}|^2 + {}^t\mathbf{x}\Omega\left(I_{4n} - \frac{s^2P^2}{4\pi^2m^2}\right)^{-1} {}^t\Omega\mathbf{x}\right)\right] \\ &\quad \cdot \det\left(I_{4n} - \frac{sP}{2\pi m}\right)^{-1} d\mathbf{x}. \end{aligned}$$

It is easy to verify that

$$\det\left(I_{4n} - \frac{sP}{2\pi m}\right) = \left(1 + \frac{4s^2\rho^2}{\pi^2 m^2}\right)^{2n},$$

$${}^t\Omega\mathbf{x} = ({}^t\omega\mathbf{x}_1, \ldots, {}^t\omega\mathbf{x}_n),$$

$$({}^t\omega\mathbf{x}_r)_0^2 + ({}^t\omega\mathbf{x}_r)_1^2 = \frac{1}{\rho^2}\left[\rho^2 x_{r0}^2 + \left(\sum_{i=1}^{3} v_i x_{ri}\right)^2\right],$$

$$\begin{aligned} & ({}^t\omega\mathbf{x}_r)_2^2 + ({}^t\omega\mathbf{x}_r)_3^2 \\ &= \frac{1}{\rho^2}\left[\left(-\sigma x_{r1} + \frac{v_1v_2}{\sigma}x_{r2} + \frac{v_1v_3}{\sigma}x_{r3}\right)^2 + \left(\frac{-\rho v_3}{\sigma}x_{r2} + \frac{\rho v_2}{\sigma}x_{r3}\right)^2\right] \\ &= \frac{1}{\rho^2}\left[\rho^2\sum_{i=1}^{3} x_{ri}^2 - \left(\sum_{i=1}^{3} v_i x_{ri}\right)^2\right], \end{aligned}$$

and hence

$$\sum_{i=0}^{3}({}^t\omega\mathbf{x}_r)_i^2 = \sum_{i=0}^{3} x_{ri}^2,\ \ |{}^t\Omega\mathbf{x}|^2 = |\mathbf{x}|^2.$$

From the expression of matrix P we have

$$\begin{aligned}
&\exp\left[\frac{1}{s}\left(-|\mathbf{x}|^2+{}^t\mathbf{x}\Omega\left(I_{4n}-\frac{s^2P^2}{4\pi^2m^2}\right)^{-1}{}^t\Omega\mathbf{x}\right)\right]\\
&=\exp\left\{\frac{1}{s}\sum_{r=1}^{n}\left(-|\mathbf{x}_r|^2+{}^t\mathbf{x}_r\omega\left(I_4-\frac{s^2p^2}{4\pi^2m^2}\right)^{-1}{}^t\omega\mathbf{x}_r\right)\right\}\\
&=\exp\left\{\frac{1}{s}\sum_{r=1}^{n}\sum_{i=0}^{3}\left(-1+\frac{\pi^2m^2}{\pi^2m^2+4s^2\rho^2}\right)x_{ri}^2\right\},
\end{aligned}$$

since

$$\frac{u}{\sinh u}=\prod_{m=1}^{\infty}\left(1+\frac{u^2}{\pi^2m^2}\right)^{-1},$$

and

$$\coth u=\frac{1}{u}+\sum_{m=1}^{\infty}\frac{2u}{\pi^2m^2+u^2},$$

therefore

$$\begin{aligned}
&\prod_{m=1}^{\infty}\exp\left[\frac{1}{s}\left(-|\mathbf{x}|^2+{}^t\mathbf{x}\Omega\left(I-\frac{s^2P^2}{4\pi^2m^2}\right)^{-1}{}^t\Omega\mathbf{x}\right)\right]\det\left(I-\frac{sP}{2\pi m}\right)^{-1}\\
&=\exp\left(\frac{1}{s}\sum_{r=1}^{n}\sum_{i=0}^{3}\sum_{m=1}^{\infty}\frac{-4s^2\rho^2}{\pi^2m^2+4s^2\rho^2}x_{ri}^2\right)\prod_{m=1}^{\infty}\left(1+\frac{4s^2\rho^2}{\pi^2m^2}\right)^{-2n}\\
&=\exp\left(\frac{|\mathbf{x}|^2}{2s}(1-2s\rho\cosh 2s\rho)\right)\left(\frac{2s\rho}{\sinh 2s\rho}\right)^{2n}.
\end{aligned}$$

Substituting this equality into the right-hand side of (1) and taking the Euclidean Fourier transform, we obtain at once the desired result.

3. The Green functions of the quaternionic Heisenberg groups.

It is known that the Green function can be derived from the heat kernel by the formula

$$G(\mathbf{g})=\int_0^{+\infty}p_s(e,\mathbf{g})ds.$$

So using Theorem 2.3, the Green function of the quaternionic Heisenberg group can be written as

$$
\begin{aligned}
G(\mathbf{x},\mathbf{u}) &= \int_0^{+\infty} p_s(e,(\mathbf{x},\mathbf{u}))ds \\
&= (2\pi)^{-(6n+3)} \int_0^{+\infty} s^{-2n} ds \int_{\mathbf{R}^3} \exp\left(-\left(\sqrt{-1}\mathbf{v}.\mathbf{u} + |\mathbf{x}|^2 \rho \coth 2s\rho\right)\right) \\
&\quad \cdot \left(\frac{2s\rho}{\sinh 2s\rho}\right)^{2n} \prod_{j=1}^{3} dv_j.
\end{aligned}
$$

First by performing the change of variables $v_j \to s v_j$, then it follows that $\rho \to s\rho$, and

$$
\begin{aligned}
&G(\mathbf{x},\mathbf{u}) \\
&= (2\pi)^{-(6n+3)} \int_0^{+\infty} s^{-(2n+3)} ds \int_{\mathbf{R}^3} \exp\left\{\frac{1}{s}(\sqrt{-1}\mathbf{v}.\mathbf{u} + |\mathbf{x}|^2 \rho \coth 2\rho)\right\} \\
&\quad \cdot \left(\frac{2\rho}{\sinh 2\rho}\right)^{2n} \prod_{j=1}^{3} dv_j \\
&= (2\pi)^{-(6n+3)} \Gamma(2n+2) \int_{\mathbf{R}^3} (|\mathbf{x}|^2 \coth 2\rho + \sqrt{-1}\mathbf{v}.\mathbf{u})^{-(2n+2)} \\
&\quad \cdot \left(\frac{2\rho}{\sinh 2\rho}\right)^{2n} \prod_{j=1}^{3} dv_j,
\end{aligned}
$$

where $\Gamma(.)$ is Euler's Gamma-function.

In Euclidean space $\mathbf{R}^3$, we use polar coordinates and let the positive direction of the z-axis coincide with that of vector $\mathbf{u}$, i.e., we set

$$
\mathbf{u} = |\mathbf{u}|(0,\ 0,\ 1),\ \mathbf{v} = \rho(\sin\theta\cos\varphi,\ \sin\theta\sin\varphi,\ \cos\theta).
$$

Thus

(2)

$$
\begin{aligned}
&G(\mathbf{x},\mathbf{u})\\
&= 2^n(2\pi)^{-(6n+3)}\Gamma(2n+2)\int\limits_0^{+\infty}\rho^2\left(\frac{\rho}{\sinh 2\rho}\right)^{2n}d\rho\\
&\quad\cdot\int_0^{2\pi}\int_0^{\pi}(|\mathbf{x}|^2\rho\coth 2\rho+\sqrt{-1}\rho|\mathbf{u}|\cos\theta)^{-(2n+2)}\sin\theta d\theta d\phi\\
&= 2^{2n}(2\pi)^{-(6n+2)}\frac{\Gamma(2n+2)}{2n+1}\int_0^{+\infty}\left(\frac{1}{\sinh 2\rho}\right)^{2n}\frac{1}{\sqrt{-1}|\mathbf{u}|}\\
&\quad\cdot\left[(|\mathbf{x}|^2\coth 2\rho-\sqrt{-1}|\mathbf{u}|)^{-(2n+1)}-(|\mathbf{x}|^2\coth 2\rho+\sqrt{-1}|\mathbf{u}|)^{-(2n+1)}\right]d\rho.
\end{aligned}
$$

For $n=1$ it is easy to complete the last integration, and we obtain:

Proposition 3.1. *The Green function $G(\mathbf{x},\mathbf{u})$ of the quaternionic Heisenberg group $\mathbf{H}H^1$ is*

$$G(\mathbf{x},\mathbf{u}) = 4(2\pi)^{-8}(|\mathbf{x}|^4+|\mathbf{u}|^2)^{-2}.$$

Proof. When $n=1$, performing the integral in the right-hand side of Equation (2) gives us

$$
\begin{aligned}
G(\mathbf{x},\mathbf{u}) &= (2\pi)^{-8}\frac{1}{\sqrt{-1}|\mathbf{u}||\mathbf{x}|^2}\\
&\qquad\cdot\left[(|\mathbf{x}|^2\coth 2\rho-\sqrt{-1}|\mathbf{u}|)^{-2}-(|\mathbf{x}|^2\coth 2\rho+\sqrt{-1}|\mathbf{u}|)^{-2}\right]_{\rho=0}^{\rho=+\infty}\\
&= 4(2\pi)^{-8}(|\mathbf{x}|^4+|\mathbf{u}|^2)^{-2}.
\end{aligned}
$$

For general $n\in\mathbf{N}$, there is some difficulty to evaluate the integration in Equation (2), while the above proposition and the corresponding results for the Heisenberg groups [**5**] motivate us to pose:

Theorem 3.2. *The Green functions $G(\mathbf{x},\mathbf{u})$ of the quaternionic Heisenberg groups $\mathbf{H}H^n$ are*

$$G(\mathbf{x},\mathbf{u}) = c_n(|\mathbf{x}|^4+|\mathbf{u}|^2)^{-(n+1)},$$

where

$$c_n^{-1} = 4(n+1)(n+2)\int_{\mathbf{H}H^n}|\mathbf{x}|^2(|\mathbf{x}|^4+|\mathbf{u}|^2+1)^{-(n+3)}d(\mathbf{x},\mathbf{u}).$$

The method of proof is completely analogous to that for the Heisenberg groups [**5**].

4. Riesz transforms on the quaternionic Heisenberg groups.

In this section we shall study the uniformly boundedness of the Riesz transforms with respect to the dimensions of the quaternionic Heisenberg groups.

Stein [**11**] first studied the Riesz transforms on Euclidean spaces. Afterward various authors investigated the Riesz transform on Riemannian manifolds. Although the boundedness of the Riesz transform on every nilpotent Lie group is well-known [**10**]. It was Coulhon [**3**] et al. who first showed the uniform boundedness of Riesz transforms with respect to the dimensions of the Heisenberg groups.

In our investigation there are many properties analogous to that in [**3**], and for completeness we quote briefly these points. One part which differs, however, is the so called "main estimate" in [**3**], so this part is presented in detail.

4.1. The vector of the Riesz transforms. For the quaternionic Heisenberg groups, the skew-adjoint Riesz transforms are defined analogously to that in [**3**] by

$$\widetilde{R}_{ri} = X_{ri}\triangle^{-1/2} + \triangle^{-1/2}X_{ri},\ 1\le r\le n,\ 0\le i\le 3.$$

In [**3**] Coulhon et al. proved the following results:

Lemma 4.1. *The skew-adjoint Riesz transform has the expression*

$$\widetilde{R}_{ri}f(\mathbf{x},\mathbf{u}) = \int_{\mathbf{H}H^n} X_{ri}p_1(e,(\mathbf{y},\mathbf{v}))\mathcal{H}_{(\mathbf{y},\mathbf{v})}f(\mathbf{x},\mathbf{u})d(\mathbf{y},\mathbf{v}),$$

where

$$\mathcal{H}_{(\mathbf{y},\mathbf{v})}f(\mathbf{x},\mathbf{u}) = \int_0^{+\infty}[f((\mathbf{x},\mathbf{u})\delta_t((\mathbf{y},\mathbf{v})^{-1})) - f((\mathbf{x},\mathbf{u})\delta_t(\mathbf{y},\mathbf{v}))]\frac{dt}{t}.$$

In fact, this formula is valid for every stratified Lie group.

Lemma 4.2. *For every $p\in(1,+\infty)$, there exists $c>0$ depending only on p, such that*

$$\|\mathcal{H}_{(\mathbf{y},\mathbf{v})}\|_{p\to p\le}c,\ \forall n\in\mathbf{N},\ \forall(\mathbf{y},\mathbf{v})\in\mathbf{H}H^n.$$

Lemma 4.3. *There exists $c'>0$ such that*

$$\|X_{ri}p_1\|_{L^1(\mathbf{H}H^n)}\le c',\ \forall n\in\mathbf{N},\ \forall 1\le r\le n,\ 0\le i\le 3.$$

This follows from the fact that one can express the heat kernels for the quaternionic Heisenberg groups $\mathbf{H}H^n$ as p_s^n, then it is easy to verify that

$$p_1^n(\mathbf{x}_1,\dots,\mathbf{x}_n,\mathbf{u}) = [p_1^1(\mathbf{x}_1,.)*\cdots*p_1^1(\mathbf{x}_n,.)](\mathbf{u}).$$

For $f\in C_0^\infty(\mathbf{H}H^n)$ and $(\mathbf{x},\mathbf{u})\in\mathbf{H}H^n$, we define a vector field

$$\widetilde{R}f(\mathbf{x},\mathbf{u}) = (\widetilde{R}_{10}f(\mathbf{x},\mathbf{u}),\dots,\widetilde{R}_{n3}f(\mathbf{x},\mathbf{u}))$$

and its Euclidean length

$$|\widetilde{R}f(\mathbf{x},\mathbf{u})| = \left(\sum_{r=1}^{n}\sum_{i=0}^{3}|\widetilde{R}_{ri}f(\mathbf{x},\mathbf{u})|^2\right)^{1/2}.$$

Let S^n be the unit sphere in the quaternionic Heisenberg group $\mathbf{H}H^n$, i.e.,

$$S^n = \left\{\mathbf{h} = (\mathbf{z},\mathbf{w}) \in \mathbf{H}H^n \;\middle|\; \mathbf{z} = (\mathbf{z}_1,\dots,\mathbf{z}_n) \in \mathbf{H}^n,\ \mathbf{w} \in \mathbf{R}^3, \left(\sum_{r=1}^{n}\sum_{i=0}^{3} z_{ri}^2\right)^2 + \sum_{j=1}^{3} w_j^2 = 1\right\}.$$

As in [**3**], we can deduce that

$$|\widetilde{R}f(\mathbf{x},\mathbf{u})| = \int_{S^n}\int_0^{+\infty}\sum_{r=1}^{n}\sum_{i=0}^{3}\lambda_{ri}X_{ri}p_1(\delta_t(\mathbf{h}))\mathcal{H}_{(\mathbf{h})}f(\mathbf{x},\mathbf{u})t^{4n+5}dtd\sigma(\mathbf{h}),$$

where $d\sigma(\mathbf{h})$ is the surface element on S^n, the λ_{ri} are dependent on f and $(\mathbf{x},\mathbf{u})$, and $\sum_{r=1}^{n}\sum_{i=0}^{3}\lambda_{ri}^2 = 1$(we have also used the fact that $\mathcal{H}_{\delta_t(\mathbf{y},\mathbf{v})} = \mathcal{H}_{(\mathbf{y},\mathbf{v})}$). Applying Hölder's inequality with respect to $d\sigma$, we get, for $1/p + 1/q = 1$,

(3)

$$|\widetilde{R}f(\mathbf{x},\mathbf{u})| \leq \left\|\int_0^{+\infty}\sum_{r=1}^{n}\sum_{i=0}^{3}\lambda_{ri}X_{ri}p_1(\delta_t(\mathbf{h}))t^{4n+5}dt\right\|_{L^q(d\sigma(\mathbf{h}))}\|\mathcal{H}_{\mathbf{h}}f(\mathbf{x},\mathbf{u})\|_{L^p(d\sigma(\mathbf{h}))}.$$

After a rotation on $S^n\bigcap\mathbf{H}^n$, one can send $\sum_{r=1}^{n}\sum_{i=0}^{3}\lambda_{ri}X_{ri}$ to X_{10}, and it is clear that the heat kernel $p_1(\delta_t(\mathbf{h}))$ is invariant under this rotation, hence

(4)
$$\left\|\int_0^{+\infty}\sum_{r=1}^{n}\sum_{i=0}^{3}\lambda_{ri}X_{ri}p_1(\delta_t(\mathbf{h}))t^{4n+5}dt\right\|_{L^q(d\sigma(\mathbf{h}))} = \left\|\int_0^{+\infty}X_{10}p_1(\delta_t(\mathbf{h}))t^{4n+5}dt\right\|_{L^q(d\sigma(\mathbf{h}))}.$$

Via Lemma 4.2, the argument analogous to that in [**3**] implies that there exists $c(p) > 0$ such that

(5)
$$\|\|\mathcal{H}_{\mathbf{h}}f(\mathbf{x},\mathbf{u})\|_{L^p(d\sigma(\mathbf{h}))}\|_{L^pd(\mathbf{x},\mathbf{u})} \leq c(p)[\sigma(S^n)]^{1/p}\|f\|_{L^p(d\mathbf{x},\mathbf{u})}.$$

4.2. The main estimate. Let

$$\Phi(\mathbf{h}) = \int_0^{+\infty} X_{10} p_1(\delta_t(\mathbf{h})) t^{4n+5} dt, \ \mathbf{h} \in S^n.$$

Theorem 2.3 gives us

$$p_1(\mathbf{x}, \mathbf{u}) = \frac{1}{(2\pi)^{6n+3}} \int_{\mathbf{R}^3} \exp\left\{-(\sqrt{-1}\mathbf{v}.\mathbf{u} + |\mathbf{x}|^2 \rho\coth 2\rho)\right\} \left(\frac{2\rho}{\sinh 2\rho}\right)^{2n} \prod_{j=1}^{3} dv_j.$$

Since $X_{10} = \frac{\partial}{\partial x_{10}} - \sum_{i=1}^{3} 2x_{1i} \frac{\partial}{\partial u_i}$, it follows that

$$\begin{aligned}
(6)\quad X_{10} p_1(\mathbf{x}, \mathbf{u}) &= \frac{1}{(2\pi)^{6n+3}} \int_{\mathbf{R}^3} \left(-2x_{10}\rho\coth 2\rho + 2\sqrt{-1} \sum_{j=1}^{3} v_j x_{1j}\right) \\
&\qquad \cdot \exp\{-(\sqrt{-1}\mathbf{v}.\mathbf{u} + |\mathbf{x}|^2 \rho\coth 2\rho)\} \left(\frac{2\rho}{\sinh 2\rho}\right)^{2n} \prod_{j=1}^{3} dv_j \\
&= \frac{2}{(2\pi)^{6n+3}} (-F_1(\mathbf{x}, \mathbf{u}) + \sqrt{-1} F_2(\mathbf{x}, \mathbf{u})),
\end{aligned}$$

where

$$\begin{aligned}
F_1(\mathbf{x}, \mathbf{u}) &= x_{10} \int_{\mathbf{R}^3} \exp\{-(\sqrt{-1}\mathbf{v}.\mathbf{u} + |\mathbf{x}|^2 \rho\coth 2\rho)\} \\
&\qquad \cdot \left(\frac{2\rho}{\sinh 2\rho}\right)^{2n} \rho\coth 2\rho \prod_{j=1}^{3} dv_j, \\
F_2(\mathbf{x}, \mathbf{u}) &= \int_{\mathbf{R}^3} \exp\{-(\sqrt{-1}\mathbf{v}.\mathbf{u} + |\mathbf{x}|^2 \rho\coth 2\rho)\} \\
&\qquad \cdot \left(\frac{2\rho}{\sinh 2\rho}\right)^{2n} \sum_{j=1}^{3} v_j x_{1j} \prod_{j=1}^{3} dv_j.
\end{aligned}$$

Thus

$$\begin{aligned}
(7)\qquad \Phi(\mathbf{h}) &= \frac{2}{(2\pi)^{6n+3}} \int_0^{+\infty} [-F_1(\delta_t(\mathbf{h})) + \sqrt{-1} F_2(\delta_t(\mathbf{h}))] t^{4n+5} dt \\
&= \frac{2}{(2\pi)^{6n+3}} (-\Phi_1 + \sqrt{-1}\Phi_2)(\mathbf{h}).
\end{aligned}$$

Let $\mathbf{h} = (\mathbf{z}, \mathbf{w}) = (z_{10}, \ldots, z_{n3}, w_1, w_2, w_3) \in S^n$. So we get

$$\begin{aligned}\Phi_1(\mathbf{h}) &= \int_0^{+\infty} F_1(\delta_t(\mathbf{h}))t^{4n+5}dt \\ &= z_{10}\int_{\mathbf{R}^3}\int_0^{+\infty} \exp\{-t^2(\sqrt{-1}\mathbf{v}.\mathbf{w} + |\mathbf{z}|^2\rho\text{coth}2\rho)\} \\ &\quad \cdot \left(\frac{2\rho}{\text{sinh}2\rho}\right)^{2n} \rho\text{coth}2\rho t^{4n+6}dt \prod_{j=1}^{3} dv_j \\ &= \frac{1}{2}\Gamma\left(\frac{n+7}{2}\right) z_{10}\int_{\mathbf{R}^3} (|\mathbf{z}|^2\rho\text{coth}2\rho + \sqrt{-1}\mathbf{v}.\mathbf{w})^{-\frac{4n+7}{2}} \\ &\quad \cdot \left(\frac{2\rho}{\text{sinh}2\rho}\right)^{2n} \rho\text{coth}2\rho \prod_{j=1}^{3} dv_j.\end{aligned}$$

Taking the polar coordinates in $\mathbf{R}^3$ as in the proof of Proposition 3.1, the above function becomes

$$\begin{aligned}\Phi_1(\mathbf{h}) &= \frac{1}{2}\Gamma\left(\frac{4n+7}{2}\right) z_{10}\int_0^{+\infty}\int_0^{\pi}\int_0^{2\pi} (|\mathbf{z}|^2\text{coth}2\rho + \sqrt{-1}|\mathbf{w}|\text{cos}\theta)^{-\frac{4n+7}{2}} \\ &\quad \cdot \left(\frac{2\rho}{\text{sinh}2\rho}\right)^{2n} \rho^{-\frac{4n+1}{2}}\text{coth}2\rho\text{sin}\theta d\theta d\phi d\rho \\ &= \frac{2^{2n}}{4n+5}2\pi\Gamma\left(\frac{4n+7}{2}\right) z_{10}\int_0^{+\infty} \rho\text{cosh}2\rho\left(\frac{\rho}{\text{sinh}2\rho}\right)^{-3/2}\left(\frac{1}{\sqrt{-1}|\mathbf{w}|}\right) \\ &\quad \cdot \Big[(|\mathbf{z}|^2\text{cosh}2\rho - \sqrt{-1}|\mathbf{w}|\text{sinh}2\rho)^{-\frac{4n+5}{2}} \\ &\qquad - (|\mathbf{z}|^2\text{cosh}2\rho + \sqrt{-1}|\mathbf{w}|\text{sinh}2\rho)^{-\frac{4n+5}{2}}\Big] d\rho.\end{aligned}$$

On S^n we perform the change of variables given by $|\mathbf{z}|^2 = \text{cos}\psi$, $|\mathbf{w}| = \text{sin}\psi$. Then

(8)

$$\begin{aligned}&(|\mathbf{z}|^2\text{cosh}2\rho - \sqrt{-1}|\mathbf{w}|\text{sinh}2\rho)^{-\frac{4n+5}{2}} - (|\mathbf{z}|^2\text{cosh}2\rho + \sqrt{-1}|\mathbf{w}|\text{sinh}2\rho)^{-\frac{4n+5}{2}} \\ &= \text{cosh}^{-\frac{4n+5}{2}}(2\rho - \sqrt{-1}\psi) - \text{cosh}^{-\frac{4n+5}{2}}(2\rho + \sqrt{-1}\psi) \\ &= 2(4n+5)\sqrt{-1}\text{cosh}^{-\frac{4n+7}{2}}2\rho\text{sinh}2\rho\psi + O(|\psi|).\end{aligned}$$

Noting that $\frac{|\psi|}{|\mathbf{w}|} = \frac{|\psi|}{|\text{sin}\psi|} \le c$, we obtain the inequality:

$$\begin{aligned}|\Phi_1(\mathbf{h})| &\le 2^{2n+1}(2\pi)\Gamma\left(\frac{4n+7}{2}\right)|z_{10}| \\ &\quad \cdot \int_0^{+\infty} \text{cosh}^{-\frac{4n+5}{2}}2\rho\,\text{sinh}^2 2\rho\left(\frac{\rho}{\text{sinh}2\rho}\right)^{-1/2} d\rho.\end{aligned}$$

Since $\frac{\sinh 2\rho}{\rho} = 2\cosh(2\theta\rho) \leq 2\cosh 2\rho$, with $0 < \theta < 1$, we finally obtain an estimation for $\Phi_1(\mathbf{h})$:

$$\begin{aligned}
(9) \qquad |\Phi_1(\mathbf{h})| &\leq 2^{2n+3/2}(2\pi)\Gamma\left(\frac{4n+7}{2}\right)|z_{10}|\int_0^{+\infty}\cosh^{-2n}2\rho d\rho \\
&\leq c.2^{4n}(2\pi)\Gamma\left(\frac{4n+7}{2}\right)|z_{10}|B(n,n),
\end{aligned}$$

where $B(n,n)$ is Euler's Beta-function, since $\int_0^{+\infty}\cosh^{-2n} 2\rho\, d\rho = 2^{2n-2}B(n,n)$.

Now we begin to estimate $\Phi_2(\mathbf{h})$. This is different from the case of the Heisenberg groups. In the present situation the method to evaluate $\Phi_2(\mathbf{h})$ is not analogous to that for $\Phi_1(\mathbf{h})$, so we record its details. Let $\mathbf{z}_1' = (z_{11}, z_{12}, z_{13})$. Then

$$\begin{aligned}
\Phi_2(\mathbf{h}) &= \int_0^{+\infty} F_2(\delta_t(\mathbf{h}))t^{4n+5}dt \\
&= -\sqrt{-1}\int_0^{+\infty}\int_{\mathbf{R}^3}\exp\{-t^2(\sqrt{-1}\mathbf{v}.\mathbf{w} + |\mathbf{z}|^2\rho\cosh\rho)\} \\
&\quad\cdot\left(\frac{2\rho}{\sinh 2\rho}\right)^{2n} t^{4n+6}\sum_{j=1}^{3} v_j z_{1j} dt \prod_{j=1}^{3} dv_j \\
&= -\frac{\sqrt{-1}}{2}\Gamma\left(\frac{4n+7}{2}\right)\int_{\mathbf{R}^3}(|\mathbf{z}|^2\rho\coth 2\rho + \sqrt{-1}\mathbf{v}.\mathbf{w})^{-\frac{4n+7}{2}} \\
&\quad\cdot\left(\frac{2\rho}{\sinh 2\rho}\right)^{2n}\mathbf{z}_1'.\mathbf{v}\prod_{j=1}^{3} dv_j.
\end{aligned}$$

In the $\mathbf{w}$-space we take polar coordinates, and let the positive direction of the z-axis coincide with that of the vector $\mathbf{w}$, i.e.,

$$\begin{aligned}
\mathbf{w} &= |\mathbf{w}|(0,\ 0,\ 1), \\
\mathbf{v} &= \rho(\sin\theta\cos\phi,\ \sin\theta\sin\phi,\ \cos\theta), \\
\mathbf{z}_1' &= |\mathbf{z}_1'|(\sin\theta'\cos\phi',\ \sin\theta'\sin\phi',\ \cos\theta'),
\end{aligned}$$

where θ' is the angle between vectors $\mathbf{w}$ and $\mathbf{z}_1'$. Therefore

$$\Phi_2(\mathbf{w}) = -\frac{\sqrt{-1}}{2}\Gamma\left(\frac{4n+7}{2}\right)|\mathbf{z}_1'|$$
$$\cdot \int_0^{+\infty}\int_0^{\pi}\int_0^{2\pi}(|\mathbf{z}|^2\rho\coth 2\rho + \sqrt{-1}\rho|\mathbf{w}|\cos\theta)^{-\frac{4n+7}{2}}$$
$$\cdot (\sin\theta\sin\theta'\cos\phi\cos\phi' + \sin\theta\sin\theta'\sin\phi\sin\phi' + \cos\theta\cos\theta')\rho^3\sin\theta$$
$$\cdot \left(\frac{2\rho}{\sinh 2\rho}\right)^{2n} d\phi\, d\theta\, d\rho.$$

First, integrating with respect to ϕ yields

$$\Phi_2(\mathbf{h}) = -2^{2n-1}(2\pi)\sqrt{-1}\Gamma\left(\frac{4n+7}{2}\right)|\mathbf{z}_1'|\cos\theta'$$
$$\cdot \int_0^{+\infty}\int_0^{\pi}(|\mathbf{z}|^2\cosh 2\rho + \sqrt{-1}|\mathbf{w}|\sinh 2\rho\cos\theta)^{-\frac{4n+7}{2}}$$
$$\cdot \left(\frac{\rho}{\sinh 2\rho}\right)^{-7/2}\rho^3\sin\theta\cos\theta\, d\theta\, d\rho.$$

Let

$$K(\rho) = \int_0^{\pi}(|\mathbf{z}|^2\cosh 2\rho + \sqrt{-1}|\mathbf{w}|\sinh 2\rho\cos\theta)^{-\frac{4n+7}{2}}\sin\theta\cos\theta d\theta.$$

Then a simple calculation gives us

$$K(\rho) = \frac{1}{\sqrt{-1}|\mathbf{w}|\sinh 2\rho}\int_0^{\pi}\Big[(|\mathbf{z}|^2\cosh 2\rho + \sqrt{-1}|\mathbf{w}|\sinh 2\rho\cos\theta)^{-\frac{4n+5}{2}}$$
$$-|\mathbf{z}|^2\cosh 2\rho(|\mathbf{z}|^2\cosh 2\rho + \sqrt{-1}|\mathbf{w}|\sinh 2\rho\cos\theta)^{-\frac{4n+7}{2}}\Big]\sin\theta d\theta$$
$$= \frac{1}{(\sqrt{-1}|\mathbf{w}|\sinh 2\rho)^2}\Bigg\{\frac{2}{4n+3}\Big[(|\mathbf{z}|^2\cosh 2\rho - \sqrt{-1}|\mathbf{w}|\sinh 2\rho)^{-\frac{4n+3}{2}}$$
$$-(|\mathbf{z}|^2\cosh 2\rho + \sqrt{-1}|\mathbf{w}|\sinh 2\rho)^{-\frac{4n+3}{2}}\Big]$$
$$-\frac{2}{4n+5}|\mathbf{z}|^2\cosh 2\rho\Big[(|\mathbf{z}|^2\cosh 2\rho - \sqrt{-1}|\mathbf{w}|\sinh 2\rho)^{-\frac{4n+5}{2}}$$
$$-(|\mathbf{z}|^2\cosh 2\rho + \sqrt{-1}|\mathbf{w}|\sinh 2\rho)^{-\frac{4n+5}{2}}\Big]\Bigg\}.$$

Similar to (8), we have

$$
\begin{aligned}
&K(\rho)\\
&= -\frac{1}{\sqrt{-1}|\mathbf{w}|^2\sinh^2 2\rho}\left\{\frac{2}{4n+3}\right.\\
&\quad\cdot\left[\cosh^{-\frac{4n+3}{2}}(2\rho-\sqrt{-1}\psi)-\cosh^{-\frac{4n+3}{2}}(2\rho+\sqrt{-1}\psi)\right]\\
&\quad-\frac{2}{4n+5}|\mathbf{z}|^2\cosh 2\rho\\
&\quad\left.\cdot\left[\cosh^{-\frac{4n+5}{2}}(2\rho-\sqrt{-1}\psi)-\cosh^{-\frac{4n+5}{2}}(2\rho+\sqrt{-1}\psi)\right]+O(\psi)\right\}\\
&= -\frac{4\sqrt{-1}}{|\mathbf{w}|^2\sinh^2 2\rho}\left\{\cosh^{-\frac{4n+5}{2}}2\rho\sinh 2\rho\psi(1-|\mathbf{z}|^{\mathbf{2}})+\mathbf{O}(\psi)\right\},
\end{aligned}
$$

and hence we obtain the estimation

$$
\begin{aligned}
|K(\rho)| &\le c.\frac{1}{|\mathbf{w}|\sinh^2 2\rho}\left|\cosh^{-\frac{4n+5}{2}}2\rho\sinh 2\rho\left(\frac{\sqrt{-1}\psi}{\sin\psi}\right)\right.\\
&\qquad\left.-|\mathbf{z}|^2\cosh^{-\frac{4n+5}{2}}2\rho\sinh 2\rho\left(\frac{\sqrt{-1}\psi}{\sin\psi}\right)\right|\\
&\le c\frac{(1-|\mathbf{z}|^2)}{|\mathbf{w}|\sinh 2\rho}\cosh^{-\frac{4n+5}{2}}2\rho \le \frac{c}{\sinh 2\rho}\cosh^{-\frac{4n+5}{2}}2\rho.
\end{aligned}
$$

The last inequality follows from the fact that $1-|\mathbf{z}|^2=\frac{|\mathbf{w}|^2}{1+|\mathbf{z}|^2}\le|\mathbf{w}|^2\le 1$, since $(\mathbf{z},\mathbf{w})\in S^n$. Substituting the estimation of $K(\rho)$ into the expression of $\Phi_2(\mathbf{h})$, we finally obtain

(10)

$$
\begin{aligned}
|\Phi_2(\mathbf{h})| &\le c.2^{2n-1}(2\pi)\Gamma\left(\frac{4n+7}{2}\right)|\mathbf{z}_1'|\int_0^{+\infty}\cosh^{-\frac{4n+5}{2}}2\rho\left(\frac{\rho}{\sinh 2\rho}\right)^{-5/2}d\rho\\
&\le c.2^{2n}(2\pi)\Gamma\left(\frac{4n+7}{2}\right)|\mathbf{z}_1'|\int_0^{+\infty}\cosh^{-2n}2\rho d\rho\\
&\le c.2^{4n}(2\pi)\Gamma\left(\frac{4n+7}{2}\right)|\mathbf{z}_1'|B(n,n).
\end{aligned}
$$

Lemma 4.4. *The surface measure of the unit sphere S^n of the quaternionic Heisenberg group $\mathbf{H}H^n$ is*

$$\sigma(S^n)=2\pi^{2n+3/2}\frac{\Gamma(n)}{\Gamma(2n)\Gamma(n+3/2)}.$$

Proof. Let $f\in L^1(\mathbf{H}H^n)$ and

$$\begin{aligned}I(f) &= \int_{\mathbf{H}H^n} f(\mathbf{x},\mathbf{u})d(\mathbf{x},\mathbf{u})\\ &= \int_0^{+\infty}\int_0^{+\infty}\int_{\Sigma^{4n-1}}\int_{\Sigma^2} f(R\mathbf{z},\rho\mathbf{w})R^{4n-1}\rho^2 d\mathbf{w}d\mathbf{z}\,d\rho\,dR,\end{aligned}$$

where

$$\Sigma^{n-1} = \left\{\mathbf{x} = (x_1,\ldots,x_n)\in\mathbf{R}^n \;\Big|\; \sum_{i=1}^n x_i^2 = 1\right\}$$

is the unit sphere in Euclidean space $\mathbf{R}^n$. Performing the change of variables $R^2 = l^2\cos\theta$, $\rho = l^2\sin\theta$, we get

$$\begin{aligned}I(f) = \int_0^{+\infty}\int_0^{\pi/2}\int_{\Sigma^{4n-1}}\int_{\Sigma^2} & f(l\mathbf{z}\cos^{1/2}\theta, l^2\mathbf{w}\sin\theta)\\ & \cdot l^{4n+5}\cos^{2n-1}\theta\sin^2\theta\, d\mathbf{w}\, d\mathbf{z}\, d\theta\, dl.\end{aligned}$$

Hence

(11)

$$\int_{S^n} f(\mathbf{h})d\sigma(\mathbf{h}) = \int_0^{\pi/2}\int_{\Sigma^{4n-1}}\int_{\Sigma^2} f(\mathbf{z}\cos^{1/2}\theta, \mathbf{w}\sin\theta)\cos^{2n-1}\theta\sin^2\theta d\mathbf{w}d\mathbf{z}d\theta.$$

In particular,

$$\begin{aligned}\sigma(S^n) &= \int_0^{\pi/2}\int_{\Sigma^{4n-1}}\int_{\Sigma^2}\cos^{2n-1}\theta\sin^2\theta d\mathbf{w}d\mathbf{z}d\theta\\ &= |\Sigma^{4n-1}|.|\Sigma^2|\int_0^{\pi/2}\cos^{2n-1}\theta\sin^2\theta d\theta = \frac{1}{2}|\Sigma^{4n-1}|.|\Sigma^2|B(n,3/2).\end{aligned}$$

Then Lemma 4.4 follows from the expression of the surface measure $|\Sigma^{n-1}| = \frac{2\pi^{n/2}}{\Gamma(n/2)}$ of the unit sphere in Euclidean space $\mathbf{R}^n$.

Now we turn to evaluate the following integral:

Lemma 4.5. *We have*

$$\int_{S^n} |z_{10}|^q d\sigma(\mathbf{h}) = \frac{2\pi^{2n+1}\Gamma(n+q/4)\Gamma(\frac{q+1}{2})}{\Gamma(n+q/4+3/2)\Gamma(2n+q/2)}.$$

Proof. Let $\Sigma^{4n-1} = \{\mathbf{z} = (\mathbf{z}_1,\mathbf{z}')\in\mathbf{R}^4\times\mathbf{R}^{4n-4} \mid |\mathbf{z}_1|^2 + |\mathbf{z}'|^2 = 1\}$. We introduce new polar coordinates by setting $\mathbf{z}_1 = \mathbf{a}\cos\phi$, $\mathbf{z}' = \mathbf{b}\sin\phi$ with $\mathbf{a}\in\Sigma^3$, $\mathbf{b}\in\Sigma^{4n-5}$, $0\le\phi\le\pi/2$. Then formula (11) reads as

$$\begin{aligned}&\int_{S^n} f(\mathbf{h})d\sigma(\mathbf{h})\\ &= \int_0^{\pi/2}\int_0^{\pi/2}\int_{\Sigma^3}\int_{\Sigma^{4n-5}}\int_{\Sigma^2} f(\mathbf{a}\cos^{1/2}\theta\cos\phi, \mathbf{b}\cos^{1/2}\theta\sin\phi, \mathbf{w}\sin\theta)\\ &\qquad\cdot\cos^{2n-1}\theta\sin^2\theta\cos^3\phi\sin^{4n-5}\phi d\mathbf{w}d\mathbf{b}d\mathbf{a}d\theta d\phi.\end{aligned}$$

In particular, we have

$$\int_{S^n}|z_{10}|^q d\sigma(\mathbf{h})=\int_0^{\pi/2}\int_0^{\pi/2}\int_{\Sigma^3}\int_{\Sigma^{4n-5}}\int_{\Sigma^2}|a_0\cos^{1/2}\theta\cos\phi|^q \cdot\cos^{2n-1}\theta\sin^2\theta\cos^3\phi\sin^{4n-5}\phi d\mathbf{w}d\mathbf{b}d\mathbf{a}d\theta d\phi. \tag{12}$$

Furthermore, on the unit sphere Σ^3 of the $\mathbf{a}$-space, we employ the spherical coordinates, i.e.,

$$\mathbf{a}=\begin{pmatrix}a_0\\a_1\\a_2\\a_3\end{pmatrix}=\begin{pmatrix}\sin\psi_1\\ \cos\psi_1\sin\psi_2\\ \cos\psi_1\cos\psi_2\sin\phi_3\\ \cos\psi_1\cos\psi_2\cos\psi_3\end{pmatrix},$$

with $-\pi/2\le\psi_1,\ \psi_2\le\pi/2$ and $-\pi\le\psi_3\le\pi$. Then the integral on Σ^3 in (12) becomes

$$\int_{\Sigma^3}|a_0|^q d\mathbf{a}$$
$$=\int_{-\pi/2}^{\pi/2}\int_{-\pi/2}^{\pi/2}\int_{-\pi}^{\pi}|\sin\psi_1|^q\cos^2\psi_1\cos\psi_2 d\psi_3 d\psi_2 d\psi_1=4\pi B\left(\frac{q+1}{2},\ \frac{3}{2}\right).$$

Finally we obtain

$$\int_{S^n}|z_{10}|^q d\sigma(\mathbf{h})$$
$$=4\pi B\left(\frac{q+1}{2},\ \frac{3}{2}\right)$$
$$\cdot\int_0^{\pi/2}\int_0^{\pi/2}\int_{\Sigma^{4n-5}}\int_{\Sigma^2}\cos^{2n+q/2-1}\theta\sin^2\theta\cos^{q+3}\phi\sin^{4n-5}\phi d\mathbf{w}d\mathbf{b}d\phi d\theta$$
$$=\pi|\Sigma^{4n-5}||\Sigma^2|B\left(\frac{q+1}{2},\ \frac{3}{2}\right)B\left(n+\frac{q}{4},\ \frac{3}{2}\right)B\left(\frac{q+4}{2},\ 2n-2\right),$$

as required.

Corollary 4.6. *We also have*

$$\int_{S^n}|\mathbf{z}_1'|^q d\sigma(\mathbf{h})=\frac{q+1}{2}\left(\frac{\pi^{2n+1}\Gamma(n+\frac{q}{4})\Gamma(\frac{q+1}{2})}{\Gamma(n+\frac{q+6}{4})\Gamma(2n+\frac{q}{2})}\right).$$

Proof. Since

$$|\mathbf{z}_1'|=\left(\sum_{i=1}^{3}a_{1i}^2\right)^{1/2}\cos^{1/2}\theta\cos\phi=\cos\psi_1\cos^{1/2}\theta\cos\phi,$$

an analogous calculation to that in the proof of Lemma 4.5 gives:

$$\int_{S^n} |\mathbf{z}_1'|^q d\sigma(\mathbf{h})$$
$$= \pi|\Sigma^{4n-5}||\Sigma^2|B\left(\frac{q+3}{2}, \frac{1}{2}\right) B\left(n+\frac{q}{4}, \frac{1}{2}\right) B\left(\frac{q+4}{2}, 2n-2\right),$$

which is exactly the conclusion of Corollary 4.6.

Lemma 4.7. *There exists a constant* $c = c(q) > 0$*, such that,* $\forall n \in \mathbf{N}$*,*

$$\left(\int_{S^n} |\Phi(\mathbf{h})|^q d\sigma(\mathbf{h})\right)^{1/q} \leq c[\sigma(S^n)]^{-1/p}.$$

Proof. We know ([**10**]) that when $x \to +\infty$,

$$\frac{\Gamma(x)}{\Gamma(x+a)} \sim x^{-a}, \quad a > 0, \tag{13}$$

and Stirling's formula

$$\Gamma(x) \sim \sqrt{2\pi} x^{x-1/2} \exp(-x). \tag{14}$$

From Equations (7), (9), (10), it follows that

$$|\Phi(\mathbf{h})| \leq c.\frac{2^{4n}}{(2\pi)^{6n+2}} \Gamma(2n+7/2) B(n,\ n) \max\{|z_{10}|,\ |\mathbf{z}_1'|\}.$$

Via Lemmas 4.4 and 4.5 and Corollary 4.6 we get

$$
\begin{aligned}
&[\sigma(S^n)]^{1/p}.\left(\int_{S^n}|\Phi(\mathbf{h})|^q d\sigma(\mathbf{h})\right)^{1/q}\\
&\le c.\frac{2^{4n}}{(2\pi)^{6n+2}}\Gamma(2n+7/2)B(n,\ n)\\
&\quad.\left(\frac{\pi^{2n+1}\Gamma(n+q/4)\Gamma(\frac{q+1}{2})}{\Gamma(n+\frac{q+6}{4})\Gamma(2n+q/2)}\right)^{1/q}\left(\frac{2\pi^{2n+3/2}\Gamma(n)}{\Gamma(n+3/2)\Gamma(2n)}\right)^{1/p}\\
&\le c.\frac{2^{4n}}{(2\pi)^{4n+1}}\Gamma(2n+7/2)B(n,\ n)\\
&\quad.\left(\frac{\Gamma(n+q/4)}{\Gamma(n+q/4+3/2)\Gamma(2n+q/2)}\right)^{1/q}\left(\frac{\Gamma(n)}{\Gamma(n+3/2)\Gamma(2n)}\right)^{1/p}\\
&\le c.\Gamma(2n+7/2)B(n,\ n)\left(\frac{1}{\Gamma(2n+q/2)}\right)^{1/q}\left(\frac{1}{\Gamma(2n)}\right)^{1/p}\\
&\le c.B(n,\ n)\left(\frac{\Gamma(2n+7/2)}{\Gamma(2n)}\right)\left(\frac{\Gamma(2n)}{\Gamma(2n+q/2)}\right)^{1/q}\\
&\le c.(2n)^3B(n,\ n)\\
&\le c.(2n)^3\left(\frac{(\sqrt{2\pi}n^{n-1/2}\exp(-n))^2}{\sqrt{2\pi}(2n)^{2n-1/2}\exp(-2n)}\right)\\
&\le c,
\end{aligned}
$$

which completes the proof of Lemma 4.7. (In the proof of this Lemma, we have used formulas (13) and (14).)

Due to Lemma 4.7 and the inequalities (3), (4), (5) we have proved:

Proposition 4.8. *For every* $p\in(0,\ +\infty)$*, there exsists a constant* $c(p)>0$*, such that*

$$\|\widetilde{R}f\|_{L^p(\mathbf{H}H^n)}\le c(p)\|f\|_{L^p(\mathbf{H}H^n)},\ \forall f\in C_0^\infty(\mathbf{H}H^n),\ \forall n\in\mathbf{N}.$$

4.3. The full Riesz transforms. We consider the full Riesz transform $\mathbf{R}f(\mathbf{x},\mathbf{u})=(R_{10}f(\mathbf{x},\mathbf{u}),\dots,R_{n3}f(\mathbf{x},\mathbf{u}))$, where

$$
\begin{aligned}
R_{ri}f(\mathbf{x},\mathbf{u})&=(X_{ri}\triangle^{-1/2}f)(\mathbf{x},\mathbf{u})\\
&=\lim_{\varepsilon\to0^+}\int_{\mathbf{H}H^n}X_{ri}p_1(\mathbf{y},\mathbf{v})\int_\varepsilon^{1/\varepsilon}f((\mathbf{x},\mathbf{u})\delta_t((\mathbf{y},\mathbf{v})^{-1}))\frac{dt}{t}d(\mathbf{y},\mathbf{v}).
\end{aligned}
$$

Analogous to (6), we have

$$X_{ri}p_1(\mathbf{y},\mathbf{v})=\frac{2}{(2\pi)^{6n+3}}(-F_1+\sqrt{-1}F_2)(\mathbf{y},\mathbf{v})=\frac{2}{(2\pi)^{6n+3}}F(\mathbf{y},\mathbf{v}),$$

hence

$$
\begin{aligned}
&(X_{ri}\triangle^{-1/2}f)(\mathbf{x},\mathbf{u})\\
&= -\frac{2}{(2\pi)^{6n+3}}\lim_{\varepsilon\to 0^+}\int_{\mathbf{H}H^n}(F_1-\sqrt{-1}F_2)(\mathbf{y},\mathbf{v})\\
&\quad\cdot\int_\varepsilon^{1/\varepsilon} f((\mathbf{x},\mathbf{u})\delta_t((\mathbf{y},\mathbf{v})^{-1}))\frac{dt}{t}d(\mathbf{y},\mathbf{v}),
\end{aligned}
$$

where

$$
\begin{aligned}
F_1(\mathbf{x},\mathbf{u}) &= x_{10}\int_{\mathbf{R}^3}\exp\{-(\sqrt{-1}\mathbf{v}.\mathbf{u}+|\mathbf{x}|^2\rho\coth 2\rho)\}\\
&\quad\cdot\left(\frac{2\rho}{\sinh 2\rho}\right)^{2n}\rho\coth 2\rho\prod_{j=1}^{3}dv_j,\\
F_2(\mathbf{x},\mathbf{u}) &= \int_{\mathbf{R}^3}\exp\{-(\sqrt{-1}\mathbf{v}.\mathbf{u}+|\mathbf{x}|^2\rho\coth 2\rho)\}\\
&\quad\cdot\left(\frac{2\rho}{\sinh 2\rho}\right)^{2n}\sum_{j=1}^{3}x_{1j}v_j\prod_{j=1}^{3}dv_j.
\end{aligned}
$$

From these equations it is easy to see that

$$F(-\mathbf{x},\mathbf{u}) = -F(\mathbf{x},\mathbf{u}),$$

and hence we can write

$$
\begin{aligned}
&(X_{ri}\triangle^{-1/2}f)(\mathbf{x},\mathbf{u})\\
&= \frac{1}{(2\pi)^{6n+3}}\int_{\mathbf{H}H^n}F(\mathbf{y},\mathbf{v})\\
&\quad\cdot\int_0^{+\infty}[f((\mathbf{x},\mathbf{u})\delta_t(-\mathbf{y},-\mathbf{v}))-f((\mathbf{x},\mathbf{u})\delta_{\mathbf{t}}(\mathbf{y},-\mathbf{v}))]\frac{\mathbf{dt}}{\mathbf{t}}\mathbf{d}(\mathbf{y},\mathbf{v}).
\end{aligned}
$$

Similar arguments to those in [**3**] show that:

Theorem 4.9. *For every $p\in(1,+\infty)$, there exists a constant $c=c(p)>0$, such that for all $n\in\mathbf{N}$,*

$$\frac{1}{c}\|f\|_{L^p(\mathbf{H}H^n)}\le\||\mathbf{R}f|\|_{L^p(\mathbf{H}H^n)}\le c\|f\|_{L^p(\mathbf{H}H^n)},\ \forall f\in C_0^\infty(\mathbf{H}H^n).$$

References

[1] D. Allcock, *An isoperimetric inequality for the Heisenberg groups*, GAFA, Geom. Funct. Anal., **8** (1998), 219-233, MR 99g:53038, Zbl 0907.22009.

[2] S. Barker and S. Salamon, *Analysis on a generalized Heisenberg group*, J. London Math. Soc., **28** (1983), 184-192, MR 85d:22022, Zbl 0522.22008.

[3] T. Coulhon, D. Műller and J. Zienkiewicz, *About Riesz transforms on the Heisenberg groups*, Math. Ann., **305** (1996), 369-379, MR 97f:22015, Zbl 0859.22006.

[4] J. Cygan, *Heat kernels for class* 2 *nilpotent groups*, Studia Math., **64** (1979), 227-238, MR 82b:22016, Zbl 0404.35027.

[5] G. Folland, *A fundamental solution for a subelliptic operator*, Bull. Amer. Math. Soc., **29** (1973), 373-376, MR 47 #3816, Zbl 0256.35020.

[6] N. Garofalo and F. Segala, *Estimates of the fundamental solution and Wiener's criterion for the heat equation on the Heisenberg group*, Indiana Univ. Math. Jour., **38**(**4**) (1989), 1155-1196, MR 91k:35138, Zbl 0808.35046.

[7] B. Gaveau, *Principle de moindre action, propogation de la chaleur et estimees sous-elliptiques sur certains groups nilpotents*, Acta Math., **139** (1997), 95-153, MR 57 #1574, Zbl 0366.22010.

[8] H. Hueber and D. Műller, *Asymptotics for some Green kernels on the Heisenberg group and the Martin boundary*, Math. Ann., **283** (1989), 97-119, MR 90b:58257, Zbl 0639.31005.

[9] A. Hulanicki, *The distribution of energy of the Brownian motion in Gaussian field and analytic hypoellipticity of certain subelliptic operators on the Heisenberg group*, Studia Math., **56** (1976), 165-173, MR 54 #6298, Zbl 0336.22007.

[10] N. Lohoué and N. Varopoulos, *Remarques sur les transformees de Riesz sur les groupes nilpotents*, C. R. Acad. Sci. Paris, **301**(**11**) (1985), 559-560, MR 87b:43008, Zbl 0582.43003.

[11] E. Stein, *Some results in harmonic analysis in* $\mathbf{R}^n$ *for* $n \to +\infty$, Bull. Amer. Math. Soc., **9** (1983), 71-73, MR 84g:42019, Zbl 0515.42018.

[12] E. Titchmarsh, *The Theory of Functions*, Oxford University Press, Oxford, 1952, MR 33 #5850, Zbl 0336.30001.

Received August 9, 2000 and revised June 20, 2002. This work is partially supported by Chinese National Natural Science Foundation Grant-19871063.

Institute of Mathematics
Wuhan University
Wuhan, 430072
P.R. China
E-mail address: flzhu@whu.edu.cn

Guidelines for Authors

Authors may submit manuscripts at pjm.math.berkeley.edu/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 Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

PACIFIC JOURNAL OF MATHEMATICS

Volume 209 No. 1 March 2003

0030-8730(200303)209:1;1-O